



REVIEW

A Review of Deep Learning-Based Precision Weed Detection in Farmland Environments

Peng Shen¹, Tenglong Li^{1,2}, Yongpeng Sun^{1,2}, Hao Cui^{1,2} and Guoqing Zhang^{3,*}

¹School of Aeronautics and Astronautics, North China Institute of Aerospace Engineering, Langfang, China

²Collaborative Innovation Center of Micro & Nano Satellites, North China Institute of Aerospace Engineering, Langfang, China

³School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China

*Corresponding Author: Guoqing Zhang. Email: zgq15134067170@163.com

Received: 26 May 2026; Accepted: 03 August 2026; Published: 15 September 2026

ABSTRACT: Agriculture plays an important role in food security and social development. With the rapid development of the Fourth Agricultural Revolution, also known as Agriculture 4.0, traditional weed control methods that rely on manual experience and uniform herbicide application can no longer meet the demands for efficient, precise, and environmentally friendly production. Farmland weeds compete with crops for light, water, and nutrients, thereby seriously affecting crop yield and quality. Therefore, the development of efficient weed detection and recognition technologies is of great significance. In recent years, the rapid progress of deep learning in computer vision has provided new technical approaches for automatic weed recognition, precise localization, and intelligent weed control in farmland environments. This paper presents a review of deep learning-based precision weed detection technologies in farmland. Unlike previous reviews that mainly focus on weed recognition algorithms, datasets, or agricultural robotic systems separately, this review provides a complementary Agriculture 4.0-oriented and deployment-aware perspective by linking data foundations, deep learning models, field robustness, edge deployment, and precision weeding applications within a unified technical framework. First, mainstream weed image datasets, data preprocessing methods, and data augmentation strategies are summarized. Then, recent advances in the application of convolutional neural networks to weed image classification, object detection, semantic segmentation, and instance segmentation are analyzed. In addition, the application value of semi-supervised, unsupervised, and weakly supervised learning in small-sample scenarios is discussed, and research progress on the deployment of weed detection models on edge devices is reviewed. Finally, this paper summarizes the key challenges associated with complex field environments, visual similarity between crops and weeds, occlusion and overlap, and real-time deployment, providing a reference for the subsequent design of weed detection models and the application of precision weeding equipment.

KEYWORDS: Agriculture 4.0; precision agriculture; precision weed detection; computer vision; deep learning; smart agricultural machinery; sustainable agriculture

1 Introduction

With the continuing growth of the global population and the increasing constraint of arable land resources, the contradiction between food demand and land availability has become increasingly prominent. The continuing growth of the global population has increased the pressure on agricultural production, making yield improvement and loss reduction important pathways for ensuring food security. In this context, increasing yield per unit area and reducing yield losses during agricultural production are important pathways for ensuring food security. The Fourth Agricultural Revolution, also known as Agriculture 4.0,

represents a new stage following agricultural mechanization, automation, and informatization. Its core objective is to integrate the Internet of Things, artificial intelligence, big data, cloud computing, and robotics into the entire agricultural production process, thereby enabling precise perception of production environments, intelligent decision-making, and autonomous field operations. These technologies provide important support for balancing food security with ecological protection [1]. Within the technological framework of Agriculture 4.0, precision prevention and control of diseases, pests, and weeds is an important application direction. Among these factors, weed control is particularly critical because weeds are widely distributed, persist over long periods, and are difficult to manage, making them a key factor affecting the productivity and ecological benefits of field crop production [2].

Weeds refer to plants that grow in farmland without intentional cultivation. They have strong environmental adaptability and reproductive capacity. In farmland ecosystems, weeds compete with target crops for light, water, nutrients, and growing space. They may also serve as intermediate hosts for pests and pathogens, thereby intensifying the combined damage caused by crop diseases, insect pests, and weeds. The yield losses caused by weeds far exceed those caused by insect pests, diseases, and other natural hazards.

Traditional farmland weed control methods mainly include chemical control, manual weeding, and mechanical weeding. Chemical control refers to the uniform application of herbicides across the entire field. Owing to its simple operation and relatively high control efficiency, it remains one of the most widely used approaches in agricultural weed management. However, the spatial distribution of weeds in farmland is highly heterogeneous, and uniform spraying causes a large amount of herbicide to be applied to weed-free areas, resulting in pesticide waste. Meanwhile, excessive and uniform herbicide application may accelerate the development of herbicide resistance in weeds and increase ecological risks to farmland ecosystems. Manual weeding provides high control accuracy and causes almost no chemical pollution. Nevertheless, it is labor-intensive, inefficient, and costly, making it more suitable for small-scale farmland or high-value crop production rather than large-scale field operations [3]. Mechanical weeding mainly removes inter-row weeds through cultivation or sensor-guided equipment, thereby reducing the use of chemical herbicides [4]. However, conventional mechanical weeding has limited capability in controlling intra-row weeds. Under complex terrain, uneven crop growth, and high-speed operation conditions, crop damage and restricted operational efficiency remain major problems [5].

The core of precision weed control is “on-demand herbicide application and targeted weeding,” which means applying an appropriate dose of herbicide only to weed-infested areas or removing weeds at specific locations using mechanical, laser-based, or other physical methods. The prerequisite for achieving this objective is the real-time and accurate detection and recognition of weed species, spatial location, density, and growth stage in the field [6]. The development of precision weed detection technology in farmland has generally progressed in parallel with the evolution of computer vision, and can be broadly divided into three stages, as shown in Fig. 1. The first stage, from the 1990s to 2010, mainly relied on traditional machine vision and handcrafted feature-based methods. During this period, color, shape, texture, and other features were designed based on domain knowledge and combined with classical machine learning classifiers to distinguish crops from weeds. These methods achieved satisfactory performance in laboratory environments and simple field scenarios, but their generalization ability in complex real field environments was limited [7]. The second stage, from 2012 to 2018, was mainly represented by convolutional neural networks and transfer learning. With the development of deep learning in general computer vision tasks, CNNs were introduced into agricultural vision, overcoming the strong dependence on handcrafted features and the weak adaptability of traditional methods, and improving recognition accuracy in complex field environments [8]. The third stage, from 2019 to the present, has shifted toward full-scenario weed detection technologies for engineering applications under Agriculture 4.0. At this stage, research no longer focuses only on algorithmic accuracy, but

also emphasizes robustness in real field environments, real-time deployment on edge devices, multi-source information fusion, and coordination between weed recognition results, herbicide application decisions, and weeding actuators. These developments are promoting the transition of weed detection technology from laboratory validation to field-scale application [9].

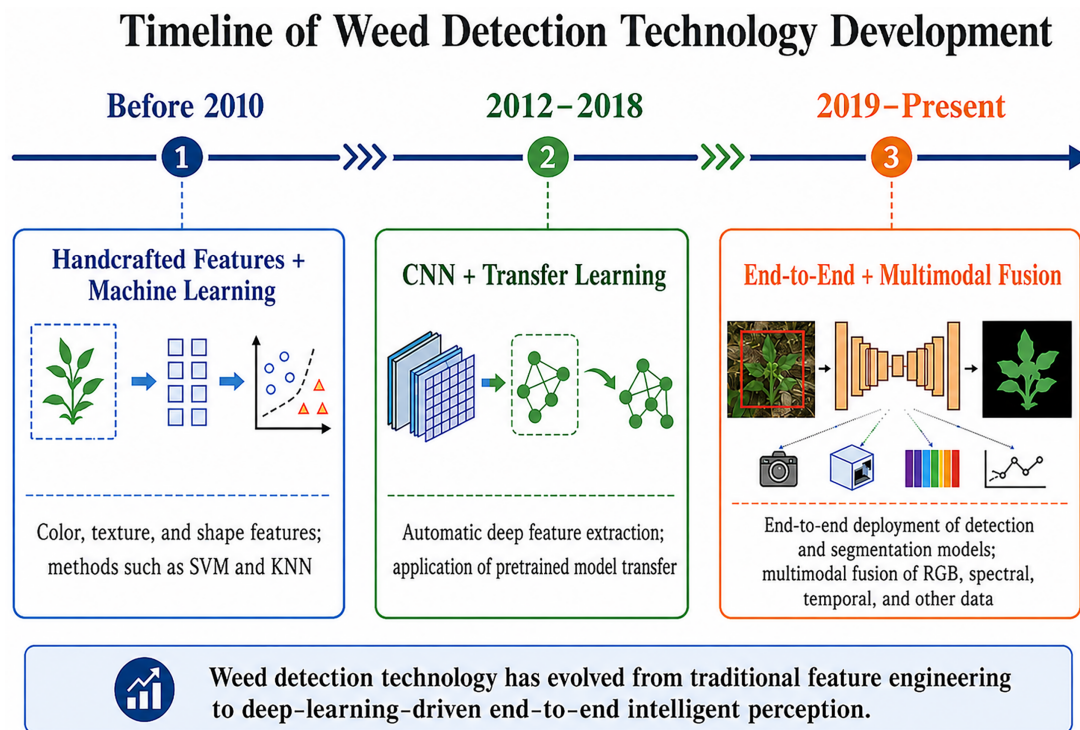


Figure 1: Timeline of weed detection technology development. The representative studies corresponding to each stage are cited as follows: traditional machine vision and handcrafted feature-based methods before 2010 [10]; CNN- and transfer learning-based weed recognition from 2012 to 2018 [11,12]; and full-scenario, end-to-end, and deployment-oriented weed detection technologies under Agriculture 4.0 from 2019 to the present [13,14].

Although several reviews have summarized deep learning-based weed detection, most existing studies tend to focus on specific aspects such as visual recognition algorithms, dataset construction, performance evaluation, or robotic weed management. In contrast, this review provides a complementary and scenario-oriented synthesis under the framework of Agriculture 4.0. Rather than treating weed detection as an isolated computer vision task, this paper organizes recent progress along the complete technical chain of precision weed management, including data foundations, deep learning model categories, label-efficient learning, lightweight edge deployment, field robustness, and the integration of detection outputs with precision weeding equipment. Therefore, this review does not claim to be the first review on deep learning-based weed detection; instead, it complements existing studies by emphasizing the transition from laboratory-oriented model evaluation to real-field agricultural applications.

The main contributions of this review are summarized as follows:

(1) This review summarizes the data foundation of deep learning-based weed detection, including public weed datasets, data preprocessing, annotation strategies, data augmentation methods, and dataset partitioning principles, thereby clarifying the data conditions required for reliable model development.

(2) This review compares mainstream deep learning approaches for weed detection, including image classification, object detection, semantic segmentation, instance segmentation, and semi-supervised, unsupervised, and weakly supervised learning, with emphasis on their strengths, limitations, robustness, and applicable agricultural scenarios.

(3) This review discusses lightweight model design and edge deployment from the perspective of real-time agricultural operation requirements, including inference latency, model complexity, hardware constraints, and deployment feasibility on field platforms.

(4) This review connects weed detection models with practical precision weeding applications, including variable-rate herbicide spraying, mechanical weeding, laser weeding, UAV-based monitoring, and intelligent agricultural machinery, thereby highlighting the perception–decision–execution chain required for Agriculture 4.0.

This paper reviews farmland weed detection under the background of Agriculture 4.0 from three perspectives: data foundations, deep learning methods, and engineering deployment. The review covers the limitations of traditional weed detection methods, the applications of deep learning models in weed classification, object detection, and image segmentation, mainstream public weed datasets, and the remaining challenges in edge deployment and field applications.

Review Methodology and Article Structure

To improve the transparency and reproducibility of this review, a structured literature retrieval and screening process was adopted. Relevant studies were collected from Web of Science Core Collection, Scopus, IEEE Xplore, ScienceDirect, SpringerLink, MDPI, and Google Scholar. The main publication period covered in this review was from January 2019 to May 2026, because this period corresponds to the rapid development of deep learning-based weed detection, edge deployment, and intelligent weeding systems. Meanwhile, earlier representative studies were also included when they provided important background information on traditional machine vision, convolutional neural networks, or the historical development of weed detection technologies.

The search strategy combined terms related to weed detection, deep learning, precision agriculture, and deployment. The main search string was as follows: (“weed detection” OR “weed recognition” OR “crop weed detection” OR “crop-weed classification” OR “weed segmentation”) AND (“deep learning” OR “convolutional neural network” OR “CNN” OR “YOLO” OR “object detection” OR “semantic segmentation” OR “instance segmentation” OR “Transformer”) AND (“precision agriculture” OR “smart agriculture” OR “agricultural robot” OR “UAV” OR “edge deployment”). In addition, the reference lists of highly relevant publications were manually screened to identify representative studies that may have been missed during database retrieval.

Studies were included if they met the following criteria: (1) the study focused on crop–weed classification, weed object detection, weed segmentation, or deployment of weed detection models; (2) deep learning was used as the main technical approach; (3) the study was conducted in farmland, orchard, or UAV-based agricultural environments; and (4) the dataset, model architecture, evaluation metrics, and experimental results were clearly reported. Studies were excluded if they: (1) relied only on handcrafted features or classical machine learning methods without using deep learning, except when used for historical comparison; (2) focused on non-agricultural plant recognition or general vegetation mapping rather than weed detection; (3) lacked quantitative experimental results; (4) were duplicate records, patents, editorials, or non-research materials; or (5) provided insufficient methodological details for comparison. Peer-reviewed journal and conference papers were prioritized, while a small number of highly relevant preprints were considered as supplementary sources when they represented recent technical developments in this field.

After duplicate records were removed, the titles and abstracts of the retrieved studies were screened according to the above criteria. Subsequently, the full texts of potentially relevant articles were reviewed to determine final eligibility. For each included study, information on the dataset, crop and weed categories, sensing modality, task type, model architecture, evaluation metrics, deployment platform, and key findings was extracted and summarized. This process provided the basis for the subsequent analysis of public weed datasets, weed image classification, object detection, image segmentation, label-efficient learning, edge deployment, and practical precision weeding applications.

As shown in Fig. 2, the remainder of this review is organized as follows. Section 2 summarizes deep learning-based weed detection methods, including public datasets, data processing, image classification, object detection, image segmentation, and label-efficient learning. Section 3 discusses lightweight model design and edge deployment. Section 4 analyzes practical applications of weed detection in precision weeding systems. Section 5 discusses current challenges and future development trends. Finally, Section 6 concludes the review.

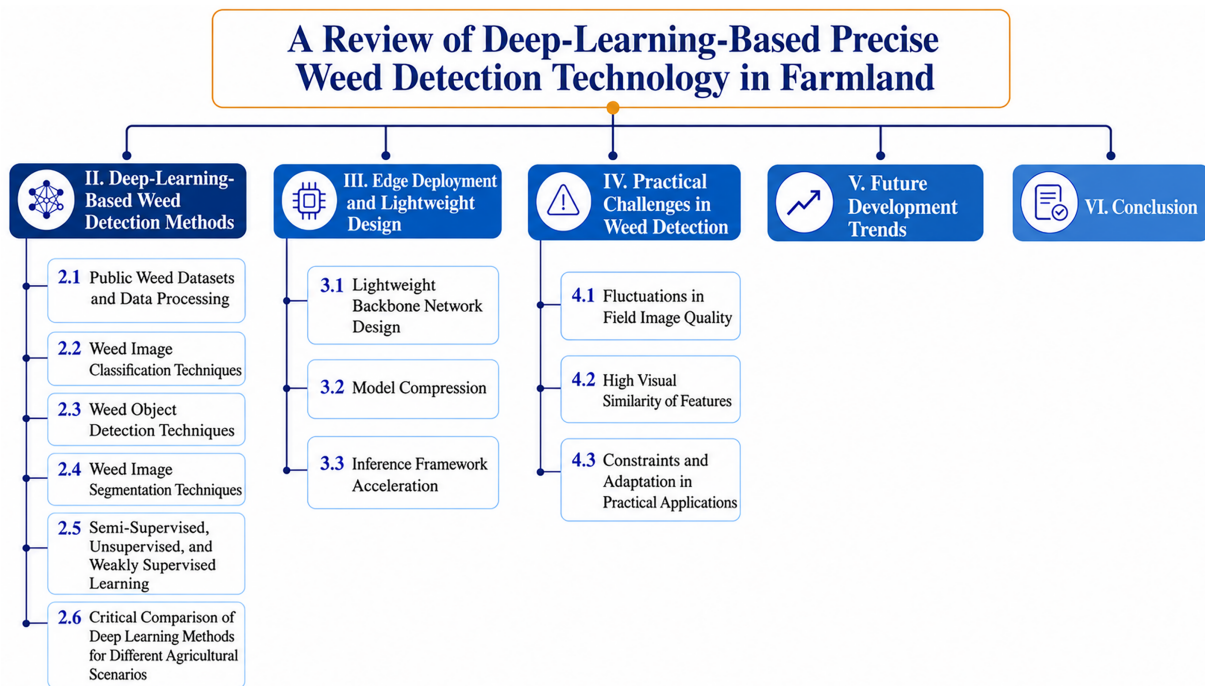


Figure 2: Overall structure of this review article. The figure illustrates the organization of this review from data foundations and deep learning methods to edge deployment, practical challenges, future trends, and conclusions.

2 Deep Learning-Based Methods for Weed Detection

Deep learning methods have been widely applied to farmland weed detection. Compared with traditional methods that rely on manually designed features, deep learning models can directly extract discriminative features of crops and weeds from raw images, thereby reducing the limitations associated with handcrafted color, texture, and shape descriptors. In complex field environments, deep learning models show more stable performance in weed recognition, localization, and segmentation tasks, and have therefore become the dominant technical approach in current research. The overall workflow for training and applying deep learning-based weed detection models is shown in Fig. 3.

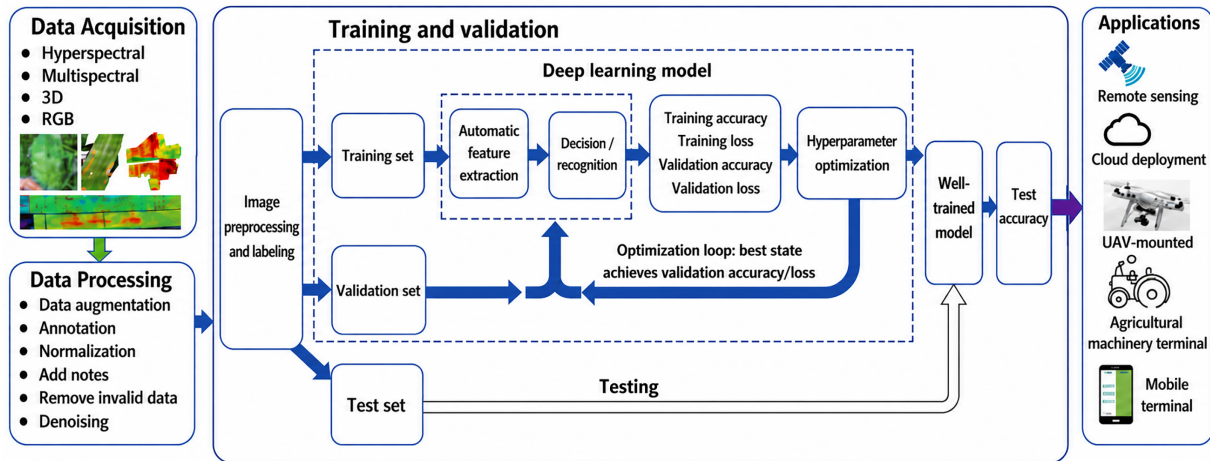


Figure 3: Workflow for training and applying deep learning-based weed detection models. The workflow shows the complete process from data acquisition and preprocessing to model training, validation, testing, and field application.

2.1 Public Weed Datasets and Data Processing Methods

2.1.1 Public Weed Datasets

In recent years, the rapid development of precision agriculture has promoted the release of numerous public weed datasets. These datasets cover different crop types, weed species, growth stages, imaging conditions, and sensor types, providing data support for the development, training, and evaluation of weed detection algorithms. The mainstream datasets in this field are listed in [Table 1](#).

Table 1: Representative publicly available weed datasets.

Datasets	Task Type	Annotation Type	Number of Classes	Download Link
DeepWeeds [15]	Classification/ recognition	Image-level class labels	8 classes of weeds	GitHub
CWFID (Crop/Weed Field Image Dataset) [16]	Semantic segmentation	Segmentation Annotation	Crop and weed	GitHub
CropAndWeed Dataset [17]	Detection/classification/ segmentation	Multimodal annotations	Multiple weed species	GitHub
ACRE Crop-Weed Dataset [18]	Detection/classification/ segmentation	detection boxes, semantic masks, instance masks	Multiple weed species	Zenodo
WeedCrop/Dataset of Annotated Food Crops and Weed Images [19]	Object detection	Bounding box annotations	14 classes (6 crops + 8 weeds)	ScienceDirect
Weed-crop RGB dataset in precision agriculture [20]	Object detection	YOLO-format annotations	13 classes (8 crops + 5 weeds)	Mendeley Data
CottonWeedDet12 [21]	Object detection	Bounding boxes	12 classes of weeds	ScienceDirect
3SeasonWeedDet10 [22]	Object detection/cross-season	Bounding boxes	10 classes of weeds	ScienceDirect
Chicory dataset [23]	Object detection	Multiple formats	Crop and weed	Universe

(Continued)

Table 1 (continued)

Datasets	Task Type	Annotation Type	Number of Classes	Download Link
Maize dataset [24]	Object detection/Different growth stages	PascalVOC	3 classes of weeds	Zenodo
MH-Weed16 [25]	Object detection	Bounding boxes	16 classes of weeds	Mendeley Data
Multi-format open-source weed image dataset [26]	Object detection	Multiple formats	Crop and weed	ScienceDirect
Chilli crop-weed dataset [27]	Object detection	Pascal VOC + YOLO	Crop and weed	GitHub
CWD30 [28]	Object detection	Image data	30 classes (10 crops + 20 weeds)	ScienceDirect
SorghumWeedDataset_Segmentation [29]	Detection/classification/segmentation	Pixel-level annotations	Weed	Mendeley Data
Crop and weed dataset (greenhouse) [30]	Object detection	Image data	Crop and weed	Mendeley Data
Dataset for weed detection in fruit orchards [31]	Orchard weed detection	Multispectral and multi-temporal	6 weed species and 1 ground disturbance class	ScienceDirect

Although the number of publicly available weed datasets has increased in recent years, current benchmark datasets still suffer from dataset bias and annotation-related limitations. These problems directly affect the generalization ability of deep learning models and partially explain why models that achieve high accuracy on a specific dataset may perform poorly when transferred to different crops, regions, seasons, or imaging platforms.

First, class imbalance is a common problem in weed datasets. In many field scenarios, dominant weed species appear frequently, whereas rare weed species, early-growth weeds, partially occluded weeds, and small weed instances are under-represented. As a result, deep learning models may become biased toward majority classes and may show reduced recall for rare or visually similar weed species. This problem is particularly serious in multi-class weed recognition and detection tasks, where the distribution of weed species is often affected by crop type, field management practice, season, soil condition, and local climate. Therefore, dataset evaluation should report not only overall accuracy or mAP, but also per-class precision, recall, F1-score, and confusion matrices.

Second, geographic and environmental diversity remains insufficient in many existing datasets. Some datasets are collected from a single country, region, crop type, growth stage, or acquisition platform. Although such datasets are useful for model development under specific conditions, they may not fully represent the diversity of real agricultural environments. Differences in soil background, illumination, weed community structure, crop variety, planting density, field management practice, and sensor configuration can lead to domain shifts. A model trained on one dataset may therefore fail when applied to another region or crop system. Cross-season datasets and multi-temporal datasets are valuable because they can partly evaluate robustness under temporal changes, but broader cross-domain benchmark suites are still needed.

Third, annotation cost is another major limitation. Image-level classification labels are relatively easy to obtain, but they cannot provide spatial information for precision spraying or mechanical weeding. Bounding box annotation is more informative and supports object detection, but bounding boxes cannot accurately describe irregular weed boundaries or overlapping plant structures. Pixel-level semantic segmentation and

instance segmentation annotations provide more precise spatial information, but they are time-consuming, labor-intensive, and require trained annotators. In dense crop–weed scenes, severe leaf overlap, ambiguous plant boundaries, and visual similarity between crops and weeds further increase annotation difficulty and may introduce label inconsistency.

These limitations indicate that current benchmark datasets are not sufficient for fully evaluating model robustness in real field applications. Future benchmark construction should move toward cross-domain benchmark suites that include multiple crops, weed species, geographic regions, seasons, growth stages, soil backgrounds, illumination conditions, sensor types, and annotation formats. Such benchmark suites should include standardized training, validation, and test protocols, as well as cross-dataset evaluation settings, for example training on one region or crop type and testing on another. In addition, benchmark reports should include per-class metrics, cross-domain performance, robustness under illumination and occlusion, annotation quality information, and real-time deployment indicators. This would help researchers better understand the limits of current models and develop weed detection systems with stronger field generalization ability.

2.1.2 Data Preprocessing

Datasets are the foundation of deep learning model training. High-quality, large-scale, and diverse public datasets can support the training, validation, and evaluation of farmland weed detection models, and are also an important condition for promoting the development of this field. Before model training, raw field images need to be preprocessed so that their image format, size, pixel distribution, and background information meet the input requirements of the network. Appropriate data preprocessing can improve the consistency of training data, reduce interference from irrelevant background information, and enhance the stability of the model training process.

Common data preprocessing methods include image resizing, denoising, color space conversion, illumination correction, image normalization, and standardization. For example, Raza et al. standardized cotton–weed images to a fixed input size when benchmarking YOLO-family models, ensuring that different detection models were trained and evaluated under a unified preprocessing protocol [32]. Ramesh et al. resized input images to match the network input size, normalized pixel values by scaling them to a fixed range, and applied image transformations such as flipping, rotation, brightness adjustment, and blur to improve model generalization [33]. Shen et al. further processed weed images by adding Gaussian noise and adjusting the color space, which helped enhance the robustness of YOLOv8-MLD weed detection under complex agricultural environments [34]. Image resizing is used to unify the input size, enabling field images from different sources and with different resolutions to adapt to the network structure. Denoising and illumination correction are used to reduce the effects of sensor noise, local shadows, and brightness variations. Deep learning models also require image normalization and standardization, in which pixel values are scaled and centered to meet the requirements of model training, accelerate network convergence, and improve training stability. In agricultural image processing, the Excess Green (ExG) color index is often used to enhance vegetation regions, and background removal methods are combined to filter non-green background pixels such as soil and crop residues. Such processing can reduce interference from redundant background information and improve the efficiency with which the model learns crop and weed features. It should be noted that the main purpose of data preprocessing is not to expand the number of samples, but to improve the quality of raw images and the standardization of input data, thereby providing a stable data basis for subsequent annotation, augmentation, and model training. Representative image augmentation operations and the corresponding bounding-box annotation example are shown in Fig. 4.

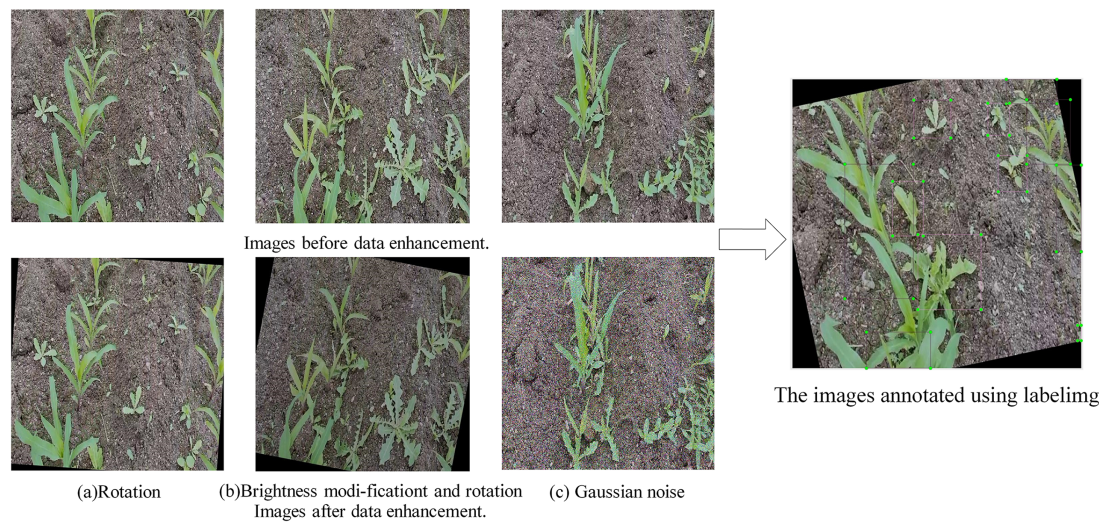


Figure 4: Images after data augmentation. (a) Rotation (b) brightness modification and rotation (c) Gaussian noise. Green rectangles are labeled as target bounding boxes. The left panel presents augmented images generated by rotation, brightness modification with rotation, and Gaussian noise, while the right panel shows bounding box annotation of weed targets using LabelImg [34].

2.1.3 Data Annotation

Data annotation is a key step in the training of artificial intelligence models, and its quality directly affects model recognition accuracy and generalization ability. The more accurate the annotations are and the larger the number of annotated samples is, the easier it is for the model to learn stable target features. The main function of data annotation is to provide clear category or location labels for target objects in images, enabling the model to establish the correspondence between input images and target information during training, and ultimately to achieve automatic recognition of unseen samples.

Common image recognition tasks include image classification, object detection, and image segmentation, and different tasks require different annotation methods. Image classification is mainly used to determine the target category of an entire image, and its annotation process is relatively simple. Classification annotation usually assigns one or more predefined category labels to an image. For example, in a crop–weed classification task, an image of a maize seedling can be labeled as “maize”, while an image of a weed can be labeled as “weed”.

Object detection requires not only identifying the target category but also determining the spatial location of the target in the image. Therefore, bounding box annotation is commonly used. Rectangular bounding box annotation is the most widely applied annotation method in object detection, as it allows target objects such as crops and weeds to be rapidly enclosed in images or video frames, as shown in Fig. 5a. Image segmentation requires assigning category labels to pixels in an image. Common forms include semantic segmentation and instance segmentation. Semantic segmentation assigns a corresponding category label to each pixel, thereby obtaining pixel-level region information for crops, weeds, and the background, as shown in Fig. 5b.

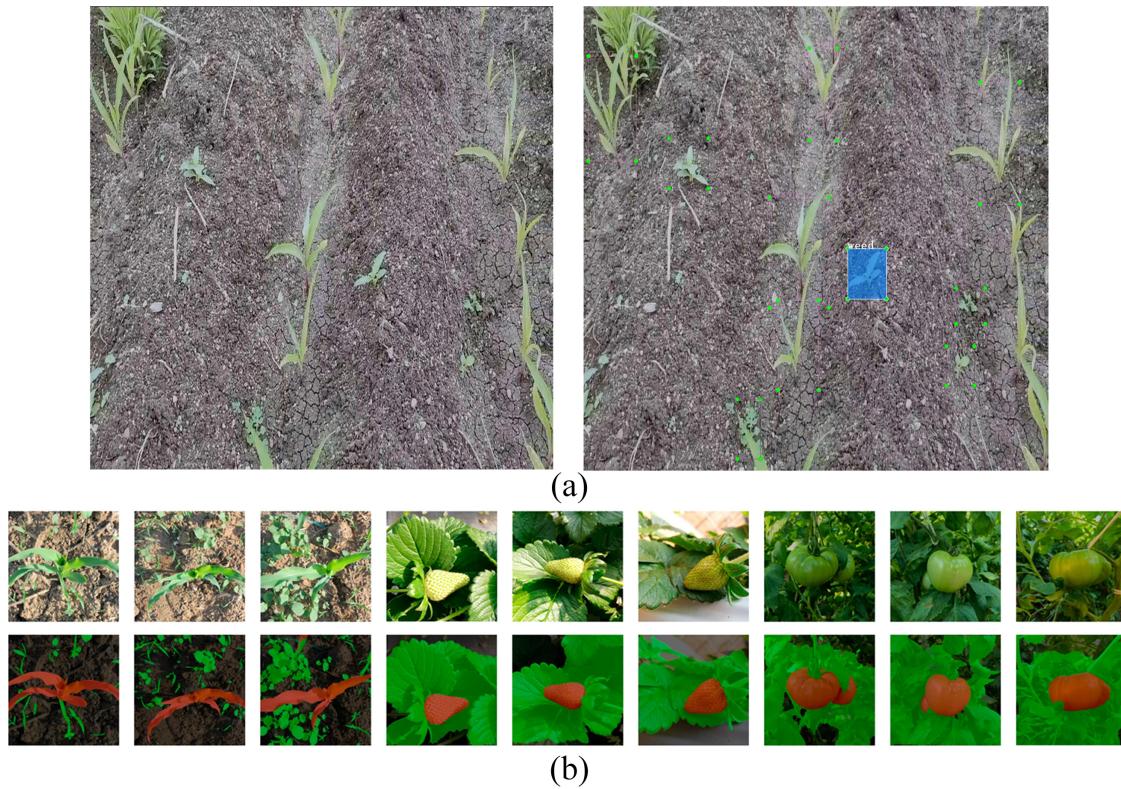


Figure 5: Rectangular bounding box annotation and pixel-level annotation. Panel (a) presents bounding box annotation for weed detection, while panel (b) presents pixel-level annotation for semantic segmentation [35].

2.1.4 Data Augmentation

Data augmentation is usually performed after preprocessing and annotation to expand the training samples. This process requires synchronous transformations of images and their corresponding labels to generate samples with different poses, scales, illumination conditions, and background variations. By doing so, the coverage of training scenarios can be expanded, and the model's dependence on a single data acquisition condition can be reduced. Unlike data preprocessing, the main purpose of data augmentation is not to standardize the input data, but to improve the generalization ability of the model in complex field environments by constructing more diverse samples. Considering the imaging characteristics of field images, commonly used data augmentation methods can be divided into the following three categories.

(1) Basic spatial transformation augmentation is commonly used. This type of method simulates variations in shooting angle, target position, and imaging distance in field environments through spatial geometric transformations. It is one of the most widely used data augmentation strategies in the training of weed detection models. Typical operations include random horizontal or vertical flipping, random rotation, random scaling, random cropping, random translation, and random shearing. Spatial transformation augmentation does not change image pixel values or the intrinsic target features. Instead, it mainly changes the position and scale of targets in the image, thereby improving the model's adaptability to spatial variations and alleviating the influence of target size changes caused by fluctuations in field shooting angle and distance.

(2) Pixel-level and illumination augmentation methods are used to simulate variations in field illumination, imaging noise, and color fluctuation by randomly adjusting image pixel values. Common operations include random adjustment of brightness, contrast, saturation, and hue, as well as Gaussian noise, Gaussian

blur, motion blur, random shadow, and illumination perturbation. Pixel-level and illumination augmentation can expand the variation range of training samples in terms of brightness, color, and imaging quality, improve the model's adaptability to complex field imaging conditions, and reduce the degradation of detection performance caused by drastic illumination changes.

(3) Synthetic data generation and virtual simulation augmentation methods use generative models or virtual simulation techniques to construct samples that are insufficient in real field scenarios. These methods are often used to supplement scarce data under extreme illumination, severe occlusion, and different growth stages. Among them, synthetic data generation based on generative adversarial networks (GANs) has been widely applied. Related studies have generated weed images containing complex illumination, occlusion relationships, and growth-stage differences by learning the feature distribution of real field weed images. The synthetic images are then combined with real images for model training, thereby improving detection accuracy and reducing the cost of data acquisition and manual annotation. Sapkota et al. generated synthetic images using different strategies based on plant instances extracted from UAV images, and further combined them with fake plant instances generated by GANs to train a Mask R-CNN model for the detection and segmentation of morningglory and grass weeds in cotton fields. This approach effectively alleviated the problem of insufficient training data [36]. Xu et al. introduced a weather-driven domain adaptation method based on CycleGAN to generate realistic synthetic images, enhancing the model's adaptability under different weather conditions and providing a reference for weed detection with limited datasets and unpredictable environmental conditions [37].

2.1.5 Dataset Partitioning

Dataset partitioning is a key step in the construction and optimization of deep learning models, and it directly affects model training performance and evaluation results. As shown in Fig. 6, the original dataset is usually divided into training, validation, and test sets according to a certain ratio. The training set is used for model parameter learning, the validation set is used to adjust the network structure, hyperparameters, and model complexity, and the test set is used to evaluate the performance of the final model on data that have not been involved in training.

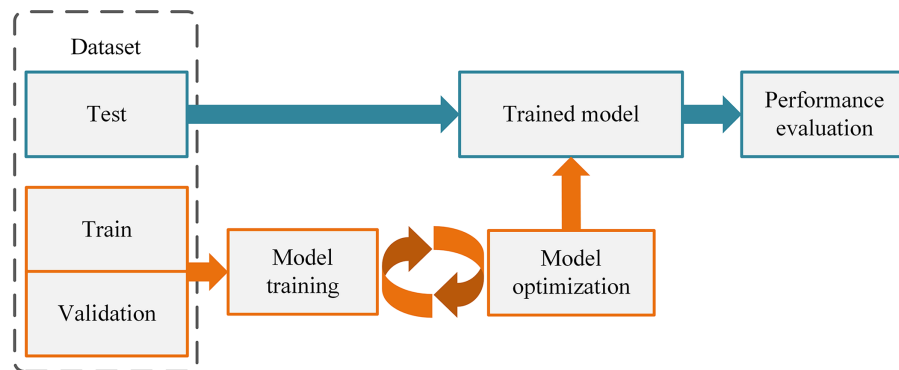


Figure 6: Application of the training, validation, and test sets in model training. The training set is used for parameter learning, the validation set supports model selection and hyperparameter optimization, and the independent test set evaluates final generalization performance.

Several basic principles should be followed during dataset partitioning. First, the training set should be sufficiently large to ensure that the model can learn stable target features and improve its generalization ability. Second, the validation set and test set should remain independent of each other to avoid overestimated

performance caused by duplicated data or overlapping scenarios. Finally, dataset partitioning should be performed randomly whenever possible to reduce the influence of specific acquisition scenarios, crop growth stages, or image backgrounds on training and testing results. When the dataset size is limited, cross-validation can be adopted. In this method, the dataset is divided into several non-overlapping subsets. In each round, one subset is selected as the test set, while the remaining subsets are used as the training and validation sets. The model performance is then evaluated by averaging the results of multiple experiments.

To provide an overview of representative studies discussed in the following sections, recent deep learning studies on weed detection are summarized in . Based on the studies summarized in [Table 2](#), the strengths, weaknesses, and suitable application scenarios of representative deep learning methods are further compared in [Table 3](#).

Table 2: Summary of deep learning techniques and models for weed detection over the past five years.

Authors and Year	Task	Dataset/Reproducibility Availability	Method	Deep Learning Model	Key Findings
Saleh et al., 2023 [38]	Image classification	CottonWeedID15, DeepWeeds	Contrastive learning with class-optimized loss	CNNs, ViTs, GAN, self-supervised learning	Proposed the WeedCLR contrastive learning method, which significantly improved accuracy without requiring large amounts of labeled data and showed particular advantages under class-imbalanced conditions.
Rathore, 2023 [39]	Image classification	DeepWeeds Code: Available Weights: Available	Quantized deep neural network	ResNet-50, InceptionV3	Proposed the QWID model, which compresses deep neural networks through 8-bit quantization, reducing model size and computational cost while maintaining high accuracy, making it suitable for resource-constrained devices.
Rahman et al., 2025 [40]	Image classification	Custom dataset	Incremental learning framework	CNN (DenseNet), ViTs	Proposed ADeepWeeD, an adaptive incremental learning framework that can dynamically update with the emergence of new species while maintaining high accuracy.
Dheeraj and Chand, 2025 [41]	Image classification	Custom dataset	Transfer learning-based deep classification system	DenseNet169, MobileNetV2	Achieved an accuracy of 97.14% using DenseNet169, demonstrating its effectiveness in early-stage weed and crop classification, particularly at the early growth stages of plants.
Rozendo et al., 2024 [42]	Image classification	DeepWeeds, CottonWeedID15	CNNs, Vision Transformers, PVT, and ensemble strategy	CNNs, ViTs, Pyramid Vision Transformers	The combination of CNNs and Vision Transformers significantly improved accuracy, while the ensemble learning strategy enhanced model robustness, achieving a classification accuracy of 99.17%.
Deng et al., 2024 [43]	Object detection	Various datasets Code: Available Weights: Available	Unsupervised domain adaptation methods	YOLOX/YOLOv8 + I ³ Net	The author conducted comparisons of seasonal and cross-seasonal detection using three object detectors—YOLOX, YOLOv8, and DINO—and further integrated the I ³ Net unsupervised domain adaptation method into YOLOX and YOLOv8 to mitigate performance degradation across seasons.

(Continued)

Table 2 (continued)

Authors and Year	Task	Dataset/Reproducibility Availability	Method	Deep Learning Model	Key Findings
Li et al. (2025) [44]	Object detection	CottonWeedDet12	Multi-scale feature fusion with a dynamic detection head	YOLOv8n, PD-YOLO	Proposed PD-YOLO, which significantly improved weed detection accuracy by introducing multi-scale feature fusion and a dynamic detection head, increasing mAP by 1.7%–1.8% and making it suitable for complex agricultural backgrounds.
Zhu et al. (2024) [45]	Object detection	Custom maize seedling–weed dataset	YOLO-based multi-class detection method	YOLOx	A lightweight attention module was integrated into the deep Darknet backbone of YOLOx, reducing channel noise effects in residual computation and improving the F1-score and AP by 0.07% and 1.16%, respectively.
Khan et al. (2025) [46]	Object detection	CottonWeedDet12	Improved YOLOv8 for multi-class weed detection	YOLOv8	Used YOLOv8 to improve multi-class weed detection performance in complex farmland environments, achieving higher accuracy and robustness than previous versions.
Li and Zhang (2024) [47]	Object detection	Farmland crop and weed dataset	Lightweight detection algorithm based on YOLOv7-tiny	YOLOv7-tiny, DC-YOLO	Proposed DC-YOLO, which enhanced feature extraction and semantic context modeling, achieving detection accuracies of 95.7% mAP@0.5 and 81.7% mAP@0.5:0.95, making it particularly suitable for resource-constrained devices.
Deng and Lu (2025) [48]	Object detection	Farmland weed dataset (19,931 images) Code: Available	Semi-supervised learning + YOLOv8 object detection framework	YOLOv8, WeedTeacherSemi-supervised learning	Proposed the WeedTeacher semi-supervised learning framework, which improved weed detection accuracy by using a small amount of labeled data and pseudo-label generation, increasing mAP@50 by 2.6% and mAP@50:95 by 3.1% while reducing annotation costs.
Li et al. (2024) [49]	Object detection	CottonWeedDet3, CottonWeedDet12 Code: Available	Semi-supervised learning framework with a joint network	FCOS, Faster R-CNN, Semi-supervised learning	Through the semi-supervised learning framework, approximately 96% of the fully supervised performance was achieved using only 10% of the labeled data, demonstrating the potential for label-efficient utilization.
Silva et al. (2024) [50]	Image segmentation	Farmland image dataset and UAV image dataset	Deep learning-based semantic segmentation of UAV images	YOLOv8, Mask R-CNN, U-Net	Evaluated the effectiveness of different architectures, including YOLOv8, Mask R-CNN, and U-Net, for weed detection and segmentation, with YOLOv8s achieving the best accuracy, reaching an mAP50 of 97%.
Pai et al. (2025) [51]	Image segmentation	Peanut field crop dataset Code: Available	Edge-aware semantic segmentation with boundary-aware loss	PSPEdgeWeedNet	Proposed PSPEdgeWeedNet, which incorporates an edge-aware loss for weed segmentation in peanut crops, significantly improving segmentation accuracy, particularly under complex background conditions.

(Continued)

Table 2 (continued)

Authors and Year	Task	Dataset/Reproducibility Availability	Method	Deep Learning Model	Key Findings
Sarker et al. (2025) [52]	Image segmentation	16-weed-species dataset Code: Available	Multi-task learning for weed segmentation, growth-stage classification, and height estimation	WeedSense (CNN + Transformer)	Proposed WeedSense, which integrates segmentation, estimation, and classification tasks and uses multi-task learning to improve segmentation accuracy, achieving a high IoU of 89.78% on weed images.
Gupta et al. (2023) [53]	Image segmentation	Custom weed dataset	Multi-class weed semantic segmentation with transfer learning	U-Net, Inception-ResNetV2	Used transfer learning to improve weed segmentation accuracy, with the combination of U-Net and Inception-ResNetV2 achieving an F1-score of 96.78% and enabling effective multi-class weed segmentation.
Garibaldi-Márquez et al. (2025) [54]	Image segmentation	Maize field weed dataset	Deep semantic segmentation combining Transformers and CNNs	Swin-UNet, Segmenter, Reformer, CNN	Combined Transformer-based architectures, such as Swin-UNet and SegFormer, with CNN structures to perform efficient semantic segmentation of weeds in maize fields, outperforming traditional methods.

Table 3: Strengths and weaknesses of deep learning methods summarized in Table 2.

Ref.	Method/Model	Strengths	Weaknesses	Suitable Application Scenarios
[38]	WeedCLR/contrastive learning with class-optimized loss	Reduces dependence on large labeled datasets; improves feature representation under long-tailed and class-imbalanced conditions; useful for weed datasets with uneven class distributions	Mainly designed for classification rather than localization; performance depends on the quality and diversity of pretraining data; does not directly provide weed position or boundary information	Weed species classification, long-tailed weed datasets, label-scarce recognition tasks
[39]	QWID/quantized deep neural network	Uses low-bit quantization to reduce model size and inference time; suitable for resource-constrained platforms such as mobile or embedded devices	Quantization may cause accuracy degradation if calibration is insufficient; mainly addresses classification rather than detection or segmentation; robustness under complex field conditions still requires validation	Embedded weed classification, low-power devices, rapid field inference
[40]	ADeepWeeD/adaptive incremental learning framework	Supports incremental learning and adaptation to newly emerging weed species; reduces the need to retrain the full model from scratch; suitable for dynamic weed communities	Incremental learning may suffer from catastrophic forgetting if old knowledge is not well preserved; requires careful update strategy and representative new-class samples	Long-term weed monitoring, new weed species identification, adaptive weed recognition systems
[41]	Transfer learning with DenseNet169 and MobileNetV2	Transfer learning improves classification performance with limited agricultural datasets; DenseNet169 provides high accuracy, while MobileNetV2 offers better efficiency	Classification-only approach lacks spatial localization; DenseNet169 is heavier for edge deployment; model generalization depends on similarity between source and target domains	Early-stage crop-weed classification, small or moderate datasets, transfer-learning-based recognition

(Continued)

Table 3 (continued)

Ref.	Method/Model	Strengths	Weaknesses	Suitable Application Scenarios
[42]	CNN, ViT, PVT, and ensemble strategy	Compares convolutional and Transformer-based models; ensemble strategy can improve classification robustness; ViT/PVT can capture global image relationships	Transformer and ensemble models usually require more computation and larger datasets; deployment cost is higher than lightweight CNNs	High-accuracy weed classification, comparative model evaluation, complex visual scenes
[43]	Public weed dataset survey and cross-season adaptation	Summarizes public weed datasets and highlights cross-season detection issues; useful for understanding dataset bias, domain shift, and benchmark limitations	Limited dataset scale and seasonal diversity; adaptation gains remain insufficient to fully solve cross-season performance degradation	Dataset selection, benchmark design, cross-season robustness analysis
[44]	PD-YOLO/YOLOv8n with multi-scale feature fusion	Enhances multi-scale feature fusion; improves detection of weeds under complex backgrounds and scale variation; maintains YOLO-style real-time potential	Added modules may increase model complexity; improvements may be dataset-specific; bounding boxes cannot provide precise weed contours	Real-time weed detection, small weed detection, field robot and variable-rate spraying systems
[45]	Improved YOLOX with lightweight attention module	Attention module helps suppress channel noise and enhance useful weed features; YOLOX structure supports efficient detection	Attention modules may increase computational burden; detection performance may still degrade under dense occlusion or severe crop–weed similarity	Multi-class weed detection, real-time field detection, lightweight attention-enhanced models
[46]	YOLOv8 for multi-class weed detection	YOLOv8 provides strong accuracy–speed balance; suitable for multi-class weed detection in cotton fields; easy to deploy compared with two-stage detectors	Bounding-box detection cannot describe irregular boundaries; performance may decline for small, overlapping, or visually similar weeds	Cotton field weed detection, real-time variable-rate spraying, UAV or mobile robot monitoring
[47]	DC-YOLO/improved YOLOv7-tiny	Based on lightweight YOLOv7-tiny; improves key feature extraction and upsampling; suitable for resource-constrained field detection	Lightweight design may sacrifice accuracy in complex scenes; still limited by bounding-box representation; needs validation across crops and regions	Edge-deployable weed detection, corn seedling and weed detection, real-time agricultural robots
[48]	WeedTeacher/semi-supervised YOLOv8 framework	Uses unlabeled data and pseudo-labeling to reduce annotation cost; improves in-domain detection when labeled data are limited	Cross-domain improvement is limited; pseudo-label noise may reduce reliability; performance depends on teacher model quality and domain consistency	Vegetable field weed detection, limited labeled data, semi-supervised field model training
[49]	Semi-supervised learning framework with FCOS and Faster R-CNN	Achieves a large proportion of fully supervised performance using limited labeled data; evaluates both one-stage and two-stage detectors under semi-supervised settings	Pseudo-label quality strongly affects results; two-stage models are slower; limited labels may still be insufficient for rare weed classes	Label-efficient weed detection, benchmark comparison of semi-supervised frameworks
[50]	YOLOv8, Mask R-CNN, and U-Net for UAV weed detection and segmentation	Compares detection and segmentation models on UAV imagery; YOLOv8 supports fast detection, while Mask R-CNN and U-Net provide finer spatial information	UAV images are affected by scale variation, motion blur, and changing illumination; segmentation models require more annotation and computation	UAV-based weed mapping, field-scale monitoring, comparison of detection and segmentation methods
[51]	PSPEdgeWeedNet	Edge-aware design improves boundary segmentation; useful for separating crops and weeds in peanut cultivation; suitable for precise weed region extraction	Requires pixel-level labels; may be crop-specific; computational cost and cross-crop generalization need further validation	Peanut field weed segmentation, boundary-sensitive segmentation, precision spraying

(Continued)

Table 3 (continued)

Ref.	Method/Model	Strengths	Weaknesses	Suitable Application Scenarios
[52]	WeedSense/multi-task segmentation, height estimation, and growth-stage classification	Jointly performs segmentation, height estimation, and growth-stage classification; multi-task learning improves information sharing; useful for comprehensive weed monitoring	Requires multi-task annotations, including masks, height, and growth-stage labels; training is more complex; dataset construction cost is high	Weed phenotyping, growth-stage monitoring, precision weed management beyond detection
[53]	Semantic segmentation with U-Net and Inception-ResNetV2	Pixel-level segmentation provides weed boundaries and coverage information; transfer learning can improve feature extraction for multi-class weed identification	Pixel-level annotation is costly; segmentation models are computationally heavier than detectors; boundary accuracy depends on label quality	Precision spraying, weed coverage estimation, multi-class segmentation
[54]	Transformer-based segmentation architectures, such as Swin-UNet, Segmenter, and Reformer	Transformer-based models capture global context and long-range dependencies; useful for complex scenes with occlusion and visual similarity between crops and weeds	Higher memory consumption and computational cost than CNN-based backbones; less suitable for resource-limited edge devices without optimization	High-accuracy weed segmentation, complex field scenes, offline or high-resource analysis

2.1.6 Standardized Evaluation Metrics for Precision Weed Management

Current studies on deep learning-based weed detection report different evaluation metrics, including accuracy, precision, recall, F1-score, mAP, IoU, mIoU, FPS, inference latency, model size, and FLOPs. Although these metrics are useful for evaluating algorithmic performance, they are not always sufficient for precision weed management. In practical field applications, a weed detection system should not only recognize weeds accurately, but also support correct spraying or physical weeding decisions. To connect algorithmic evaluation with practical precision weeding requirements, a minimal set of evaluation metrics is summarized in [Table 4](#).

Table 4: Minimal evaluation metrics for precision weed detection systems.

Evaluation Dimension	Recommended Metrics	Purpose
Classification/recognition	Per-class precision, recall, F1-score, confusion matrix	Evaluate class-level recognition performance and minority-class errors
Object detection	mAP@0.5, mAP@0.5:0.95, per-class weed recall	Evaluate weed localization and detection accuracy
Segmentation	mIoU, Dice coefficient, pixel-level precision and recall	Evaluate weed boundary, coverage, and crop–weed separation accuracy
Weed management errors	Weed miss rate, crop false-positive rate	Evaluate untreated weeds and potential crop damage
Herbicide application impact	False-positive herbicide use, erroneous sprayed-area ratio, weed coverage estimation error	Evaluate unnecessary spraying and variable-rate prescription reliability
Real-time deployment	FPS, latency, end-to-end delay, model size, parameters, FLOPs, memory usage, hardware platform	Evaluate whether the method can support real-time field operation

(Continued)

Table 4 (continued)

Evaluation Dimension	Recommended Metrics	Purpose
Robustness and generalization	Cross-dataset, cross-season, cross-region, and cross-sensor performance drop	Evaluate field robustness and domain transferability

Task-level vision metrics should be reported according to the model type. Classification studies should report per-class precision, recall, and F1-score rather than only overall accuracy. Object detection studies should report mAP@0.5 and preferably mAP@0.5:0.95, together with per-class recall for weed categories. Segmentation studies should report mIoU, Dice coefficient, and pixel-level precision and recall for crop, weed, and background classes. These metrics are necessary for evaluating whether a model can correctly identify and localize weeds.

Weed-management-oriented error metrics should be included. In precision weeding, false negatives mean that weeds are missed and remain untreated, whereas false positives may lead to unnecessary herbicide application or crop damage. Therefore, weed miss rate, crop false-positive rate, and false-positive herbicide use should be reported whenever possible. False-positive herbicide use can be defined as the proportion of spraying actions or sprayed area triggered by incorrectly detected weeds in crop or weed-free regions. In addition, for segmentation-based systems, the error of estimated weed coverage or weed-to-crop area ratio should be reported, because this directly affects variable-rate herbicide prescription and weed density assessment.

Deployment metrics are required for real-time field operation. FPS should be reported together with single-frame inference latency, input resolution, model size, parameters, FLOPs, memory usage, and the specific hardware platform. More importantly, end-to-end latency should be considered, including image acquisition, preprocessing, model inference, decision-making, communication, and actuator response. A model with high FPS may still be unsuitable for precision spraying if the complete perception–decision–execution delay exceeds the allowable response time of the field machine.

Robustness and generalization metrics should be reported. Weed detection models should be evaluated not only on a random split from the same dataset, but also under cross-domain conditions, such as different crop types, growth stages, seasons, geographic regions, illumination conditions, soil backgrounds, sensor types, and occlusion levels. Performance degradation across domains, such as the drop in mAP, mIoU, F1-score, or weed recall, should be reported to indicate field robustness.

Therefore, the minimum recommended evaluation set for precision weed detection should include: (1) per-class precision, recall, and F1-score; (2) mAP for detection or mIoU/Dice for segmentation; (3) weed miss rate and crop false-positive rate; (4) false-positive herbicide use or erroneous sprayed-area ratio; (5) FPS, latency, model size, FLOPs, and hardware platform; and (6) cross-domain performance degradation. This metric set can better connect model evaluation with the practical requirements of precision herbicide application, mechanical weeding, laser weeding, and agricultural robot deployment.

2.2 Weed Image Classification Techniques

Image classification is one of the earliest and most fundamental applications of deep learning in weed detection. Its task is to extract global features from the entire input image and determine whether weeds are present and, if so, identify their species. Classification models can provide basic data support for assessing weed occurrence density in fields and issuing crop management warnings, making them an important technical approach for large-scale agricultural weed management.

Convolutional neural networks (CNNs) are the mainstream architecture for weed image classification. By combining convolutional layers, pooling layers, and fully connected layers, CNNs enable hierarchical feature learning from low-level textures to high-level semantic representations. Standard CNN backbones, such as ResNet-50, Inception-v3, VGG series, and DenseNet, have been widely used in various public weed image recognition tasks. In deep learning studies, transfer learning is commonly adopted. For example, model weights pretrained on large-scale natural image datasets such as ImageNet are used for initialization, and the model is then fine-tuned on agricultural scene datasets. This strategy effectively reduces the dependence on large-scale annotated data and improves classification robustness [55].

For weed image classification in real field environments, the DeepWeeds dataset is one of the most representative public datasets. DeepWeeds consists of 17,509 field images collected from multiple rangeland environments in Australia. It includes eight important weed species and a negative class containing weed-free samples. The dataset was designed to simulate real challenges in robotic weed control and has laid a foundation for subsequent research. Hu et al. proposed Graph Weeds Net (GWN), which represents field RGB weed images as multi-scale graph structures to model fine-grained features and their relationships from local to global levels. The method achieved a top-1 accuracy of 98.1% on the DeepWeeds dataset [56].

In addition, recent studies have shown that different CNN architectures perform differently in weed classification tasks. For example, in classification tasks based on DeepWeeds or similar real farmland weed image datasets, certain lightweight models, such as MobileNet and EfficientNet, can also achieve high recognition accuracy after transfer learning and are more suitable for deployment in embedded systems. The performance of these models in large-scale and complex environments provides more possibilities for the practical application of weed image classification in agricultural automation. However, their overall performance is still affected by dataset types and acquisition conditions [57]. The basic structure of a CNN-based weed classification model is illustrated in Fig. 7.

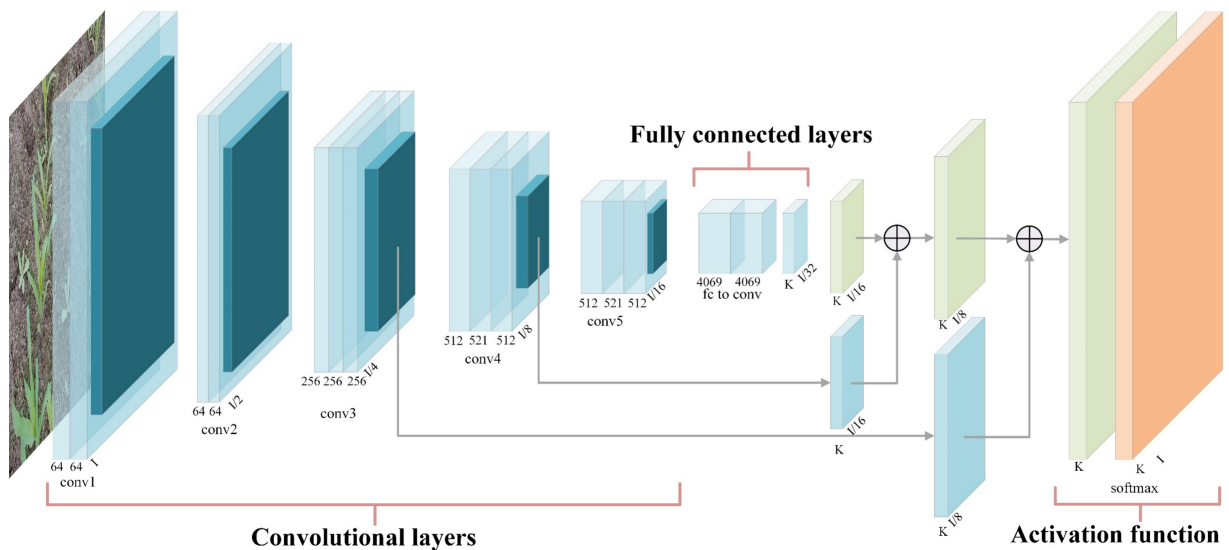


Figure 7: Basic structure of a CNN-based model. The convolutional layers extract low-level and high-level visual features from weed images, such as edges, textures, shapes, and leaf morphology, while the fully connected layers integrate these features for final classification through the softmax activation function.

Current research trends are no longer limited to standard CNN models themselves, but are also moving toward more refined structural improvements, attention mechanism enhancement, and multi-scale feature fusion to better address weed classification tasks characterized by complex scenes and subtle inter-class differences. Classifiers based on these improved methods have achieved performance comparable to or even better than the ResNet-50 baseline in multiple tasks, demonstrating the strong adaptability and optimizability of deep learning in weed image classification.

Therefore, weed image classification is more suitable for coarse-grained weed species identification and field-level monitoring, but it cannot meet the spatial localization requirements of precision herbicide spraying or mechanical weeding. In complex scenes where crops and weeds coexist, classification results may also be strongly affected by background interference and class imbalance.

2.3 Weed Object Detection Techniques

Object detection technology can simultaneously determine weed categories and localize their spatial coordinates. By outputting rectangular bounding boxes containing weed targets, it provides explicit information on weed location, size, and category, thereby offering direct decision-making support for variable-rate herbicide application systems. It is therefore a key technical foundation for precision chemical weeding in field environments. According to detection architecture, weed object detection algorithms can generally be divided into two categories: two-stage detectors and one-stage detectors.

Two-stage object detection algorithms are represented by Faster R-CNN. Their basic workflow usually consists of two stages. First, a Region Proposal Network (RPN) generates candidate regions. Then, RoI feature extraction is performed on these candidate regions, followed by target category classification and bounding box regression. Compared with one-stage detectors, two-stage methods generally provide more stable localization accuracy in scenarios involving complex field backgrounds, crop–weed occlusion, and small-scale target recognition. Therefore, they remained an important technical route in the early stage of farmland weed detection research [58]. The typical workflow of a two-stage object detection framework is illustrated in Fig. 8.

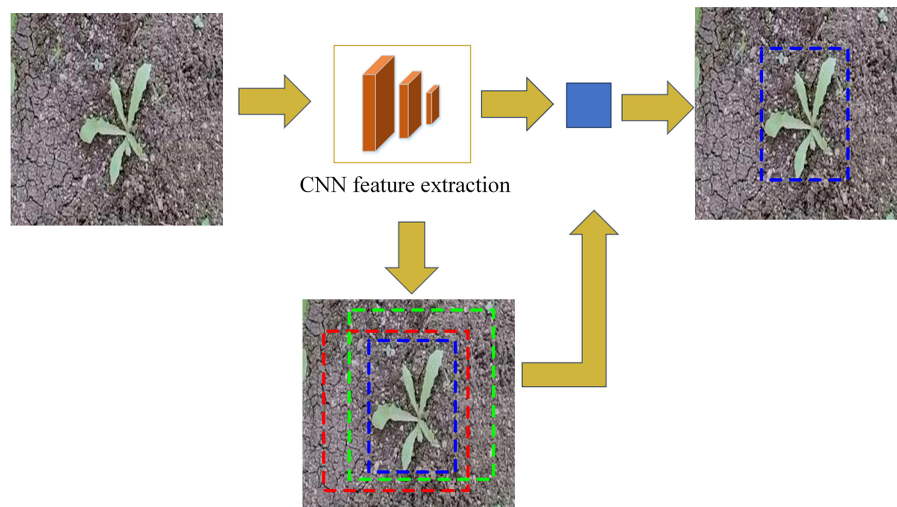


Figure 8: Schematic diagram of the two-stage object detection framework. The figure shows the typical workflow of two-stage detectors, in which candidate regions are first generated and then classified and refined through bounding box regression.

In empirical studies, Mu et al. integrated a feature pyramid network (FPN) into Faster R-CNN and adopted ResNeXt-101 as the backbone network for weed seedling recognition in crop fields. The results showed that the model achieved an accuracy, recall, F1-score, and MIoU of 95.61%, 87.26%, 91.24%, and 93.7%, respectively, outperforming YOLOv3 and SSD in the same study. However, its detection time for a single image was 330 ms, indicating that although this method achieved high accuracy, its real-time performance was limited [59]. In a study on soybean seedlings under more natural field conditions, Zhang et al. introduced the convolutional block attention module (CBAM) into Faster R-CNN and optimized the training process. The final average accuracy reached 99.16%, and the average accuracy in external validation under natural environments was 90.31%. However, the average recognition time was approximately 0.59 s per image, further indicating that two-stage methods are more suitable for low-speed or accuracy-prioritized scenarios, rather than high-speed continuous field operations [60].

Mask R-CNN adds a parallel mask branch to Faster R-CNN, enabling it to generate instance-level pixel masks while outputting object detection boxes. Therefore, it can not only localize weeds, but also further extract morphological information such as weed contours, area, and coverage. This makes it more suitable for applications that require precise spatial boundary constraints, such as mechanical weeding and laser weeding. Osorio et al. compared HOG-SVM, YOLO, and R-CNN methods using multispectral images of lettuce fields. The results showed that the R-CNN method achieved a precision of 98% and an F1-score of 94%, indicating that region-based detection and segmentation frameworks can achieve high recognition accuracy in multispectral crop scenarios [61]. In addition, in potato fields with severe occlusion and complex backgrounds, Goyal et al. compared Mask R-CNN with YOLOv8. Although the overall mAP@0.5 of Mask R-CNN was 79%, slightly lower than the 83.4% achieved by YOLOv8, its precision, recall, and F1-score for the weed class reached 0.83, 0.76, and 0.91, respectively. This suggests that instance segmentation frameworks still have strong weed recognition capability under complex occlusion conditions [62].

Overall, the main advantages of two-stage detection algorithms lie in their high localization accuracy, stronger robustness to complex backgrounds and occlusion, and natural compatibility with instance segmentation tasks. However, their limitations are also clear: the model structure is relatively complex and the inference speed is generally slow. Sharma et al. compared several weed detection models in 2024 and found that the inference time of Faster R-CNN was 63.8 ms, whereas that of YOLOv11 was only 13.5 ms. A 2025 study on inter-row and intra-row weed detection in maize fields also showed that Faster R-CNN achieved an mAP of 0.919, which was lower than the 0.975 achieved by YOLOv11. The authors further reported that Faster R-CNN produced more false detections and missed detections for smaller weeds [63]. Therefore, at the current stage, two-stage methods are more suitable for offline analysis, low-speed autonomous weeding platforms, or robotic systems that prioritize detection accuracy. For high-speed real-time field operations, they usually need to be combined with more lightweight one-stage detectors or edge deployment strategies.

One-stage detection algorithms are represented by the YOLO series and SSD. These methods omit the candidate region generation step used in two-stage detectors and integrate feature extraction, category classification, and bounding box regression into a unified network. As a result, they usually achieve higher inference efficiency and are more suitable for deployment on real-time field operation equipment. The typical structure of a one-stage object detection framework is illustrated in Fig. 9.

In weed detection research, the YOLO series is currently one of the most widely used one-stage detection frameworks. From YOLOv3 and YOLOv5 to YOLOv7 and YOLOv8, research has mainly focused on multi-scale feature fusion, improvement of small-object detection capability, incorporation of attention mechanisms, and optimization for real-time deployment on edge devices. Taking sugar beet fields as an example, Gao et al. constructed a convolutional neural network based on tiny YOLOv3 and trained it using both synthetic images and real field images. The model improved the mAP from 0.751 to 0.829 and achieved

a high inference speed of 6.48 ms per image on a high-performance GPU, demonstrating its potential for deployment on embedded mobile platforms [64].



Figure 9: Schematic diagram of the one-stage object detection framework. The figure illustrates how one-stage detectors perform feature extraction, classification, and bounding box regression within a unified network.

To address challenges such as small targets, occlusion, and scale variation in field environments, Deng et al. proposed HAD-YOLO by further improving YOLOv5 from three aspects: feature extraction, feature fusion, and detection head design. A P2 feature layer was also added to enhance the model's ability to recognize multi-scale and small targets. The model achieved an mAP of 96.2% and a detection speed of 30.6 fps on field test data. Although good detection performance was obtained in real occlusion scenarios, the confidence scores in occluded regions remained relatively low, indicating that occlusion is still one of the key challenges in field weed detection [65].

Furthermore, YOLOv7, YOLOv8, and their improved variants have become important benchmarks for current field weed detection. Li et al. proposed YOLOv7-FWeed for weed detection in soybean fields, achieving a precision of 0.9496, recall of 0.9125, F1-score of 0.9307, and mAP of 0.9662, which demonstrates its high accuracy and robustness in complex field environments [66]. Another study by Liu et al. improved YOLOv8s and proposed YOLOv8-MBM. By introducing the lightweight vision transformer MobileViTv3 and integrating input features, local CNN features, and global ViT features, the improved model achieved an accuracy of 92.7% in wheat field weed detection. Compared with the original YOLOv8s, its precision, recall, mAP1, and mAP2 increased by 10.6%, 8.9%, 9.7%, and 9.3%, respectively [67]. An example of YOLOv8-MBM detection results for wheat field weeds is shown in Fig. 10.

In contrast, SSD is also a one-stage detector and has relatively fast inference speed. However, in farmland weed detection tasks, especially in scenarios with dense weeds, large scale variations, and a high proportion of small targets, its detection performance is often inferior to that of specifically improved YOLO-series models. For example, in the comparative experiment conducted by Wang et al. on crop and weed detection in sugar beet fields, SSD achieved an mAP@0.5 of 0.814, which was lower than the 0.900 achieved by the improved TIA-YOLOv5. The study also clearly indicated that SSD performed weakly in small-object detection [68]. Similarly, Chen et al. reported in a sesame field weed detection study that the improved YOLO-sesame outperformed SSD and several mainstream detectors, achieving an mAP of 96.16% and a detection speed of 36.8 fps [69]. In addition, Wu et al. further pointed out that YOLO-based structural improvements can significantly improve the AP of small weeds and the overall mAP in field small-weed detection tasks [70]. Therefore, in application scenarios that require both real-time field performance and small-object recognition capability, improved YOLO-series models generally have greater advantages than SSD.

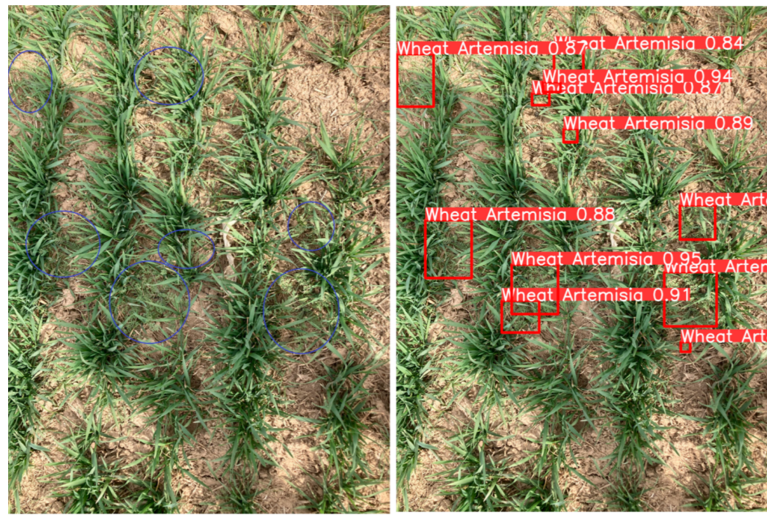


Figure 10: Detection results of the YOLOv8-MBM model proposed by Liu et al. for wheat field weeds [67]. The left panel shows the original wheat field image, and the right panel shows the detection result with red bounding boxes and confidence scores.

At the current stage, research on weed object detection mainly focuses on small and occluded target recognition, false detection suppression, and lightweight deployment on edge devices. Multi-scale feature fusion, adaptive anchor boxes, small-object enhanced training, and crop-row spatial priors are commonly used methods for improving detection performance in complex field environments. In general, object detection methods provide a practical compromise between localization accuracy and computational efficiency. YOLO-based detectors are more suitable for real-time field robots and UAV platforms, whereas two-stage detectors are more suitable for accuracy-prioritized or low-speed applications. However, bounding-box-based detection remains limited when weeds are small, densely distributed, or severely overlapped with crops.

2.4 Weed Image Segmentation Techniques

Semantic segmentation performs category classification for each pixel in an image, thereby achieving pixel-level separation of crops, weeds, and background. It can therefore provide more refined spatial localization information than object detection and is suitable for scenarios that require precise boundary constraints, such as intra-row precision weeding, variable-rate spraying, and mechanical or laser weeding. Instance segmentation further distinguishes individual weed plants on the basis of semantic segmentation, enabling independent recognition, localization, and contour extraction of single weeds. Therefore, it is more suitable for high-precision applications such as single-plant targeted treatment. Compared with detection methods that only output bounding boxes, segmentation methods have higher application value in complex field backgrounds and in scenarios involving crop–weed adhesion and occlusion. However, they also face the problems of higher model complexity and greater real-time deployment cost [71].

Fully convolutional networks (FCNs) are an important foundational architecture for modern semantic segmentation. Their principle is to replace the fully connected layers in traditional classification networks with convolutional operations, thereby enabling end-to-end pixel-level prediction. In recent years, FCNs have more often been used as baseline or comparative models in farmland weed recognition studies. Researchers usually build upon FCNs by introducing encoder–decoder structures, multi-scale feature extraction, and attention mechanisms to improve segmentation performance for small weeds, boundary details,

and occluded regions [72]. Therefore, rather than emphasizing the absolute performance of FCNs in a single scenario, it is more appropriate to position FCNs as an early foundational framework for agricultural visual segmentation, while summarizing current mainstream methods as encoder–decoder paradigms represented by U-Net, DeepLab series, and their improved variants [73].

In practical applications, U-Net and its improved variants have become one of the most commonly used technical routes for semantic segmentation of farmland weeds. Through its symmetric encoder–decoder structure and skip connections, U-Net integrates shallow detailed information with deep semantic information, allowing it to balance target localization and boundary restoration effectively. As a result, it shows stable performance in agricultural images with limited samples, complex backgrounds, and obvious target scale variations. Joy et al. compared several segmentation models using UAV images of sugar beet fields and found that U-Net with ResNet-34 as the backbone achieved the best performance in semantic segmentation. The IoU values for sugar beet and weeds reached 0.85 and 0.72, respectively, indicating that U-Net remains an important benchmark model for field weed segmentation [74]. In complex field environments at the maize seedling stage, Cui et al. proposed RDS_UNet based on U-Net by introducing ResNeXt-50, deformable convolution, and the scSE attention module. Compared with the original U-Net, RDS_UNet improved MIOU by 5.91% and also enhanced the segmentation of small weeds and regions with blurred boundaries [75]. These studies show that although standard U-Net has a simple structure and is easy to train, there is still room for improvement in complex occlusion, small-scale target segmentation, and boundary refinement. Therefore, enhanced U-Net variants are more practical in agricultural scenarios.

To address problems such as large scale variation, slender leaves, and severe mutual occlusion among weeds in field environments, recent improved models have mainly focused on multi-scale modeling and attention mechanisms. Liao et al. proposed SC-Net for weed segmentation at the seedling stage in paddy fields. Based on U-Net, SC-Net incorporates a multi-scale strip convolution structure to specifically enhance the representation of slender targets and scale-varying objects. This indicates that multi-scale structures are helpful for improving segmentation performance when rice seedlings and weeds have similar textures but obvious size differences [76]. Li et al. further proposed an improved U-Net that integrates a MaxViT encoder and the CBAM attention module for sugar beet and weed segmentation. This method enhances both global and local feature modeling as well as boundary detail recovery, thereby improving segmentation accuracy and robustness in complex field environments [77]. Similarly, Gao et al. proposed EDM-UNet for UAV-based weed segmentation in soybean fields, improving weed boundary extraction through edge enhancement and attention guidance. This reflects that “U-Net plus attention modules” has become an important optimization direction in current agricultural weed segmentation [78]. Therefore, multi-scale convolution, atrous convolution, Transformer encoders, and attention mechanisms are the main approaches used in recent years to improve the field weed segmentation performance of U-Net-based models.

Imran Moazzam proposed a new method to improve pixel-level inter-class classification accuracy between crop and weed pixels. This technique applies semantic segmentation in two stages. In the first stage, a binary pixel-level classifier based on U-Net with a Vanilla Mini CNN backbone is used to segment vegetation from the background. In the second stage, a three-class pixel-level classifier is designed using U-Net with VGG16 as the backbone to classify background, weeds, and tobacco. The output of the first stage is used as the input of the second stage. This method increased the IoU of tobacco crops from 0.67 to 0.85 and the IoU of weeds from 0.76 to 0.91 [79]. The two-stage semantic segmentation results reported by Moazzam et al. are shown in Fig. 11.

In instance segmentation, research has gradually expanded toward encoder–decoder-based instance segmentation methods that are more suitable for edge deployment and high-density weed scenarios. Xu et al. proposed a novel method for soybean fields that combines visible color indices with an instance

segmentation approach based on an encoder–decoder architecture. In this encoder–decoder framework, the integration of ResNet101_v and DSASPP helps enhance the extraction of rich multi-scale semantic information, thereby improving the accuracy of weed patch boundary segmentation. The method achieved a weed segmentation accuracy of 0.905, an IoU score of 0.959, an overall aAcc of 0.978, an mIoU of 0.939, and an mAcc of 0.972. These results indicate that the method can effectively address the challenge of accurate weed detection and segmentation in densely distributed soybean–weed field environments [80]. Representative weed segmentation maps produced by the model of Xu et al. are shown in Fig. 12.

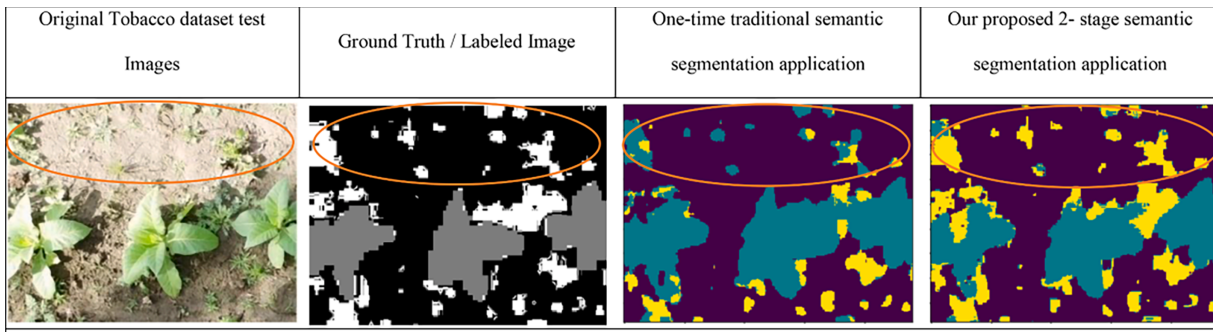


Figure 11: Two-stage semantic segmentation results for tobacco and weed pixels in aerial images reported by Moazzam et al. [79]. From left to right, the panels show the original image, the ground-truth label, the result of a one-time traditional semantic segmentation method, and the result of the proposed two-stage semantic segmentation method. The comparison demonstrates that the two-stage method provides more accurate crop–weed–background separation under complex field conditions.



Figure 12: Examples of weed segmentation maps produced by the proposed model from Xu et al. [80]. From left to right, the panels show the original field image, the manually labeled ground truth, and the model prediction.

In addition to the U-Net series, SegNet and DeepLab series are also commonly used models for semantic segmentation of farmland weeds. SegNet adopts an encoder–decoder structure and uses the pooling indices stored during the encoding stage to guide upsampling in the decoding stage. Therefore, it achieves a certain balance between segmentation accuracy and computational efficiency. Asad and Bais compared SegNet, U-Net, and other models using high-resolution images of canola fields. The results showed that SegNet with ResNet-50 as the backbone performed best, achieving a mean intersection over union (mIoU) of 0.8288 and a frequency-weighted intersection over union (FWIoU) of 0.9869. This indicates that SegNet has good application potential in binary segmentation tasks involving crops and weeds in field environments [81].

The DeepLab series extracts contextual features at multiple scales through atrous convolution and the atrous spatial pyramid pooling (ASPP) module. This design expands the receptive field and enhances the representation of targets at different scales. Therefore, DeepLab models have clear advantages in field images

where weeds and crops differ greatly in scale and the background is complex. Hashemi-Beni et al. compared several semantic segmentation models on two public UAS weed datasets and found that DeepLabV3+ achieved an overall classification accuracy of 84.3% on the CWFID dataset, which was higher than the 77.9% achieved by U-Net. Its pixel classification accuracy for the weed class reached 90.5%, demonstrating its advantage in extracting weed regions [82]. In a study more closely related to practical maize field applications, Fu et al. proposed DSC-DeepLabv3+. By replacing the heavy backbone network with MobileNetV2 and introducing lightweight improvements to the ASPP and decoder modules, the number of model parameters was reduced from 54.714M to 2.89M. The model achieved an mIoU of 85.57% and an inference speed of 42.89 FPS in maize field weed segmentation, indicating that the DeepLab series can achieve a good balance between accuracy and real-time performance [83]. Owing to its multi-scale contextual modeling capability, the DeepLab series usually shows strong robustness in multi-class field weed segmentation tasks and has become an important technical route beyond U-Net.

In instance segmentation, Mask R-CNN remains one of the most widely used classical models in agricultural scenarios. This model adds a mask branch to Faster R-CNN and can simultaneously output target categories, bounding boxes, and pixel-level instance masks. Therefore, it is particularly suitable for precision weeding scenarios that require the contour and spatial position of individual weeds to be distinguished. Quan et al. proposed an instance segmentation-based method for weed phenotyping in complex maize field environments. By performing pixel-level segmentation of weeds, the method further extracted key phenotypic information such as leaf age and plant center. The results showed that instance segmentation can not only achieve precise weed localization and recognition, but also provide more biologically meaningful decision support for targeted herbicide application and field management [84]. In a potato field study based on UAV orthophotos, Lu et al. compared the instance segmentation performance of YOLOv8 and Mask R-CNN. The results showed that Mask R-CNN achieved an mAP50 of 0.920, slightly higher than the 0.902 achieved by YOLOv8, reflecting the fine localization advantage of two-stage instance segmentation models in complex field images [85].

In recent years, Transformer-based segmentation models have become a new research hotspot in weed segmentation. These methods use self-attention mechanisms to model long-range dependencies and can more fully exploit global contextual information among leaves and plants. Therefore, they show considerable potential in field environments with severe occlusion, target overlap, and similar inter-class textures. Based on real farmland UAV RGB images, Zhao et al. proposed a 1DTransformer model for weed semantic segmentation. The model performed pixel-level segmentation of weeds, crops, and background under complex field conditions across 15 cereal fields over three consecutive years. The results showed that the model achieved an overall accuracy of 98.694% on the validation set, and its weed coverage estimation showed good agreement with ground reference data, with an R^2 of 0.598. This indicates that Transformer architectures can effectively capture global features and cross-scale information in real field environments, thereby improving the robustness of weed segmentation and mapping. Such models can provide reliable support for precision spraying and variable-rate weeding [86].

Although Transformer-based segmentation models show strong potential in weed segmentation, their computational cost and deployment suitability should be evaluated carefully. Compared with CNN-based backbones such as U-Net, DeepLab, SegNet, and MobileNet-based segmentation networks, Transformer-heavy models usually require higher memory consumption and more computational resources because self-attention operations need to model relationships among image tokens or feature patches. This is beneficial for capturing long-range dependencies and global contextual information, especially in scenes with severe occlusion, dense leaf overlap, irregular weed distribution, and high visual similarity between

crops and weeds. However, this advantage is often obtained at the cost of higher inference latency and more difficult deployment on edge devices.

Different Transformer-based segmentation structures also show different deployment characteristics. Pure Transformer models or Transformer-heavy architectures, such as Segmenter-like models, are more suitable for offline analysis or high-performance computing platforms when segmentation accuracy and global context modeling are prioritized. Swin-UNet reduces the computational burden of global self-attention by using window-based attention, making it more efficient than standard global-attention Transformers, but it may still be heavier than lightweight CNN-based segmentation networks. Hybrid CNN–Transformer models provide a more practical compromise: CNN modules can efficiently extract local texture, edge, and leaf-shape features, while Transformer modules can enhance global context modeling. Therefore, hybrid models may be more suitable than pure Transformer models for field weed segmentation when both accuracy and deployability are required.

For real-time edge applications, CNN-based backbones still have clear advantages in computational efficiency, hardware compatibility, and deployment maturity. Lightweight CNN-based segmentation models are generally more suitable for UAVs, agricultural robots, and tractor-mounted edge platforms with limited computing resources. In contrast, Transformer-heavy segmentation models are more suitable when hardware resources are sufficient or when the application requires robust segmentation under complex occlusion and long-range spatial dependencies. Therefore, future studies should not report only segmentation accuracy, mIoU, or IoU, but should also provide parameters, FLOPs, FPS, memory usage, input resolution, and edge-platform latency. This would allow a fairer evaluation of whether Transformer-based segmentation models are practically deployable in real agricultural scenarios.

Overall, semantic segmentation and instance segmentation methods can provide more refined spatial location and boundary information than object detection. Therefore, they are more suitable for refined field operations that require accurate target contours and treatment regions, such as intra-row weed recognition, variable-rate precision spraying, mechanical weeding, and laser weeding [87]. However, their practical performance still depends strongly on crop type, imaging platform, image resolution, occlusion degree, dataset scale, annotation quality, model complexity, and available hardware resources. Compared with CNN-based segmentation backbones, Transformer-heavy models have stronger ability to capture long-range dependencies and global contextual information, which is beneficial for complex scenes with dense leaf overlap, severe occlusion, and high crop–weed visual similarity. Nevertheless, these advantages are usually accompanied by higher memory consumption, computational cost, and inference latency. Therefore, CNN-based segmentation models remain more suitable for real-time deployment on resource-limited edge devices, whereas Transformer-heavy models are more appropriate for high-accuracy segmentation tasks when sufficient computing resources are available. Hybrid CNN–Transformer models may provide a compromise between local feature extraction, global context modeling, and deployment feasibility. Future weed segmentation research should balance segmentation accuracy, annotation cost, computational efficiency, field robustness, and edge deployment feasibility rather than pursuing model complexity alone.

2.5 Semi-Supervised, Unsupervised, and Weakly Supervised Learning

The annotation cost of weed images in agriculture is relatively high. This is particularly evident in pixel-level segmentation and instance segmentation tasks, where crops, weeds, and backgrounds must be distinguished at a fine-grained level. Moreover, annotation quality depends on an understanding of crop growth stages, weed morphology, and field scenarios. Therefore, the construction of high-quality datasets remains an important bottleneck for the practical application of deep learning in precision weeding. Recent reviews generally suggest that dataset scarcity, high annotation cost, and insufficient cross-scenario generalization

are among the key factors limiting the field deployment of agricultural vision models. Consequently, label-efficient learning methods, including semi-supervised learning, weakly supervised learning, self-supervised or unsupervised pretraining, transfer learning, and domain adaptation, have become important research directions in agricultural intelligent perception [55,88].

2.5.1 Semi-Supervised Learning

Semi-supervised learning is one of the most mature and widely used low-annotation-cost solutions in current agricultural vision. Its basic idea is to train a model using a small amount of labeled data together with a large amount of unlabeled data, thereby reducing dependence on manual annotation. In weed segmentation, Nong et al. proposed SemiWeedNet, which incorporates both labeled and unlabeled UAV images into a unified semantic segmentation framework and combines selective kernel attention with consistency constraints. On the WeedMap dataset, the model achieved segmentation performance close to that obtained using fully labeled training when only 20% of the images were labeled and unlabeled data were introduced. This demonstrates that semi-supervised learning can substantially alleviate the burden of pixel-level annotation [89]. In multi-class weed object detection, Li et al. evaluated Faster R-CNN and FCOS based on a teacher–student pseudo-labeling framework. On the CottonWeedDet3 dataset, when only 20% of the data were labeled, the mAP@ [0.5:0.95] of semi-supervised Faster R-CNN increased from 42.65 under fully supervised training to 49.20, while that of semi-supervised FCOS increased from 52.42 to 55.79. These results indicate that semi-supervised learning can not only reduce annotation requirements, but may also improve detection robustness under certain settings [49]. In addition, Liu et al. compared the Π -model, Mean Teacher, and FixMatch for turfgrass weed detection. The results showed that FixMatch still achieved validation and test accuracies above 0.953 and F1-scores above 0.951 when only 50 labeled images were available for each class. This further demonstrates the practical potential of mainstream semi-supervised algorithms in agricultural weed recognition [90].

2.5.2 Unsupervised Learning

In unsupervised or self-supervised learning, recent studies have mainly followed the strategy of first learning visual representations from unlabeled agricultural images and then fine-tuning downstream weed classification or detection models with a small amount of labeled data, rather than directly performing reliable automatic weed classification. Mishra et al. proposed a self-supervised weed recognition method, PSPUSegNet, which learns visual representations of crop and weed leaves under complex illumination conditions using datasets such as DeepWeeds, CWFID, and MMIDDWF. The method achieved a classification accuracy of 98.96% on the DeepWeed dataset, indicating that self-supervised representation learning can improve the robustness and performance of weed recognition in low-annotation scenarios [91]. Self-supervised or unsupervised pretraining can reduce the dependence of models on large amounts of manually annotated data by learning domain-relevant feature representations. However, its effectiveness usually still needs to be verified through downstream fine-tuning, and its practical validation in weed detection and segmentation remains less extensive than that of semi-supervised learning.

2.5.3 Weakly Supervised Learning

Weakly supervised learning reduces manual annotation costs by lowering the granularity of annotations. For example, point annotations, scribble annotations, bounding boxes, or rule-generated labels can be used instead of complete pixel-level masks. In maize seedling and weed segmentation, Zhao et al. proposed SL-Net, a weakly supervised semantic segmentation model based on scribble annotations. By using a small number of scrawl labels instead of complete pixel-level annotations, the generated pseudo-labels achieved

an IoU of 83.32% with the ground-truth annotations. The final model achieved an mIoU of 87.30% and a mean accuracy of 94.06% on the test set, outperforming the weakly supervised and some fully supervised baselines compared in the same study [35]. In perennial weed detection in barley fields, Thomas et al. designed rule-generated weak labels using NDVI and spatial structure, and then trained U-Net with these labels. The model achieved an F1-score of 0.82–0.83 on a manually annotated test set, indicating that rule-based weak label generation is also a feasible low-cost strategy in agricultural scenarios [92]. Current weakly supervised research on weeds mainly focuses on scribble annotations, rule-generated labels, and pseudo-labeling mechanisms. These methods can obtain segmentation performance with practical application value while substantially reducing the burden of pixel-level annotation.

In addition to the above methods, transfer learning remains one of the most commonly used low-sample solutions in agricultural weed recognition. Chen et al. evaluated deep transfer learning for the recognition of 15 common weed species in cotton fields. They constructed a dataset containing 5187 images and compared 35 pretrained models. The results showed that most models achieved high recognition performance through transfer learning. ResNet101 achieved an F1-score of 99.1%, and 14 models obtained F1-scores above 98.0%. These results indicate that fine-tuning pretrained weights remains one of the most effective and stable basic strategies when annotated samples are limited [93]. Although transfer learning has been widely adopted, models are still susceptible to domain shifts across different fields, imaging conditions, and cropping systems, which can lead to performance degradation in cross-scenario applications.

To address domain shift, unsupervised domain adaptation is gradually becoming an important direction for improving the cross-field generalization ability of weed detection models. Huang et al. proposed an unsupervised domain adaptation framework based on teacher–student self-training. The framework jointly trains a segmentation model using labeled source-domain data and pseudo-labeled target-domain data, and incorporates local dynamic weighting and class dynamic weighting to alleviate pseudo-label noise and class imbalance. The method showed good generalization ability on datasets with differences in robotic platforms and crop growth stages, indicating that unsupervised domain adaptation can improve the cross-domain performance of weed segmentation without relying on manual annotations in the target domain [94]. By using feature alignment and pseudo-label self-training, unsupervised domain adaptation can alleviate domain shifts caused by different fields, devices, and growth stages. It is a potentially effective approach for improving field generalization, although large-scale field validation in weed recognition remains limited.

2.6 Critical Comparison of Deep Learning Methods for Different Agricultural Scenarios

Although deep learning-based methods have significantly improved the accuracy of weed detection, different technical routes show distinct strengths and limitations across agricultural scenarios. Therefore, model selection should not be based only on a single accuracy metric. Instead, it should comprehensively consider annotation cost, robustness, real-time performance, deployment platform, and the requirements of downstream weeding equipment.

Image classification methods are relatively simple and computationally efficient, making them suitable for image-level weed species recognition, field monitoring, and early warning tasks. However, classification models can only provide global category information and cannot localize individual weeds or estimate their spatial coverage. Therefore, they are difficult to apply directly to precision spraying or mechanical weeding systems that require target-level spatial information. In addition, when crops and weeds coexist in the same image or when the background is complex, the robustness of classification models is limited.

Object detection methods, especially YOLO-based models, provide a better balance between detection accuracy and inference speed. By outputting weed categories and bounding boxes, they can directly support variable-rate herbicide spraying and mobile robot-based weed monitoring. However, bounding boxes only

provide approximate spatial information and cannot accurately describe irregular weed boundaries. In addition, small weeds, dense weed clusters, and severe crop–weed occlusion remain challenging. Two-stage detectors usually have stronger localization ability in complex scenes, but their slower inference speed limits their application in high-speed field operations. In contrast, one-stage detectors are more suitable for real-time deployment, but they may sacrifice localization accuracy under dense occlusion or small-target conditions. Therefore, among the reviewed method categories, lightweight one-stage YOLO-based detectors currently provide the most practical balance between accuracy and speed for real-time weed detection, especially in variable-rate spraying, UAV monitoring, and mobile field robot applications. However, this conclusion is scenario-dependent: segmentation or instance segmentation methods are still preferable when accurate weed boundaries or individual plant contours are required.

Semantic segmentation and instance segmentation methods provide finer spatial information than classification and object detection. Semantic segmentation is suitable for weed coverage estimation, weed distribution mapping, and precision spraying, while instance segmentation can distinguish individual weed plants and is more appropriate for mechanical weeding, laser weeding, and weed phenotyping. However, these methods require more expensive pixel-level or instance-level annotations and usually have higher computational costs. Their performance is also strongly affected by image resolution, illumination variation, leaf overlap, crop growth stage, and weed density.

Semi-supervised, weakly supervised, and unsupervised learning methods have great potential for reducing annotation costs in weed detection. These methods are particularly valuable when only a small number of labeled samples are available or when new weed species and crop scenarios are introduced. However, their reliability largely depends on the quality of pseudo-labels, the consistency of data distributions, and the robustness of teacher models or pretrained models. If pseudo-label noise accumulates during training, model performance may degrade, especially in complex field environments with severe occlusion, class imbalance, and domain shift.

In practical agricultural applications, no single approach can be regarded as universally optimal. Classification methods are suitable for coarse monitoring tasks, object detection methods are appropriate for real-time localization and variable-rate spraying, segmentation methods are preferable when precise weed boundaries or coverage estimation are required, and lightweight models are necessary for embedded deployment. Therefore, future weed detection systems should be designed according to specific operational scenarios, including crop type, weed density, imaging platform, field operating speed, hardware resources, and the precision requirements of the weeding actuator. A critical comparison of these deep learning approaches for precision weed detection is summarized in [Table 5](#).

Table 5: Critical comparison of deep learning approaches for precision weed detection.

Approach	Strengths	Limitations	Robustness	Suitable Scenarios
Image classification	Simple annotation, low computational cost, suitable for image-level weed identification	Cannot provide precise weed location or boundary	Sensitive to background complexity and mixed crop–weed images	Weed species recognition, field-level monitoring, early warning
Object detection	Provides weed category and location, good balance between accuracy and speed	Bounding boxes cannot describe irregular weed contours; small or overlapping weeds may be missed	YOLO variants are robust for real-time use, but cross-season generalization remains limited	Variable-rate spraying, UAV monitoring, mobile field robots

(Continued)

Table 5 (continued)

Approach	Strengths	Limitations	Robustness	Suitable Scenarios
Semantic segmentation	Provides pixel-level weed regions and coverage estimation	Requires expensive pixel-level annotation; high computational cost	More robust for dense weeds and irregular shapes, but affected by illumination and occlusion	Weed coverage mapping, intra-row weed detection, precision spraying
Instance segmentation	Distinguishes individual weed plants and provides contours	More complex model structure; slower inference; difficult annotation	Strong for overlapping plants, but deployment is challenging	Mechanical weeding, laser weeding, phenotyping
Semi/weakly supervised learning	Reduces annotation cost and improves data utilization	Pseudo-label noise and unstable training may reduce reliability	Useful under limited labeled data, but less stable under large domain shifts	Small-sample weed datasets, new crop/weed scenarios
Lightweight and edge models	Better real-time performance and lower hardware cost	Accuracy degradation after pruning/quantization/compression	Depends strongly on hardware, input resolution, and inference framework	Embedded agricultural robots, UAVs, real-time field operations

3 Edge Deployment and Lightweight Optimization

The engineering application of Agriculture 4.0 requires weed detection models to achieve high-accuracy and low-latency inference on field edge computing platforms. Field robots, UAVs, and portable spraying systems often rely on resource-constrained devices such as Jetson and Raspberry Pi. Therefore, weed detection models must balance detection accuracy and inference speed under limited computing power, storage capacity, and power consumption. This issue has become a key factor in the industrial application of precision weeding technologies. The perception–decision–execution workflow of a practical weed control system is shown in [Fig. 13](#).

To meet the demand for real-time detection on edge devices, existing studies mainly focus on lightweight structural design, model compression, and inference framework acceleration. Lightweight structures reduce the number of network parameters and computational complexity at the architectural level, while model compression methods such as pruning, quantization, and knowledge distillation further reduce storage cost and inference latency. However, for farmland weed detection, compression should not be evaluated only by model size, FLOPs, or FPS. Field images are affected by variable illumination, shadows, crop–weed occlusion, motion blur, and sensor noise. Excessive compression may weaken the representation of small weeds, boundary details, low-contrast targets, and occluded plant structures, thereby reducing robustness in real field environments. Therefore, the suitability of pruning, quantization, and knowledge distillation should be assessed according to the target task, deployment platform, and environmental variability. Inference frameworks such as TensorRT can improve execution efficiency through computational graph optimization and hardware adaptation, but their practical value should also be evaluated together with accuracy degradation, field robustness, and end-to-end perception–decision–execution latency.

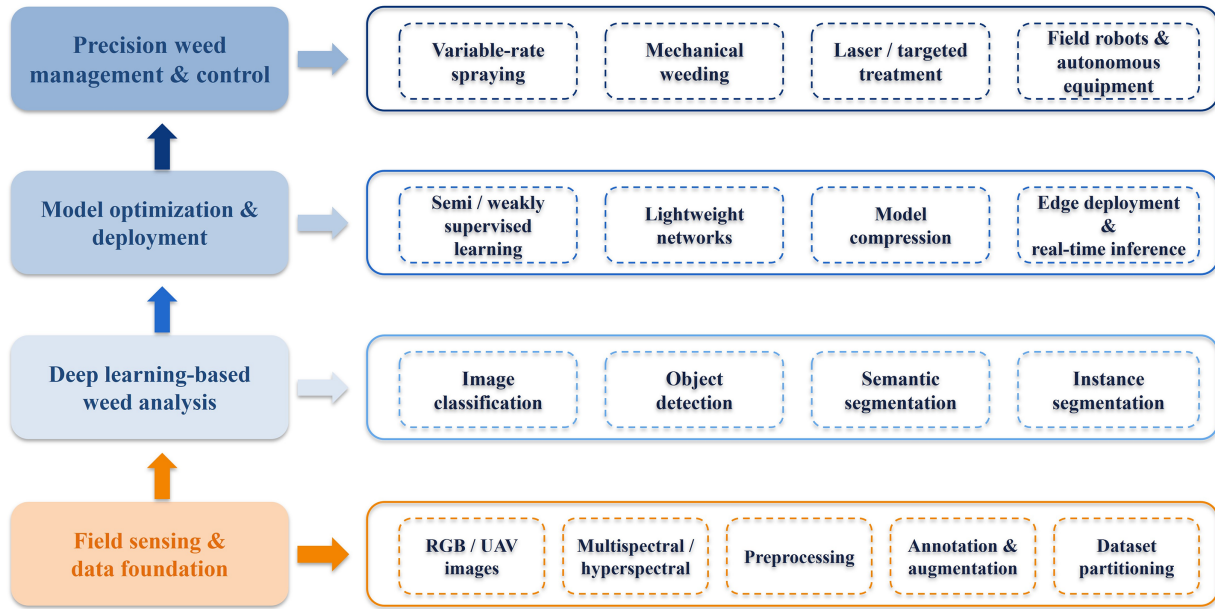


Figure 13: Perception, decision-making, and execution workflow of the weed control system. The figure summarizes the workflow from field data acquisition and dataset preparation to weed analysis, model deployment, and precision weed control.

3.1 Lightweight Backbone Network Design

In lightweight backbone network design, researchers usually employ mechanisms such as depthwise separable convolution, bottleneck structures, lightweight attention, and efficient feature fusion to substantially reduce the number of parameters and computational cost while maintaining detection accuracy as much as possible. Islam et al. replaced the original backbones of SSD and DeepLabv3+ with MobileNetV3 using UAV images from maize and soybean fields, and further proposed MobileNetV4-Seg. The results showed that such lightweight models can achieve real-time performance on resource-constrained edge devices. In particular, MobileNetV4-Seg achieved 44 FPS on Jetson Orin Nano with an input size of 256×256 , while obtaining IoU values of 69.9% and 76.8% on the maize and soybean datasets, respectively. This indicates that lightweight backbone networks are highly feasible for real-time field weed segmentation [95].

In object detection, lightweight modification is also a mainstream research direction. Huang et al. proposed an improved YOLOv8s model for weed detection in maize fields. The results showed that the computational cost was reduced to 12.7 GFLOPs and the model size to 9.1 MB, representing reductions of 57.4% and 59.2%, respectively, compared with the original model. Meanwhile, mAP@0.5 increased from 90.8% to 94.5%, indicating that lightweight structural design in field object detection does not necessarily lead to a significant loss of accuracy [96]. Similarly, Wang et al. proposed YOLO-Weed Nano for cotton field weed detection. By replacing the backbone network with DS_HGNetV2, introducing BiFPN, and designing a lightweight detection head, the model increased mAP by 1% while reducing the number of parameters, computational cost, and weight file size by 63.8%, 42.0%, and 60.7%, respectively. These results show that lightweight detection backbones and lightweight detection heads have become important design directions for edge-based weed recognition systems [97]. Overall, lightweight convolutional backbones and lightweight detection heads have been widely used for edge adaptation of agricultural weed detection models, with the goal of significantly reducing parameters, model size, and computational burden with only minor accuracy degradation.

3.2 Model Compression

In model compression optimization, pruning, quantization, and knowledge distillation are common approaches for compressing high-accuracy large models and adapting them to the resource constraints of edge devices. Existing reviews generally suggest that these methods can reduce redundant model parameters, decrease memory usage, and improve deployment latency. However, their specific benefits are jointly affected by network architecture, task type, and hardware platform. Therefore, it is inappropriate to draw a unified conclusion without considering specific experimental scenarios. For agricultural weed recognition, knowledge distillation has begun to be used to improve the accuracy of lightweight models. Saltık et al. proposed Context-Weighted Distillation (CWD) for sugar beet weed detection. Without increasing the complexity of the student model, CWD achieved an absolute mAP50 improvement of 2.5% on a real-world dataset and verified deployment feasibility on Jetson Orin Nano and Raspberry Pi 5. This shows that distillation can effectively alleviate the problem of accuracy degradation after model lightweighting in edge deployment scenarios [98].

As shown in Table 6, pruning is more suitable when network redundancy is obvious and structured acceleration is required, but the pruning ratio should be controlled to avoid losing small-target and occlusion-related features. Quantization is advisable when the hardware has strong low-bit inference support and the calibration set contains sufficient variations in illumination, background, crop growth stage, and sensor noise. Knowledge distillation is particularly useful when the goal is to maintain the accuracy of a compact model, because it can transfer semantic and contextual information from a stronger teacher model to a lightweight student model. However, all three methods should be validated on real field images rather than only on clean test sets, because model compression may reduce tolerance to shadows, occlusion, motion blur, and domain shifts.

Table 6: Appraisal of model compression methods for edge-based weed detection.

Compression Method	Main Mechanism	Advisable When	Potential Impact on Field Robustness	Possible Degradation Risk
Pruning	Removes redundant channels, filters, layers, or weights from a trained model	The original model has structural redundancy and the deployment platform has strict memory or latency constraints	Moderate pruning can preserve major weed features while improving speed; structured pruning is more hardware-friendly for edge devices	Excessive pruning may remove weak but useful feature channels, reducing sensitivity to small weeds, shaded weeds, partially occluded weeds, and low-contrast targets
Quantization	Reduces numerical precision, such as FP32 to FP16, INT8, or lower-bit representation	The target hardware supports low-bit acceleration and the model has been calibrated using representative field images	Well-calibrated quantization can provide large speed and memory benefits with limited accuracy loss	Poor calibration may amplify errors under illumination shifts, sensor noise, and low-contrast images; small weed targets and boundary localization may be degraded
Knowledge Distillation	Transfers knowledge from a large teacher model to a compact student model	A lightweight model is required, but direct compression causes unacceptable accuracy loss	Distillation can help the student model preserve high-level semantic and contextual information, improving robustness compared with training the student alone	Its effectiveness depends on teacher quality and training data diversity; if the teacher is biased under occlusion, shadows, or sensor noise, these errors may be transferred to the student model

(Continued)

Table 6 (continued)

Compression Method	Main Mechanism	Advisable When	Potential Impact on Field Robustness	Possible Degradation Risk
Combined Compression	Uses pruning, quantization, distillation, and inference acceleration together	Real-time operation is required on embedded platforms such as Jetson, Raspberry Pi, UAV processors, or tractor-mounted edge devices	Can achieve a better balance between speed and accuracy when each step is carefully calibrated and validated	Compression errors may accumulate; field validation is necessary because FPS improvement does not necessarily mean stable detection under complex field conditions

3.3 Inference Framework Acceleration

In inference acceleration, dedicated inference frameworks play an important role in improving the deployment feasibility of weed detection models on edge devices. Frameworks such as TensorRT can optimize computational graphs, fuse operators, reduce redundant memory access, and adapt models to the parallel computing characteristics of embedded GPUs. These techniques can substantially reduce inference latency and improve the real-time performance of weed detection systems. However, inference acceleration should not be evaluated only by FPS or latency. In field environments, weed images are affected by variable illumination, shadows, occlusion, motion blur, small target size, and sensor noise. Therefore, the practical value of inference acceleration should be assessed together with detection accuracy, segmentation quality, and robustness under real agricultural conditions.

For agricultural weed recognition, knowledge distillation has begun to be used to recover the accuracy of compressed lightweight models. Zhou et al. combined channel pruning with knowledge distillation and proposed a coordinated distillation training strategy for lightweight weed detection. Their optimized YOLOv5s model substantially reduced model size while maintaining high detection accuracy and achieved real-time deployment on Raspberry Pi, indicating that distillation can help alleviate accuracy degradation after model compression [99].

Mwitta et al. evaluated real-time detection models such as YOLOv4, EfficientDet, and CenterNet on Jetson Xavier AGX. The results showed that YOLOv4 achieved an mAP@0.5 of 93.4% at 14 FPS on the embedded platform, whereas YOLOv4-tiny achieved 52 FPS with a lower mAP@0.5 of 89%. This comparison indicates that lightweight models can significantly improve inference speed, but the improvement is often accompanied by measurable accuracy degradation [100]. For high-speed spraying or UAV monitoring, YOLOv4-tiny-like lightweight detectors may be more suitable because their high FPS can satisfy real-time operation requirements. However, for scenarios involving small weeds, dense weed clusters, severe occlusion, or high crop–weed visual similarity, the accuracy loss caused by lightweighting may increase missed detections and false detections.

More recent studies on UAV edge systems have also shown that lightweight edge models for Palmer amaranth detection and instance segmentation can achieve detection and segmentation inference times of 3–5 and 4 ms, respectively, under image tiling inference. This suggests that near-real-time or even real-time processing is technically feasible for specific weed detection tasks and image resolutions when model design, tiling strategy, and inference optimization are jointly considered [101]. Nevertheless, image tiling may introduce additional issues such as boundary discontinuity, repeated detection, and missed small weeds near tile borders. Therefore, tiled inference should also be evaluated under real field conditions rather than only using clean benchmark images.

Overall, the edge deployment of weed detection models has shifted from single-objective accuracy optimization toward a comprehensive balance among detection accuracy, inference speed, resource consumption, and field robustness. Existing studies do not support the conclusion that any single lightweight backbone, compression method, or inference acceleration framework is universally superior. A more reasonable judgment is that the deployment strategy should be selected according to the specific agricultural scenario and hardware platform. For real-time variable-rate spraying and UAV monitoring, lightweight one-stage detectors combined with inference acceleration often provide the most practical balance between accuracy and speed. For mechanical weeding, laser weeding, or weed coverage estimation, segmentation or instance segmentation models may still be necessary, even if they require stronger hardware support and careful optimization.

Therefore, future deployment-oriented evaluations should report not only FPS, latency, model size, and FLOPs, but also accuracy degradation after acceleration and robustness under changing illumination, shadows, crop–weed occlusion, small weed size, motion blur, and sensor noise. In addition, end-to-end delay from image acquisition, model inference, decision-making, and actuator response should be considered, because real-time weed control depends on the complete perception–decision–execution chain rather than inference speed alone.

4 Practical Challenges in Weed Detection

Unlike image recognition tasks conducted under controlled laboratory conditions, farmland weed detection under Agriculture 4.0 must directly adapt to open, dynamic, and unstructured real field environments. It must also meet the requirements of real-time performance, stability, and low power consumption for mobile operations of intelligent agricultural machinery. The challenges involved cover the entire process of image acquisition, feature extraction, model inference, and engineering deployment, and they remain key obstacles that must be overcome before weed detection technologies can move from laboratory research to large-scale practical application.

4.1 Fluctuations in Field Image Quality

Field image acquisition is affected by natural illumination, operation time, weather conditions, crop canopy structure, and crop growth status, resulting in substantial uncertainty in imaging quality. This uncertainty is a fundamental challenge for visual weed detection in field environments and is also an important reason why traditional fixed-threshold and handcrafted feature-based methods have insufficient generalization ability in real field scenarios [102]. First, natural illumination in field environments is dynamically variable. Under different weather conditions and at different times of day, light intensity, incident angle, and spectral composition all change, which can cause shifts in image brightness, contrast, and color representation. These changes weaken the effectiveness of vegetation–background segmentation or crop–weed discrimination based on color features [103]. In real field images, complex changes in illumination intensity, crop–weed occlusion, and uneven target size distribution can reduce the robustness of crop–weed segmentation methods [104]. Even when deep learning methods are used, shadows, occlusion, and complex backgrounds can still cause shifts in feature representation and limit cross-scenario generalization. Therefore, recent studies have attempted to improve model stability and deployability in real field environments through image enhancement, illumination correction, high dynamic range imaging, multispectral or polarization information fusion, and more robust deep networks [105].

Leaf occlusion and target overlap are among the key difficulties in visual weed detection in field environments. As crops grow, canopy coverage increases continuously, and leaf stacking, cross-occlusion, and boundary adhesion between crops and weeds or among weeds become more frequent. This is particularly

evident in densely planted crops such as maize and wheat, or in high-density field scenarios, where the local visibility and morphological integrity of targets are significantly reduced. As a result, the discriminative capability of handcrafted methods based on color, shape, and texture is weakened. Even with deep learning methods, dense occlusion, target overlap, and scale variation still increase the risks of missed detections and false detections, thereby affecting model stability and generalization in real field environments [62]. In addition, soil, crop residues, plastic mulch, stones, water droplets, and other non-target backgrounds, as well as appearance changes in crops caused by disease spots, insect damage, water stress, or nutrient imbalance, further reduce the separability between crops and weeds and increase the complexity of the detection task.

Dynamic imaging interference during field operations should also not be ignored. For intelligent weed-ing robots, tractor-mounted variable-rate spraying equipment, and other mobile field platforms, mechanical vibration, ground bumps, and camera motion during operation can cause image jitter and motion blur. Meanwhile, terrain undulation, plant height variation, and platform attitude changes can lead to continuous fluctuations in imaging distance, viewing angle, and target scale, reducing the consistency of visual features. Recent studies on agricultural robot perception systems and image deblurring for precision agriculture have indicated that motion blur, device movement, and imaging instability in complex open environments have become important factors affecting recognition, detection, and localization accuracy. Therefore, related algorithms need stronger anti-interference capability, environmental adaptability, and the ability to integrate depth, stereo vision, or multimodal perception.

4.2 High Similarity in Visual Features

The high similarity between the visual features of crops and weeds is a major technical difficulty in weed detection at the seedling stage and is also an important factor limiting further improvements in algorithmic accuracy. First, crops and weeds at the seedling stage often show strong similarities in morphology and color. The cotyledons and true leaves of most annual weeds and field crops are mainly light green during the seedling stage, and their spectral reflectance characteristics differ only slightly. Even technicians with agricultural experience may find it difficult to distinguish them quickly based only on visual observation. For example, sugar beet seedlings are highly similar to broadleaf weed seedlings such as lambsquarters and redroot pigweed in cotyledon shape, leaf color, and texture characteristics. Wheat seedlings are also similar to grass weeds such as blackgrass and wild oat in leaf morphology and plant structure. In such scenarios, accurate recognition cannot be achieved by relying only on global visual features, and models must be able to capture subtle local differences.

The small interspecific differences and large intraspecific variations between crops and weeds further increase the difficulty of classification and recognition. For example, rice and barnyardgrass both belong to the Poaceae family, and their plant structure, leaf morphology, and leaf vein characteristics are very similar at the seedling stage. Their main differences are concentrated in millimeter-scale structures such as the ligule and auricles. These details are difficult to capture consistently under conventional field imaging conditions, which limits the improvement of weed detection accuracy in such scenarios. In addition, weed emergence time is often inconsistent in the field, and weeds at different growth stages frequently coexist within the same plot. Their leaf morphology, size, and color change significantly during growth. The same weed species may also show large morphological differences under different water, fertilizer, and illumination conditions. This intraspecific feature variation further increases the difficulty of model classification.

4.3 Constraints and Adaptation in Practical Applications

With the large-scale advancement of Agriculture 4.0, the research focus of weed detection technology is shifting from high recognition accuracy under laboratory conditions to engineering deployment capability in

complex field environments. For applications in intelligent agricultural machinery, related algorithms must not only maintain high detection accuracy, but also meet constraints such as real-time performance, low power consumption, high stability, and low cost. These factors interact with each other, making the transition of weed detection technology from laboratory research to industrial application still highly challenging.

Strict real-time requirements and edge deployment conditions are key factors limiting system implementation. During field operations, intelligent weeding robots and tractor-mounted variable-rate spraying equipment need to complete image acquisition, preprocessing, model inference, and result output within a short time, thereby supporting synchronized detection and spraying or weeding control during movement. However, field equipment usually relies on edge computing platforms such as Jetson, Raspberry Pi, and embedded chips, whose computing power, memory, and power supply capacity are limited. High-accuracy large models commonly used in laboratories often achieve good detection performance, but they are difficult to deploy for real-time inference on embedded platforms. Lightweight models run faster, but they are more likely to suffer from accuracy degradation and increased missed detections in complex field scenarios.

Harsh field environments and cost constraints also increase the difficulty of engineering deployment. Farmland operations are often accompanied by high temperature, high humidity, heavy dust, and strong vibration. These factors not only affect the operating status of hardware devices such as cameras and computing platforms, but also cause long-term changes in imaging conditions, placing higher requirements on the environmental adaptability and long-term stability of detection algorithms. Meanwhile, large-scale agricultural applications are highly sensitive to system cost. Weeding equipment usually cannot be equipped with expensive industrial cameras and high-performance computing platforms on a large scale, and instead relies more on low-cost consumer-grade cameras and embedded chips. Therefore, how to balance detection accuracy, operating efficiency, system stability, and deployment cost under limited hardware resources and complex field environments has become a core issue in the engineering application of weed detection technology.

5 Future Development Trends

Under the background of Agriculture 4.0, the research objective of farmland weed detection technology is no longer limited to improving accuracy on a single dataset. Instead, it is gradually shifting toward stable perception in complex field environments, low-cost deployment, and collaborative application with operational systems. Based on the preceding analysis of deep learning-based weed detection methods, public datasets, lightweight deployment technologies, and real-world challenges, future farmland weed detection will mainly focus on environmental robustness enhancement, fine-grained recognition, multi-source information fusion, standardized dataset construction, edge-based real-time deployment, and system-level applications.

5.1 Enhancement of Environmental Robustness

To address field-related problems such as illumination variation, shadow occlusion, crop canopy overlap, motion blur, and background interference, future weed detection models need to shift from “static image recognition” to “robust perception in complex open environments.” On the one hand, models should strengthen their ability to learn illumination-stable features, shadow-resistant features, and background-suppression features. Methods such as image enhancement, color correction, deblurring, attention mechanisms, and multi-scale feature fusion can be used to improve model adaptability to brightness variation, local shadows, and complex backgrounds. On the other hand, dataset construction should actively cover different weather conditions, acquisition times, soil types, crop growth stages, and field management

conditions, so that the training data are closer to real farmland environments and the gap between laboratory data and field applications can be reduced.

Meanwhile, future studies should not only report average accuracy on a single test set, but should also pay attention to generalization performance across fields, seasons, crop varieties, and imaging devices. For Agriculture 4.0 applications, the ability of a model to maintain stable detection performance across different field environments is more practically valuable than achieving higher accuracy on a single public dataset. Therefore, robustness evaluation under real field disturbances and cross-domain generalization testing will become important evaluation components in weed detection research.

5.2 Recognition of Similarities between Crops and Weeds

Crops and weeds often have similar color, texture, and morphological features at the seedling stage. It is especially difficult to distinguish grass crops from grass weeds and broadleaf crops from broadleaf weeds, because these cases are characterized by small inter-class differences and large intra-class variations. To address this issue, weed detection models need to gradually move from coarse-grained “crop/weed” binary classification toward fine-grained recognition based on species category, seedling morphology, growth stage, and spatial position.

In terms of model structure, local detail enhancement, multi-scale contextual modeling, attention mechanisms, Transformer architectures, and edge-aware modules can be introduced to improve the extraction of leaf edges, vein textures, plant structural features, and local organ characteristics. In scenarios involving severe occlusion and leaf overlap, relying only on object detection boxes is insufficient for precision weeding. Semantic segmentation and instance segmentation are more suitable for providing target boundary information. Future studies can develop multi-task models that combine detection and segmentation, so that weed location, contour, coverage area, density, and growth status can be obtained simultaneously. This would provide more refined spatial support for variable-rate spraying, mechanical weeding, and laser weeding.

5.3 Multi-Source Information Fusion

A single RGB image has clear limitations under drastic illumination variation, high crop–weed appearance similarity, and severe vegetation overlap. Future farmland weed detection will place greater emphasis on the fusion of multi-source sensing information, including RGB images, multispectral images, hyperspectral images, depth information, thermal infrared information, three-dimensional point clouds, agricultural machinery pose, and crop-row priors. Through multi-source information fusion, the discrimination between crops and weeds can be enhanced from multiple aspects, including color, spectrum, spatial structure, and geometric morphology.

Among these sources, multispectral and hyperspectral information can provide richer vegetation reflectance characteristics than RGB images, which helps distinguish crops and weeds with similar appearances but different physiological properties. Depth information and three-dimensional structural information can supplement the spatial morphological features of targets and alleviate recognition difficulties caused by leaf occlusion, target overlap, and scale variation. Crop-row priors, field positioning information, and operational path information can provide spatial constraints for the model and reduce false detections in complex backgrounds. Future research should not simply increase the number of sensors. Instead, it should focus on building lightweight, stable, and deployable multimodal fusion networks, so that multi-source information can truly support real-time field detection without making the system too complex for engineering application.

5.4 Construction of Standardized Datasets

Current weed detection still heavily depends on manually annotated data. Semantic segmentation and instance segmentation require pixel-level or instance-level annotations, which impose higher requirements on annotation accuracy, annotation consistency, and labor cost. Future research can focus on semi-supervised learning, weakly supervised learning, self-supervised learning, transfer learning, active learning, and unsupervised domain adaptation to reduce data annotation costs and improve model generalization across scenarios. By jointly training models with a small number of high-quality annotated samples and a large amount of unlabeled field images, the annotation burden can be reduced while enhancing model adaptability to new fields, new crops, and new weed species.

The weed detection field still needs larger-scale datasets with more unified standards and scenarios closer to real farmland conditions. Future datasets should not be limited to a single crop, region, or growth stage. Instead, they should cover multiple crops, weed species, seasons, weather conditions, sensors, and operating platforms. At the same time, category definitions, annotation formats, dataset partitioning principles, and evaluation metrics should be standardized to reduce performance comparison bias caused by differences in data partitioning and evaluation protocols. For Agriculture 4.0 applications, dataset construction is a fundamental step in the practical implementation of weed detection technology and directly affects model training, performance evaluation, and field deployment.

5.5 Edge Deployment and Real-Time Inference

For Agriculture 4.0 applications, farmland weed detection should not remain limited to improving the accuracy of a single visual algorithm. It should further develop toward edge-based real-time deployment and collaboration with operational systems. On the one hand, weed detection models eventually need to run on edge platforms such as intelligent weeding robots, UAVs, and tractor-mounted spraying systems. Therefore, future research should continue to focus on balancing accuracy, speed, power consumption, and cost. Lightweight backbone networks, lightweight detection heads, model pruning, low-bit quantization, knowledge distillation, neural architecture search, and inference engine acceleration will remain important directions for improving the engineering adaptability of models.

However, evaluation of model lightweighting should not rely only on indicators such as parameter count, computational cost, and single-frame FPS. It should also consider real operating speed, camera frame rate, nozzle response delay, detection distance, and platform computing capacity, so as to establish system-level real-time evaluation standards for field operations. For variable-rate spraying systems, detection latency directly affects spraying position accuracy. For mechanical or laser weeding systems, localization error and actuator response delay may further increase the risk of crop damage. Therefore, future evaluation systems should be extended from “single-frame inference speed” to “end-to-end latency across perception, decision-making, and execution.”

On the other hand, weed detection results should not remain limited to categories, bounding boxes, or segmentation masks. They should further support variable-rate spraying decisions, mechanical weeding path planning, laser targeting control, and field weed distribution mapping. For example, herbicide dosage can be automatically determined based on weed species, growth stage, density, and spatial distribution. The trajectory of mechanical weeding mechanisms can be adjusted according to the relative positions of weeds and crops. Long-term monitoring results can also be used to generate weed occurrence hotspot maps, providing a basis for precision field management. In the future, farmland weed detection technology will gradually evolve from the optimization of a single deep learning model toward an integrated technical system involving multi-source perception, stable recognition, lightweight deployment, and intelligent execution.

Its evaluation focus will also shift from detection accuracy on laboratory datasets to stability, real-time performance, low cost, and sustainable application capability in real field environments.

6 Conclusion

This paper reviews deep learning-based precision weed detection technologies in farmland under the background of Agriculture 4.0, covering research progress in public weed image datasets, data preprocessing and augmentation, image classification, object detection, image segmentation, label-efficient learning, and edge deployment. Existing studies show that deep learning has become the mainstream technical route for farmland weed detection. Compared with traditional methods that rely on handcrafted features such as color, texture, and shape, models such as convolutional neural networks and segmentation networks can extract more discriminative deep features from complex field images, achieving good performance in weed recognition, spatial localization, and pixel-level region extraction. These methods provide an important technical foundation for intelligent agricultural machinery operations, including variable-rate herbicide application, mechanical weeding, and laser weeding.

However, existing studies are still limited by field illumination variation, occlusion and overlap, morphological similarity between crops and weeds, difficulty in small-target recognition, inconsistent dataset standards, and insufficient edge computing capacity. The generalization ability of models across crops, regions, and seasons still needs to be improved. Future research should be oriented toward real field applications by constructing higher-quality datasets, developing lightweight detection models that balance accuracy, robustness, and real-time performance, and strengthening the coordination between detection results and intelligent agricultural machinery execution systems. These efforts will promote the development of weed control toward greater precision, sustainability, and intelligence.

Acknowledgement: Not applicable.

Funding Statement: This research was funded by the School of Aeronautics and Astronautics, North China Institute of Aerospace Engineering Fund: hkyh-2025-01.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Peng Shen and Tenglong Li; methodology, Tenglong Li and Yongpeng Sun; validation, Peng Shen and Guoqing Zhang; formal analysis, Guoqing Zhang, Hao Cui; investigation, Tenglong Li and Yongpeng Sun; writing—original draft preparation, Tenglong Li, Hao Cui; writing—review and editing, Guoqing Zhang and Peng Shen. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ehrampoosh A, Hettiarachchi P, Koirala A, Hassan J, Islam N, Ray B, et al. Intelligent weed management using aerial image processing and precision herbicide spraying: an overview. *Crop Prot.* 2025;194(7):107206. doi:10.1016/j.cropro.2025.107206.
2. Naseera M, Gupta A. A review of modern techniques for plant disease identification and weed detection in precision agriculture. *Int J Artif Intell.* 2026;15(1):998. doi:10.11591/ijai.v15.i1.pp998-1008.

3. Coleman GRY, Bender A, Hu K, Sharpe SM, Schumann AW, Wang Z, et al. Weed detection to weed recognition: reviewing 50 years of research to identify constraints and opportunities for large-scale cropping systems. *Weed Technol.* 2022;36(6):741–57. doi:10.1017/wet.2022.84.
4. Wang K, Hu X, Zheng H, Lan M, Liu C, Liu Y, et al. Weed detection and recognition in complex wheat fields based on an improved YOLOv7. *Front Plant Sci.* 2024;15:1372237. doi:10.3389/fpls.2024.1372237.
5. Das S, Upadhyay A, Sun X. Technology-driven approaches to intelligent mechanical weed control: a systematic review for sustainable weed management. *Front Plant Sci.* 2025;16:1734507. doi:10.3389/fpls.2025.1734507.
6. Liu B, Bruch R. Weed detection for selective spraying: a review. *Curr Robot Rep.* 2020;1(1):19–26. doi:10.1007/s43154-020-00001-w.
7. Wu Z, Chen Y, Zhao B, Kang X, Ding Y. Review of weed detection methods based on computer vision. *Sensors.* 2021;21(11):3647. doi:10.3390/s21113647.
8. Hasan ASMM, Sohel F, Diepeveen D, Laga H, Jones MGK. A survey of deep learning techniques for weed detection from images. *Comput Electron Agric.* 2021;184(3):106067. doi:10.1016/j.compag.2021.106067.
9. Zhang W, Miao Z, Li N, He C, Sun T. Review of current robotic approaches for precision weed management. *Curr Robot Rep.* 2022;3(3):139–51. doi:10.1007/s43154-022-00086-5.
10. Meyer GE, Neto JC. Verification of color vegetation indices for automated crop imaging applications. *Comput Electron Agric.* 2008;63(2):282–93. doi:10.1016/j.compag.2008.03.009.
11. Dyrmann M, Karstoft H, Midtby HS. Plant species classification using deep convolutional neural network. *Biosyst Eng.* 2016;151(1):72–80. doi:10.1016/j.biosystemseng.2016.08.024.
12. Milioto A, Lottes P, Stachniss C. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. In: *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*; 2018 May 21–25; Brisbane, Australia. p. 2229–35. doi:10.1109/icra.2018.8460962.
13. Su WH. Advanced machine learning in point spectroscopy, RGB- and hyperspectral-imaging for automatic discriminations of crops and weeds: a review. *Smart Cities.* 2020;3(3):767–92. doi:10.3390/smartcities3030039.
14. Zheng L, Yi J, He P, Tie J, Zhang Y, Wu W, et al. Improvement of the YOLOv8 model in the optimization of the weed recognition algorithm in cotton field. *Plants.* 2024;13(13):1843. doi:10.3390/plants13131843.
15. Olsen A, Konovalov DA, Philippa B, Ridd P, Wood JC, Johns J, et al. DeepWeeds: a multiclass weed species image dataset for deep learning. *Sci Rep.* 2019;9(1):2058. doi:10.1038/s41598-018-38343-3.
16. Haug S, Ostermann J. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In: *Computer Vision—ECCV 2014 Workshops*. Cham, Switzerland: Springer International Publishing; 2015. p. 105–16. doi:10.1007/978-3-319-16220-1_8.
17. Steininger D, Trondl A, Croonen G, Simon J, Widhalm V. The CropAndWeed dataset: a multi-modal learning approach for efficient crop and weed manipulation. In: *Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; 2023 Jan 3–7; Waikoloa, HI, USA. p. 3718–27. doi:10.1109/wacv56688.2023.00372.
18. Bertoglio R, Kalouguine A, Boffety D, Boulet M, Berducat M, Facchinetti D, et al. The ACRE crop-weed dataset (1.0.0) [Data set]. *Zenodo.* 2023. doi:10.5281/zenodo.8102217.
19. Sudars K, Jasko J, Namatevs I, Ozola L, Badaukis N. Dataset of annotated food crops and weed images for robotic computer vision control. *Data Brief.* 2020;31(1):105833. doi:10.1016/j.dib.2020.105833.
20. Upadhyay A, Sunil GC, Mahecha MV, Mettler J, Howatt K, Aderholdt W, et al. Weed-crop dataset in precision agriculture: resource for AI-based robotic weed control systems. *Data Brief.* 2025;60(6):111486. doi:10.1016/j.dib.2025.111486.
21. Dang F, Chen D, Lu Y, Li Z. YOLOWeeds: a novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. *Comput Electron Agric.* 2023;205(2):107655. doi:10.1016/j.compag.2023.107655.
22. Deng B, Lu Y. Weed image augmentation by ControlNet-added stable diffusion for multi-class weed detection. *Comput Electron Agric.* 2025;232(2):110123. doi:10.1016/j.compag.2025.110123.

23. Gallo I, Rehman AU, Dehkord RH, Landro N, La Grassa R, Boschetti M. Weed detection by UAV 416a computer vision dataset. 2022 [cited 2026 Jan 1]. Available from: <https://universe.roboflow.com/chicory-crop-weeds-5m7vo/weed-detection-by-uav-416a/dataset/1>.
24. Correa JML, Andújar D, Todeschini M, Karouta J, Begochea JM, Ribeiro A. WeedMaize dataset. Zenodo. 2021. doi:10.5281/zenodo.5106795.
25. Shinde S, Attar V. An Indian annotated weed dataset for computer vision tasks in precision farming. Data Brief. 2025;61:111691. doi:10.17632/d3n3mgjjbv.2.
26. Rai N, Mahecha MV, Christensen A, Quanbeck J, Zhang Y, Howatt K, et al. Multi-format open-source weed image dataset for real-time weed identification in precision agriculture. Data Brief. 2023;51:109691. doi:10.1016/j.dib.2023.109691.
27. Shivaram B, Ramesh TK, Kumar J. Chilli crop-weed detection dataset [Dataset]. GitHub repository; 2025. Available from: <https://github.com/sbhaktavatsala/chilli-weed-dataset>.
28. Ilyas T, Arsa DMS, Ahmad K, Lee J, Won O, Lee H, et al. CWD30: a new benchmark dataset for crop weed recognition in precision agriculture. Comput Electron Agric. 2025;229(1):109737. doi:10.1016/j.compag.2024.109737.
29. Michael J, Thenmozhi M. SorghumWeedDataset_Segmentation. Mendeley Data, V1. 2023 [cited 2026 Jan 1]. Available from: <https://data.mendeley.com/datasets/y9bmtf4xmr/1>.
30. Sunil GC, Sun X. Crop and weed dataset. Mendeley Data, V1. 2024 [cited 2026 Jan 1]. Available from: <https://data.mendeley.com/datasets/hs7d7kpd3z/1>.
31. Salcedo-Navarro A, Montalban-Faet G, Garcia-Pineda M, Segura-Garcia J. Dataset for weed detection in fruit orchards. Data Brief. 2025;63(3):112276. doi:10.1016/j.dib.2025.112276.
32. Raza H, Abu Bakr M, Khan SD, Batool H, Ullah H, Ullah M. Benchmarking YOLO models for crop growth and weed detection in cotton fields. AgriEngineering. 2025;7(11):375. doi:10.3390/agriengineering7110375.
33. Ramesh P, Prabakaran G, Nagavel V, Bino J, Shabana Parveen M, Bhuvaneswari PTV. Detection of commercial crop weeds using machine learning algorithms. Sci Rep. 2025;15(1):38791. doi:10.1038/s41598-025-22676-x.
34. Shen P, Li T, Xue H, Zhao Z, Cao H, Mei K, et al. YOLOv8-MLD: a lightweight method for weed detection in cornfields. IEEE Access. 2026;14:66787–806. doi:10.1109/access.2026.3687147.
35. Zhao L, Zhao Y, Liu T, Deng H. A weakly supervised semantic segmentation model of maize seedlings and weed images based on scrawl labels. Sensors. 2023;23(24):9846. doi:10.3390/s23249846.
36. Sapkota BB, Popescu S, Rajan N, Leon RG, Reberg-Horton C, Mirsky S, et al. Use of synthetic images for training a deep learning model for weed detection and biomass estimation in cotton. Sci Rep. 2022;12(1):19580. doi:10.1038/s41598-022-23399-z.
37. Xu B, Werle R, Chudzik G, Zhang Z. Enhancing weed detection using UAV imagery and deep learning with weather-driven domain adaptation. Comput Electron Agric. 2025;237(4):110673. doi:10.1016/j.compag.2025.110673.
38. Saleh A, Olsen A, Wood J, Philippa B, Azghadi MR. WeedCLR: weed contrastive learning through visual representations with class-optimized loss in long-tailed datasets. arXiv:2310.12465. 2023. doi:10.48550/arXiv.2310.12465.
39. Rathore PS. QWID: quantized weed identification deep neural network. arXiv:2310.18921. 2023. doi:10.48550/arXiv.2310.18921.
40. Rahman MG, Rahman MA, Parvez MZ, Patwary MAK, Ahamed T, Fleming-Muñoz DA, et al. ADeepWeeD: an adaptive deep learning framework for weed species classification. Artif Intell Agric. 2025;15(4):590–609. doi:10.1016/j.aiia.2025.04.009.
41. Dheeraj A, Chand S. Using deep learning models for crop and weed classification at early stage. In: Sentiment analysis and deep learning. Singapore: Springer Nature; 2023. p. 931–42. doi:10.1007/978-981-19-5443-6_69.
42. Rozendo GB, Roberto GF, do Nascimento MZ, Alves Neves L, Lumini A. Weeds classification with deep learning: an investigation using CNN, vision transformers, pyramid vision transformers, and ensemble strategy. In: Progress in pattern recognition, image analysis, computer vision, and applications. Cham, Switzerland: Springer Nature; 2024. p. 229–43. doi:10.1007/978-3-031-49018-7_17.

43. Deng B, Lu Y, Xu J. Weed database development: an updated survey of public weed datasets and cross-season weed detection adaptation. *Ecol Inform.* 2024;81(3):102546. doi:10.1016/j.ecoinf.2024.102546.
44. Li S, Chen Z, Xie J, Zhang H, Guo J. PD-YOLO: a novel weed detection method based on multi-scale feature fusion. *Front Plant Sci.* 2025;16:1506524. doi:10.3389/fpls.2025.1506524.
45. Zhu H, Zhang Y, Mu D, Bai L, Wu X, Zhuang H, et al. Research on improved YOLOx weed detection based on lightweight attention module. *Crop Prot.* 2024;177(4):106563. doi:10.1016/j.cropro.2023.106563.
46. Khan AT, Jensen SM, Khan AR. Advancing precision agriculture: a comparative analysis of YOLOv8 for multi-class weed detection in cotton cultivation. *Artif Intell Agric.* 2025;15(2):182–91. doi:10.1016/j.aiaa.2025.01.013.
47. Li W, Zhang Y. DC-YOLO: an improved field plant detection algorithm based on YOLOv7-tiny. *Sci Rep.* 2024;14(1):26430. doi:10.1038/s41598-024-77865-x.
48. Deng B, Lu Y. Semi-supervised weed detection in vegetable fields: in-domain and cross-domain experiments. *arXiv:2502.17673.* 2025. doi:10.48550/arXiv.2502.17673.
49. Li J, Chen D, Yin X, Li Z. Performance evaluation of semi-supervised learning frameworks for multi-class weed detection. *Front Plant Sci.* 2024;15:1396568. doi:10.3389/fpls.2024.1396568.
50. Silva JAOS, de Siqueira VS, Mesquita M, Vale LSR, do Nascimento Borges Marques T, da Silva JLB, et al. Deep learning for weed detection and segmentation in agricultural crops using images captured by an unmanned aerial vehicle. *Remote Sens.* 2024;16(23):4394. doi:10.3390/rs16234394.
51. Pai DG, Balachandra M, Kamath R. Deep learning approach for crop-weed segmentation in peanut cultivation using PSPEdgeWeedNet. *Sci Rep.* 2025;15(1):43032. doi:10.1038/s41598-025-24174-6.
52. Sarker TT, Ahmed KR, Islam T, Rankrape CB, Gage K. WeedSense: multi task learning for weed segmentation, height estimation, and growth stage classification. *arXiv:2508.14486.* 2025. doi:10.48550/arXiv.2508.14486.
53. Gupta SK, Yadav SK, Soni SK, Shanker U, Singh PK. Multiclass weed identification using semantic segmentation: an automated approach for precision agriculture. *Ecol Inform.* 2023;78(15):102366. doi:10.1016/j.ecoinf.2023.102366.
54. Garibaldi-Márquez F, Martínez-Barba DA, Montañez-Franco LE, Flores G, Valentín-Coronado LM. Enhancing site-specific weed detection using deep learning transformer architectures. *Crop Prot.* 2025;190(1):107075. doi:10.1016/j.cropro.2024.107075.
55. Rai N, Zhang Y, Ram BG, Schumacher L, Yellavajjala RK, Bajwa S, et al. Applications of deep learning in precision weed management: a review. *Comput Electron Agric.* 2023;206(7):107698. doi:10.1016/j.compag.2023.107698.
56. Hu K, Coleman G, Zeng S, Wang Z, Walsh M. Graph weeds net: a graph-based deep learning method for weed recognition. *Comput Electron Agric.* 2020;174(7):105520. doi:10.1016/j.compag.2020.105520.
57. Shackleton C, Ali RH, Ali Khan T. Enhancing rangeland weed detection through convolutional neural networks and transfer learning. *Crop Des.* 2024;3(3):100060. doi:10.1016/j.cropro.2024.100060.
58. Saleem MH, Velayudhan KK, Potgieter J, Arif KM. Weed identification by single-stage and two-stage neural networks: a study on the impact of image resizers and weights optimization algorithms. *Front Plant Sci.* 2022;13:850666. doi:10.3389/fpls.2022.850666.
59. Mu Y, Feng R, Ni R, Li J, Luo T, Liu T, et al. A faster R-CNN-based model for the identification of weed seedling. *Agronomy.* 2022;12(11):2867. doi:10.3390/agronomy12112867.
60. Zhang X, Cui J, Liu H, Han Y, Ai H, Dong C, et al. Weed identification in soybean seedling stage based on optimized faster R-CNN algorithm. *Agriculture.* 2023;13(1):175. doi:10.3390/agriculture13010175.
61. Osorio K, Puerto A, Pedraza C, Jamaica D, Rodríguez L. A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering.* 2020;2(3):471–88. doi:10.3390/agriengineering2030032.
62. Goyal R, Nath A, Niranjana U. Weed detection using deep learning in complex and highly occluded potato field environment. *Crop Prot.* 2025;187:106948. doi:10.1016/j.cropro.2024.106948.
63. Sharma A, Kumar V, Longchamps L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agric Technol.* 2024;9:100648. doi:10.1016/j.atech.2024.100648.
64. Gao J, French AP, Pound MP, He Y, Pridmore TP, Pieters JG. Deep convolutional neural networks for image-based *Convolvulus sepium* detection in sugar beet fields. *Plant Methods.* 2020;16(1):29. doi:10.1186/s13007-020-00570-z.

65. Deng L, Miao Z, Zhao X, Yang S, Gao Y, Zhai C, et al. HAD-YOLO: an accurate and effective weed detection model based on improved YOLOV5 network. *Agronomy*. 2025;15(1):57. doi:10.3390/agronomy15010057.
66. Li J, Zhang W, Zhou H, Yu C, Li Q. Weed detection in soybean fields using improved YOLOv7 and evaluating herbicide reduction efficacy. *Front Plant Sci*. 2023;14:1284338. doi:10.3389/fpls.2023.1284338.
67. Liu Y, Zeng F, Diao H, Zhu J, Ji D, Liao X, et al. YOLOv8 model for weed detection in wheat fields based on a visual converter and multi-scale feature fusion. *Sensors*. 2024;24(13):4379. doi:10.3390/s24134379.
68. Wang A, Peng T, Cao H, Xu Y, Wei X, Cui B. TIA-YOLOv5: an improved YOLOv5 network for real-time detection of crop and weed in the field. *Front Plant Sci*. 2022;13:1091655. doi:10.3389/fpls.2022.1091655.
69. Chen J, Wang H, Zhang H, Luo T, Wei D, Long T, et al. Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion. *Comput Electron Agric*. 2022;202(4):107412. doi:10.1016/j.compag.2022.107412.
70. Wu H, Wang Y, Zhao P, Qian M. Small-target weed-detection model based on YOLO-V4 with improved backbone and neck structures. *Precis Agric*. 2023;24(6):2149–70. doi:10.1007/s11119-023-10035-7.
71. Hu K, Wang Z, Coleman G, Bender A, Yao T, Zeng S, et al. Deep learning techniques for in-crop weed recognition in large-scale grain production systems: a review. *Precis Agric*. 2024;25(1):1–29. doi:10.1007/s11119-023-10073-1.
72. Singh P, Zhao B, Shi Y. Computer vision for site-specific weed management in precision agriculture: a review. *Agriculture*. 2025;15(21):2296. doi:10.3390/agriculture15212296.
73. Zhao H, Wang Y. Deep learning-based approaches for weed detection in crops. *Front Plant Sci*. 2025;16:1746406. doi:10.3389/fpls.2025.1746406.
74. Joy J, Kelvin B, Howatt K, Aderholdt W, Khan M, Peters T, et al. Edge-deployable segmentation and prescription mapping of post-emergence weeds in sugar beet crops for UAV-based precision spraying. *J Agric Food Res*. 2025;24(1):102422. doi:10.1016/j.jafr.2025.102422.
75. Cui J, Tan F, Bai N, Fu Y. Improving U-Net network for semantic segmentation of corns and weeds during corn seedling stage in field. *Front Plant Sci*. 2024;15:1344958. doi:10.3389/fpls.2024.1344958.
76. Liao J, Chen M, Zhang K, Zhou H, Zou Y, Xiong W, et al. SC-Net: a new strip convolutional network model for rice seedling and weed segmentation in paddy field. *Comput Electron Agric*. 2024;220(12):108862. doi:10.1016/j.compag.2024.108862.
77. Li Y, Guo R, Li R, Ji R, Wu M, Chen D, et al. An improved U-Net and attention mechanism-based model for sugar beet and weed segmentation. *Front Plant Sci*. 2024;15:1449514. doi:10.3389/fpls.2024.1449514.
78. Gao J, Tan F, Li X. EDM-UNet: an edge-enhanced and attention-guided model for UAV-based weed segmentation in soybean fields. *Agriculture*. 2025;15(24):2575. doi:10.3390/agriculture15242575.
79. Moazzam SI, Khan US, Qureshi WS, Nawaz T, Kunwar F. Towards automated weed detection through two-stage semantic segmentation of tobacco and weed pixels in aerial Imagery. *Smart Agric Technol*. 2023;4(13):100142. doi:10.1016/j.atech.2022.100142.
80. Xu B, Fan J, Chao J, Arsenijevic N, Werle R, Zhang Z. Instance segmentation method for weed detection using UAV imagery in soybean fields. *Comput Electron Agric*. 2023;211(3):107994. doi:10.1016/j.compag.2023.107994.
81. Asad MH, Bais A. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Inf Process Agric*. 2020;7(4):535–45. doi:10.1016/j.inpa.2019.12.002.
82. Hashemi-Beni L, Gebrehiwot A, Karimoddini A, Shahbazi A, Dorbu F. Deep convolutional neural networks for weeds and crops discrimination from UAS imagery. *Front Remote Sens*. 2022;3:755939. doi:10.3389/frsen.2022.755939.
83. Fu H, Li X, Zhu L, Pan X, Wu T, Li W, et al. DSC-DeepLabv3+: a lightweight semantic segmentation model for weed identification in maize fields. *Front Plant Sci*. 2025;16:1647736. doi:10.3389/fpls.2025.1647736.
84. Quan L, Wu B, Mao S, Yang C, Li H. An instance segmentation-based method to obtain the leaf age and plant centre of weeds in complex field environments. *Sensors*. 2021;21(10):3389. doi:10.3390/s21103389.
85. Lu C, Gehring K, Kopfinger S, Bernhardt H, Beck M, Walther S, et al. Weed instance segmentation from UAV orthomosaic images based on deep learning. *Smart Agric Technol*. 2025;11:100966. doi:10.1016/j.atech.2025.100966.
86. Zhao J, Berge TW, Geipel J. Transformer in UAV image-based weed mapping. *Remote Sens*. 2023;15(21):5165. doi:10.3390/rs15215165.

87. Gao J, Liao W, Nuyttens D, Lootens P, Xue W, Alexandersson E, et al. Cross-domain in transfer learning for weed segmentation and mapping in precision farming using ground and UAV images. *Expert Syst Appl.* 2024;246(12):122980. doi:10.1016/j.eswa.2023.122980.
88. Li J, Chen D, Qi X, Li Z, Huang Y, Morris D, et al. Label-efficient learning in agriculture: a comprehensive review. *Comput Electron Agric.* 2023;215(11):108412. doi:10.1016/j.compag.2023.108412.
89. Nong C, Fan X, Wang J. Semi-supervised learning for weed and crop segmentation using UAV imagery. *Front Plant Sci.* 2022;13:927368. doi:10.3389/fpls.2022.927368.
90. Liu T, Zhai D, He F, Yu J. Semi-supervised learning methods for weed detection in turf. *Pest Manag Sci.* 2024;80(6):2552–62. doi:10.1002/ps.7959.
91. Mishra AM, Kaur P, Singh MP, Singh SP. A self-supervised overlapped multiple weed and crop leaf segmentation approach under complex light condition. *Multimed Tools Appl.* 2024;83(27):68993–9018. doi:10.1007/s11042-024-18272-2.
92. Thomas LF, Änäkälä M, Lajunen A. Weakly supervised perennial weed detection in a barley field. *Remote Sens.* 2023;15(11):2877. doi:10.3390/rs15112877.
93. Chen D, Lu Y, Li Z, Young S. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Comput Electron Agric.* 2022;198(5):107091. doi:10.1016/j.compag.2022.107091.
94. Huang Y, Hussein AE, Wang X, Bais A, Yao S, Wilder T. Unsupervised domain adaptation with self-training for weed segmentation. *Intell Syst Appl.* 2025;25(12):200468. doi:10.1016/j.iswa.2024.200468.
95. Islam MD, Liu W, Izere P, Singh P, Yu C, Riggan B, et al. Towards real-time weed detection and segmentation with lightweight CNN models on edge devices. *Comput Electron Agric.* 2025;237(17):110600. doi:10.1016/j.compag.2025.110600.
96. Huang J, Xia X, Diao Z, Li X, Zhao S, Zhang J, et al. A lightweight model for weed detection based on the improved YOLOv8s network in maize fields. *Agronomy.* 2024;14(12):3062. doi:10.3390/agronomy14123062.
97. Wang J, Qi Z, Wang Y, Liu Y. A lightweight weed detection model for cotton fields based on an improved YOLOv8n. *Sci Rep.* 2025;15(1):457. doi:10.1038/s41598-024-84748-8.
98. Saltik AO, Voigt M, Modak S, Beckworth M, Stein A. Improving lightweight weed detection via knowledge distillation. In: *Proceedings of the 2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*; 2025 Oct 19–20; Honolulu, HI, USA. p. 7219–28. doi:10.1109/iccwv69036.2025.00744.
99. Zhou P, Zhu Y, Jin C, Gu Y, Kong Y, Ou Y, et al. A new training strategy: coordinating distillation techniques for training lightweight weed detection model. *Crop Prot.* 2025;190:107124. doi:10.1016/j.cropro.2025.107124.
100. Mwitta C, Rains GC, Prostko E. Evaluation of inference performance of deep learning models for real-time weed detection in an embedded computer. *Sensors.* 2024;24(2):514. doi:10.3390/s24020514.
101. Ram B, Joy J, Marcotte N, Urlacher S, McDonald D, Amundson JK, et al. An Edge-AI enabled UAV system for site-specific application targeting Palmer amaranth in corn and soybean fields. *J Agric Food Res.* 2026;25(1):102577. doi:10.1016/j.jafr.2025.102577.
102. Tosin MC, Merotto Júnior A, Sulzbach E, Scheeren I, Bagavathiannan M, Markus C. Weed mapping using UAV imagery and AI techniques: current trends and challenges. *Pest Manag Sci.* 2025;81(12):7625–38. doi:10.1002/ps.70151.
103. Lu Y, Chen D, Olaniyi E, Huang Y. Generative adversarial networks (GANs) for image augmentation in agriculture: a systematic review. *Comput Electron Agric.* 2022;200(7):107208. doi:10.1016/j.compag.2022.107208.
104. Yang Q, Ye Y, Gu L, Wu Y. MSFCA-net: a multi-scale feature convolutional attention network for segmenting crops and weeds in the field. *Agriculture.* 2023;13(6):1176. doi:10.3390/agriculture13061176.
105. Wang Z, Chen W, Xing J, Zhang X, Tian H, Tang H, et al. Extracting vegetation information from high dynamic range images with shadows: a comparison between deep learning and threshold methods. *Comput Electron Agric.* 2023;208:107805. doi:10.1016/j.compag.2023.107805.