Weed Science

www.cambridge.org/wsc

CD-YOLO-based deep learning method for weed detection in vegetables

Wenpeng Zhu[1,2], Qiuyu Zu[3], Jinxu Wang[3], Teng Liu[3], Aniruddha Maity[4], Jihong Sun[3], Mian Li[3], Xiaojun Jin[5] and Jialin Yu[6]

[1]Intern, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China; [2]Visiting Student, National Engineering Research Center of Biomaterials, Nanjing Forestry University, Nanjing, China; [3]Research Assistant, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China; [4]Assistant Professor, Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL, USA; [5]Associate Professor, National Engineering Research Center of Biomaterials, Nanjing Forestry University, Nanjing, China and [6]Professor and Principal Investigator, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong

Research Article

Cite this article: Zhu W, Zu Q, Wang J, Liu T, Maity A, Sun J, Li M, Jin X, Yu J (2025). CD-YOLO-based deep learning method for weed detection in vegetables. Weed Sci. **73**(e99), 1–10. doi: 10.1017/wsc.2025.10074

Received: 29 May 2025
Revised: 9 September 2025
Accepted: 11 November 2025

Associate Editor:
Nathan S. Boyd, Gulf Coast Research and Education Center

Keywords:
Detection algorithms; image processing; precision weed management

Corresponding authors: Xiaojun Jin; Email: xjin@njfu.edu.cn; Jialin Yu; Email: jialin.yu@pku-iaas.edu.cn

Abstract

Computer vision–based precision weed control has proven effective in reducing herbicide usage, lowering weed management costs, and enhancing sustainability in modern agriculture. However, developing deep learning models remains challenging due to the effort required for weed dataset annotation and the difficulty of identifying weeds at different stages and densities in complex field conditions. To address these challenges, this study introduces an indirect weed detection method that combines deep learning and image processing techniques. The proposed approach first employs an object detection network to identify and label crops within the images. Subsequently, image processing techniques are applied to segment the remaining green pixels, thereby enabling indirect detection of weeds. Furthermore, a novel detection network—CD-YOLOv10n (You Only Look Once version 10 nano)—was developed based on the YOLOv10 framework to optimize computational efficiency. Redesigning the backbone (C2f-DBB) and integrating an optimized upsampling module (DySample) permitted the network to achieve higher detection accuracy while maintaining a lightweight structure. Specifically, the model achieved a mean average precision (mAP50) of 98.1%, which is a 1.4% percentage-point increase compared with the YOLOv10n baseline, a relevant improvement given the already strong baseline performance. At the same time, compared with YOLOv10n, its GFLOPs (giga floating-point operations per second) were reduced by 22.62%, and the number of parameters decreased by 15.87%. These innovations make CD-YOLOv10n highly suitable for deployment on resource-constrained platforms.

Introduction

Vegetables play a critical role in human health by enhancing immunity, preventing chronic diseases, and addressing global "hidden hunger" through their rich nutritional content and phytochemicals (Asaduzzaman et al. 2018). Alongside staple crops, vegetables rank among the most widely cultivated and economically significant crops worldwide (Dias and Ryder 2011). However, weeds pose a major challenge, not only by diminishing vegetable quality but also by causing yield losses of up to 45% to 90% (Mennan et al. 2020). Herbicides remain effective tools for weed suppression; however, excessive application can leave chemical residues in vegetables, increase environmental risks, and accelerate the evolution of resistant weed populations, which further complicates management (Mennan et al. 2020). Manual weeding, while reducing dependence on herbicides, is becoming increasingly unsustainable for large-scale agricultural production due to rising labor costs (Jin et al. 2021). Thus, developing automated, vision-based methods for distinguishing between crops and weeds has become increasingly essential for modern weed management (Jin et al. 2021).

Computer vision has shown significant potential for precision herbicide application in modern agriculture (Yu et al. 2019a, 2020). Current approaches for distinguishing between crops and weeds primarily rely on either traditional image processing techniques or deep learning methods (Wu et al. 2021). Traditional methods rely on features such as texture (Bakhshipour et al. 2017; Ishak et al. 2009), shape (Bakhshipour and Jafari 2018; Pereira et al. 2012), spectral properties (Elstone et al. 2020; Pignatti et al. 2017), and color (Hamuda et al. 2016; Rasmussen et al. 2019). However, relying on a single handcrafted feature (such as color or texture) is often insufficient for distinguishing crops from weeds, underscoring the need for multi-feature integration or deep learning–based approaches (Wu et al. 2021). To overcome this limitation, many studies have focused on integrating multiple features to improve

© The Author(s), 2025. Published by Cambridge University Press on behalf of Weed Science Society of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

detection accuracy (Sabzi et al. 2020). Machine learning techniques, such as support vector machines and artificial neural networks, have been widely employed for crop and weed classification (Behmann et al. 2015; Tellaeche et al. 2011). While these methods can accurately identify weeds under certain conditions, their reliance on single or manually designed features often limits their robustness and generalization ability, especially in complex and diverse agricultural environments (Kong et al. 2024; Wu et al. 2021).

In recent years, deep learning has been increasingly applied to the agricultural domain (Fu et al. 2020; Too et al. 2019). Deep learning–based approaches for weed detection and classification have demonstrated promising results (Tiwari et al. 2018; Yu et al. 2019b). Commonly used deep learning methods for weed detection include convolutional neural networks (CNNs) (Dyrmann et al. 2016; Olsen et al. 2019; Yu et al. 2019c, 2020) and fully convolutional networks (Huang et al. 2018). Yu et al. (2019a) evaluated several deep convolutional neural network models, including AlexNet, Visual Geometry Group Network (VGGNet), GoogleNet, and DetectNet, for detecting dandelion (*Taraxacum officinale* F.H. Wigg.), ground ivy (*Glechoma hederacea* L.), and spotted spurge [*Euphorbia maculata* L.; syn.: *Chamaesyce maculata* (L.) Small] in annual bluegrass (*Poa annua* L.). Among these, VGGNet achieved a high F_1 score of 92.78% and a recall of 99.52% in multi-class classification tasks, while DetectNet excelled in detecting *T. officinale* with an F_1 score of 98.43% and a recall of 99.11%. Similarly, Jin et al. (2022b) evaluated DenseNet, EfficientNetV2, ResNet, RegNet, and VGGNet models for both multi-class and binary classification of weed species. In multi-class classification, VGGNet achieved an F_1 score of 95.0% for detecting *T. officinale* and performed exceptionally well in identifying goosegrass [*Eleusine indica* (L.) Gaertn.], purple nutsedge (*Cyperus rotundus* L.), and white clover (*Trifolium repens* L.) in bermudagrass [*Cynodon dactylon* (L.) Pers.], with F_1 scores $\geq$98.3%. In binary classification, where the goal was to distinguish weed-infested regions from turfgrass, the EfficientNetV2 model performed best, achieving F_1 scores $\geq$98.1%. These results underscore the efficacy of deep learning–based models in addressing the challenges of weed detection and classification across different contexts and tasks.

Direct weed identification remains highly challenging due to the visual similarity between crops and weeds and the variability across field environments (Coleman et al. 2022; Jin et al. 2022a). Such morphological overlap and ecological variability introduce instability in feature extraction and reduce detection reliability (Coleman et al. 2022; Jin et al. 2022a; Zhuang et al. 2022). Another major bottleneck lies in dataset construction. Collecting and annotating sufficient images to represent diverse weed species across regions, growth stages, and densities requires immense effort, and the lack of such comprehensive datasets limits the robustness and generalization of deep learning models (Kong et al. 2024; Wu et al. 2021). This highlights the need for innovative approaches to overcome the limitations posed by data diversity and availability in weed detection tasks. To tackle these challenges, this study proposes a novel deep learning–based approach for training a vegetable detection model. First identifying vegetables within an image allows all remaining green regions outside the detected area to be classified as weeds, providing a streamlined and efficient method for weed detection. In this study, we developed an effective two-step approach for weed detection:

1. Building upon the YOLOv10 (A Wang et al. 2024), we propose a novel detection model, CD-YOLOv10n (C2f-DBB_Dysample_YOLOv10n), specifically designed for efficient vegetable identification. This model not only demonstrates outstanding performance in terms of detection accuracy but also achieves significant lightweighting, making it suitable for deployment in resource-constrained environments.
2. Once vegetables are identified, all green pixels outside the bounding box are classified as weeds. Weed detection and segmentation are then performed using image processing techniques, ensuring a streamlined and precise approach to weed detection.

Table 1. Number of images used for training, validation, and testing.

Dataset	Number of images	Proportion
Training	1,200	80%
Validation	150	10%
Testing	150	10%

Materials and Methods

Dataset

The images of vegetable seedlings used for training, validation, and testing in this study were collected in two batches from a vegetable farm on Bagua Island, Qixia District, Nanjing, China (32.2°N, 118.8°E), during July and September 2020. Each image had an original resolution of 4,032 × 3,024 pixels, and a total of 1,500 images were obtained. To enhance the diversity and generalization ability of the neural network, the dataset included images of bok choy [*Brassica rapa* subsp. *chinensis* (L.) Hanelt.] from vegetable fields with different sowing times, captured under various lighting conditions, including sunny and overcast weather. To reduce processing time and enhance real-time performance in field applications, all sample images were standardized to 1,400 × 1,050 pixels using a custom Python script. The collected images were annotated using LabelImg software, focusing on rectangular bounding boxes around vegetable seedlings. After annotation, corresponding extensible markup language (XML) label files were generated for each image, serving as training samples for the neural network model. The dataset was split into training (80%), validation (10%), and testing (10%) sets, as detailed in Table 1.

Improvement of YOLOv10n

YOLO is a widely recognized single-stage object detection algorithm, demonstrating excellent performance across various detection tasks. Compared with two-stage object detection algorithms, single-stage algorithms like YOLO integrate object localization and candidate region generation (first stage) with the classification phase (second stage), significantly improving detection efficiency.

To meet the need for cost-effective implementation, this study selected YOLOv10n, the simplest and most lightweight model in the YOLOv10 series, for detecting vegetable seedlings. However, the YOLOv10n algorithm has certain limitations in specific application scenarios, including limited feature extraction capacity that may hinder detection accuracy under complex and variable environmental conditions. Additionally, the model's ability to differentiate between morphologically similar objects may be

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

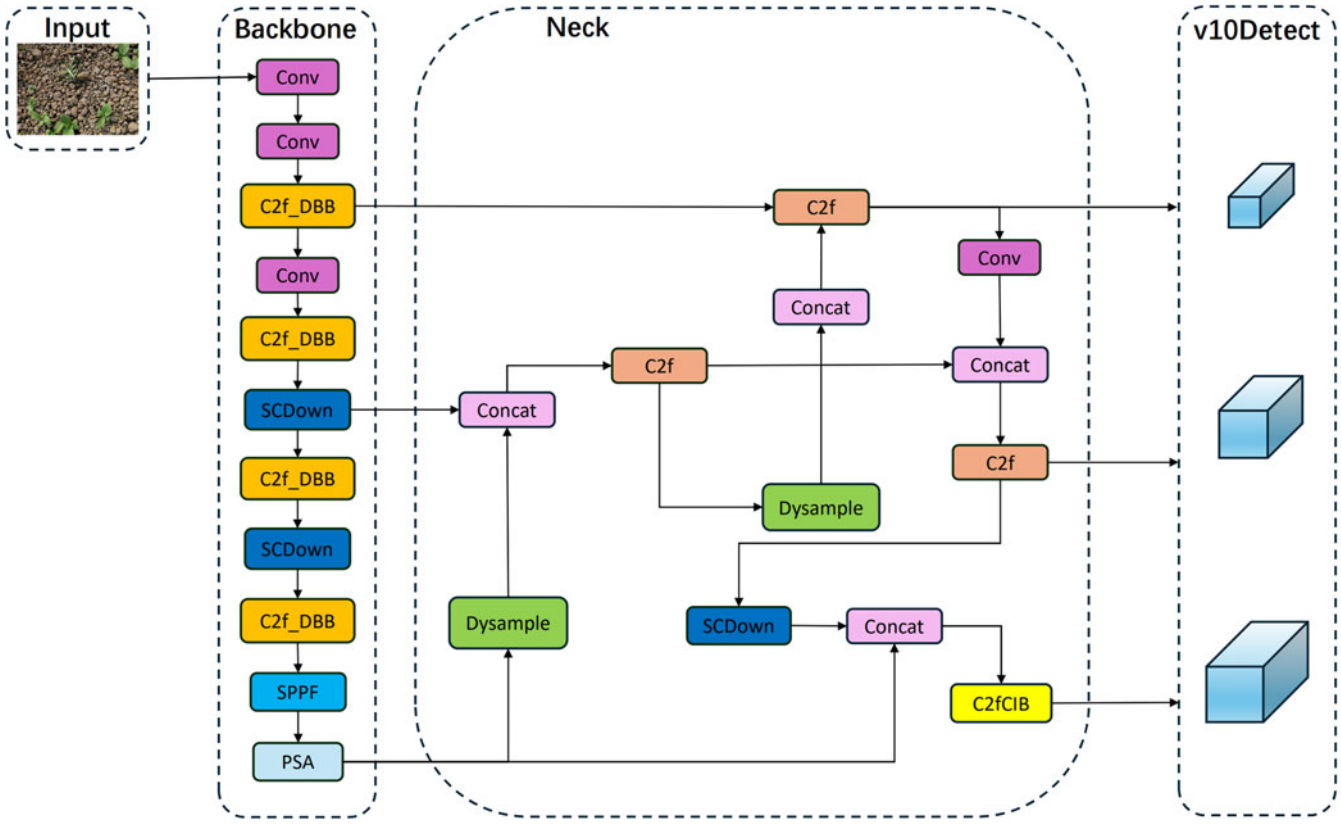


Figure 1. CD-YOLOv10n architecture.

insufficient, particularly in high-precision applications such as identifying specific plant species in diverse agricultural environments. To overcome these challenges and enhance the algorithm's performance in detecting vegetable seedlings, this study introduces several optimizations to the YOLOv10n framework. These improvements were designed to address its limitations while maintaining its efficiency and lightweight structure, ensuring better suitability for complex agricultural tasks. The algorithm was improved in two key aspects, as illustrated in Figure 1:

1. The Cross Stage Partial Network Fusion (C2f) module in YOLOv10n's backbone network was replaced with the C2f with Dynamic Block Branching (C2f-DBB) module (Zhang et al. 2024), resulting in a reduction in module size and improved efficiency.
2. The DySample module (Liu et al. 2023) was integrated into the neck network to replace the original upsampling mechanism in YOLOv10n, thereby enhancing the network's feature extraction capabilities.

C2f-DBB Module

The C2f-DBB module is an optimized version of the original C2f design, in which Dynamic Block Branching (DBB) is integrated into the bottleneck structure (Zhang et al. 2024). DBB incorporates six transformation modes, all of which can be converted into convolutions during inference, thereby improving the model's representation capability. In addition, the C2f-DBB module incorporates the Global Attention Mechanism, which strengthens feature interactions across both channel and spatial dimensions. By introducing DBB and attention mechanisms, the module significantly enhances feature extraction capabilities, leading to improved detection accuracy and model stability.

DySample Block

DySample is an ultra-lightweight and efficient dynamic upsampling technique that redefines the perspective of point sampling in the upsampling process. It avoids the high computational complexity and resource consumption inherent in traditional convolution-based methods. Notably, DySample operates without relying on high-resolution guiding features and is not constrained by additional CUDA package requirements. This results in significantly lower inference latency, memory usage, floating-point operations (FLOPs), and parameter count, optimizing both performance and resource efficiency. Liu et al. (2023) demonstrated that DySample outperforms other upsampling techniques across five major dense prediction tasks: semantic segmentation, object detection, instance segmentation, panoptic segmentation, and monocular depth estimation. In addition to its exceptional performance, DySample achieves efficiency comparable to bilinear interpolation. This makes it a reliable alternative to traditional methods like nearest neighbor or bilinear interpolation, offering a practical solution to enhance the performance and efficiency of existing dense prediction models.

Experimental Configuration

In this study, all deep neural network models were trained and tested under a consistent hardware and software environment to ensure reproducibility. The hardware setup included Ubuntu

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

Table 2. The hyperparameters for deep learning training.[a]

Hyperparameter[b]	Parameter value
Total epochs (*n*)	100
Batch size (images/iteration)	16
Number of workers (threads)	4
Learning rate (—)	0.001
Momentum (—)	0.937
Optimizer (algorithm)	SGD[c]

[a]Hyperparameter settings follow common practice for YOLO-family detectors (A Wang et al. 2024) and general principles of hyperparameter optimization (Yang and Shami 2020).
[b]A long dash (—) indicates a dimensionless parameter (no units).
[c]SGD, stochastic gradient descent.

20.04.6 LTS as the operating system, an Intel® Xeon® W-2265 CPU, and an Nvidia GeForce RTX 3080 Ti GPU. For the software environment, a configured Conda environment was utilized, which included Python 3.8, PyTorch 2.3.1, and CUDA 11.3. Details of the hyperparameter settings for the deep learning models are provided in Table 2. Additionally, to achieve better convergence, higher precision, and enhanced adaptability to real-world agricultural scenarios, Mosaic data augmentation was disabled during the final 10 training epochs.

Evaluation Metrics

This study employs multiple performance metrics, including precision, recall, mAP50, mAP50-95, and inference time, to comprehensively evaluate the performance of the deep learning models. The formulas for these metrics are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{mAP} = \text{AP} = \int_0^1 \text{Precision}(\text{Recall})\, d(\text{Recall}) \tag{3}$$

In this study, mAP is used to represent the average precision. Because our experimental case involves the identification of a single category, mAP specifically refers to the average precision (AP) for the vegetable category. True positive (TP) denotes the number of samples correctly identified as vegetables, false positive (FP) refers to the number of samples incorrectly identified as vegetables, and false negative (FN) represents the number of samples that were not correctly identified as vegetables.

Weed Segmentation

After crop bounding boxes were detected using the vegetable detector, green pixels outside the bounding boxes were segmented through a color-based image processing technique and marked as weeds. In this study, crops were detected but not segmented at the pixel level, which reduces annotation cost and computational complexity. However, this design may limit accurate identification of weeds located very close to seedlings, as bounding boxes cannot perfectly match crop boundaries. To enhance this process, the weed segmentation index proposed by Jin et al. (2021) was adopted and further optimized. Specifically, pixels were evaluated using the color index technique only if their green (G) component exceeded the red (R) or blue (B) components; otherwise, they were directly classified as non-weed pixels. This optimization improved segmentation accuracy while reducing computational complexity, resulting in a more efficient and precise weed detection workflow.

$$f(x) = \begin{cases} (\text{G} < \text{R} \| \text{G} < \text{B}) \| (-19\text{R} + 24\text{G} - 2\text{B} < 862), & \text{background} \\ \text{else}, & \text{weed} \end{cases} \tag{4}$$

Results and Discussion

Ablation Experiments of Each Module

Table 3 summarizes the performance metrics of the model after replacing individual components. The experimental results demonstrate that both the C2f-DBB and DySample strategies positively impacted model performance, with the C2f-DBB module achieving a mAP of 97.3% and the DySample module reaching 97.6% compared with 96.9% for the baseline YOLOv10n before adding these modules. However, it is noteworthy that while the DySample module improved overall detection accuracy, it led to a 1.2% reduction in recall. Both optimization components effectively reduced the model complexity, significantly decreasing the number of parameters and giga floating-point operations per second (GFLOPs). DySample exhibited the most pronounced impact on performance when replacing the neck network, whereas C2f-DBB achieved the greatest reduction in parameter count when replacing the backbone network. Both components contributed to a decrease in GFLOPs, indicating a reduction in the computational load of the neural network.

To validate the effectiveness of the proposed optimization strategies, ablation experiments were conducted for each module, with detailed results presented in Table 4. The findings reveal that the improved CD-YOLOv10n model demonstrates exceptional and efficient feature extraction capabilities. Built on a lightweight core architecture, the model achieves significant improvements in operational efficiency without compromising performance. Additionally, both computational costs and parameter counts are effectively reduced.

While the DySample module enhances overall performance, it comes at the expense of a slight reduction in recall value. In contrast, the C2f-DBB module improved performance without sacrificing recall. By leveraging the strengths of both components, the proposed model achieved a balanced trade-off, reducing overall parameters while enhancing performance. The proposed model outperformed YOLOv10n across key metrics, achieving improvements of +1.2% in mAP50, +2.1% in mAP50-95, +1.4% in precision, and +1.6% in recall. Meanwhile, the parameter counts and GFLOPs were reduced by 15.87% and 22.62%, respectively. These results highlight significant advances in both model performance and lightweight design, indicating the effectiveness of the proposed optimization strategies.

Training Performance of the Proposed Method Compared with YOLOv10n

As shown in Figure 2, CD-YOLOv10n achieved higher accuracy than YOLOv10n throughout training, with a consistently superior mAP50 curve. First, the model exhibited a faster convergence rate, with its mAP50 value surpassing that of YOLOv10n during the early training stages (around the 7th to 10th epoch). This indicates that CD-YOLOv10n is capable of extracting effective features at an

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

Table 3. Performance improvements achieved through the replacement of each component.

Model	mAP50	mAP50-95	Precision	Recall	Parameter quantity	GFLOPs[b]
	%					G
YOLOv10n	96.9	73.1	91.5	91.8	2,707,430	8.4
+C2f-DBB[a]	97.3	74.8	94.4	91.8	2,265,363	6.5
+DySample	97.6	73.9	94.1	90.6	2,277,715	6.5

[a]C2f-DBB, C2f with Dynamic Block Branching.
[b]GFLOPs, giga floating-point operations per second.

Table 4. Results of ablation experiments.

Method[a]	C2f-DBB[b]	DySample	mAP50	mAP50-95	Precision	Recall	Parameter quantity	GFLOPs[c]
			%					G
YOLOv10n			96.9	73.1	91.5	91.8	2,707,430	8.4
PM1	√		97.3	74.8	94.4	91.8	2,265,363	6.5
PM2		√	97.6	73.9	94.1	90.6	2,277,715	6.5
Ours	√	√	98.1	75.2	92.9	93.4	2,277,715	6.5

[a]PM, proposed method.
[b]C2f-DBB, C2f with Dynamic Block Branching.
[c]GFLOPs, giga floating-point operations per second.

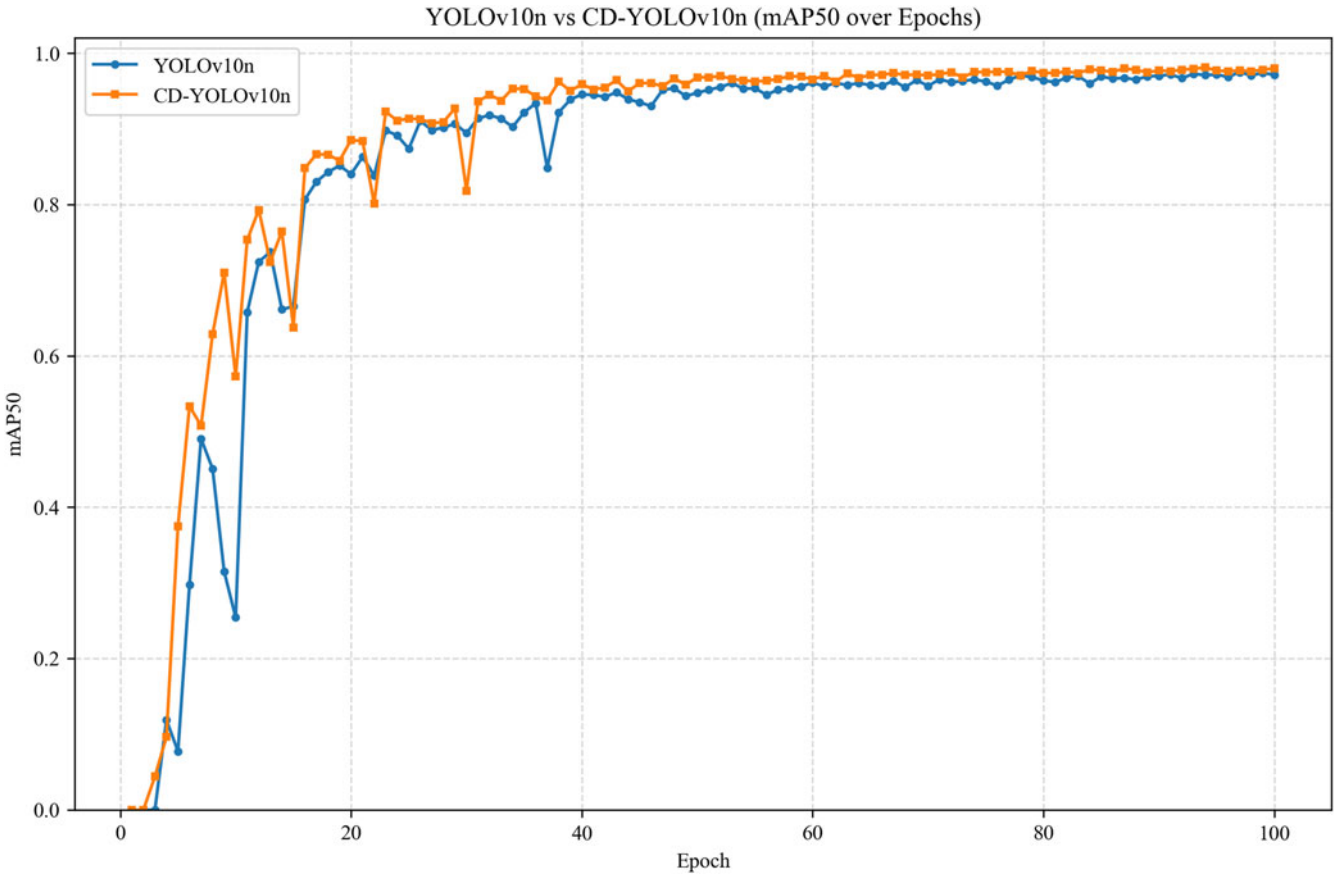


Figure 2. Training accuracy (mAP50) versus epoch (0–100) for YOLOv10n and CD-YOLOv10n. The *x* axis shows training epochs, and the *y* axis shows mAP50. Curves are averaged across epochs and smoothed with a three-epoch moving average for clarity.

earlier stage. Second, the training process of CD-YOLOv10n was notably more stable: as the mAP50 value approached 1, the curve showed smaller fluctuations and greater smoothness, suggesting improved reliability in later-stage predictions. Finally, although the final mAP50 values of the two models were very close, CD-YOLOv10n consistently maintained a slight advantage throughout most of the training process, consistently outperforming YOLOv10n. Overall, CD-YOLOv10n demonstrates superior performance in terms of convergence speed, stability, and training efficiency, highlighting the effectiveness of the proposed optimizations.

CD-YOLOv10n demonstrated significant advantages in training loss performance. The training loss curve reflects the optimization process of bounding box regression, where lower values indicate more accurate localization of objects. As shown in

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

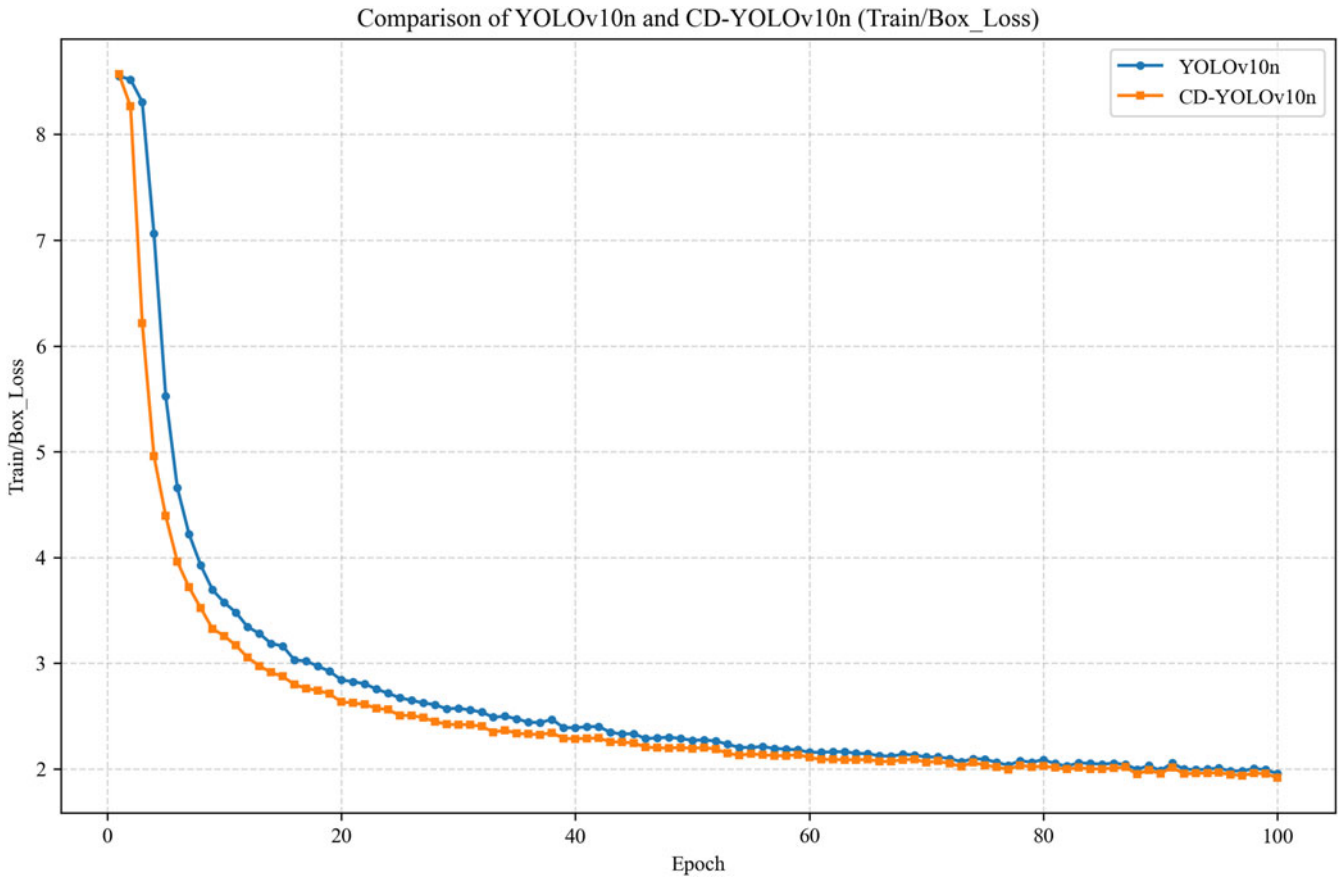


Figure 3. Training loss versus epoch (0–100) for YOLOv10n and CD-YOLOv10n. Loss represents the weighted sum of box regression, objectness, and classification components. Lower values indicate more accurate bounding box regression.

the training loss curve in Figure 3, CD-YOLOv10n exhibited a faster decline in loss during the initial training stages. Although its initial loss was slightly higher than that of YOLOv10n, it quickly surpassed YOLOv10n, indicating higher efficiency in the early learning phase. Further analysis revealed that the loss curve of CD-YOLOv10n remained smoother throughout the training process. In particular, during the mid- to late stages of training, the fluctuation amplitude was noticeably reduced, reflecting greater stability in the training process. Training stability refers to the smoothness and consistency of the optimization process, wherein fewer oscillations in the loss curve indicate more reliable convergence and reduced risk of overfitting. By the end of training, the loss value of CD-YOLOv10n was slightly lower than that of YOLOv10n, suggesting superior performance in bounding box precision and optimization. Overall, these findings highlight that CD-YOLOv10n outperforms YOLOv10n in both localization accuracy and training stability.

Comparison of CD-YOLOv10n and YOLOv10n in Vegetable Detection

As illustrated in Figure 4, the experimental results clearly demonstrate that CD-YOLOv10n achieved higher accuracy compared with YOLOv10n, particularly in terms of bounding box localization. Moreover, CD-YOLOv10n exhibited significantly greater robustness in handling complex scenarios, such as occlusion and overlapping targets. These improvements highlight the model's enhanced capability for precise detection under challenging conditions.

Weed Detection

As shown in Figure 5, the proposed method accurately detected vegetable seedlings and effectively segmented weeds, even in visually complex agricultural scenes. The pipeline maintained robustness under challenging conditions such as occlusion, illumination variation, and overlapping plants, highlighting its potential for practical weed management in real field environments.

Model Comparison Experiments

To comprehensively evaluate the efficiency, accuracy, and superiority of the CD-YOLOv10n model, this study compared its performance with various object detection models for vegetable detection. The results are summarized in Table 5. To gain deeper insights into the performance of each model, a comparative analysis was conducted using key metrics, including mAP, recall, parameter count, and GFLOPs. The reported values correspond to the worst-case results among 10 independent runs for each model, thereby providing a conservative evaluation of performance.

Although some models, such as YOLOX-Tiny (mAP50= 98.2%) and Real-Time Detection Transformer ResNet-18 (RT-DETR-R18) (mAP50 98.2%), achieved slightly higher mAP50 values compared with CD-YOLOv10n (mAP50 98.1%), they exhibited clear disadvantages in other critical areas. Specifically, YOLOX-Tiny had a 13.8% lower recall, along with higher parameter counts and GFLOPs, thereby limiting its suitability for resource-constrained environments. RT-DETR-R18, despite its

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

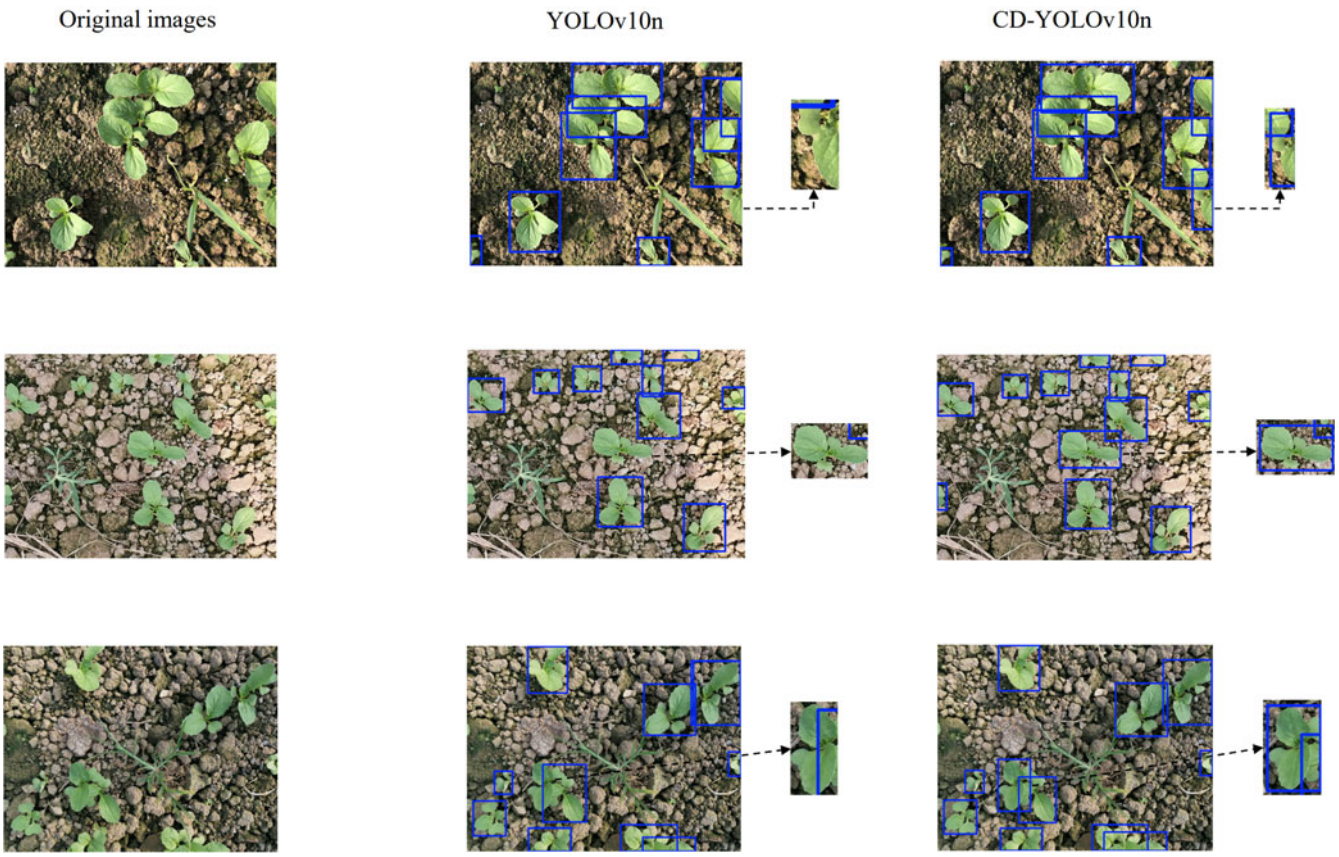


Figure 4. Vegetable detection results of YOLOv10n and CD-YOLOv10n on challenging field scenes. Columns show the original image, YOLOv10n output, and CD-YOLOv10n output. Smaller inset boxes highlight regions where the two models differ, with bounding boxes indicating predictions in difficult areas.

impressive mAP50 and recall (94.7%), required significantly higher parameter counts and GFLOPs than CD-YOLOv10n, leading to increased computational burdens and reduced efficiency for real-time applications. For other models, the performance gap with CD-YOLOv10n was even more pronounced, particularly in terms of recall, parameter count, and GFLOPs, further underscoring the advantages of CD-YOLOv10n.

In summary, although models like YOLOX-Tiny and RT-DETR-R18 achieve slightly higher mAP50, their lower recall or significantly higher computational demands make them less suitable for lightweight and deployment solutions. Considering all performance indicators, and given that even the worst-case results of CD-YOLOv10n remain competitive, the model achieves an optimal balance between accuracy, recall, and resource efficiency, making it the most robust and practical detection model for precision agriculture applications.

The proposed method demonstrates strong robustness and adaptability by combining deep learning–based vegetable detection with a color-based segmentation approach for weed identification. By leveraging bounding box information to isolate non-crop areas, the method narrows the scope of weed segmentation, effectively avoiding the challenges associated with directly recognizing diverse weed species. This design reduces reliance on large-scale annotated weed datasets, simplifying the data-collection process and improving the method's practicality in various agricultural scenarios.

Furthermore, the two-stage framework reduces potential errors in weed identification by focusing segmentation efforts on non-crop areas. This targeted approach enhances the method's reliability and ensures its applicability across diverse farming environments. By streamlining the weed identification process and eliminating the need for extensive weed datasets, the proposed method provides an efficient and practical solution for precision agriculture, addressing critical challenges in weed management with high accuracy.

It should be noted that a limitation of the proposed color-based segmentation approach is its sensitivity to illumination variations and the presence of non-weed green objects such as crop residues or algae. While the current implementation incorporates a green-dominance guard condition to mitigate some false positives, further improvements are needed. Future work could explore adaptive thresholding or color normalization in hue-saturation-value and hue-saturation-lightness color spaces to enhance robustness under varying light conditions. In addition, integrating multispectral or near-infrared information may further help discriminate weeds from non-weed vegetation in complex field environments.

This study developed an efficient vegetable recognition model, CD-YOLOv10n, which demonstrated exceptional performance with a mAP50 of 98.1% and a recall of 93.4%. The model also significantly reduced computational costs, as demonstrated by a notable decrease in GFLOPs, improving its resource efficiency and suitability for practical deployment. Furthermore, an innovative indirect weed detection strategy was introduced, requiring only crop annotations during training. By combining crop identification with image processing techniques, this approach effectively detected weeds in non-crop regions, addressing limitations of traditional direct weed detection methods that rely heavily on

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

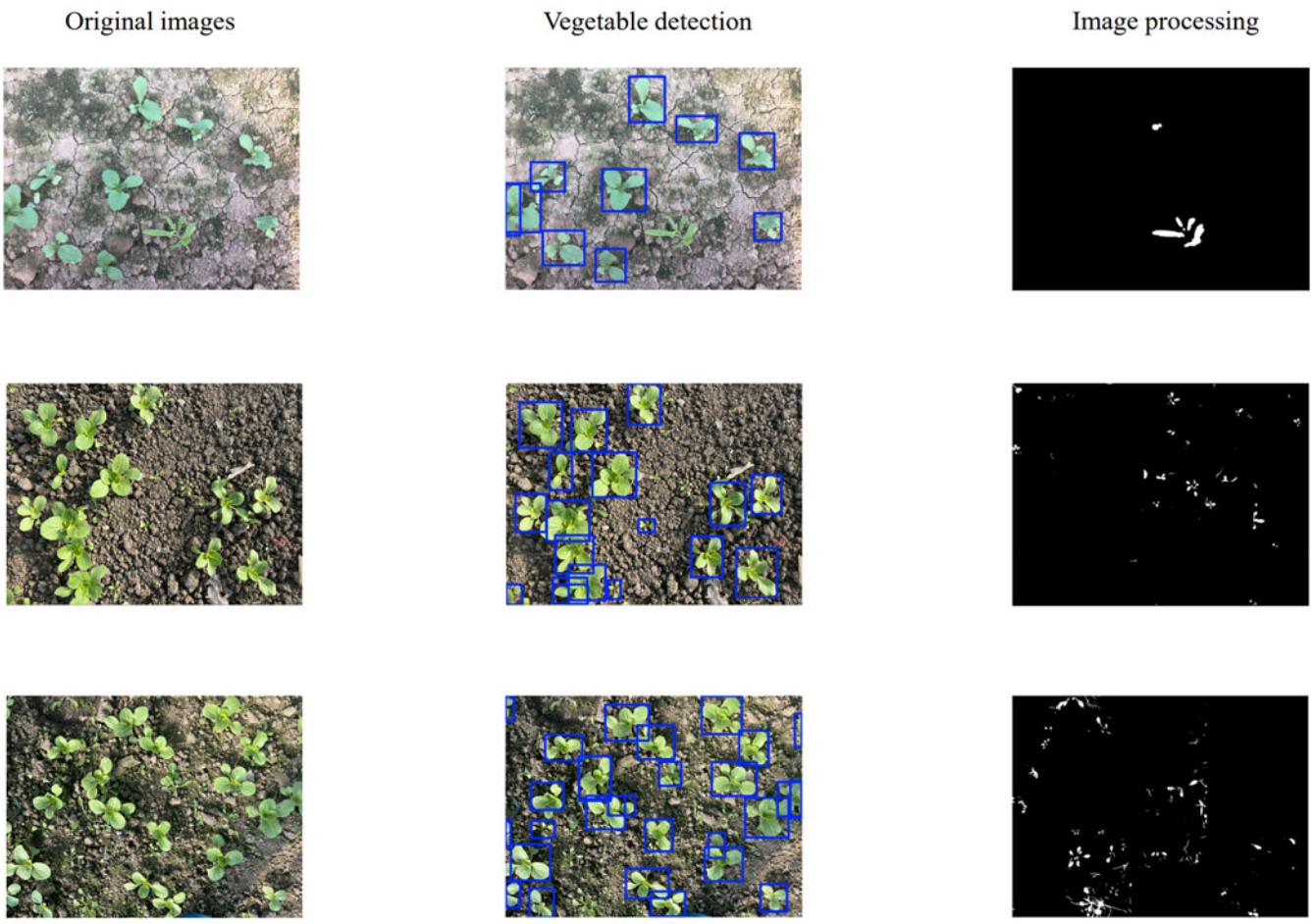


Figure 5. Vegetable detection and weed segmentation results. Columns show the original image, CD-YOLOv10n detection (vegetable bounding boxes), and the segmentation output where green pixels outside the boxes are classified as weeds.

Table 5. Performance comparison of detection models.[a]

Model[b]	mAP50	mAP50-95	Recall	Parameter quantity	GFLOPs[c]
		%			G
ATSS-R50 (S Zhang et al. 2020)	95.4	63.6	72.1	32,113,000	171
AutoAssign-R50 (Zhu et al. 2020)	94.1	59.1	68.1	36,244,000	168
Cascade R-CNN-R50-FPN (Cai and Vasconcelos 2019)	96.0	68.0	74.3	69,152,000	205
DINO (Zhang et al. 2022)	87.8	66.9	84.0	47,540,000	235
Dynamic R-CNN-R50 (H Zhang et al. 2020)	93.3	66.8	73.8	5,033,000	7.571
Faster R-CNN-R50-FPN-CIOU (Ren et al. 2016)	96.5	67.1	73.6	41,394,000	178
FreeAnchor-R50 (Zhang et al. 2019)	94.5	59.3	66.4	36,330,000	173
FCOS-R50 (Tian et al. 2019)	79.2	40.4	55.7	32,113,000	167
FSAF (Zhu et al. 2019)	95.8	59.5	66.7	36,238,000	172
GFL (Li et al. 2020)	95.1	64.8	73.7	32,258,000	174
Grid R-CNN (Lu et al. 2019)	94.9	62.2	68.8	64,467,000	291
TOOD-R50 (Feng et al. 2021)	95.1	64.8	72.8	32,018,000	168
YOLOX-Tiny (Ge et al. 2021)	98.2	75.3	79.6	5,033,000	7.571
YOLOv9t (CY Wang et al. 2024)	97.9	76.5	93.8	1,970,979	7.6
RT-DETR-R18 (Jun et al. 2023)	98.2	78.8	94.7	19,873,044	56.9
YOLOv10 (A Wang et al. 2024)	96.9	73.1	91.8	2,707,430	8.4
CD-YOLOv10	98.1	75.2	93.4	2,277,715	6.5

[a]Values represent the worst results from 10 independent runs for each model, providing a conservative assessment of performance.
[b]ATSS, adaptive training sample selection; R50, ResNet-50; R-CNN, region-based convolutional neural network; FPN, Feature Pyramid Network; DINO, DETR with improved denoising anchor boxes for end-to-end object detection; CIOU, complete intersection over union; FCOS, fully convolutional one-stage; FSAF, feature selective anchor-free; GFL, generalized focal loss; TOOD, task-aligned one-stage object detection; YOLO, You Only Look Once; RT-DETR-R18, Real-Time Detection Transformer ResNet-18.
[c]GFLOPs, giga floating-point operations per second.

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

labor-intensive annotations. This proposed strategy improved robustness and adaptability to varying field conditions, effectively handling challenges such as species diversity, weed density, and growth stages. Future research should focus on validating the method across diverse agricultural scenarios to further enhance its practical applicability, including other vegetable species and mixed cropping systems, as differences in morphology, planting patterns, and canopy structure may affect model performance.

This study proposed CD-YOLOv10n, a lightweight detection model for indirect weed identification. By integrating C2f-DBB and DySample, the model achieved superior accuracy (mAP50 98.1%, recall 93.4%) while reducing parameters and GFLOPs compared with YOLOv10n. The indirect weed detection strategy, based on crop detection followed by optimized color-index segmentation, reduced annotation costs and improved robustness under complex conditions.

While this study validated the approach on bok choy, the proposed pipeline has broader applicability. Because the method relies on detecting crops rather than classifying diverse weed species, it is less sensitive to the variability of weeds across environments. For adaptation to other vegetables or mixed cropping systems, the model would only require retraining on the limited set of crop classes relevant to the target field. Once the crop(s) are reliably identified, all non-crop vegetation can still be indirectly classified as weeds, regardless of species composition. This design reduces the need for extensive weed-specific annotations and highlights the scalability of the method. Nonetheless, additional validation across different vegetables and cropping patterns is necessary to confirm robustness under more diverse agronomic conditions.

Funding statement. This work was supported by the Weifang Science and Technology Development Plan Project (grant no. 2024ZJ1097), Shandong Provincial Natural Science Foundation (grant no. SYS202206), the National Natural Science Foundation of China (grant no. 32072498), Taishan Scholar Program of Shandong Province, and the Yuandu Scholar, Program of Weifang, Shandong China.

Competing interests. The authors declare no conflicts of interest.

References

Asaduzzaman M, Asao T, Amao I (2018) Vegetables—importance of quality vegetables to human health. Pages 1–18 *in* Asaduzzaman M, Asao T, eds. Vegetables—Importance of Quality Vegetables to Human Health. London: IntechOpen

Bakhshipour A, Jafari A (2018) Evaluation of support vector machine and artificial neural networks in weed detection using shape features. Comput Electron Agric 145:153–160

Bakhshipour A, Jafari A, Nassiri SM, Zare D (2017) Weed segmentation using texture features extracted from wavelet sub-images. Biosyst Eng 157:1–12

Behmann J, Mahlein A-K, Rumpf T, Römer C, Plümer L (2015) A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. Precis Agric 16:239–260

Cai Z, Vasconcelos N (2019) Cascade R-CNN: high quality object detection and instance segmentation. Pages 1483–1492 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society

Coleman GR, Bender A, Hu K, Sharpe SM, Schumann AW, Wang Z, Bagavathiannan MV, Boyd NS, Walsh MJ (2022) Weed detection to weed recognition: reviewing 50 years of research to identify constraints and opportunities for large-scale cropping systems. Weed Technol 36:741–757

Dias JS, Ryder E (2011) World vegetable industry: production, breeding, trends. Pages 299–356 *in* Janick J, ed. Horticultural Reviews. Volume 38. Hoboken, NJ: Wiley

Dyrmann M, Karstoft H, Midtiby HS (2016) Plant species classification using deep convolutional neural network. Biosyst Eng 151:72–80

Elstone L, How KY, Brodie S, Ghazali MZ, Heath WP, Grieve B (2020) High speed crop and weed identification in lettuce fields for precision weeding. Sensors 20:455

Feng C, Zhong Y, Gao Y, Scott MR, Huang W (2021) TOOD: task-aligned one-stage object detection. Pages 3490–3499 *in* Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos, CA: IEEE Computer Society

Fu L, Gao F, Wu J, Li R, Karkee M, Zhang Q (2020) Application of consumer RGB-D cameras for fruit detection and localization in field: a critical review. Comput Electron Agric 177:105687

Fu X, Qu H (2017) Research on semantic segmentation of high-resolution remote sensing image based on full convolutional neural network. Pages 1–4 *in* Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Los Alamitos, CA: IEEE Computer Society

Ge Z, Liu S, Wang F, Li Z, Sun J (2021) YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430

Hamuda E, Glavin M, Jones E (2016) A survey of image processing techniques for plant extraction and segmentation in the field. Comput Electron Agric 125:184–199

Huang H, Deng J, Lan Y, Yang A, Deng X, Wen S, Zhang H, Zhang Y (2018) Accurate weed mapping and prescription map generation based on fully convolutional networks using UAV imagery. Sensors 18:3299

Ishak AJ, Hussain A, Mustafa MM (2009) Weed image classification using Gabor wavelet and gradient field distribution. Comput Electron Agric 66:53–61

Jin X, Bagavathiannan M, Maity A, Chen Y, Yu J (2022a) Deep learning for detecting herbicide weed control spectrum in turfgrass. Plant Methods 18:94

Jin X, Bagavathiannan M, McCullough PE, Chen Y, Yu J (2022b) A deep learning-based method for classification, detection, and localization of weeds in turfgrass. Pest Manag Sci 78:4809–4821

Jin X, Che J, Chen Y (2021) Weed identification using deep learning and image processing in vegetable plantation. IEEE Access 9:10940–10950

Jun ELT, Tham M-L, Kwan BH (2023) A comparative analysis of RT-DETR and YOLOv8 for urban zone aerial object detection. Pages 340–345 *in* Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos, CA: IEEE Computer Society

Kong X, Liu T, Chen X, Jin X, Li A, Yu J (2024) Efficient crop segmentation net and novel weed detection method. Eur J Agron 161:127367

Li X, Wang W, Wu L, Chen S, Hu X, Li J, Tang J, Yang J (2020) Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. Adv Neural Inf Process Syst 33:21002–21012

Liu W, Lu H, Fu H, Cao Z (2023) Learning to upsample by learning to sample. Pages 6027–6037 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society

Lu X, Li B, Yue Y, Li Q, Yan J (2019) Grid R-CNN. Pages 7363–7372 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society

Mennan H, Jabran K, Zandstra BH, Pala F (2020) Non-chemical weed management in vegetables by using cover crops: a review. Agronomy 10:257

Olsen A, Konovalov DA, Philippa B, Ridd P, Wood JC, Johns J, Banks W, Girgenti B, Kenny O, Whinney J (2019) DeepWeeds: a multiclass weed species image dataset for deep learning. Sci Rep 9:2058

Pereira LA, Nakamura RY, De Souza GF, Martins D, Papa JP (2012) Aquatic weed automatic classification using machine learning techniques. Comput Electron Agric 87:56–63

Pignatti S, Casa R, Harfouche A, Huang W, Palombo A, Pascucci S (2017) Maize crop and weeds species detection by using uav vnir hyperpectral data. Pages 7235–7238 *in* Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Los Alamitos, CA: IEEE Computer Society

Rasmussen J, Nielsen J, Streibig J, Jensen J, Pedersen K, Olsen S (2019) Pre-harvest weed mapping of *Cirsium arvense* in wheat and barley with off-the-shelf UAVs. Precis Agric 20:983–999

Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press

Sabzi S, Abbaspour-Gilandeh Y, Arribas JI (2020) An automatic visible-range video weed detection, segmentation and classification prototype in potato field. Heliyon 6:e03787

Tellaeche A, Pajares G, Burgos-Artizzu XP, Ribeiro A (2011) A computer vision approach for weeds identification through support vector machines. Appl Soft Comput 11:908–915

Tian Z, Shen C, Chen H, He T (2019) FCOS: fully convolutional one-stage object detection. Pages 9627–9636 *in* Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos, CA: IEEE Computer Society

Tiwari O, Goyal V, Kumar P, Vij S (2018) An experimental set up for utilizing convolutional neural network in automated weed detection. Pages 1–6 *in* Proceedings of the IEEE International Conference on Intelligent Systems. Los Alamitos, CA: IEEE Computer Society

Too EC, Yujian L, Njuki S, Yingchun L (2019) A comparative study of fine-tuning deep learning models for plant disease identification. Comput Electron Agric 161:272–279

Wang A, Chen H, Liu L, Chen K, Lin Z, Han J (2024) Yolov10: real-time end-to-end object detection. Adv Neural Inf Process Syst 37:107984–108011

Wang CY, Yeh IH, Liao HYM (2024) Yolov9: learning what you want to learn using programmable gradient information. Pages 1–21 *in* Computer Vision—ECCV 2022. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland.

Wu Z, Chen Y, Zhao B, Kang X, Ding Y (2021) Review of weed detection methods based on computer vision. Sensors 21:3647

Yang L, Shami A (2020) On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415:295–316

Yu J, Schumann AW, Cao Z, Sharpe SM, Boyd NS (2019a) Weed detection in perennial ryegrass with deep learning convolutional neural network. Front Plant Sci 10:1422

Yu J, Schumann AW, Sharpe SM, Li X, Boyd NS (2020) Detection of grassy weeds in bermudagrass with deep convolutional neural networks. Weed Sci 68:545–552

Yu J, Sharpe SM, Schumann AW, Boyd NS (2019b) Deep learning for image-based weed detection in turfgrass. Eur J Agron 104:78–84

Yu J, Sharpe SM, Schumann AW, Boyd NS (2019c) Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. Pest Manag Sci 75:2211–2218

Zhang H, Chang H, Ma B, Wang N, Chen X (2020) Dynamic R-CNN: towards high quality object detection via dynamic training. Pages 260–275 *in* Computer Vision—ECCV 2020. Lecture Notes in Computer Science. Volume 12349. Cham: Springer International Publishing.

Zhang H, Li F, Liu S, Zhang L, Su H, Zhu J, Ni LM, Shum HY (2022) DINO: DETR with improved denoising anchor boxes for end-to-end object detection. arXiv:2203.03605

Zhang S, Chi C, Yao Y, Lei Z, Li SZ (2020) Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. Pages 9759–9768 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society

Zhang X, Wan F, Liu C, Ji R, Ye Q (2019) Freeanchor: learning to match anchors for visual object detection. Adv Neural Inf Process Syst 32:14789–14800

Zhang Z, Wang X, Wang L, Xia X (2024) Surface defect detection method for discarded mechanical parts under heavy rust coverage. Sci Rep 14:7963

Zhu B, Wang J, Jiang Z, Zong F, Liu S, Li Z, Sun J (2020) Autoassign: differentiable label assignment for dense object detection. arXiv:2007.03496

Zhu C, He Y, Savvides M (2019) Feature selective anchor-free module for single-shot object detection. Pages 840–849 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society

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

https://doi.org/10.1017/wsc.2025.10074 Published online by Cambridge University Press