



OPEN ACCESS

EDITED BY

Zhou Zhang,
University of Wisconsin-Madison,
United States

REVIEWED BY

Lorena Parra,
Universitat Politècnica de València, Spain
Paulo Flores,
North Dakota State University, United States
Sushopti Gawade,
Vidyalankar Institute of Technology, India

*CORRESPONDENCE

Zhaojian Li

✉ lizhaoj1@egr.msu.edu

RECEIVED 05 March 2024

ACCEPTED 24 July 2024

PUBLISHED 20 August 2024

CITATION

Li J, Chen D, Yin X and Li Z (2024)
Performance evaluation of semi-supervised
learning frameworks for multi-class
weed detection.
Front. Plant Sci. 15:1396568.
doi: 10.3389/fpls.2024.1396568

COPYRIGHT

© 2024 Li, Chen, Yin and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Performance evaluation of semi-supervised learning frameworks for multi-class weed detection

Jiajia Li¹, Dong Chen², Xunyu Yin³ and Zhaojian Li^{4*}

¹Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, United States, ²Environmental Institute, University of Virginia, Charlottesville, VA, United States, ³School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore, Singapore, ⁴Department of Mechanical Engineering, Michigan State University, East Lansing, MI, United States

Precision weed management (PWM), driven by machine vision and deep learning (DL) advancements, not only enhances agricultural product quality and optimizes crop yield but also provides a sustainable alternative to herbicide use. However, existing DL-based algorithms on weed detection are mainly developed based on supervised learning approaches, typically demanding large-scale datasets with manual-labeled annotations, which can be time-consuming and labor-intensive. As such, label-efficient learning methods, especially semi-supervised learning, have gained increased attention in the broader domain of computer vision and have demonstrated promising performance. These methods aim to utilize a small number of labeled data samples along with a great number of unlabeled samples to develop high-performing models comparable to the supervised learning counterpart trained on a large amount of labeled data samples. In this study, we assess the effectiveness of a semi-supervised learning framework for multi-class weed detection, employing two well-known object detection frameworks, namely FCOS (Fully Convolutional One-Stage Object Detection) and Faster-RCNN (Faster Region-based Convolutional Networks). Specifically, we evaluate a generalized student-teacher framework with an improved pseudo-label generation module to produce reliable pseudo-labels for the unlabeled data. To enhance generalization, an ensemble student network is employed to facilitate the training process. Experimental results show that the proposed approach is able to achieve approximately 76% and 96% detection accuracy as the supervised methods with only 10% of labeled data in CottonWeedDet3 and CottonWeedDet12, respectively. We offer access to the source code (<https://github.com/JiajiaLi04/SemiWeeds>), contributing a valuable resource for ongoing semi-supervised learning research in weed detection and beyond.

KEYWORDS

precision weed management, precision agriculture, label-efficient learning, computer vision, deep learning

1 Introduction

Weeds pose a significant risk to global crop production, with potential losses attributed to these unwelcome plants estimated at 43% worldwide (Oerke, 2006). Specifically, in the context of cotton farming, inefficient management of weeds can result in a staggering 90% reduction in yield (Manalil et al., 2017). Traditional weed control methods typically involve the use of machinery, manual weeding, or application of herbicides. These weed management approaches, while commonly utilized, require significant labor and cost considerations. Manual and mechanical weeding methods are especially labor-intensive, a predicament that has been intensified by recent global labor shortages triggered by public health crises (e.g., the COVID-19 pandemic) and geopolitical conflicts (e.g., the Russia-Ukraine War) (Laborde et al., 2020; Ben Hassen and El Bilali, 2022). Furthermore, the use of herbicides brings about significant environmental harm and potential risks to human health, and contributes to the emergence of herbicide-resistant weed species (Norsworthy et al., 2012; Chen et al., 2022b).

PWM, integrating sensors, computer systems, and robotics into agricultural practices, has emerged as a promising and sustainable approach for efficient weed management (Young et al., 2013). It allows for targeted treatment based on specific site conditions and weed species, thereby significantly minimizing the use of herbicides and other resources (Gerhards and Christensen, 2003). To achieve successful implementation of PWM, it is essential to accurately identify, localize, and monitor weeds, which requires robust machine vision algorithms for weed recognition (Chen et al., 2022b). Traditional image processing techniques, often encompassing edge detection, color analysis, and texture feature extraction, along with subsequent steps such as thresholding or supervised modeling, are widely utilized in the field of weed classification and detection (Meyer and Neto, 2008; Wang et al., 2019). For instance, a weed classification algorithm that relies on extracted texture features was developed by (Bawden et al., 2017). Ahmad et al. (2018) used local shape and edge orientation features to differentiate between monocot and dicot weeds. However, despite promising results, these conventional machine vision techniques often necessitate manual feature engineering for specific weed detection or classification tasks, which requires extensive domain knowledge and can be error-prone and time-consuming. Moreover, these methods may struggle with complex visual tasks and be sensitive to variations in lighting conditions and occlusion (O'Mahony et al., 2020).

Recently, DL-based advanced computer vision has been recognized as a promising approach for sustainable weed management (Farooq et al., 2019; Yu et al., 2019; Parra et al., 2020; Chen et al., 2022b; Coleman et al., 2023; Rahman et al., 2023; Rai et al., 2023; Sportelli et al., 2023). For example, four different YOLO (You Only Look Once) object detectors were evaluated for weed detection in different turfgrass scenarios in Sportelli et al. (2023). Additionally, in Chen et al. (2022b), 35 state-of-the-art deep neural networks (DNNs) were examined and benchmarked for multi-class weed classification within cotton production systems, with nearly all models attaining high classification accuracy, reflected by F1 scores exceeding 95%. Despite their proven

effectiveness, these DL-based approaches are notoriously data-hungry, and their performance is heavily dependent on large-scale and accurately labeled image datasets (Lu and Young, 2020; Rai et al., 2023), whereas manually labeling such large-scale image datasets is often error-prone, tedious, expensive, and time-consuming (Li et al., 2023).

To address these challenges, label-efficient learning algorithms (Li et al., 2023) have emerged as promising solutions to reduce the high labeling costs by harnessing the potential of unlabeled samples. Specifically, in dos Santos Ferreira et al. (2019), the efficacy of two popular unsupervised learning algorithms, namely Joint Unsupervised Learning of Deep Representations and Image Clusters (JULE, Yang et al. (2016)) and Deep Clustering for Unsupervised Learning of Visual Features (DeepCluster, Caron et al. (2018)), were evaluated in the context of weed recognition utilizing two publicly available weed datasets. In addition, the semi-supervised learning for weed classification was studied in (Liu et al., 2023, 2024; Benchallal et al., 2024). Furthermore, a semi-supervised learning strategy called SemiWeedNet was introduced in Nong et al. (2022); this method was designed for the segmentation of weeds and crops in challenging environments characterized by complex backgrounds. Moreover, the study presented in Hu et al. (2021) employed the cut-and-paste image synthesis approach and semi-supervised learning to address the issue of insufficient training data for weed detection. This approach was evaluated on an image dataset consisting of 500 images across four categories: "cotton", "morningglory", "grass", and "other", which culminated in an mAP of 46.0. Although the results were intriguing, their methodology was tested only on a two-stage object detector [i.e., Faster-RCNN (Ren et al., 2015)] and a four-category image dataset, which does not sufficiently substantiate the efficacy of semi-supervised learning for weed detection. Therefore, our research aims to further probe the potential of semi-supervised learning in weed detection, and comparatively assess a variety of object detectors and multi-class weed species. The key contributions of this study are as follows:

- We rigorously evaluate the semi-supervised learning framework utilizing two open-source cotton weed datasets. These datasets include 3 and 12 weed classes commonly found in U.S. cotton production systems.
- We further analyze and compare the performance of one-stage and two-stage object detectors within the semi-supervised learning framework.
- In the spirit of reproducibility, we make all our training and evaluation codes¹ freely accessible.

The remainder of this paper is organized as follows: Section 2 details the dataset and technical aspects pertinent to this study. Section 3 presents experimental results and provides a comprehensive analysis, followed by further discussions and limitations in Section 4. Lastly, Section 5 offers concluding remarks and outlines potential future research directions.

¹ <https://github.com/JijiaLi04/SemiWeeds>

2 Materials and methods

In this section, we begin by introducing the two datasets employed in our study. Then, we provide an overview of two representative object detectors: the two-stage Faster R-CNN and the one-stage FCOS detector, along with the details of our semi-supervised framework. Lastly, we present the evaluation metrics and describe the experimental setups.

2.1 Weed datasets

To assess the performance and efficacy of our semi-supervised framework, we conducted evaluations on two publicly available weed datasets tailored specifically to the U.S. cotton production systems: CottonWeedDet3 (Rahman et al., 2023) and CottonWeedDet12 (Dang et al., 2023).

CottonWeedDet3² (Rahman et al., 2023) comprises 848 high-resolution images (4442 × 4335 pixels) annotated with 1532 bounding boxes. It contains three distinct classes of weeds commonly found in southern U.S. cotton fields, primarily in North Carolina and Mississippi. These images include three types of weeds: carpetweed (*mollugo verticillata*), morning glory (*ipomoea genus*), and palmer amaranth (*amaranthus palmeri*). For adaptability, the annotations for each image were saved in both YOLO and COCO formats. Notably, around 99% of the images contain less than 10 bounding boxes, with only a small portion (9 out of the 848 images) containing a more substantial quantity of bounding boxes, even up to 93 in some cases. Additionally, carpetweed is the most frequently annotated, while palmer amaranth is the least. Visual examples of the three-class weed images can be found in Figure 1.

CottonWeedDet12 dataset³ (Dang et al., 2023) contains 5648 images of 12 weed classes, annotated with a total of 9370 bounding boxes (saved in both YOLO and COCO formats). These images, with a resolution exceeding 10 megapixels, were captured under natural lighting conditions and across various weed growth stages in cotton fields. Each weed class is represented by more than 140 bounding boxes. Moreover, waterhemp and morning glory have the highest number of bounding boxes while goose grass and cutleaf ground cherry have the least. In terms of image volume, the CottonWeedDet12 dataset surpasses the CottonWeedDet3 dataset (Rahman et al., 2023) by more than tenfold. Moreover, it represents the most extensive public dataset currently available for weed detection in cotton production systems. Figure 2 shows sample annotated images where a single weed class in each image is present, despite that each image may include multiple weed classes in the dataset.

2.2 DL-based object detectors

DL-based object detectors are typically structured around two primary components: a backbone and a detection head (Bochkovskiy et al., 2020). The backbone is responsible for extracting features from high-dimensional inputs and is commonly pre-trained on ImageNet data (Deng et al., 2009). Conversely, the head is leveraged to predict the classes and bounding boxes of objects. Existing detectors consist of anchor-based detectors (Ren et al., 2015; Cai et al., 2016; Lin et al., 2017) and anchor-free detectors (Law and Deng, 2018; Tian et al., 2022; Zhou et al., 2019). Anchor-based detectors utilize pre-defined anchor boxes, adjusting them for position shifts and scaling to align with the ground-truth boxes, primarily based on their intersection-over-union (IoU) scores. Conversely, the pre-defined anchor boxes are discarded in the detection head for the anchor-free object detection models.

2.2.1 Anchor-based detectors

Anchor-based object detectors utilize pre-defined anchor boxes to efficiently localize and classify objects in images, being a representative approach in object detection methodologies. These methods have led to significant advancements and impressive outcomes in object detection (Ren et al., 2015; Cai et al., 2016; Lin et al., 2017). The most notable embodiment of this framework is Faster-RCNN (Ren et al., 2015), which was built upon the earlier Fast RCNN model (Girshick, 2015). Deviating from the selective search methods utilized in Fast RCNN, Faster RCNN employs CNNs to generate region proposals via an efficient Region Proposal Network (RPN). The features from the final shared convolutional layer are then harnessed for both RPN's region proposal task and Fast RCNN's region classification task. In this study, we use Faster RCNN as one of the detectors in our semi-supervised framework.

2.2.2 Anchor-free detectors

While anchor-based detectors have demonstrated impressive outcomes, their application to novel datasets necessitates expertise in tuning hyperparameters (Jiao et al., 2019) associated with anchor boxes. This constraint limits the adaptability of these detectors to new datasets or environments (Zhang et al., 2020). Furthermore, anchor-based approaches are often proved to be computationally expensive for current mobile/edge devices used in agricultural applications, which typically have constrained storage and computational capacity. Alternatively, these limitations are addressed in anchor-free detectors by getting rid of the need for pre-defined anchor boxes in detection models. These methods can directly predict class probabilities and bounding box offsets from full images using a single feed-forward CNN without necessitating the generation of region proposals or subsequent classification/feature resampling, thereby encapsulating all computation within a single network (Liu et al., 2020). YOLO (Redmon et al., 2016), one of the most representative one-stage detectors, transforms the task of object detection into a regression problem by directly mapping image pixels to spatially separated bounding boxes and corresponding class probabilities. YOLO is designed for speed, capable of operating in real-time at 45 frames per second (FPS) by eliminating the region proposal generation process. On the other

² CottonWeedDet3 dataset: <https://www.kaggle.com/datasets/yuzhenlu/cottonweeddet3>

³ CottonWeedDet12 dataset: <https://zenodo.org/record/7535814>

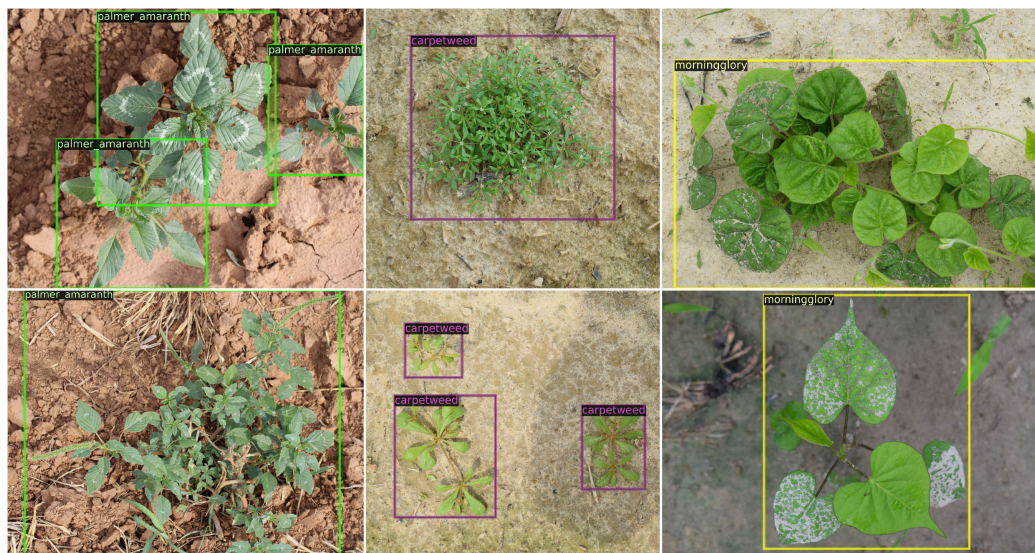


FIGURE 1 Weed samples in the CottonWeedDet3 dataset (Rahman et al., 2023). Each column represents the image samples for one weed class.

hand, FCOS (Tian et al., 2022) is an anchor box-free and proposal-free one-stage object detector. By eliminating the anchor box designs, FCOS avoids the complicated computation related to anchor boxes such as calculating overlapping during training and all hyper-parameters related to anchor boxes. In this study, FCOS serves as one of our base object detection models, chosen for its accessibility and extensive adoption within the field as evidenced by previous research (Zhang et al., 2020; Li et al., 2021).

2.3 Semi-supervised Learning

Semi-supervised learning, a form of label-efficient learning, leverages unlabeled samples to augment the learning process (Van

Engelen and Hoos, 2020; Li et al., 2023). Most existing semi-supervised learning works (Tarvainen and Valpola, 2017; Berthelot et al., 2019; Xie et al., 2020; Sohn et al., 2020a; Xu et al., 2021) can be categorized into consistency regularization where the prediction is consistent with different perturbations, and self-training that involves an iterative update process.

The teacher-student framework is one of the mainstream ways for semi-supervised object detection (Sohn et al., 2020a; Xu et al., 2021; Liu et al., 2021b; Li et al., 2022; Chen et al., 2022a) using the self-training approach, which is illustrated in Figure 3. Initially, a “teacher” model is trained on the labeled samples using supervised learning. This trained “teacher” model is duplicated into a “student” model and employed to generate pseudo-labels for the unlabeled samples. Subsequently, a mixture of the most confidently selected

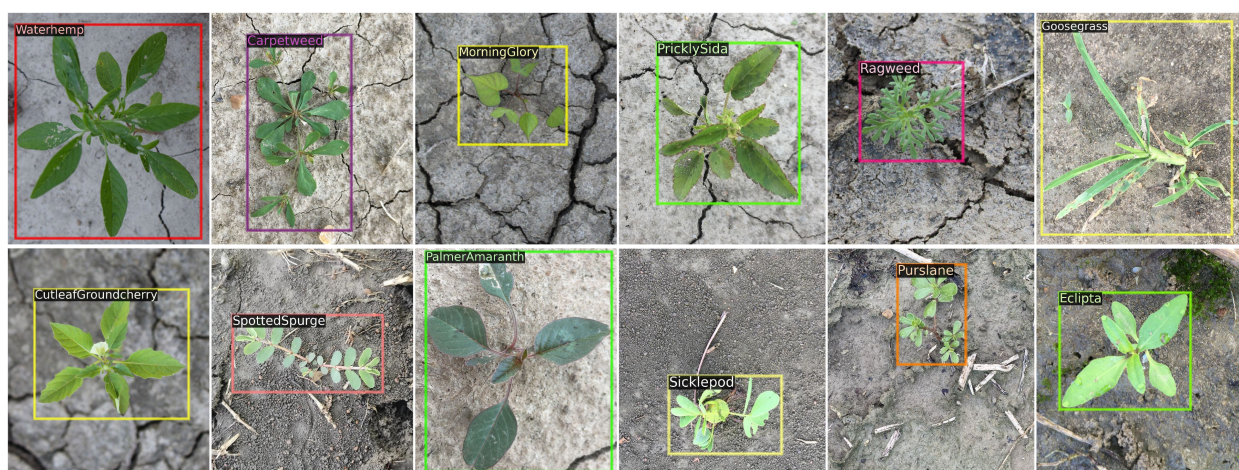
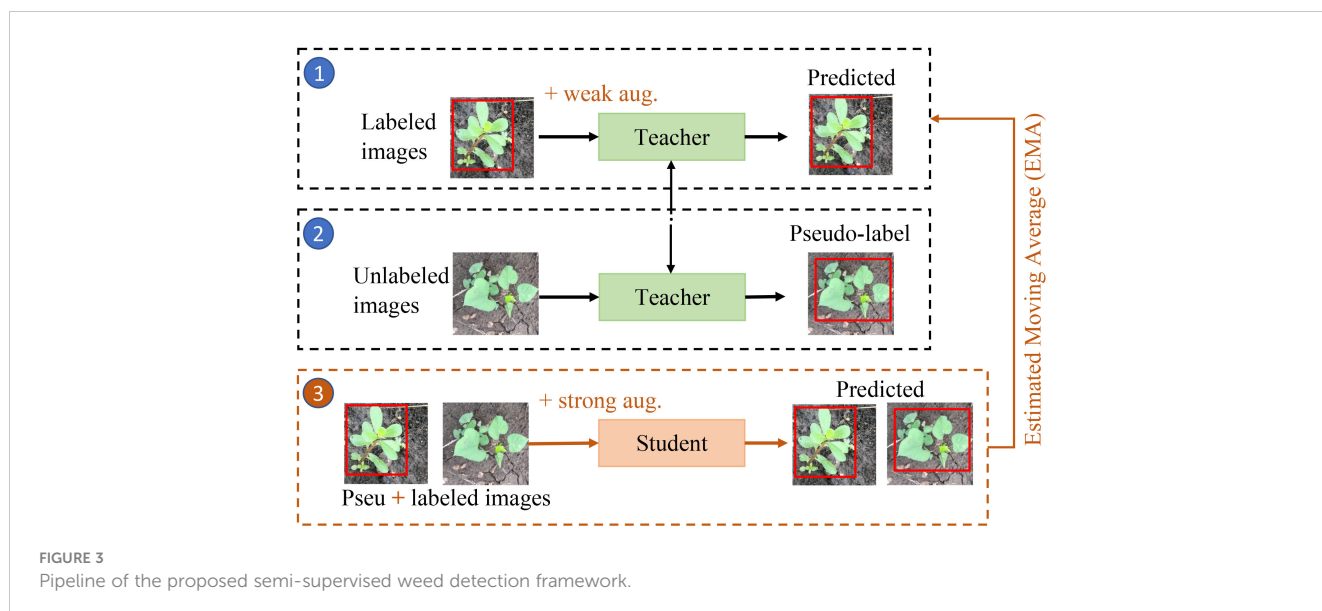


FIGURE 2 Weed samples in the CottonWeedDet12 dataset (Dang et al., 2023).



pseudo-labeled samples and the original labeled samples are utilized to train a “student” model. Subsequently, the “teacher” model is updated with the “student” model using an Estimated Moving Average (EMA) strategy (Tarvainen and Valpola, 2017) according to the Equation 1:

$$\theta_{\text{teacher}} = \alpha \cdot \theta_{\text{teacher}} + (1 - \alpha) \cdot \theta_{\text{student}} \quad (1)$$

where θ_{teacher} and θ_{student} represent the parameters of the “teacher” and “student” models, respectively. The factor α determines the extent of the update. An α of 1 retains the original “teacher” model parameters, while an α of 0 fully replaces the “teacher” model with the “student” model. In this study, we use cross-validations and find that $\alpha = 0.99$ is the optimal choice for the designed semi-supervised learning framework. The EMA strategy serves as a crucial mechanism to reduce variance (Tarvainen and Valpola, 2017). We apply weak augmentation approaches (e.g., horizontal flip, multi-scale training with a shorter size range [400, 1200], and scale jittering) to the student learning process and strong augmentation methods (e.g., randomly added gray scale, Gaussian blur, cutout patches (DeVries and Taylor, 2017)) to the teacher learning processes, respectively, to enhance the performance during training process (Xie et al., 2020; Xu et al., 2021). Figure 3 provides a visual representation of the described process.

This iterative process (steps 1-3) is repeated until the model achieves satisfactory performance. Upon completion of the model training, the “student” model is discarded, and only the “teacher” model is retained for inference. The versatility of self-training methods allows them to be integrated with any supervised learning-based approach, including one-stage and two-stage object detectors. In this study, we employ a self-training-based semi-supervised learning framework and assess two representative object detectors, Faster RCNN (Ren et al., 2015) and FCOS (Tian et al., 2022).

2.3.1 Pseudo-labeling on detectors

It is important to obtain the most confident and accurate pseudo-labels in semi-supervised learning. Published works (Sohn et al., 2020b; Zhou et al., 2021a; Liu et al., 2021b) exploit the pseudo-labeling method to address semi-supervised object detection, and the majority of them concentrated on anchor-based detectors. Our focus, however, lies in introducing the generalization approach for both anchor-free and anchor-based detectors, drawing inspiration from (Liu et al., 2021b, 2022).

We take the widely used FCOS model (Tian et al., 2022) as an example to demonstrate the semisupervised object detection tasks. FCOS comprises three prediction branches, classifier, centerness, and regressor, where the centerness score/branch dominates the bounding boxes score. However, the reliability of centerness scores in distinguishing foreground instances is questionable, particularly under conditions of limited label availability, as there is no supervision mechanism to suppress the centerness score for background instances within the centerness branch (Li et al., 2020; Liu et al., 2022). Consequently, although the centerness branch improves the anchor-free detector performance for the supervised training, it proves ineffective or even counterproductive for semi-supervised training scenarios (Li et al., 2020; Liu et al., 2022). To address this issue, our approach prioritizes pseudo-boxes based solely on classification scores (Liu et al., 2022). The classifier is trained with the hard labels (i.e., one-hot vector) with the box localization weighting. Finally, we use the standard label assignment method instead of center-sampling, which designates all elements within the bounding boxes as foreground and everything outside as background.

2.3.2 Unsupervised regression loss

Confidence thresholding has proven effective in prior studies (Tarvainen and Valpola, 2017; Sohn et al., 2020b; Liu et al., 2021b). However, depending solely on box confidence is insufficient for effectively eliminating misleading instances in box regression, since

the “teacher” may still provide a contradictory regression to the ground-truth direction (Chen et al., 2017; Saputra et al., 2019). To address this challenge, we categorize the pseudo-labels into two groups: beneficial instances and misleading instances. We then leverage the relative prediction information between the “student” and the “teacher” to identify beneficial instances and filter out misleading ones during the training of the regression branch. We define the unsupervised regression loss by selecting beneficial instances where the “teacher” exhibits lower localization uncertainty than the Student by a margin of σ , as shown in the Equation 2:

$$L_{reg}^{unsup} = \begin{cases} \sum_i \|\tilde{d}_t^i - \tilde{d}_s^i\|, & \text{if } \delta_t^i + \sigma \leq \delta_s^i \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The parameter $\sigma \geq 0$ represents a margin between the localization uncertainties of the “teacher” and the “student”, where the localization uncertainty is loosely associated with the deviation from the ground-truth labels. Specifically, δ_t^i represents the teacher’s localization uncertainty, while δ_s^i represents the student’s localization uncertainty. Furthermore, \tilde{d}_t^i and \tilde{d}_s^i are the regression predictions for “teacher” and “student”, respectively. For more details of the design for the unsupervised regression loss, please refer to Liu et al. (2022).

2.4 Performance evaluation metrics

In this evaluation, we rely on Average Precision (AP) as a primary metric, a measure derived from precision (P) and recall (R). AP summarizes the P(R) Curve to one scalar value. However, since AP is traditionally evaluated for each object category separately, we employ the mean Average Precision (mAP) metric (Liu et al., 2020) to provide a comprehensive assessment across all object categories. The mAP is calculated as the average of AP scores over all object categories, and both AP and mAP are determined using the following Equations 3, 4:

$$AP = \int_0^1 P(R)dR, \quad (3)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i, \quad (4)$$

where n represents the number of weed classes, and mAP signifies the average AP across these classes. A higher area under the Precision-Recall (PR) curve indicates improved object detection accuracy. Moreover, we consider mAP@[0.5:0.95], reflecting the mean average precision across IoU thresholds ranging from 0.5 to 0.95. These metrics collectively offer a representative evaluation of the model’s performance across varying detection thresholds, ensuring a comprehensive understanding of its object detection capabilities.

2.5 Experimental setups

In the process of model development and evaluation, the cotton weed dataset was partitioned into three subsets randomly.

Specifically, for a comprehensive evaluation, the CottonWeedDet3 dataset was randomly partitioned into training, validation, and testing sets following a ratio of 65%, 20%, and 15%, resulting in subsets comprising 550, 170, and 128 images. Similarly, the CottonWeedDet12 dataset was also divided into training, validation, and testing subsets, with a distribution ratio of 65%, 20%, and 15%, respectively. This results in subsets comprising 3670, 1130, and 848 images. The validation set is used to select the optimal trained model, while the test set is utilized to evaluate the model’s performance.

To expedite the model training process, we leveraged transfer learning (Zhuang et al., 2020) for all object detectors backbone, fine-tuning them with pre-trained weights obtained from the ImageNet dataset (Deng et al., 2009). The model was implemented based on Detectron2 (Wu et al., 2019). All models underwent training for 80k iterations, a duration deemed sufficient for effective modeling of the weed data. Stochastic Gradient Descent (SGD) was adopted as the optimizer, maintaining a momentum of 0.9 throughout the training process. The learning rate was selected as 0.01, and each batch contains 4 labeled images and 4 unlabeled images. We adopted the weak augmentation (horizontal flip, multi-scale training with a shorter size range [400, 1200] and scale jittering) for the “student”, and randomly add gray scale, Gasussian blur, cutout patches (DeVries and Taylor, 2017), and color jittering as the strong augmentation for the “teacher”. The computational setup included a server running Ubuntu 20.04, equipped with two Geforce RTX 2080Ti GPUs, each with 12 GB of memory, ensuring efficient model training and testing.

3 Results

In this section, we first evaluate the performance of various object detectors within the context of a semisupervised learning framework. Subsequently, we will delve into a detailed analysis of the performance exhibited by individual weed classes.

3.1 Semi-supervised object detector comparison

Figure 4 illustrates the training curves for FCOS and Faster RCNN, utilizing various proportions of labeled samples on the two cotton weed datasets: CottonWeedDet3 and CottonWeedDet12. We evaluated each algorithm in both supervised and semi-supervised learning contexts. For example, the configuration represented as Faster RCNN-sup-5% refers to the Faster RCNN trained with supervised learning using 5% of labeled samples. Conversely, Faster RCNN-semi-5% is the same detector trained with semi-supervised learning using 5% of the labeled samples and 95% of the unlabeled samples.

It is evident from the results that semi-supervised learning outperforms its supervised counterparts on both datasets, given the exploitation of a large volume of unlabeled samples to bolster the training process. As an example, Faster RCNN-semi-5% achieves superior training performance compared to Faster RCNNsup-5%.

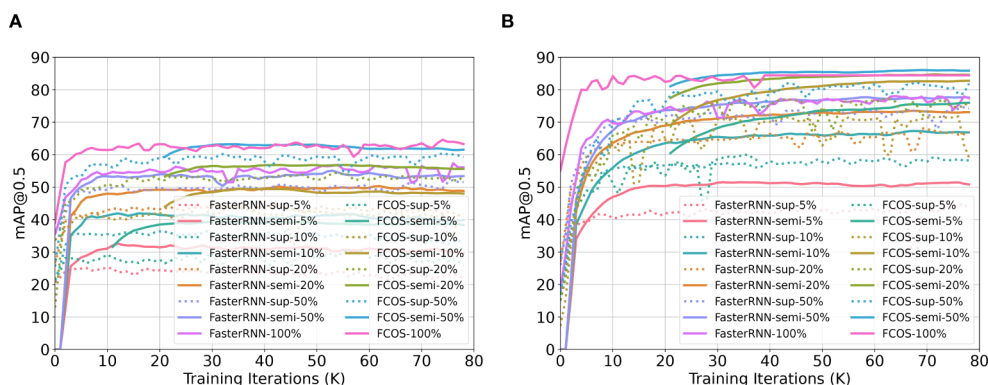


FIGURE 4 Training curves for FCOS and Faster RCNN with different proportions of labeled samples for two cotton weed datasets: CottonWeedDet3 and CottonWeedDet12. (A) Training curves for CottonWeedDet3 dataset (B) Training curves for CottonWeedDet12 dataset.

Moreover, it is noteworthy that FCOS-semi-50% manages to attain performance comparable to that of FCOS-100% (where all samples are labeled) on the CottonWeedDet3 dataset. FCOS-semi-50% even surpasses FCOS-100% on the CottonWeedDet12 dataset, suggesting that with only half the labeling effort, we can achieve improved performance, which also showcases that semi-supervised learning can be more robust compared with the supervised learning (Liu et al., 2021a). Furthermore, CottonWeedDet12 shows significant performance superiority over CottonWeedDet3, largely due to the latter’s smaller image dataset and the greater complexity of scenes within each image.

Tables 1, 2 summarize the test performance (measured by mAP@[0.5:0.95]) comparison between the supervised and semi-supervised learning approaches based on the Faster-RCNN and FCOS models on the CottonWeedDet3 and CottonWeedDet12 datasets, respectively. Across both datasets, FCOS consistently outperforms Faster-RCNN in both the semi-supervised and supervised learning contexts. These findings are in agreement with the observations drawn from the training curves illustrated in Figure 4. For any given proportion of labeled samples, the semi-supervised learning approaches are found to enhance the test performance. For instance, on the CottonWeedDet3 dataset, the Faster RCNN model using a semisupervised learning approach attains 86.70% and 93.73% of the performance of its supervised approach with only 20% and 50% of the samples labeled, respectively. Furthermore, it is worth highlighting that on the CottonWeedDet12 dataset, the FCOS model trained using semi-

supervised learning with only 50% of labeled samples outperforms the test performance of the fully supervised approach, which uses 100% of the samples manually labeled. That is because semi-supervised learning can effectively leverage the vast amount of unlabeled samples, which may capture the inherent distribution of the data better than a limited set of labeled samples.

Figures 5, 6 show selected images predicted using both supervised and semi-supervised FCOS for CottonWeedDet3 and CottonWeedDet12, respectively. In both figures, only 5% and 10% of labeled samples are utilized for training. Remarkably, the semi-supervised FOCS exhibits visually compelling predictions, especially for images featuring diverse and/or cluttered backgrounds, as well as those with densely populated weed instances. Notably, the semi-supervised learning approach demonstrates superior performance compared to the supervised learning approach. For instance, in Figure 5, the semi-supervised FOCS with 5% labeled samples produces better predictions than the supervised learning approach with only 5% labeled samples. This underscores the ability of semi-supervised learning to leverage valuable information from a large volume of unlabeled data.

3.2 Class-specific performance

Tables 3, 4 present the class-specific performance of the FCOS model on the CottonWeedDet3 and CottonWeedDet12 datasets, respectively. The instance count reflects the number of

TABLE 1 Testing performance (mAP@[0.5:0.95]) comparison between the supervised and semi-supervised based on Faster-RCNN and FCOS models on the CottonWeedDet3 dataset.

Algorithms	Supervision type	Proportion of labeled data for training				
		5%	10%	20%	50%	100%
Faster-RCNN	Supervised	21.14	31.90	42.65	50.51	56.75
	Semi-supervised	29.33	40.14	49.20	53.19	
FCOS	Supervised	27.37	42.20	52.42	59.84	62.80
	Semi-supervised	38.17	47.93	55.79	61.32	

TABLE 2 Testing performance (mAP@[0.5:0.95]) comparison between the supervised and semi-supervised based on Faster RCNN and FCOS models and CottonWeedDet12 dataset.

Algorithms	Supervision type	Proportion of labeled data for training				
		5%	10%	20%	50%	100%
Faster-RCNN	Supervised	45.02	61.18	68.29	75.97	80.47
	Semi-supervised	53.08	70.21	75.15	78.83	
FCOS	Supervised	62.28	72.99	79.14	83.87	86.47
	Semi-supervised	76.91	83.43	85.28	87.26	

bounding boxes associated with each weed category within the test images. It is evident that the CottonWeedDet12 dataset exhibits a considerable imbalance, as indicated by the significantly uneven distribution of instances across various weed classes.

On the CottonWeedDet3 dataset, the semi-supervised learning approaches demonstrate promising performance. Notably, the semi-supervised model trained with 50% of the labeled samples surpasses the performance of the fully supervised learning model,

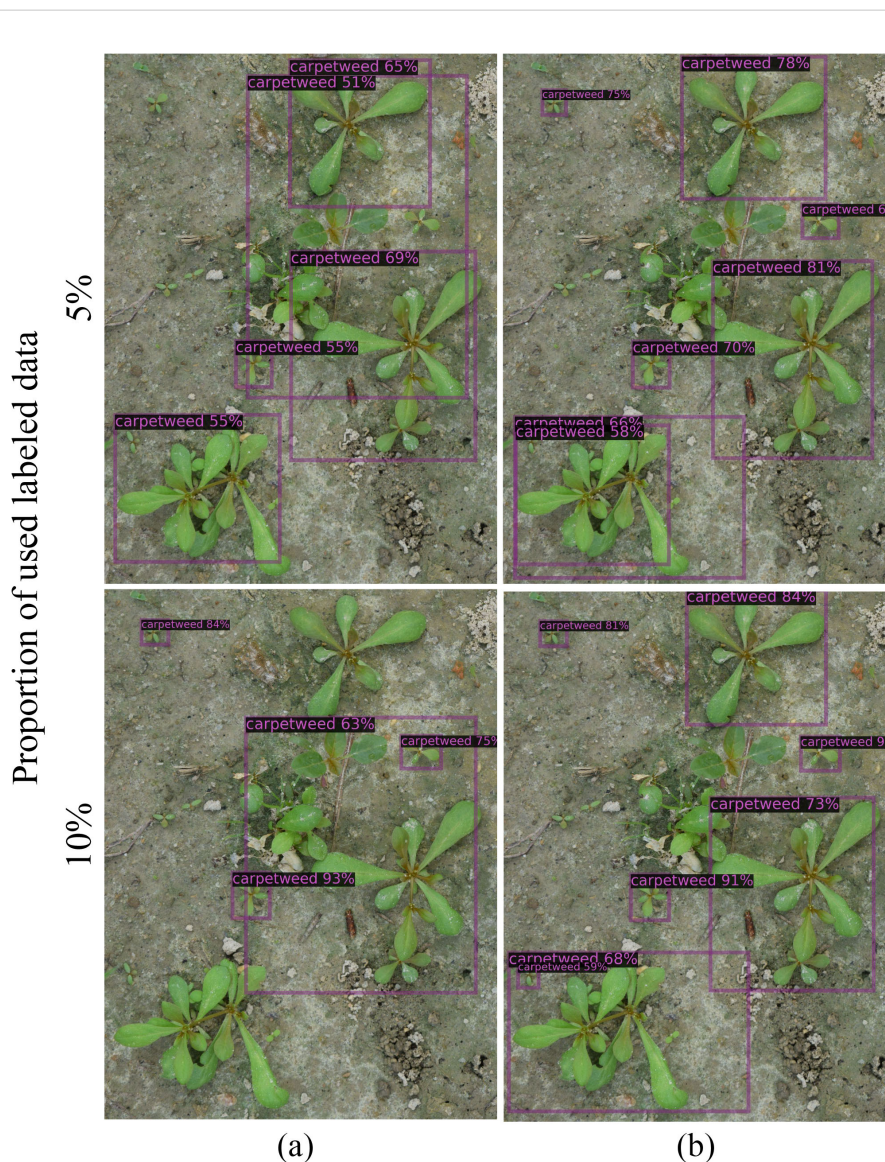


FIGURE 5 Examples of images annotated with ground truth labels (A) and predicted labels (B) using semi-supervised FCOS for CottonWeedDet3.

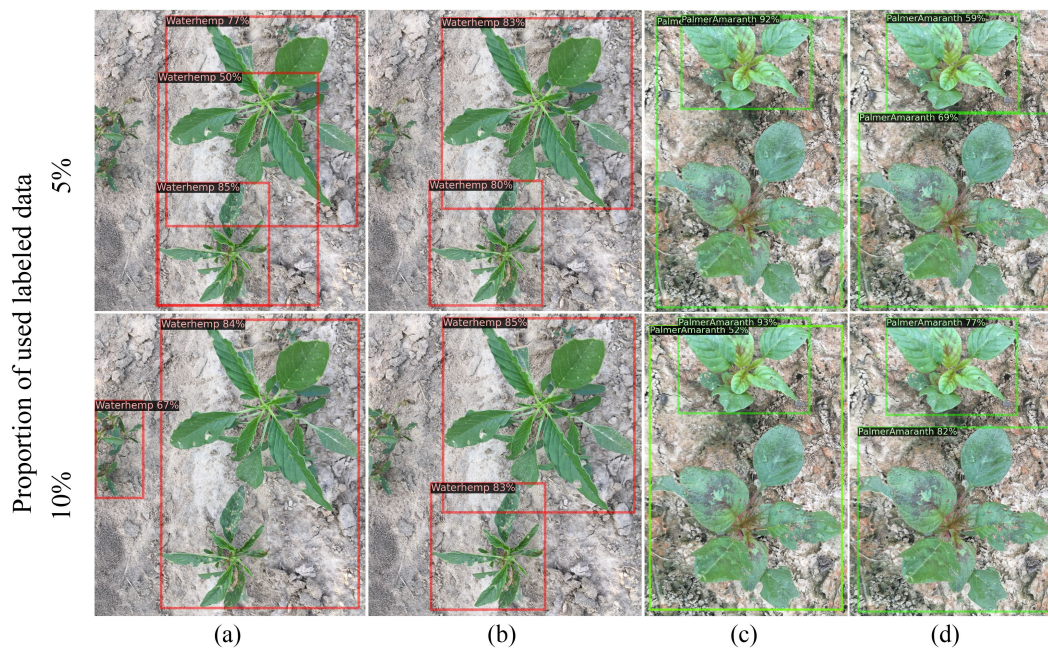


FIGURE 6 Comparing method results on CottonWeedDet12: (A, C) - supervised baseline, (B, D) semi-supervised FCOS.

TABLE 3 Test performance (mAP@[0.5:0.95]) on a specific category of weeds on CottonWeedDet3.

Weeds	# of instances	Proportion of labeled data for training				
		5%	10%	20%	50%	100%
PalmerAmaranth	100	48.33	52.18	60.82	63.25	62.70
MorningGlory	101	46.82	55.47	63.56	65.97	70.83
Carpetweed	93	19.38	36.14	42.99	54.73	54.87

TABLE 4 Test performance (mAP@[0.5:0.95]) on the specific category of weeds on CottonWeedDet12.

Weeds	# of instances	Proportion of labeled data for training				
		5%	10%	20%	50%	100%
Waterhemp	352	85.25	86	88.52	89.78	88.10
MorningGlory	201	83.82	85.75	87.40	89.71	88.03
Purslane	161	74.89	78.58	80.10	81.40	83.68
SpottedSpurge	122	77.12	81.49	83.59	85.79	82.78
Carpetweed	137	63.04	69.62	68.64	71.34	68.18
Ragweed	144	78.08	78.26	81.88	83.11	81.83
Eclipta	117	90.28	90.69	91.34	93.58	95.19
PricklySida	60	78.26	82.91	83.75	83.48	84.01
PalmerAmaranth	42	86.76	89.09	87.82	91.31	93.55
Sicklepod	31	94.27	96.56	97.09	97.01	96.43
Goosegrass	31	78.83	81.69	85.58	90.02	85.31
CutleafGroundcherry	15	32.33	80.59	87.70	90.59	90.50

particularly for palmer amaranth weeds. However, the detection accuracy for carpetweed remains relatively low, attributed to its small size which poses an inherent challenge for recognition. A similar trend is observed in the performance metrics presented in Table 4 for the CottonWeedDet12 dataset.

Remarkably, on the CottonWeedDet12 dataset, the semi-supervised FCOS model trained with 50% and 20% of labeled samples outperforms the fully supervised model for 8 out of 12 and 6 out of 12 weed classes, respectively. Impressively, for the top 3 minority weed classes — cutleaf groundcherry, goosegrass, and sicklepod — the FCOS model delivers superior performance even with only 50% of the labeling costs compared to the supervised learning approach. This underscores the potential of semi-supervised learning models to effectively address class imbalance and provide superior performance even with fewer labeled samples.

3.3 Comparative analysis: semi-supervised learning vs. ground truth inaccuracies

In the preceding discussions, we demonstrate the remarkable performance improvement achieved by semi-supervised learning, even with a limited number of labeled samples, surpassing the results of traditional supervised learning approaches. In Figure 7, we present image samples from CottonWeedDet12, showcasing both ground truth annotations and the predicted results obtained through the semi-supervised FCOS-10%. Notably, a discernible observation is the presence of inaccuracies and mislabels in the ground truth annotations, highlighting the challenges associated with manual labeling by human experts, including instances of noise and incorrect labels. The application of a semi-supervised learning approach demonstrates to be a potent solution in

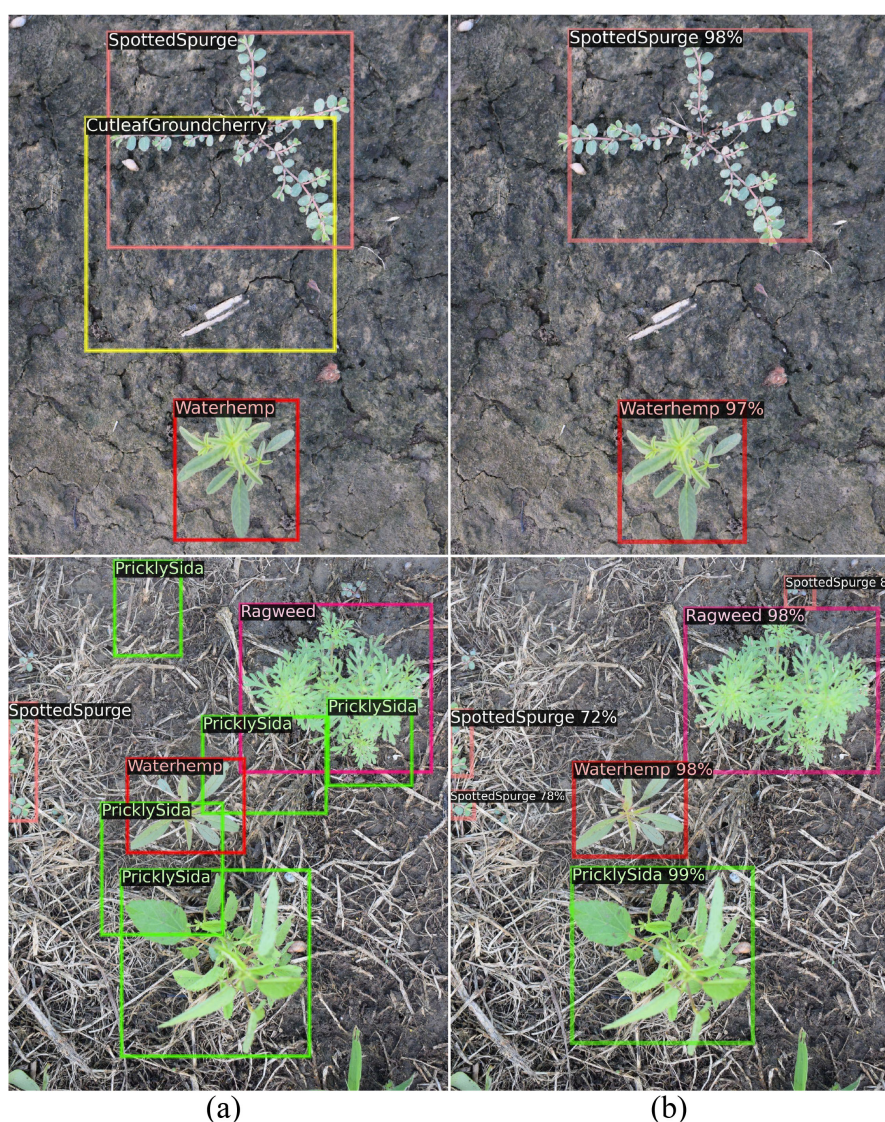


FIGURE 7 Image samples from CottonWeedDet12 with ground truth annotations (A) and predicted results with semi-supervised FCOS-10% (B).

mitigating the above challenges, and effectively enhancing accuracy and rectifying ground truth inaccuracies.

4 Discussions

4.1 Key contributions

The field of multi-class weed detection and localization remains largely unexplored in the existing literature (Dang et al., 2023; Rai et al., 2023). In the transition to the next-generation machine vision-based weeding systems, the focus is progressively shifting towards attaining higher precision and instituting weed-specific controls. Concurrently, the capability to differentiate between various weed species and identify individual weed instances emerges as an increasingly critical requirement within these vision tasks. While significant progress has been made in the development of DL-based weed detection (dos Santos Ferreira et al., 2017; Wang et al., 2019; Wu et al., 2021; Dang et al., 2022, 2023), these approaches typically rely heavily on expansive and manually-labeled image datasets, which makes these processes costly, prone to human error, and laboriously time-consuming. In our previous review on label-efficient learning in agriculture (Li et al., 2023), we presented various techniques aiming at reducing labeling costs and their respective applications in agricultural applications, including crop and weed management. Nevertheless, label-efficient technologies remain largely unexplored in the field of multi-class weed detection and localization. In this regard, this study stands as a unique contribution to the research community, specifically in the area of weed detection and control. By implementing semi-supervised learning, we introduce an innovative approach to alleviate the burden of labor-intensive labeling costs. Our evaluation includes both one-stage and two-stage object detectors on two open-source weed datasets, demonstrating that semi-supervised learning can significantly reduce labeling costs without substantially compromising performance. Additionally, it can even generate enhanced performance metrics.

The results of this study have positive implications for the use of phytosanitary products and precision agriculture. By improving the efficiency and accuracy of weed detection and localization, our approach can contribute to more targeted and effective use of phytosanitary products, thereby enhancing overall agricultural productivity and sustainability.

4.2 Limitations

While this research provides valuable insights, it does acknowledge certain limitations that pave the way for potential future enhancements. Although the primary objective of this research is not to evaluate all DL-based object detectors for weed detection within the semi-supervised learning framework, there are indeed several high-performing object detectors that are not

evaluated in this study. These include one-stage detectors such as SSD (Liu et al., 2016), RetinaNet (Lin et al., 2017), EfficientDet (Tan et al., 2020) and YOLO series (Dang et al., 2023; Terven and Cordova-Esparza, 2023), as well as two-stage detectors like DINO (Zhang et al., 2022), CenterNetv2 (Zhou et al., 2021b), RTMDet (Lyu et al., 2022), and etc. We intend to test and incorporate these models into our continually updated benchmark as we refine and improve the semi-supervised learning framework through future efforts.

In the scope of this study, we work under the assumption that all unlabeled samples are drawn from the same distribution as the labeled samples. It is important to acknowledge that unlabeled data might include instances from unknown or unseen classes, presenting a challenge commonly known as the open-set challenge (Chen et al., 2020). This scenario may substantially compromise the efficacy of label-efficient learning. Consequently, we highlight a future investigation to delve into addressing out-of-distribution (OOD) issues, employing advanced sample-specific selection strategies. The aim is to identify and subsequently downplay the significance or utilization of OOD samples (Guo et al., 2020). This planned exploration intends to enhance the generalization and robustness of our approach, ensuring its effectiveness in scenarios where the dataset contains samples from classes not encountered during the training phase, thereby contributing to a more resilient and versatile semi-supervised learning framework.

5 Conclusion

In this study, we conducted an extensive evaluation of semi-supervised learning in the context of multi-class weed detection. Leveraging a set of labeled data alongside the unlabeled data for model training, our investigation focused on evaluating the efficacy of both one-stage and two-stage object detectors. The two datasets, CottonWeedDet3 and CottonWeedDet12, chosen for our study were meticulously curated to align with U.S. cotton production systems, ensuring the relevance of our findings to real-world agricultural scenarios. By leveraging semi-supervised learning, the labeling costs were significantly reduced, while only minimal impacts on the detection performance were observed. Additionally, by using the abundant unlabeled samples, the semi-supervised learning approach produced a more robust and accurate model, and it demonstrated the capability of mitigating noise and incorrect labels in the ground-truth annotations. The outcomes underscore the potential of semi-supervised learning as a cost-effective and efficient alternative approach for developing agricultural applications, particularly those requiring extensive data annotations.

In our future work, we will refine and improve the semi-supervised learning framework for weed detection by testing and incorporating more high-performing object detectors into our continually updated benchmark. In addition, we will address the open-set challenge, where unlabeled data may include instances

from unknown or unseen classes, potentially compromising the efficacy of label-efficient learning. Future investigations will delve into addressing out-of-distribution (OOD) issues by employing advanced sample-specific selection strategies.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

JL: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. DC: Writing – original draft, Formal analysis, Conceptualization. XY: Writing – review & editing, Investigation. ZL: Writing – review & editing, Supervision, Resources.

References

- Ahmad, J., Muhammad, K., Ahmad, I., Ahmad, W., Smith, M. L., Smith, L. N., et al. (2018). Visual features based boosted classification of weeds for real-time selective herbicide sprayer systems. *Comput. Industry* 98, 23–33. doi: 10.1016/j.compind.2018.02.005
- Bawden, O., Kulk, J., Russell, R., McCool, C., English, A., Dayoub, F., et al. (2017). Robot for weed species plant-specific management. *J. Field Robotics* 34, 1179–1199. doi: 10.1002/rob.21727
- Benchallal, F., Hafiane, A., Ragot, N., and Canals, R. (2024). Convnext based semi-supervised approach with consistency regularization for weeds classification. *Expert Syst. Appl.* 239, 122222. doi: 10.1016/j.eswa.2023.122222
- Ben Hassen, T., and El Bilali, H. (2022). Impacts of the Russia-Ukraine war on global food security: towards more sustainable and resilient food systems? *Foods* 11, 2301. doi: 10.3390/foods11152301
- Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., and Raffel, C. A. (2019). Mixmatch: A holistic approach to semi-supervised learning. *Adv. Neural Inf. Process. Syst.* 32. doi: 10.5555/3454287.3454741
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *ArXiv paper*. arXiv preprint arXiv:2004.10934. doi: 10.48550/arXiv.2004.10934
- Cai, Z., Fan, Q., Feris, R. S., and Vasconcelos, N. (2016). “A unified multi-scale deep convolutional neural network for fast object detection,” in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. (Springer), 354–370.
- Caron, M., Bojanowski, P., Joulin, A., and Douze, M. (2018). “Deep clustering for unsupervised learning of visual features,” in *Proceedings of the European conference on computer vision (ECCV)*. (Munich, Germany: Springer Link), 132–149.
- Chen, B., Chen, W., Yang, S., Xuan, Y., Song, J., Xie, D., et al. (2022a). “Label matching semisupervised object detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (New Orleans, Louisiana: IEEE), 14381–14390. doi: 10.1109/CVPR52688.2022.01398
- Chen, D., Lu, Y., Li, Z., and Young, S. (2022b). Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Comput. Electron. Agric.* 198, 107091. doi: 10.1016/j.compag.2022.107091
- Chen, G., Choi, W., Yu, X., Han, T., and Chandraker, M. (2017). Learning efficient object detection models with knowledge distillation. *Adv. Neural Inf. Process. Syst.* 30. doi: 10.5555/3294771.3294842
- Chen, Y., Zhu, X., Li, W., and Gong, S. (2020). “Semi-supervised learning under class distribution mismatch,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, USA, Vol. 34(04), 3569–3576.
- Coleman, G. R., Bender, A., Walsh, M. J., and Neve, P. (2023). Image-based weed recognition and control: Can it select for crop mimicry? *Weed Res.* 63, 77–82. doi: 10.1111/wre.12566
- Dang, F., Chen, D., Lu, Y., and Li, Z. (2023). Yoloweeds: A novel benchmark of yolo object detectors for multi-class weed detection in cotton production systems. *Comput. Electron. Agric.* 205, 107655. doi: 10.1016/j.compag.2023.107655
- Dang, F., Chen, D., Lu, Y., Li, Z., and Zheng, Y. (2022). “Deepcottonweeds (dcw): a novel benchmark of yolo object detectors for weed detection in cotton production systems,” in *2022 ASABE Annual International Meeting*. (Houston, Texas: American Society of Agricultural and Biological Engineers), 1.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. (Miami, Florida: IEEE), 248–255.
- DeVries, T., and Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*. doi: 10.48550/arXiv.1708.04552
- dos Santos Ferreira, A., Freitas, D. M., da Silva, G. G., Pistori, H., and Folhes, M. T. (2017). Weed detection in soybean crops using convnets. *Comput. Electron. Agric.* 143, 314–324. doi: 10.1016/j.compag.2017.10.027
- dos Santos Ferreira, A., Freitas, D. M., da Silva, G. G., Pistori, H., and Folhes, M. T. (2019). Unsupervised deep learning and semi-automatic data labeling in weed discrimination. *Comput. Electron. Agric.* 165, 104963. doi: 10.1016/j.compag.2019.104963
- Farooq, A., Jia, X., Hu, J., and Zhou, J. (2019). “Knowledge transfer via convolution neural networks for multi-resolution lawn weed classification,” in *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. (Amsterdam, Netherlands: IEEE), 01–05.
- Gerhards, R., and Christensen, S. (2003). Real-time weed detection, decision making and patch spraying in maize, sugarbeet, winter wheat and winter barley. *Weed Res.* 43, 385–392. doi: 10.1046/j.1365-3180.2003.00349.x
- Girshick, R. (2015). “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*. (Santiago, Chile: IEEE), 1440–1448.
- Guo, L.-Z., Zhang, Z.-Y., Jiang, Y., Li, Y.-F., and Zhou, Z.-H. (2020). “Safe deep semi-supervised learning for unseen-class unlabeled data,” in *International Conference on Machine Learning*. (Vienna, Austria: PMLR), 3897–3906.
- Hu, C., Thomasson, J. A., and Bagavathiannan, M. V. (2021). A powerful image synthesis and semisupervised learning pipeline for site-specific weed detection. *Comput. Electron. Agric.* 190, 106423. doi: 10.1016/j.compag.2021.106423
- Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., et al. (2019). A survey of deep learning-based object detection. *IEEE Access* 7, 128837–128868. doi: 10.1109/ACCESS.2019.2939201
- Laborde, D., Martin, W., Swinnen, J., and Vos, R. (2020). Covid-19 risks to global food security. *Science* 369, 500–502. doi: 10.1126/science.abc4765
- Law, H., and Deng, J. (2018). “Cornernet: Detecting objects as paired keypoints,” in *Proceedings of the European conference on computer vision (ECCV)*, Munich, Germany, 734–750.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, H., Wu, Z., Shrivastava, A., and Davis, L. S. (2022). Rethinking pseudo labels for semi-supervised object detection. *Proc. AAAI Conf. Artif. Intell.* 36, 1314–1322. doi: 10.1609/aaai.v36i2.20019
- Li, J., Chen, D., Qi, X., Li, Z., Huang, Y., Morris, D., et al. (2023). Label-efficient learning in agriculture: A comprehensive review. *arXiv preprint arXiv:2305.14691*. 215, 108412. doi: 10.1016/j.compag.2023.108412
- Li, X., Wang, W., Hu, X., Li, J., Tang, J., and Yang, J. (2021). “Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Nashville, Tennessee: IEEE), 11632–11641.
- Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., et al. (2020). Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* 33, 21002–21012. doi: 10.5555/3495724.3497487
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2017). “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*. (Venice, Italy: IEEE), 2980–2988.
- Liu, H., HaoChen, J. Z., Gaidon, A., and Ma, T. (2021a). Self-supervised learning is more robust to dataset imbalance. *arXiv preprint arXiv:2110.05025*. doi: 10.48550/arXiv.2110.05025
- Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., et al. (2020). Deep learning for generic object detection: A survey. *Int. J. Comput. Vision* 128, 261–318. doi: 10.1007/s11263-019-01247-4
- Liu, T., Jin, X., Zhang, L., Wang, J., Chen, Y., Hu, C., et al. (2023). Semi-supervised learning and attention mechanism for weed detection in wheat. *Crop Prot.* 174, 106389. doi: 10.1016/j.cropro.2023.106389
- Liu, T., Zhai, D., He, F., and Yu, J. (2024). Semi-supervised learning methods for weed detection in turf. *Pest Manage. Sci.* doi: 10.1002/ps.7959
- Liu, W., Angelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). “Ssd: Single shot multibox detector,” in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. (Springer), 21–37.
- Liu, Y.-C., Ma, C.-Y., He, Z., Kuo, C.-W., Chen, K., Zhang, P., et al. (2021b). Unbiased teacher for semi-supervised object detection. *arXiv preprint arXiv:2102.09480*.
- Liu, Y.-C., Ma, C.-Y., and Kira, Z. (2022). “Unbiased teacher v2: Semi-supervised object detection for anchor-free and anchor-based detectors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (New Orleans, Louisiana: IEEE), 9819–9828.
- Lu, Y., and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Comput. Electron. Agric.* 178, 105760. doi: 10.1016/j.compag.2020.105760
- Lyu, C., Zhang, W., Huang, H., Zhou, Y., Wang, Y., Liu, Y., et al. (2022). Rtmddet: An empirical study of designing real-time object detectors. *arXiv preprint arXiv:2212.07784*. doi: 10.48550/arXiv.2212.07784
- Manalil, S., Coast, O., Werth, J., and Chauhan, B. S. (2017). Weed management in cotton (*Gossypium hirsutum* L.) through weed-crop competition: A review. *Crop Prot.* 95, 53–59. doi: 10.1016/j.cropro.2016.08.008
- Meyer, G. E., and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* 63, 282–293. doi: 10.1016/j.compag.2008.03.009
- Nong, C., Fan, X., and Wang, J. (2022). Semi-supervised learning for weed and crop segmentation using uav imagery. *Front. Plant Sci.* 13, 927368. doi: 10.3389/fpls.2022.927368
- Norsworthy, J. K., Ward, S. M., Shaw, D. R., Llewellyn, R. S., Nichols, R. L., Webster, T. M., et al. (2012). Reducing the risks of herbicide resistance: best management practices and recommendations. *Weed Sci.* 60, 31–62. doi: 10.1614/WS-D-11-00155.1
- O’Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., et al. (2020). “Deep learning vs. traditional computer vision,” in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC)*, Vol. 11 (Las Vegas, USA: Springer), 128–144.
- Oerke, E.-C. (2006). Crop losses to pests. *J. Agric. Sci.* 144, 31–43. doi: 10.1017/S0021859605005708
- Parra, L., Marin, J., Yousfi, S., Rincón, G., Mauri, P. V., and Lloret, J. (2020). Edge detection for weed recognition in lawns. *Comput. Electron. Agric.* 176, 105684. doi: 10.1016/j.compag.2020.105684
- Rahman, A., Lu, Y., and Wang, H. (2023). Performance evaluation of deep learning object detectors for weed detection for cotton. *Smart Agric. Technol.* 3, 100126. doi: 10.1016/j.atech.2022.100126
- Rai, N., Zhang, Y., Ram, B. G., Schumacher, L., Yellavajjala, R. K., Bajwa, S., et al. (2023). Applications of deep learning in precision weed management: A review. *Comput. Electron. Agric.* 206, 107698. doi: 10.1016/j.compag.2023.107698
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, Nevada: IEEE), 779–788.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28. doi: 10.5555/2969239.2969250
- Saputra, M. R. U., De Gusmao, P. P., Almalioglu, Y., Markham, A., and Trigoni, N. (2019). “Distilling knowledge from a deep pose regressor network,” in *Proceedings of the IEEE/CVF international conference on computer vision*. (Seoul, Korea: IEEE), 263–272.
- Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., et al. (2020a). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Adv. Neural Inf. Process. Syst.* 33, 596–608. doi: 10.5555/3495724.3495775
- Sohn, K., Zhang, Z., Li, C.-L., Zhang, H., Lee, C.-Y., and Pfister, T. (2020b). A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*.
- Sportelli, M., Apolo-Apolo, O. E., Fontanelli, M., Frascioni, C., Raffaelli, M., Peruzzi, A., et al. (2023). Evaluation of yolo object detectors for weed detection in different turfgrass scenarios. *Appl. Sci.* 13, 8502. doi: 10.3390/app13148502
- Tan, M., Pang, R., and Le, Q. V. (2020). “Efficientdet: Scalable and efficient object detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Seattle, Washington: IEEE), 10781–10790.
- Tarvainen, A., and Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* 30. doi: 10.5555/3294771.3294885
- Terven, J., and Cordova-Esparza, D. (2023). A comprehensive review of yolo: From yolov1 to yolov8 and beyond. *Mach. Learn. Knowl. Extr.* 5(4), 1680–1716. doi: 10.3390/make5040083
- Tian, Z., Chu, X., Wang, X., Wei, X., and Shen, C. (2022). “Fully convolutional one-stage 3d object detection on lidar range images. *Adv. Neural Inf. Process. Syst.* (New Orleans, LA, USA: Curran Associates Inc.) 35, 34899–34911.
- Van Engelen, J. E., and Hoos, H. H. (2020). A survey on semi-supervised learning. *Mach. Learn.* 109, 373–440. doi: 10.1007/s10994-019-05855-6
- Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* 158, 226–240. doi: 10.1016/j.compag.2019.02.005
- Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. Available online at: <https://github.com/facebookresearch/detectron2>.
- Wu, Z., Chen, Y., Zhao, B., Kang, X., and Ding, Y. (2021). Review of weed detection methods based on computer vision. *Sensors* 21, 3647. doi: 10.3390/s21113647
- Xie, Q., Luong, M.-T., Hovy, E., and Le, Q. V. (2020). “Self-training with noisy student improves imagenet classification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Seattle, Washington: IEEE), 10687–10698.
- Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., et al. (2021). “End-to-end semi-supervised object detection with soft teacher,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (Montreal, BC, Canada: IEEE), 3060–3069.
- Yang, J., Parikh, D., and Batra, D. (2016). “Joint unsupervised learning of deep representations and image clusters,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, Nevada: IEEE), 5147–5156.
- Young, S. L., Meyer, G. E., and Woldt, W. E. (2013). “Future directions for automated weed management in precision agriculture,” in *Automation: The future of weed control in cropping systems*. (Springer), 249–259.
- Yu, J., Schumann, A. W., Cao, Z., Sharpe, S. M., and Boyd, N. S. (2019). Weed detection in perennial ryegrass with deep learning convolutional neural network. *Front. Plant Sci.* 10, 1422. doi: 10.3389/fpls.2019.01422
- Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., et al. (2022). Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*. doi: 10.48550/arXiv.2203.03605
- Zhang, S., Chi, C., Yao, Y., Lei, Z., and Li, S. Z. (2020). “Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Seattle, Washington: IEEE), 9759–9768.
- Zhou, Q., Yu, C., Wang, Z., Qian, Q., and Li, H. (2021a). “Instant-teaching: An end-to-end semi-supervised object detection framework,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Nashville, Tennessee: IEEE), 4081–4090. doi: 10.1109/CVPR46437.2021.00407
- Zhou, X., Koltun, V., and Krähenbühl, P. (2021b). Probabilistic two-stage detection. *arXiv preprint arXiv:2103.07461*. doi: 10.1109/CVPR46437.2021.00407
- Zhou, X., Zhuo, J., and Krahenbuhl, P. (2019). “Bottom-up object detection by grouping extreme and center points,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Long Beach, CA: IEEE), 850–859.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proc. IEEE* 109, 43–76. doi: 10.1109/PROC.5