Crop Protection 179 (2024) 106626

Contents lists available at ScienceDirect

Crop Protection

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

Semi-supervised learning for detection of sedges in sod farms

Xin Chen [1], Teng Liu [1], Kang Han, Xiaojun Jin [*], Jialin Yu [**]

Peking University Institute of Advanced Agricultural Sciences/Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong, China

ARTICLE INFO

Keywords:
Computer vision
Deep learning
Machine learning
Sod farm
Weed detection

ABSTRACT

Deep learning-based methods for weed detection and precise herbicide application are promising for reducing herbicide input and weed control costs. However, training the neural network to recognize weeds requires annotating a large number of training images, which is time-consuming and labor-intensive. In addition, in turf sod farms, sods need to be periodically harvested, leading to varying turfgrass and bare soil areas, which increase the complexity of weed detection. To solve this problem, this research explored semi-supervised learning (SSL) methods to train image classification neural networks. The experiments were conducted to compare the training results using different SSL strategies with 100 and 200 labeled images at three image sizes of 240 × 240, 360 × 360, or 480 × 480 pixels. The training dataset images mainly contained purple nutsedge (*Cyperus rotundus* L.) and green kyllinga (*Kyllinga brevifolia*) at the pre-flowering or seedhead stage. The F_1 score, precision, and recall were used to evaluate the performance of the trained neural networks. The results showed that the network based on the FixMatch SSL strategy trained with the input images of 240 × 240 pixels exhibited the highest F_1 score, reaching 98.1% when trained with 100 labeled images and 98.2% when trained with 200 labeled images. To summarize, these results suggest that SSL achieved a great training performance with a small number of annotations. FixMatch SSL was the most effective neural network training strategy evaluated. For the weed detection task, it was observed that neural networks trained using an input image size of 240 × 240 pixels exhibited superior performance compared to the networks trained with other image sizes. In addition, employing the SSL method with only 200 labeled images enhanced the performance of the neural network, surpassing that of fully supervised learning (FSL) approaches.

1. Introduction

Turfgrass is a ubiquitous vegetation cover in urban landscapes, such as golf courses, home lawns, parks, sport fields, school playgrounds, and roadsides (Pincetl et al., 2019). Weed control is a constant issue in turf management, as weeds compete with turfgrass for resources, including sunlight, water, and nutrients, and reduce turf aesthetic and functionality (Busey, 2003). Cultural techniques, such as irrigation and mowing, help reduce weed infestation, but the most effective way to control weeds is broadcast-spraying synthetic herbicides (McElroy and Martins, 2013). However, herbicide application without a site-specific program often results in excessive herbicide usage. Many herbicides presently registered for weed control in turfgrass are considered problematic for the environment (USEPA, 2023a, 2023b). For instance, atrazine, a photosystem II-inhibiting herbicide currently used in warm-season turfgrass for preemergence and postemergence (POST) control of weeds, has been reported to be one of the most frequently detected pesticides in underground water in the United States (USEPA, 2023a).

Machine vision-based automated precision herbicide application has the potential to reduce herbicide usage and lower weed control expenditures (Gerhards et al., 2022; Jin et al., 2023a; Monteiro and Santos, 2022; Zhang et al., 2022). Previous researchers explored a variety of sensing methods for weed detection, such as multi-spectral imaging (Rosle et al., 2021; Wu et al., 2021), visible or near-infrared spectroscopy (Liang et al., 2018; Wu et al., 2008), and fluorescence (Su et al., 2019). Current mainstream detection methods involve the use of deep learning convolutional neural networks (DCNNs) to extract features from weeds (Hasan et al., 2021; Wang et al., 2019; Zhang et al., 2022).

* Corresponding author. Peking University Institute of Advanced Agricultural Sciences/Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong, China.
** Corresponding author. Peking University Institute of Advanced Agricultural Sciences/Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong, China.
E-mail addresses: xiaojun.jin@pku-iaas.edu.cn (X. Jin), jialin.yu@pku-iaas.edu.cn (J. Yu).
[1] The first and second authors contributed equally to this work.

https://doi.org/10.1016/j.cropro.2024.106626
Received 8 December 2023; Received in revised form 3 February 2024; Accepted 7 February 2024
Available online 11 February 2024

Given that a variety of weed species at varying densities, growth stages, mowing height, and surface conditions may present in a turfgrass site, training a neural network based on large-scale and high-quality image datasets is crucial for realizing effective weed detection (Olsen et al., 2019). However, obtaining and annotating a large amount of training images is time-consuming and labor-intensive, and thus is fairly challenging.

Recent improvements in Graphics Processing Unit (GPU) computing capability have greatly advanced neural network modeling technology based on deep learning (Alzubaidi et al., 2021). Numerous innovative concepts, such as alternative activation functions, regularization, parameter optimization, and architectural advances, have been investigated to improve the performances of DCNNs (Sun et al., 2022). The primary concept of utilizing DCNNs for weed detection is to extract various features through convolutional operations. This approach aims to identify weeds based on the plant's morphological characteristics, texture, and color of the leaves (Jin et al., 2021, 2022b; Le et al., 2020; Liu and Bruch, 2020; Sharpe et al., 2019; Zhuang et al., 2022). DCNNs have been successfully applied to detect weeds in real-time across various cropping systems, such as cool- and warm-season turfgrasses (Jin et al., 2022b; Yu et al., 2019a, 2019b), vegetables (Sharpe et al., 2020a, 2020b), soybean (*Glycine Max* L. *Merrill*) (dos Santos Ferreira et al., 2017), wheat (*Triticum aestivum* L.) (Liu et al., 2023), and small-fruiting crops (Sharpe et al., 2019, 2020a). Previous studies have demonstrated the excellent performance of DCNNs in detecting multiple weed species, particularly in actively growing or dormant bermudagrass (*Cynodon dactylon* (L.) Pers.) (Yu et al., 2019b), actively growing perennial ryegrass (*Lolium perenne* L.) (Yu et al., 2019a), as well as drought-stressed or non-stressed bahiagrass (*Paspalum notatum Flueggé*) (Zhuang et al., 2022).

Unfortunately, detecting weeds remains challenging in turf landscapes due to various weed species with varying visual characteristics and lacking high-quality and large-scale labeled annotated image datasets. In previous research, Yu et al. (2019b) reported that VGGNet achieved high F_1 scores ($\geq$0.9278), with high recall values ($\geq$0.9952) for detecting dandelion (*Taraxacum officinale* Web.), ground ivy (*Glechoma hederacea* L.), and spotted spurge (*Euphorbia maculata* L.) in perennial ryegrass turfgrass; however, the authors manually annotated a total of 15,486 negative (images without the target weeds) and 17,600 positive images (images contained the target weeds). In order to develop a neural network to detect broadleaf weed seedlings growing in wheat, Zhuang et al. (2022) utilized a large training dataset consisting of 11, 000 negative (without weeds) and 11,000 positive images (with weeds) to train image classification neural networks. Collecting and labeling such large datasets requires considerable resources, time, and effort, thus significantly hindering the development of effective neural network models.

An SSL strategy, leveraging both labeled and unlabeled data of weed images (Enguehard et al., 2019; Reddy et al., 2018), may be employed to address the issue of limited training datasets with labels. The methodology of SSL combines supervised and unsupervised learning by leveraging a limited dataset of labeled data and a large amount of unlabeled data to train the neural networks and make predictions for previously unseen instances (Ding et al., 2017). The basic process involves using the existing labeled data to label the remaining unlabeled data, thereby augmenting the training data (Berthelot et al., 2019; Laine and Aila, 2016; Tarvainen and Valpola, 2017). Turfgrass sod undergoes periodic harvesting, leading to fluctuations in turfgrass coverage and the presence of bare soil area, which may affect the DCNNs for detecting weeds. Exploring the SSL approach for detecting weeds in bermudagrass sod farms may improve weed detection performance with a limited amount of manually annotated data, thus considerably improving the efficiency of developing effectual neural network models.

Previous research documented that training image size affects the performance of DCNNs for weed detection. For example, Zhuang et al. (2022) reported that increasing training image sizes from 200 × 200 pixels to 400 × 400 pixels decreased the performances of AlexNet (Krizhevsky et al., 2017) and VGGNet (Simonyan and Zisserman, 2014) but generally improved DenseNet (Iandola et al., 2014) and ResNet (He et al., 2016) for detecting weeds growing in wheat. However, the impact of training image sizes on the detection of weeds growing in turfgrass sod farms has not been documented. Therefore, the objectives of this research were (1) to comprehensively evaluate the effectiveness of SSL strategies when applied to the weed detection task using DCNNs across multiple datasets characterized by varying image sizes, (2) to undertake a comparative assessment between SSL and Fully Supervised Learning (FSL) methodologies with particular image sizes, and (3) to investigate the influence of varying turf coverage and bare soil area in sod farms on the performance of SSL for weed detection.

2. Material and methods

2.1. Image acquisition

A total of 1900 training dataset images were captured from sod farms and city parks in JiangNiang District, Nanjing, Jiangsu, China (31°37′-32°07′N, 118°28′-119°06′W). These training images, featuring a low density of purple nutsedge and green kyllinga at pre-flowering growth stage. They were captured multiple times from April to May 2021. The testing dataset (TD) 1 images, featuring a low density of purple nutsedge at pre-flowering with visually estimated turf ground coverage exceeding >90% and bare soil ground coverage <10%, were captured from a sod farm in Suqian, Jiangsu, China (118°3′N, 33°96′W) in May 2019. The TD 2 images, featuring a low density of purple nutsedge and green kyllinga at the pre-flowering stage with visually estimated turf ground coverage ranging from 30% to 40% and the bare soil coverage ranging from 60% to 70%, were captured in Jurong, Jiangsu, China (119°77′N, 31°95′W) in May 2021. The TD 3 images, featuring a low density of purple nutsedge at the pre-flowering stage with visually estimated turfgrass coverage ranging from 70% to 80% and bare soil coverage ranging from 20% to 30%, were captured in Jurong, Jiangsu, China (119°77′N, 31°95′W) in May 2021. The TD 4 images were taken in July 2018 at the turfgrass research facility located at the University of Georgia Griffin Campus in Georgia, United States (33°78′N, −84°40′W). These images showcase a low density of annual sedge at the pre-flowering stage, with a clump growth habit. The TD 5 images were taken at a golf course rough in Tampa, Florida, United States (27°95′N, −82°46′W). These images consisted of a mixture of annual nutsedge, yellow nutsedge, and green kyllinga at the flowering or seedhead stage, growing alongside smooth crabgrass, doveweed, and/or various broadleaves such as dollar weeds, Florida pusley, and old-world diamond flower at high densities. The images also included newly cutted nutsedge leaf blades.

The Training Dataset, TD 1, TD 2, and TD 3 images were taken using a Panasonic® DMC-ZS110 (Xiamen, Fujian, China) at a resolution of 4300 × 2418 pixels. The TD 4 and TD 5 images were captured with a Sony® Cyber-Shot (SONY Corporation, Minato, Tokyo, Japan) at a resolution of 1920 × 1080 pixels. All training and TD images were captured between 9:00 a.m. and 5:00 p.m. at local time with varying outdoor lighting conditions, including sunny, cloudy, and partially cloudy weather. Detailed information regarding the training and TD images are presented in Table 1.

2.2. Establishing training, validation, and testing datasets

In order to train and test the neural network, the captured images presented in Table 1 were cropped into sub-images of 240 × 240, 360 × 360, or 480 × 480 pixels utilizing Irfanview (Version 5.5, Irfan Skijan, Jaice, Bosnia). For each image size, the cropped images were grouped into two classes: positive (with weeds) and negative images (without weeds). Using the cropped images of 360 × 360 pixels as an example, the cropped images displayed the training and testing images consisting of weeds growing in turfgrass sod farms with varying turfgrass coverage

Table 1
Neural network training and testing dataset specifications.

Dataset	Location	Image acquisition date	Turfgrass	Weed species
Training Dataset	Jiangning District, Nanjing, Jiangsu, China	Apr to May 2021	Sod farm, city park	Low density of purple nutsedge and green kyllinga at pre-flowering stage.
TD1	Suqian, Jiangsu, China	May 2019	Sod farm (>90% turf ground cover)	Low density of purple nutsedge at pre-flowering stage.
TD2	Jurong, Jiangsu, China	May 2021	Sod farm (30%–40% turf ground cover)	Low density of purple nutsedge and green kyllinga at pre-flowering stage.
TD3	Jurong, Jiangsu, China	May 2021	Sod farm (70–80% turf ground cover)	Low density of purple nutsedge at pre-flowering stage.
TD4	University of Georgia Griffin Campus, Georgia, United States	July 2018	Turfgrass research facility	Low density of annual sedge at pre-flowering stage with a clump growth habit.
TD5	Tampa, Florida, United States	Aug 2018	Golf course rough	Annual nutsedge, yellow nutsedge, and green kyllinga at flowering or seedhead stage growing with smooth crabgrass, doveweed, and/or various broadleaves at high densities. Cutted nutsedge leaves also exist.

Abbreviations: TD, testing dataset.

and bare soil area, as illustrated in Fig. 1.

For each image size, a total of 100 or 200 images per class were used for training the FSL and three SSL methods. As a result, the Resnet50 network trained under 24 different conditions was trained, as shown in Tables 2 and 3. For each condition, 100 or 200 labeled images per class and 10,000 unlabeled images were used to train the neural network. To ensure a balanced number of positive and negative samples in the training dataset, approximately half of the unlabeled 10,000 images were positive, while the other half were negative. For each image size, each validation and TD was constituted by randomly selecting a total of 500 positive and 500 negative images.

2.3. Training sets and environment

The training, validation, and testing were conducted on a workstation fitted with an Intel® Core™ i9-10920× CPU and an NVIDIA RTX 3080Ti GPU (12 GB GPU memory capacity). The operating system (OS) used was Ubuntu version 20.04.1. The training was performed using the Pytorch Deep Learning Framework based on Python 3.8 (version 1.8.1, Facebook in San Jose, CA, United States). The hyperparameters used for training were standardized to eliminate the influence that hyperparameters have on the final model's performances, as outlined below.

- Optimizer: AdamW [betas:0.9,0.999; Weight decay:0.0005]
- Base learning rate: 5e-5
- Learning rate policy: StepLR
- Batch size: 8
- Training epochs: 100

2.4. Experimental training strategy and neural network structure

Three SSL strategies, including Meanteacher (Tarvainen and Valpola, 2017), Pi-model (Laine and Aila, 2016), and FixMatch (Sohn et al., 2020), were selected and compared with the FSL method.

The Pi-Model was a simplified version of the Γ-Model used in Ladder Networks (Valpola, 2015). It eliminated the corrupted encoder and utilized the same neural network to generate predictions for corrupted and uncorrupted input data. This approach leveraged the inherent randomness of the prediction function $f\theta$ in neural networks, which was a consequence of regularization methods such as data augmentation and dropout. These techniques, as described by Kostopoulos et al. (2018), played a significant role in regularizing the model without significantly altering its predictions (Kostopoulos et al., 2018). However, the Pi-model had certain limitations. The model could only be updated once per epoch due to the extensive amount of data being used. This infrequent update interval may not have allowed for optimal model learning. In addition, using the same model as both the teacher and the student could result in unsupervised loss of weight, outweighing the supervised loss of weight, as reported by Laine and Aila (2016). This imbalance could potentially hinder the effective learning of new information by the model.

Meanteacher strategy was implemented in the present study in order to tackle the aforementioned issues. This strategy updated the weights of the student model using gradient backpropagation and separately updated the weights of the teacher model through exponential moving average (EMA). By employing this approach, the limitations of the Pi-

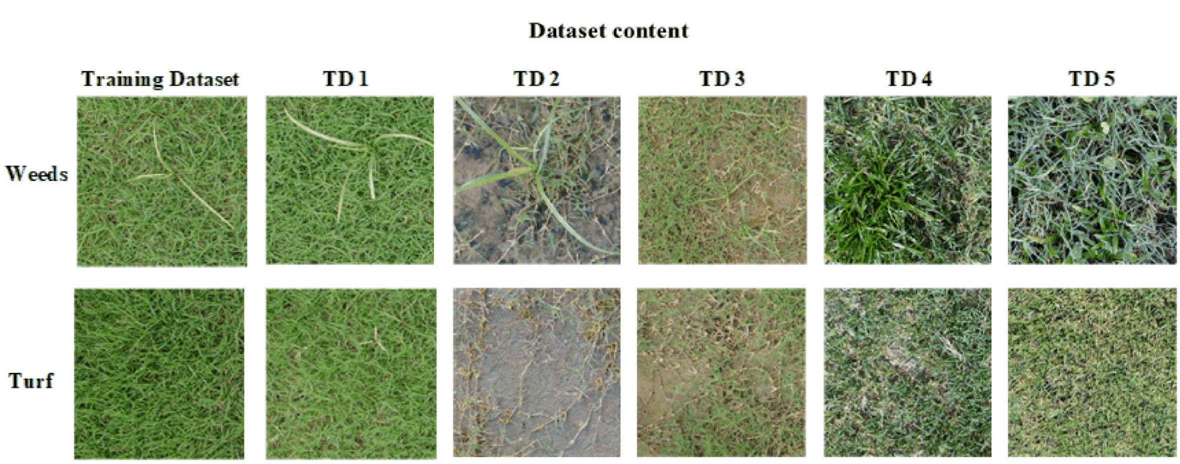

Fig. 1. Example images used for training and testing image classification neural networks. Abbreviation: TD, testing dataset.

Table 2
Weed detection validation results.

Models	Labeled images per class	Image size (pixels)	Validation dataset			
			ACC	Precision	Recall	F_1 score
Fullysupervised	100	240 × 240	0.9440	0.9458	0.9440	0.9439
	100	360 × 360	0.9160	0.9184	0.9160	0.9159
	100	480 × 480	0.9340	0.9380	0.9340	0.9338
	200	240 × 240	0.9650	0.9655	0.9650	0.9650
	200	360 × 360	0.9250	0.9288	0.9250	0.9248
	200	480 × 480	0.9530	0.9538	0.9530	0.9530
Meanteacher	100	240 × 240	0.9200	0.9210	0.9200	0.9200
	100	360 × 360	0.8360	0.8420	0.8360	0.8353
	100	480 × 480	0.8470	0.8513	0.8470	0.8465
	200	240 × 240	0.9660	0.9661	0.9660	0.9660
	200	360 × 360	0.9180	0.9237	0.9180	0.9177
	200	480 × 480	0.9730	0.9734	0.9730	0.9730
Pi-model	100	240 × 240	0.9670	0.9677	0.9670	0.9670
	100	360 × 360	0.9220	0.9228	0.9220	0.9220
	100	480 × 480	0.9540	0.9559	0.9540	0.9540
	200	240 × 240	0.9760	0.9766	0.9760	0.9760
	200	360 × 360	0.9490	0.9515	0.9490	0.9489
	200	480 × 480	0.9640	0.9645	0.9640	0.9640
FixMatch	100	240 × 240	0.9770	0.9770	0.9770	0.9770
	100	360 × 360	0.9607	0.9600	0.9600	0.9607
	100	480 × 480	0.9665	0.9660	0.9660	0.9665
	200	240 × 240	0.9860	0.9861	0.9860	0.9860
	200	360 × 360	0.9670	0.9678	0.9670	0.9670
	200	480 × 480	0.9790	0.9790	0.9790	0.9790

Abbreviation: ACC, accuracy.

Table 3
Weed detection testing results.

Models	Labeled images per class	Image size (pixels)	Testing dataset			
			ACC	Precision	Recall	F_1 score
Fullysupervised	100	240 × 240	0.940	0.998	0.882	0.936
	100	360 × 360	0.784	0.990	0.574	0.727
	100	480 × 480	0.697	1.000	0.394	0.565
	200	240 × 240	0.957	0.987	0.926	0.956
	200	360 × 360	0.814	0.979	0.642	0.775
	200	480 × 480	0.747	0.992	0.498	0.663
Meanteacher	100	240 × 240	0.904	0.895	0.916	0.905
	100	360 × 360	0.823	0.997	0.648	0.785
	100	480 × 480	0.673	0.994	0.348	0.516
	200	240 × 240	0.959	0.9890	0.9280	0.9580
	200	360 × 360	0.846	0.884	0.796	0.838
	200	480 × 480	0.906	1.000	0.812	0.896
Pi-model	100	240 × 240	0.966	0.996	0.936	0.965
	100	360 × 360	0.852	0.994	0.708	0.827
	100	480 × 480	0.850	0.989	0.708	0.825
	200	240 × 240	0.974	0.992	0.956	0.974
	200	360 × 360	0.936	0.984	0.886	0.933
	200	480 × 480	0.897	0.995	0.798	0.886
FixMatch	100	240 × 240	0.981	0.996	0.966	0.981
	100	360 × 360	0.913	0.995	0.830	0.905
	100	480 × 480	0.924	0.993	0.854	0.918
	200	240 × 240	0.982	0.992	0.972	0.982
	200	360 × 360	0.940	0.998	0.800	0.888
	200	480 × 480	0.926	0.991	0.860	0.921

Abbreviation: ACC, accuracy.

Model were overcome, resulting in more effective model updates and preventing the adverse effects of poor target quality on the learning process (Tarvainen and Valpola, 2017).

FixMatch was initiated by generating pseudo-labels through the model's predictions on weakly augmented unlabeled images. For a given image, the pseudo-label was retained only if the model exhibited a high-confidence prediction. Subsequently, the model was trained to predict these pseudo-labels when provided with a strongly augmented version of the same image (Sohn et al., 2020).

The Resnet50 (He et al., 2016) was chosen as the training network for the experiment due to its availability of pre-training weights specially designed for the Imagenet dataset (Jia et al., 2009). Additionally, Resnet50 was widely recognized as one of the outstanding basic networks for visual processing.

2.5. Performance metrics

The validation and testing results are arranged using a confusion matrix under four outcomes: true positive (tp), false positive (fp), true negative (tn), and false negative (fn). In the context of this study, tp represents the images where target weeds are correctly identified. tn represents the images where turfgrasses without target weeds are correctly identified. fp represents the images where images without target weeds are incorrectly identified as containing target weeds. fn represents the images where target weeds are incorrectly not identified as turfgrasses. Based on the confusion matrix results, accuracy (Equation (1)), precision (Equation (2)), recall (Equation (3)), and F_1 score (Equation (4)) were calculated to evaluate the performance of weed detection.

The precision, recall, and F_1 score are metrics that measure the predictive ability of a neural network. These values range from 0 to 1 and serve as unitless indices. A higher value indicates a better predictive ability of the network. High precision signifies that the neural network has a high success rate in detecting areas without weeds in the turfgrass. On the other hand, high recall indicates a high success rate in detecting the target weeds.

Accuracy (ACC) is a measurement of how many predictions were correct out of the total number of observations. It was calculated using the following equation (Sokolova and Lapalme, 2009):

$$\mathrm{ACC} = \frac{\mathrm{tp} + \mathrm{tn}}{\mathrm{tp} + \mathrm{fp} + \mathrm{tn} + \mathrm{fn}} \tag{1}$$

When there is an imbalance between positive and negative samples in the dataset, ACC may not accurately reflect the performance of the model. In such cases, it is possible for the ACC to be high even if specific classes of samples were incorrectly classified. Therefore, it is recommended to consider other metrics, such as precision, recall, and F_1 score, to assess the model's performance.

Precision measured the efficacy of the neural network in accurately identifying the targets and was computed using the following formula (Sokolova and Lapalme, 2009):

$$\mathrm{Precision} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}} \tag{2}$$

Recall offered an estimation of the ability of the trained neural network to correctly identify its targets. It was calculated using the following formula (Sokolova and Lapalme, 2009):

$$\mathrm{Recall} = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn}} \tag{3}$$

Furthermore, the F_1 score of the neural network, the harmonic means of precision and recall, was calculated using the following formula (Sokolova and Lapalme, 2009):

$$F_1 = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{4}$$

The F_1 score incorporates the aforementioned performance metrics of recall and precision, resulting in a more comprehensive and objective evaluation metric for assessing the performance of the neural networks.

2.6. Result and discussion

The experimental results presented in this study provide evidence for the viability of employing the SSL strategy to detect multiple sedge species at different densities and growth stages within turfgrass sod farms, even when confronted with a scarcity of labeled data. The obtained test results encompassed diverse turfgrass fields characterized by varying levels of turfgrass coverage and bare soil area and thus reflected the generalizability of the trained neural networks.

According to the findings presented in Tables 3 and it was observed that the most favorable outcomes were achieved when employing the smallest training image size of 240 × 240 pixels, irrespective of the selected training strategy. The superior performance can be attributed to the chosen input image size of 240 × 240 pixels, which provides the network with an optimally sized field of view. This ensures that the field of view encompasses the target area while providing sufficient detailed information to the network. Conversely, as the image pixel size increased, the performance of the model decreased. Our hypothesis is that small-sized images often contain fewer details and information, making it easier for the model to focus on learning the basic features of the image. During the training process, the model can pay more attention to learning and extracting key features of the image, such as edges, shapes, colors, etc., thereby improving the classification ability of the model.

In the context of the Mean Teacher and Pi model training strategies, the majority of the metrics for the trained models showed improvements greater than 1% when trained with 200 labeled images as opposed to when trained with 100 labeled images. This observation indicates that increasing the number of labeled images used for training has a positive impact on the performance of weed detection. Conversely, when employing the FixMatch strategy, the observed model improvement was not pronounced, resulting in a minor gain of approximately 0.1% for both ACC and F_1 score metrics.

The Meanteacher-trained model utilizing 200 labeled images, which attained an ACC of 0.959 and an F_1 score of 0.958, showed the capability of approximating the performance of the model trained through FSL. In addition, the same model trained under the same condition utilizing 100 labeled images only attained an ACC of 0.904 and an F_1 score of 0.905. This suggests that increasing the number of labeled images for the SSL strategy significantly contributes to bridging the performance gap between the two training methods. The Pi-model-trained model exhibited superior performance in comparison to the FSL-trained model under identical conditions, irrespective of whether it was trained on 100 or 200 labeled images. Moreover, the FixMatch method, surpassing even the Pi-model, demonstrated superior performance compared to the FSL, regardless of the number or pixel size of labeled images. This observation highlights the efficacy and versatility of the FixMatch approach in detecting weeds in turfgrass, even when confronted with varying turfgrass coverage and areas of bare soil.

In the study, among the trained neural networks, the testing results indicated that FSL and FixMatch, both trained with input images of 240 × 240 pixels, exhibited the highest performance. Specifically, when 100 labeled data were utilized, FSL demonstrated ACC, precision, recall, and F_1 scores of 0.940, 0.998, 0.882, and 0.936, respectively. In addition, when trained with an input image size of 240 × 240 pixels, the Resnet50 network model trained using the FixMatch method achieved ACC, precision, recall, and F_1 scores of 0.981, 0.996, 0.966, and 0.981, respectively. The optimal performance was attained using 200 labeled images when trained with FixMatch. Consequently, all subsequent comparisons were conducted between the FixMatch and the FSL strategy.

Tables 4 and 5 showed the performance of the optimized model trained via FixMatch compared to the FSL strategy across TD 1 to 5. Irrespective of the quantity of labeled data or varying input image sizes, the FixMatch demonstrated superior performance in comparison to the FSL strategy. Furthermore, the models trained using the FixMatch method effectively detected sedges across the TD 1 to 5. Especially when using the smallest image input size of 240 × 240 pixels, the FixMatch-trained model achieved high F_1 scores all above 0.838. In contrast, the FSL strategy yielded F_1 scores approximately equal to or greater than 0.864 across the TD 1, TD 2, and TD 5 only when trained with the smallest size images. Notably, the FixMatch-trained model, with only 100 or 200 annotated images, consistently outperformed the FSL-trained model across all conditions. These results suggest that the FixMatch models had superior performance and required fewer labeled samples compared to the FSL strategy.

To assess the impact of varying turf coverage within a sod farm on the model's efficacy for detecting weeds, the study was designed to investigate the performance of weed detection across TD 1, TD 2, and TD 3. These datasets featured different proportions of turfgrass coverage and bare soil areas. Specifically, TD 1, TD 2, and TD 3 exhibited turfgrass coverage >90%, 30%–40%, and 70%–80%, respectively (refer to Table 1 for details). Both FSL and FixMatch methods showed the worst results when tested with TD 3 (Tables 4 and 5). After a comprehensive comparison, the FixMatch method using 200 labeled images with pixels

Table 4
Neural network (trained with 100 labeled images) testing results for detecting sedges growing in bermudagrass.

Dataset	Models	Image size (pixels)	ACC	Percision	Recall	F_1 score
TD 1	Fullysupervised	240 × 240	0.898	0.984	0.812	0.890
	Fullysupervised	360 × 360	0.773	1.000	0.557	0.716
	Fullysupervised	480 × 480	0.681	1.000	0.397	0.569
	FixMatch	240 × 240	0.942	0.985	0.899	0.940
	FixMatch	360 × 360	0.948	0.979	0.919	0.948
	FixMatch	480 × 480	0.902	1.000	0.814	0.888
TD 2	Fullysupervised	240 × 240	0.912	0.954	0.862	0.906
	Fullysupervised	360 × 360	0.794	0.980	0.632	0.769
	Fullysupervised	480 × 480	0.586	1.000	0.169	0.289
	FixMatch	240 × 240	0.973	0.979	0.966	0.972
	FixMatch	360 × 360	0.969	0.870	0.955	0.970
	FixMatch	480 × 480	0.923	0.946	0.897	0.921
TD 3	Fullysupervised	240 × 240	0.745	1.000	0.462	0.632
	Fullysupervised	360 × 360	0.558	1.000	0.155	0.269
	Fullysupervised	480 × 480	0.474	0.400	0.043	0.078
	FixMatch	240 × 240	0.932	1.000	0.856	0.922
	FixMatch	360 × 360	0.823	1.000	0.662	0.797
	FixMatch	480 × 480	0.759	0.863	0.633	0.730
TD 4	Fullysupervised	240 × 240	0.813	1.000	0.634	0.776
	Fullysupervised	360 × 360	0.640	0.956	0.314	0.473
	Fullysupervised	480 × 480	0.538	1.000	0.120	0.215
	FixMatch	240 × 240	0.855	0.981	0.731	0.838
	FixMatch	360 × 360	0.614	0.947	0.263	0.411
	FixMatch	480 × 480	0.696	1.000	0.421	0.593
TD 5	Fullysupervised	240 × 240	0.934	1.000	0.870	0.930
	Fullysupervised	360 × 360	0.892	1.000	0.761	0.864
	Fullysupervised	480 × 480	0.500	1.000	0.049	0.094
	FixMatch	240 × 240	0.973	1.000	0.945	0.972
	FixMatch	360 × 360	0.792	1.000	0.538	0.700
	FixMatch	480 × 480	0.640	1.000	0.238	0.384

Abbreviation: ACC, accuracy; TD, testing dataset.

Table 5
Neural network (trained with 200 labeled images) testing results for detection of sedges growing in bermudagrass.

Dataset	Models	Image size (pixels)	ACC	Precision	Recall	F_1 score
TD 1	Fullysupervised	240 × 240	0.939	0.971	0.906	0.938
	Fullysupervised	360 × 360	0.863	0.982	0.745	0.847
	Fullysupervised	480 × 480	0.624	1.000	0.288	0.448
	FixMatch	240 × 240	0.956	0.972	0.940	0.956
	FixMatch	360 × 360	0.873	0.983	0.765	0.860
	FixMatch	480 × 480	0.783	1.000	0.590	0.742
TD 2	Fullysupervised	240 × 240	0.942	0.957	0.924	0.940
	Fullysupervised	360 × 360	0.825	0.965	0.703	0.813
	Fullysupervised	480 × 480	0.663	1.000	0.324	0.489
	FixMatch	240 × 240	0.969	0.986	0.952	0.968
	FixMatch	360 × 360	0.962	1.000	0.929	0.963
	FixMatch	480 × 480	0.864	0.981	0.743	0.845
TD 3	Fullysupervised	240 × 240	0.867	1.000	0.720	0.837
	Fullysupervised	360 × 360	0.590	0.971	0.223	0.363
	Fullysupervised	480 × 480	0.500	0.583	0.101	0.172
	FixMatch	240 × 240	0.949	1.000	0.894	0.944
	FixMatch	360 × 360	0.686	1.000	0.399	0.570
	FixMatch	480 × 480	0.567	0.762	0.230	0.354
TD 4	Fullysupervised	240 × 240	0.791	1.000	0.593	0.745
	Fullysupervised	360 × 360	0.569	0.958	0.168	0.286
	Fullysupervised	480 × 480	0.561	1.000	0.165	0.284
	FixMatch	240 × 240	0.869	0.982	0.759	0.856
	FixMatch	360 × 360	0.727	1.000	0.467	0.637
	FixMatch	480 × 480	0.628	0.976	0.301	0.460
TD 5	Fullysupervised	240 × 240	0.969	1.000	0.938	0.968
	Fullysupervised	360 × 360	0.764	1.000	0.479	0.647
	Fullysupervised	480 × 480	0.581	1.000	0.115	0.206
	FixMatch	240 × 240	0.924	1.000	0.849	0.919
	FixMatch	360 × 360	0.799	1.000	0.556	0.714
	FixMatch	480 × 480	0.659	1.000	0.279	0.436

Abbreviation: ACC, accuracy; TD, testing dataset.

of 480 × 480 as the training data showed the worst results when tested with the TD3, with an F_1 score of only 0.354. Additionally, the FSL strategy exhibited relatively worse performance on the TD 3 and TD 4 datasets with 200 labeled images at the pixel size of 480 × 480. The primary contributor to the observed decline in the F_1 scores with the TD 3 and TD 4 datasets for both strategies was identified as a significant reduction in recall.

In the context of the five validation datasets, the model exhibited a reduced F_1 score as the image size enlarged, predominantly attributable to low recall rather than precision, indicating that a large proportion of the actual weeds present in the trained neural network were not being detected. Such an outcome is highly undesirable, as it has the potential to result in missed spraying instances, thereby compromising the efficacy of weed control when utilizing the model in smart sprayers for precision herbicide application.

In a recent investigation conducted by Jin et al. (2022a), a software application was devised, integrating an image classification neural network model with OpenCV-Python. This innovative software automatedly partitioned a testing image measuring 1920 × 1080 pixels into 40 grid cells of equal dimensions, employing OpenCV. Subsequently, the image classification neural network was deployed to categorize the contents of these grid cells. Upon detection of weeds by the developed neural network within a grid cell, the corresponding area was designated as “sprayed,” whereas grid cells without weeds were classified as “non-sprayed.”

A recently developed smart sprayer prototype demonstrated its effectiveness in detecting weeds growing on dormant bermudagrass turf (Jin et al., 2023a). After the grid cells were located in the machine vision subsystem of the smart sprayer using the software described above, the nozzles above the weed-containing grid cells were activated to precisely spray the herbicide. Recently, a smart sprayer prototype, developed and reported by Jin et al. (2023b), demonstrated successful capabilities for detecting weeds growing in bermudagrass turf. The authors noted that following the detection of grid cells containing weeds within the machine vision subsystem of the developed smart sprayer, facilitated by the aforementioned software, the nozzles positioned above grid cells containing weeds were activated, thereby realizing precision herbicide application. Integrating the developed smart sprayer reported by Jin et al. (2023b) with the SSL-trained neural networks proposed in this paper, for precise herbicide application in turf sod farms, merits further investigation.

3. Conclusion

This research validated the effectiveness of the SSL strategy for weed detection with a small amount of labeled data. Experimental results revealed that the SSL strategy employed in FixMatch outperformed FSL. The optimal performance of the trained model was achieved when utilizing input images at a resolution of 240 × 240 pixels. The experiment revealed that the advantage derived from employing 200 labeled samples was not notably superior to that observed with 100 labeled samples when utilizing the FixMatch. Consequently, the conclusion was that the ResNet50 model, trained under the SSL strategy of FixMatch using 200 labeled images at a resolution of 240 × 240 pixels, yielded the best performance. In addition, varying turf coverage and bare soil area in bermudagrass sod farms barely had any meaningful impact on the performance of SSL for weed detection. The incorporation of SSL notably improved the ability of the neural networks to detect weeds. Nevertheless, its performance proved insufficient when tested on the TD 3 and TD 4 datasets. Therefore, collecting diverse weed images from various geographical locations and under varied environmental conditions is crucial because a higher quality and more balanced dataset enhances the model’s generalizability and strengthens the training process.

CRediT authorship contribution statement

Xin Chen: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Teng Liu:** Data curation, Formal analysis. **Kang Han:** Data curation, Formal analysis, Project administration. **Xiaojun Jin:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Writing – review & editing. **Jialin Yu:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 32072498), Key R&D Program of Shandong Province, China (202211070163), and the Taishan Scholar Program.

References

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 1–74. https://doi.org/10.1186/s40537-021-00444-8.

Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A., 2019. Mixmatch: a holistic approach to semi-supervised learning. Adv. Neural Inf. Process. Syst. 32 https://doi.org/10.48550/arXiv.1905.02249.

Busey, P., 2003. Cultural management of weeds in turfgrass: a review. Crop Sci. 43, 1899–1911. https://doi.org/10.2135/cropsci2003.1899.

Ding, S., Zhu, Z., Zhang, X., 2017. An overview on semi-supervised support vector machine. Neural Comput. Appl. 28, 969–978. https://doi.org/10.1007/s00521-015-2113-7.

dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T., 2017. Weed detection in soybean crops using ConvNets. Comput. Electron. Agric. 143, 314–324. https://doi.org/10.1016/j.compag.2017.10.027.

Enguehard, J., O’Halloran, P., Gholipour, A., 2019. Semi-supervised learning with deep embedded clustering for image classification and segmentation. IEEE Access 7, 11093–11104. https://doi.org/10.1109/ACCESS.2019.2891970.

Gerhards, R., Andujar Sanchez, D., Hamouz, P., Peteinatos, G.G., Christensen, S., Fernandez-Quintanilla, C., 2022. Advances in site-specific weed management in agriculture-A review. Weed Res. 62, 123–133. https://doi.org/10.1111/wre.12526.

Hasan, A.M., Sohel, F., Diepeveen, D., Laga, H., Jones, M.G., 2021. A survey of deep learning techniques for weed detection from images. Comput. Electron. Agric. 184, 106067 https://doi.org/10.1016/j.compag.2021.106067.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Las Vegas, USA, pp. 770–778.

Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K., 2014. Densenet: Implementing Efficient Convnet Descriptor Pyramids arXiv preprint arXiv: 1404.1869. https://arxiv.org/abs/1404.1869.

Jia, D., Wei, D., Socher, R., Li, L.J., Li, K., Li, F., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Miami, FL, USA, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Jin, X., Che, J., Chen, Y., 2021. Weed identification using deep learning and image processing in vegetable plantation. IEEE Access 9, 10940–10950. https://doi.org/10.1109/ACCESS.2021.3050296.

Jin, X., Bagavathiannan, M., McCullough, P.E., Chen, Y., Yu, J., 2022a. A deep learning-based method for classification, detection, and localization of weeds in turfgrass. Pest Manag. Sci. 78, 4809–4821. https://doi.org/10.1002/ps.7102.

Jin, X., Liu, T., Chen, Y., Yu, J., 2022b. Deep learning-based weed detection in turf: a review. Agronomy 12, 3051. https://doi.org/10.3390/agronomy12123051.

Jin, X., Liu, T., Yang, Z., Xie, J., Bagavathiannan, M., Hong, X., Xu, Z., Chen, X., Yu, J., Chen, Y., 2023a. Precision Weed Control Using a Smart Sprayer in Dormant Bermudagrass Turf. Crop Prot., 106302 https://doi.org/10.1016/j.cropro.2023.106302

Jin, X., McCullough, P.E., Liu, T., Yang, D., Zhu, W., Chen, Y., Yu, J., 2023b. A smart sprayer for weed control in bermudagrass turf based on the herbicide weed control spectrum. Crop Protect. 170, 106270 https://doi.org/10.1016/j.cropro.2023.106270.

Kostopoulos, G., Karlos, S., Kotsiantis, S., Ragos, O., 2018. Semi-supervised regression: a recent review. J. Intell. Fuzzy Syst. 35, 1483–1500. https://doi.org/10.3233/JIFS-169689.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25 https://doi.org/10.1145/3065386.

Laine, S., Aila, T., 2016. Temporal Ensembling for Semi-supervised Learning arXiv preprint arXiv:1610.02242. https://arxiv.org/abs/1610.02242.

Le, V.N.T., Ahderom, S., Alameh, K., 2020. Performances of the lbp based algorithm over cnn models for detecting crops and weeds with similar morphologies. Sensors 20, 2193. https://doi.org/10.3390/s20082193.

Liang, W., Ma, X., Wan, P., Liu, L., 2018. Plant salt-tolerance mechanism: a review. Biochem. Biophys. Res. Commun. 495, 286–291. https://doi.org/10.1016/j.bbrc.2017.11.043.

Liu, B., Bruch, R., 2020. Weed detection for selective spraying: a review. Current Robotics Reports 1, 19–26. https://doi.org/10.1007/s43154-020-00001-w.

Liu, T., Jin, X., Zhang, L., Wang, J., Chen, Y., Hu, C., Yu, J., 2023. Semi-supervised Learning and Attention Mechanism for Weed Detection in Wheat. Crop Prot., 106389 https://doi.org/10.1016/j.cropro.2023.106389

McElroy, J., Martins, D., 2013. Use of herbicides on turfgrass. Planta Daninha 31, 455–467. https://doi.org/10.1590/S0100-83582013000200024.

Monteiro, A., Santos, S., 2022. Sustainable approach to weed management: the role of precision weed management. Agronomy 12, 118. https://doi.org/10.3390/agronomy12010118.

Olsen, A., Konovalov, D.A., Philippa, B., Ridd, P., Wood, J.C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J., 2019. DeepWeeds: a multiclass weed species image dataset for deep learning. Sci. Rep. 9, 2058. https://doi.org/10.1038/s41598-018-38343-3.

Pincetl, S., Gillespie, T.W., Pataki, D.E., Porse, E., Jia, S., Kidera, E., Nobles, N., Rodriguez, J., Choi, D., 2019. Evaluating the effects of turf-replacement programs in Los Angeles. Landsc. Urban Plann. 185, 210–221. https://doi.org/10.1016/j.landurbplan.2019.01.011.

Reddy, Y., Viswanath, P., Reddy, B.E., 2018. Semi-supervised learning: a brief review. Int. J. Eng. Technol. 7, 81. https://doi.org/10.14419/ijet.v7i1.8.9977.

Rosle, R., Che'Ya, N.N., Ang, Y., Rahmat, F., Wayayok, A., Berahim, Z., Fazlil Ilahi, W.F., Ismail, M.R., Omar, M.H., 2021. Weed detection in rice fields using remote sensing technique: a review. Appl. Sci. 11, 10701 https://doi.org/10.3390/app112210701.

Sharpe, S.M., Schumann, A.W., Boyd, N.S., 2019. Detection of Carolina geranium (*Geranium carolinianum*) growing in competition with strawberry using convolutional neural networks. Weed Sci. 67, 239–245. https://doi.org/10.1017/wsc.2018.66.

Sharpe, S.M., Schumann, A.W., Boyd, N.S., 2020a. Goosegrass detection in strawberry and tomato using a convolutional neural network. Sci. Rep. 10, 9548. https://doi.org/10.1038/s41598-020-66505-9.

Sharpe, S.M., Schumann, A.W., Yu, J., Boyd, N.S., 2020b. Vegetation detection and discrimination within vegetable plasticulture row-middles using a convolutional neural network. Precis. Agric. 21, 264–277. https://doi.org/10.1007/s11119-019-09666-6.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv preprint arXiv:1409.1556. https://arxiv.org/abs/1409.1556.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L., 2020. Fixmatch: simplifying semi-supervised learning with consistency and confidence. Adv. Neural Inf. Process. Syst. 33, 596–608. https://doi.org/10.48550/arXiv.2001.07685.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.

Su, W.H., Fennimore, S.A., Slaughter, D.C., 2019. Fluorescence imaging for rapid monitoring of translocation behaviour of systemic markers in snap beans for automated crop/weed discrimination. Biosyst. Eng. 186, 156–167. https://doi.org/10.1016/j.biosystemseng.2019.07.009.

Sun, Y., Zheng, W., Ren, Z., 2022. Application of convolutional neural network in image processing. In: International Conference on Multi-Modal Information Analytics. Springer, Huhehaote, China, pp. 375–383. https://doi.org/10.1007/978-3-031-05484-6_47.

Tarvainen, A., 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. https://arxiv.org/abs/1703.01780.

USEPA, 2023a. https://www.epa.gov/ingredients-used-pesticide-products/atrazine. September, 2023.

USEPA, 2023b. https://www.epa.gov/ingredients-used-pesticide-products/monosodium-methanearsonate-msma-organic-arsenica. September, 2023.

Valpola, H., 2015. From neural PCA to deep unsupervised learning. In: Bingham, E., Kaski, S., Laaksonen, J., Lampinen, J. (Eds.), Advances in Independent Component Analysis and Learning Machines. Academic Press Inc., New York, pp. 143–171. https://doi.org/10.1016/B978-0-12-802806-3.00008-7.

Wang, A., Zhang, W., Wei, X., 2019. A review on weed detection using ground-based machine vision and image processing techniques. Comput. Electron. Agric. 158, 226–240. https://doi.org/10.1016/j.compag.2019.02.005.

Wu, D., Feng, L., Zhang, C., He, Y., 2008. Early detection of Botrytis cinerea on eggplant leaves based on visible and near-infrared spectroscopy. Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 51, 1133–1139, 10.13031/2013.24504.

Wu, Z., Chen, Y., Zhao, B., Kang, X., Ding, Y., 2021. Review of weed detection methods based on computer vision. Sensors 21, 3647. https://doi.org/10.3390/s21113647.

Yu, J., Schumann, A.W., Cao, Z., Sharpe, S.M., Boyd, N.S., 2019a. Weed detection in perennial ryegrass with deep learning convolutional neural network. Front. Plant Sci. 10, 1422. https://doi.org/10.3389/fpls.2019.01422.

Yu, J., Sharpe, S.M., Schumann, A.W., Boyd, N.S., 2019b. Deep learning for image-based weed detection in turfgrass. Eur. J. Agron. 104, 78–84. https://doi.org/10.3389/fpls.2019.01422.

Zhang, W., Miao, Z., Li, N., He, C., Sun, T., 2022. Review of current robotic approaches for precision weed management. Current robotics reports 3, 139–151. https://doi.org/10.1007/s43154-022-00086-5.

Zhuang, J., Li, X., Bagavathiannan, M., Jin, X., Yang, J., Meng, W., Li, T., Li, L., Wang, Y., Chen, Y., 2022. Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. Pest Manag. Sci. 78, 521–529. https://doi.org/10.1002/ps.6656.