Weed Science

www.cambridge.org/wsc

WeedDETR: an efficient and accurate detection method for detecting small-target weeds in UAV images

Shengxian Yang[1,*], Jianwu Lin[2,*], Tomislav Cernava[3], Xiaoyulong Chen[4,5,6] and Xin Zhang[7]

[1]Master's Student, College of Big Data and Information Engineering, Guizhou University, Guiyang, China; [2]Doctoral Student, State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang, China; [3]Professor, School of Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK; [4]Professor, College of Life Sciences, Guizhou University, Guiyang, China; [5]Guizhou-Europe Environmental Biotechnology and Agricultural Informatics Oversea Innovation Center in Guizhou University, Guizhou Provincial Science and Technology Department, Guiyang, China; [6]International Jointed Institute of Plant Microbial Ecology and Resource Management in Guizhou University, China Association of Agricultural Science Societies, Guiyang, China and [7]Associate Professor, College of Big Data and Information Engineering, Guizhou University, Guiyang, China

Research Article

Cite this article: Yang S, Lin J, Cernava T, Chen X, Zhang X (2025). WeedDETR: an efficient and accurate detection method for detecting small-target weeds in UAV images. Weed Sci. **73**(e84), 1–14. doi: 10.1017/wsc.2025.10035

Received: 26 April 2025
Revised: 1 June 2025
Accepted: 13 June 2025

Associate Editor:
Julian Yu, Peking University-IAAS

Keywords:
DETR; real-time detection; site-specific weed management; UAVs; weed detection

Corresponding authors: Xiaoyulong Chen; Email: ylcx@gzu.edu.cn; Xin Zhang; Email: xzhang1@gzu.edu.cn

*These authors contributed equally to this work.

Abstract

Site-specific weed management (SSWM) provides precise weed control and reduces the use of herbicides, which not only reduces the risk of environmental damage but also improves agricultural productivity. Accurate and efficient weed detection is the foundation for SSWM. However, complex field environments and small-target weeds in fields pose challenges for their detection. To address the above limitations, we developed WeedDETR, a real-time end-to-end detection model specifically designed to enhance the detection of small-target weeds in unmanned aerial vehicle (UAV) imagery. WeedDETR incorporates RepCBNet, a backbone network optimized through structural re-parameterization, to improve fine-grained feature extraction and accelerate inference. In addition, the designed feature complement fusion module (FCFM) was used for multi-scale feature fusion to alleviate the problem of small-target weed information being ignored in the deep network. During training, varifocal loss was used to focus on high-quality weed samples. We experimented on a new dataset, GZWeed, which contains weed imagery captured by a UAV. The experimental results demonstrated that WeedDETR achieves 73.9% and 91.8% AP0.5 (average precision at 0.5 intersection over union threshold) in the weed and Chinese cabbage [*Brassica rapa* subsp. *chinensis* (L.) Hanelt] categories, respectively, while achieving an inference speed of 76.28 frames per second (FPS). In comparison to YOLOv5-L, YOLOv6-M, and YOLOv8-L, WeedDETR demonstrated superior accuracy and speed, exhibiting 3.5%, 6.3%, and 3.6% higher AP0.5 for weed categories, while FPS was 14.9%, 12.9%, and 1.4% higher, respectively. The innovative architectural design of WeedDETR significantly enhances the detection accuracy of small-target weeds, enabling efficient end-to-end weed detection. The proposed method establishes a solid technological foundation for UAV-based precision weeding systems in field conditions, advancing the development of deep learning–driven intelligent weed management.

Introduction

In agroecosystems, weeds are considered a major problem, as they compete with crops for nutrients, water, and sunlight and also provide a habitat for pests that can cause plant diseases, leading to a reduction in crop yield and quality. Site-specific weed management (SSWM) is seen as a viable solution to control weeds by precisely limiting weed growth in a specific location (Rai et al. 2023). The use of precise weed control methods such as spot spraying of herbicides can reduce the quantity of herbicides used in the field and prevent pesticide residues (Gerhards et al. 2022).

Accurate detection of weeds in real time while avoiding crop damage is essential for the realization of SSWM. Unmanned aerial vehicles (UAVs) are an ideal platform for weed detection, because they are able to acquire weed imagery without crop damage, efficiently provide information on weed location, and adapt to the spatial and temporal heterogeneity of weed distribution (Valente et al. 2022). Crop and weed morphology, which can also be subject to substantial variations depending on genetics and the environment, are characteristics that pose great challenges for weed detection algorithms (Hu et al. 2023). Moreover, weeds occupy small pixels in aerial weed images compared with proximal remote sensing, making their detection

© 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.

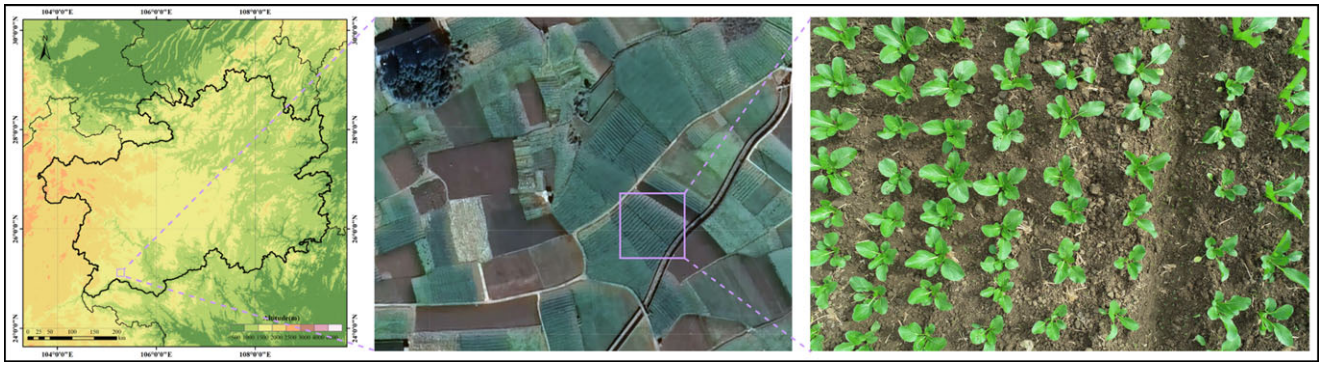


Figure 1. Location of the study site: Anlong County, Guizhou Province, China.

more difficult. It is therefore essential to develop an accurate real-time weed detection model that can capture characteristics of small targets in UAV images.

Initial weed detection algorithms were based on traditional machine learning techniques, which required manual information extraction based on the morphological and textural features of weeds, influenced by the prior knowledge of researchers (Reedha et al. 2022). An object-based image analysis algorithm enabled a three-class weed density map by processing multispectral UAV data from maize (*Zea mays* L.) fields, effectively quantifying spatial distributions of weed coverage (Peña et al. 2013). The random forest and *k*-nearest neighbors algorithms demonstrated effective detection performance when applied to calibrated and stitched UAV-derived orthophotos of weed in chili (*Capsicum annuum* L.) fields (Islam et al. 2021). Comparative assessment of four approaches demonstrated the automatic object-based classification method achieved optimal performance with 89% accuracy in oat (*Avena sativa* L.) field weed classification research (Gašparović et al. 2020). These results indicate that weeds can be identified using traditional machine learning methods, but their detection models have cumbersome steps, and most of them are based on area detection of weed density with low detection accuracy.

With the development of computer vision, deep learning methods have become widely used in agriculture (Li et al. 2023; Lin et al. 2023; Miho et al. 2024). In a soybean [*Glycine max* (L.) Merr.] field weed detection task, the object-based Faster regions with convolutional neural networks (R-CNN) achieved 65% accuracy, 68% recall, and a 66% F1 score (the harmonic mean of precision and recall), all of which outperformed the patch-based convolutional neural networks model, indicating superior performance (Veeranampalayam Sivakumar et al. 2020). A benchmark study of seven you only look once (YOLO) versions for cotton (*Gossypium hirsutum* L.) field weed detection indicated YOLOv4 exhibited optimal detection capabilities with the highest mean average precision at 0.5 intersection over union threshold (mAP0.5), whereas the YOLOv3-tiny model had a low detection accuracy (Dang et al. 2023). An enhanced YOLOv7 developed for weed detection in chicory (*Cichorium intybus* L.) fields achieved 56.6% mAP0.5, 62.1% recall, and 61.3% precision, showing improvements over baseline models (Gallo et al. 2023). The integration of the convolutional block attention module (CBAM) mechanism into YOLOv5 improves its capacity to detect weeds on a multi-granularity buffalobur (*Solanum rostratum* Dunal) field weed dataset (Wang et al. 2022). In rice (*Oryza sativa* L.) paddy weed detection research utilizing mobile platforms, RetinaNet improved recognition accuracy by combining SmoothL1 loss and achieved 94.1% mAP0.5 while retaining inference speed (Peng et al. 2022). While existing studies demonstrate the superior recognition accuracy and complex background robustness of deep learning methods compared with conventional machine learning methods, current models inadequately address the challenge of detecting small-target weeds in UAV-captured imagery.

Current deep learning object detection models can be categorized into two-stage detectors represented by the R-CNN series (Girshick et al. 2014; He et al. 2017; Ren et al. 2017) and one-stage detectors represented by the YOLO series (Bochkovskiy et al. 2020; Jocher 2020; Redmon et al. 2016; Redmon and Farhadi 2018; Ultralytics 2023) and the detection transformer (DETR) series (Carion et al. 2020; Zhang et al. 2022; Zhu et al. 2021), depending on whether the processing is required to generate region proposals or not. Compared with two-stage detectors, one-stage detectors are more computationally efficient, have faster inference, and are widely used for real-time detection. Nevertheless, the YOLO series needs to select the hyperparameter non-maximum suppression (NMS) based on experience, which has a great impact on the accuracy and speed of model detection. DETR employs the Transformer (Vaswani et al. 2017) encoder–decoder architecture, which uses bipartite matching to achieve the prediction of the target through ensemble-based global loss, avoiding the hand-designed steps of NMS and anchor generation. RT-DETR (real-time detection transformer) achieves real-time end-to-end detection through model architecture redesign and outperforms the YOLO series in terms of accuracy and inference speed on the COCO 2017 dataset (Lv et al. 2023).

Currently, the main challenges faced by SSWM are the lack of weed detection datasets acquired using UAVs and the insufficient ability of the model to detect small-target weeds (Khan et al. 2021). Inspired by this, we developed a weed detection model using DETR with end-to-end detection properties for the challenge of a large number of small-target weeds in UAV-captured weed imagery.

Materials and Methods

Data Acquisition

The study site is located in Anlong County (25.04°N, 105.25°E), Guizhou Province, China, as shown in Figure 1. The GZWeed dataset was collected on November 21, 2023, by a DJI Phantom 4 RTK (DJI, Shenzhen, China) UAV carrying a DJI FC6310R camera. The shooting angle is vertical to the ground. The undulating mountainous terrain caused the altitude above ground level of the UAV to vary from 2.42 to 3.79 m, with a mean altitude of 3.09 m corresponding to a ground coverage area of 13.92 m^2. Also, manual

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

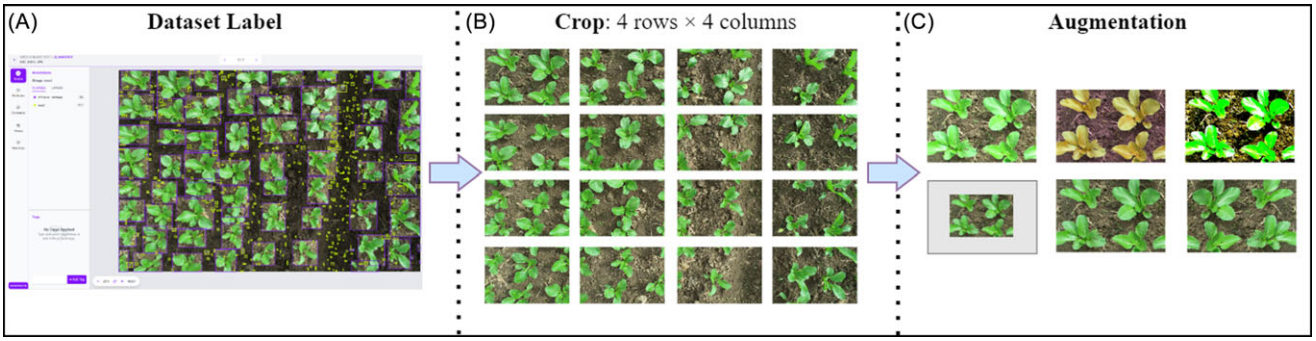


Figure 2. Flowchart of dataset preprocessing: (A) Dataset label with Roboflow, (B) image crop, and (C) augmented image.

Table 1. The number of images of object classes in GZWeed dataset.

	Object no.			
Class	Total	Training	Validation	Testing
Chinese cabbage	11,014	8,749	1,173	1,092
Weed	17,279	13,893	1,657	1,729
Total	28,293	22,642	2,830	2,821

planting irregularities during cultivation resulted in uneven Chinese cabbage [*Brassica rapa* subsp. *chinensis* (L.) Hanelt] spacing, with average row and plant spacings of 45 cm and 35 cm, respectively. We performed quality screening after image acquisition, resulting in 108 images of weeds in Chinese cabbage fields.

Image Preprocessing

The dataset was labeled with Roboflow (Roboflow 2025) to annotate the weed and Chinese cabbage locations and to generate the corresponding label files of the images used for training, as shown in Figure 2A. The original images have a raw resolution of 5,472 × 3,648 pixels. To avoid the loss of details caused by the compression of the image information on the original image during the input of the detection model, the original images were cropped 4 × 4, yielding 1,728 images, as shown in Figure 2B. The dataset was divided into a training set (1,382 images), a validation set (173 images), and a test set (173 images) according to the ratio of 8:1:1, and the number of instances is shown in Table 1. Weed species were not distinguished due to the small size of the weed target in the image.

To enhance the robustness of the model in the field environment, data augmentation methods, including proportional scaling, panning, horizontal mirroring, contrast enhancement, saturation enhancement, and brightness adjustment, are used to augment the images online. The partially augmented image is shown in Figure 2C.

The photographed weeds have variable light intensity and angles. In addition, the complex background of large quantities of dry rice straw and wet soil presented a challenge for weed detection. It can be seen from Figure 3 that there are a large number of small-target weeds, and some of the weeds are obscured by the crops, both of which present challenges for detection.

Experimental Configuration

The following experiments were performed on the GZWeed dataset. The experimental parameters used for model training are shown in Table 2.

To achieve a fair comparison of model performance, all models were trained from scratch for 250 epochs with a batch size of 16. The bounding box regression uses the GIoU (generalized intersection over union) loss (Rezatofighi et al. 2019), which is formulated as follows:

$$GIoU = \frac{(\tilde{x} \cap x)}{(\tilde{x} \cup x)} - \frac{|u/(\tilde{x} \cup \tilde{x})|}{|u|} \quad (1)$$

where x represents the ground truth box, $\tilde{x}$ represents the prediction box, and u is the smallest bounding box that contains both x and $\tilde{x}$.

Considering the graphics processing unit memory, the input images were scaled to 640 × 640 during training, and each model was tested using the model weights with the highest mAP0.5 on the validation set. We train all models using the AdamW optimizer (Loshchilov and Hutter 2019) with a 0.0001 base learning rate, 0.0001 weight decay, and 2,000 warm-up epochs.

Performance Metrics

To validate the detection performance of the proposed model, mAP0.5 and mAP0.5:0.95 are used as performance evaluation metrics. Precision is the ratio of the number of positive samples detected by the model to the number of correctly detected samples, as shown in the formula:

$$\text{Precision} = \frac{\text{TP} + \text{FP}}{\text{TP}} \quad (2)$$

Recall is the ratio of the number of positive samples correctly detected by the model to the actual number of positive samples. It is calculated by the following formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

The average precision (AP) is equal to the area under the precision-recall (PR) curve and is calculated as shown:

$$\text{AP} = \int_0^1 \text{Precision}(\text{Recall}) d(\text{Recall}) = \int_0^1 p(r) dr \quad (4)$$

Mean average precision (mAP) is the result obtained by weighted average of the AP values for all sample categories with the formula shown below:

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press



Figure 3. Representative samples from the weed dataset.

Table 2. Experimental configuration.

Name	Parameter
Central processing unit	Intel(R) Xeon (R) W-1390
Graphics processing unit	NVIDIA GeForce RTX 3090
Random access memory	64 GB
Operating system	Windows 10
Programming language	Python 3.7.13
Deep learning framework	Pytorch 1.12.1
Compute unified device architecture	12.0

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \tag{5}$$

Intersection over union (IoU) denotes the ratio of intersections and connections between the prediction box and the ground truth box. The mAP0.5 denotes the mAP when the IoU of the detection model is set to 0.5, and mAP0.5:0.95 denotes the mAP when the IoU of the detection model is set to range from 0.5 to 0.95 (taking values at intervals of 0.5). $\mathrm{AP0.5}_{\mathrm{Cabbage}}$ and $\mathrm{AP0.5}_{\mathrm{Weed}}$ represent the AP0.5 for Chinese cabbage and weed categories, respectively.

The number of model parameters, floating-point operations (FLOPs), and frames per second (FPS) are used to compare the computational complexity of the models. Additionally, Grad-CAM (Selvaraju et al. 2020) is used to generate model detection heat maps.

WeedDETR

We designed WeedDETR based on RT-DETR applied to small-target weeds, and the structure of the model is shown in Figure 4. WeedDETR contains a fine-grained feature extraction backbone (RepCBNet), the feature complement fusion module (FCFM) encoder for efficient fusion of multilevel features, and a transformer decoder module. Designed based on re-parameterization, RepCBNet provides multilevel weed features through multiple branches. FCFM achieves intra-scale interaction and cross-scale fusion of features through the complementary feature integration (CFI) module. In addition, varifocal loss is used to allow the model to focus on the difficult to detect small-target weed samples to improve weed detection performance, responding to the small-target weed problem in terms of feature extraction, feature fusion, and loss computation, respectively. The transformer decoder module is from DINO (Zhang et al. 2022), which introduces a denoising training method to accelerate the convergence of DETR. WeedDETR achieves efficient real-time end-to-end weed detection through the design of a holistic model architecture. Each component is described in detail in the following sections.

RepCBNet

The structure of RepCBNet is shown in Figure 5A, and feature downsampling was performed by setting the ConvNL and RepCBlock strides to 2 for the last layer. It is generally accepted that the deeper the network, the better the feature extraction ability of the image, which will lead to better object detection. However, when the depth of the network is too great, the features of small-target weeds tend to be lost. PadConv block adopts a two-branch structure; the branches operated by padding can provide diverse features, which enriches the information flow of the feature extraction network; the structure is shown in Figure 5C. The RepCBlock structure is shown in Figure 5D. One branch uses stacked PadConv blocks to deepen the network, and the other branch uses only one ConvNL layer for better gradient propagation and to avoid weed feature loss.

Multi-branch structure is structurally stable and easy to train, but inference speed is slow, and memory consumption is significant. Single-branch structure inference is fast and saves memory, but the feature extraction capability is relatively insufficient. By decoupling the model structure in training and inference, Ding et al. (2021) obtained both the high performance of the multi-branch structure and the speed advantage of the single-branch structure. As shown in Figure 6A, we have re-parameterized the PadConv block according to this concept. The formulation of the re-parameterization PadConv block is detailed

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

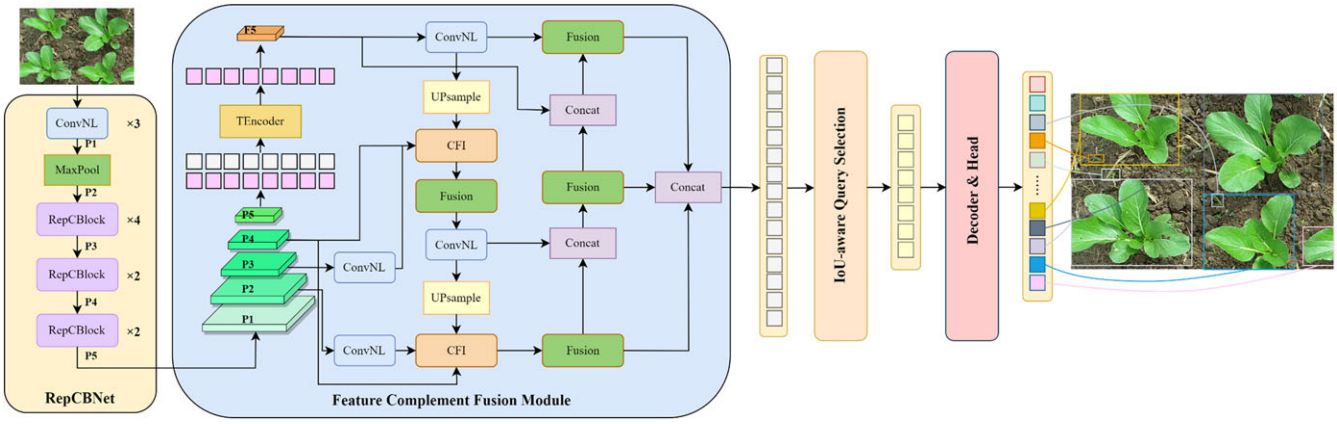


Figure 4. The structure of WeedDETR. P1–P5 represent different levels of feature maps. CFI, complementary feature integration; IoU, intersection over union; TEncoder, transform encoder.

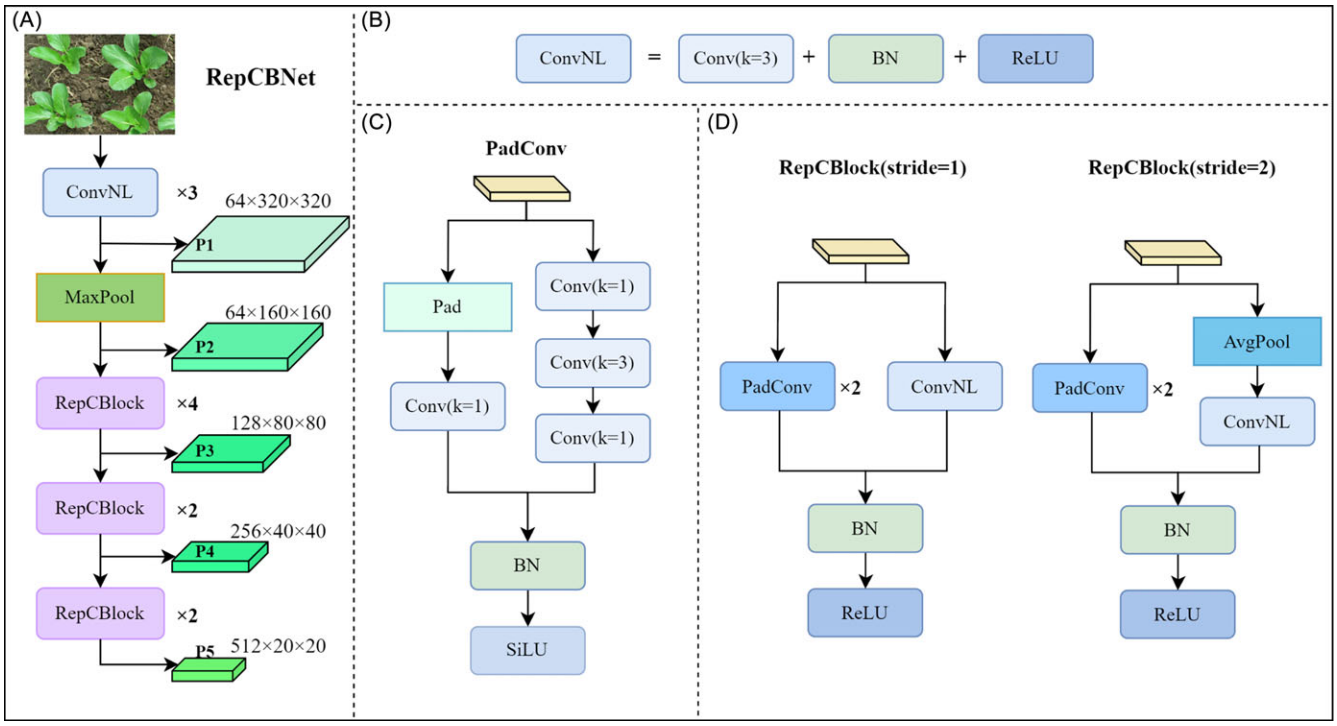


Figure 5. The structure of RepCBNet. (A) The structure of RepCBNet. P1–P5 represent different levels of feature maps. (B) The structure of ConvNL. $k = 1/3$ represents the size of the convolution kernel. (C) The structure of PadConv. (d) The structure of RepCBlock. BN, batch normalization.

in the Supplementary Material. Based on this operation, we transform the structure of the PadConv block into a succinct single-branch structure during inference, which saves computational resources and accelerates inference.

Feature Complement Fusion Module (FCFM)

Current mainstream feature fusion structures such as FPN (Lin et al. 2017a), PAN (Liu et al. 2018), and others often use the last three scales of features (P3–P5) for fusion. Using only these deep features tends to cause shallow features to be lost, which is not conducive to the detection of small-target weeds. We proposed the FCFM, which utilizes the transform encoder (TEncoder) module for intra-scale feature interaction and the CFI module for cross-scale fusion of features; its structure is shown in Figure 7A.

In the FCFM, the Fusion module is used for efficient information fusion, and its structure is shown in Figure 7B. One branch of the Fusion module increases network depth and efficiently represents features through three RepCBlocks, while the other branch effectively avoids gradient explosion. Similar to the PadConv block, RepCBlock is re-parameterized to speed up inference.

The TEncoder module is able to implement the self-attention operation by converting inputs into sequences, capturing long-range dependency between objects. To achieve a balance between accuracy and computational effort, only the last layer (P5) containing rich semantics is processed. The goal of self-attention is to capture the interactions between all entities by encoding each entity based on global contextual information, which is described in the Supplementary Material. TEncoder enables intra-scale

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

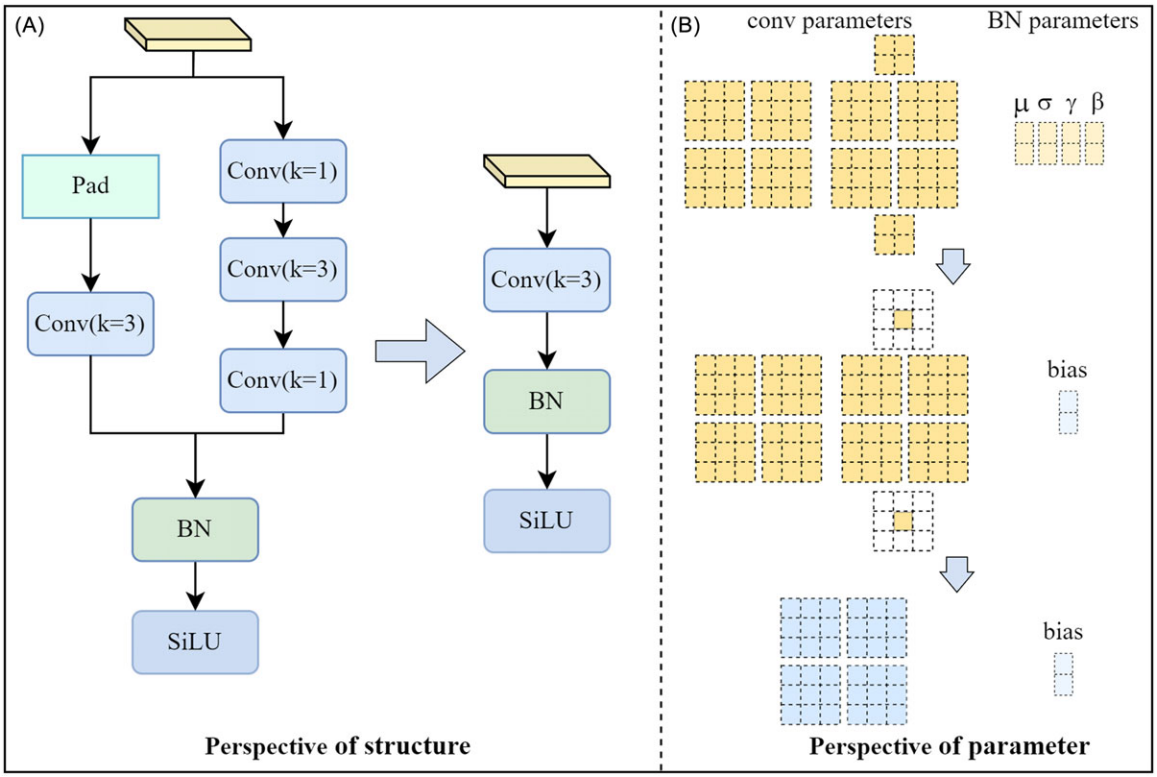


Figure 6. Re-parameterization of the PadConv block. (A) Perspective of structure. (B) Perspective of parameter. BN, batch normalization.

feature interactions to obtain connections between targets in the image for subsequent detection of weeds, and its structure is shown in Figure 7C.

The FPN-like structure lacks full utilization of shallow features and is prone to shallow feature loss, thus affecting the detection performance of small-target weeds. To address this problem, we propose the CFI module, which takes shallow features carrying positional features, neighboring mesoscale features, and transmitted mesoscale features to be fused for cross-scale feature interactions, making full use of the rich information of the shallow features. The structure of CFI modules is shown in Figure 8A.

The number of channels is adjusted to the same number of channels as the transmitted mesoscale features by a 1×1 convolution before the shallow features are input. Subsequently, shallow features are downsampled using a hybrid structure of maximum pooling and average pooling, which helps to retain the high-resolution features and diversity of the weed images. Finally, the transmitted mesoscale features are spliced with neighboring mesoscale features and downsampled large-scale features in the channel dimension. As shown in Figure 7A, the CFI-A module was used for feature fusion at the T4 and T3 feature layers, respectively, which is able to increase the richness of local features and prevent the loss of small-target feature information. Taking the CFI-A used in the T4 feature layer as an example, as the following equation:

$$P3' = Conv^{k1}(MaxPool(Conv^{k1}(P3)) + AvgPool(Conv^{k1}(P3))) \tag{6}$$

$$CFI = Concat_{chamel}(P3', P4, T4) \tag{7}$$

We have conceived upsampling deep features to mesoscale feature sizes to supplement semantic information, as shown in Figure 8B, but this is less effective compared with CFI-A, which is analyzed in detail in the subsequent experimental section.

Varifocal Loss

The IoU-Aware Classification Score (IACS) loss function varifocal loss (VFL) is used to focus the model training on small-target samples (Zhang et al. 2021). Varifocal loss is proposed on the basis of research on focal loss (FL) (Lin et al. 2017b). In this dataset, weeds only account for a small portion of the whole picture, while most of the area is the background area (negative samples). The large number of negative samples will lead to the model training effect deterioration. The focal loss balances the proportion of positive and negative samples by giving greater weight to the hard-to-detect samples, as shown in the following equation:

$$FL(\mathrm{P}, \mathrm{y}) = \begin{cases} -\alpha(1-\mathrm{P})^{\gamma}\log(\mathrm{P}) & \text{if } \mathrm{y}=1 \\ -(1-\alpha)\mathrm{P}^{\gamma}\log(1-\mathrm{P}) & \text{otherwise} \end{cases} \tag{8}$$

where $y \in \{-1,+1\}$, $y = 1$ represents the ground truth class, and $\mathrm{P} \in [0,1]$ denotes the predicted probability of the foreground class. The $(1-\mathrm{P})^{\gamma}$ and P^{γ} represent the moderating factors of the background and foreground classes, respectively.

The formula for varifocal loss is shown below:

$$VFL(\mathrm{P}, \mathrm{y}) = \begin{cases} -\mathrm{q}(\mathrm{q}\log(\mathrm{P}) + (1-\mathrm{q})\log(1-\mathrm{P})) & \mathrm{q}>0 \\ -\alpha\mathrm{P}^{\gamma}\log(1-\mathrm{P}) & \mathrm{q}=0 \end{cases} \tag{9}$$

where P is the predicted IACS and q is the target score. For the foreground class, the value of q is the IOU of the prediction box and ground truth box, and for the background class, q is zero. The varifocal loss scales the loss by the coefficient P^{γ} and will only reduce

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

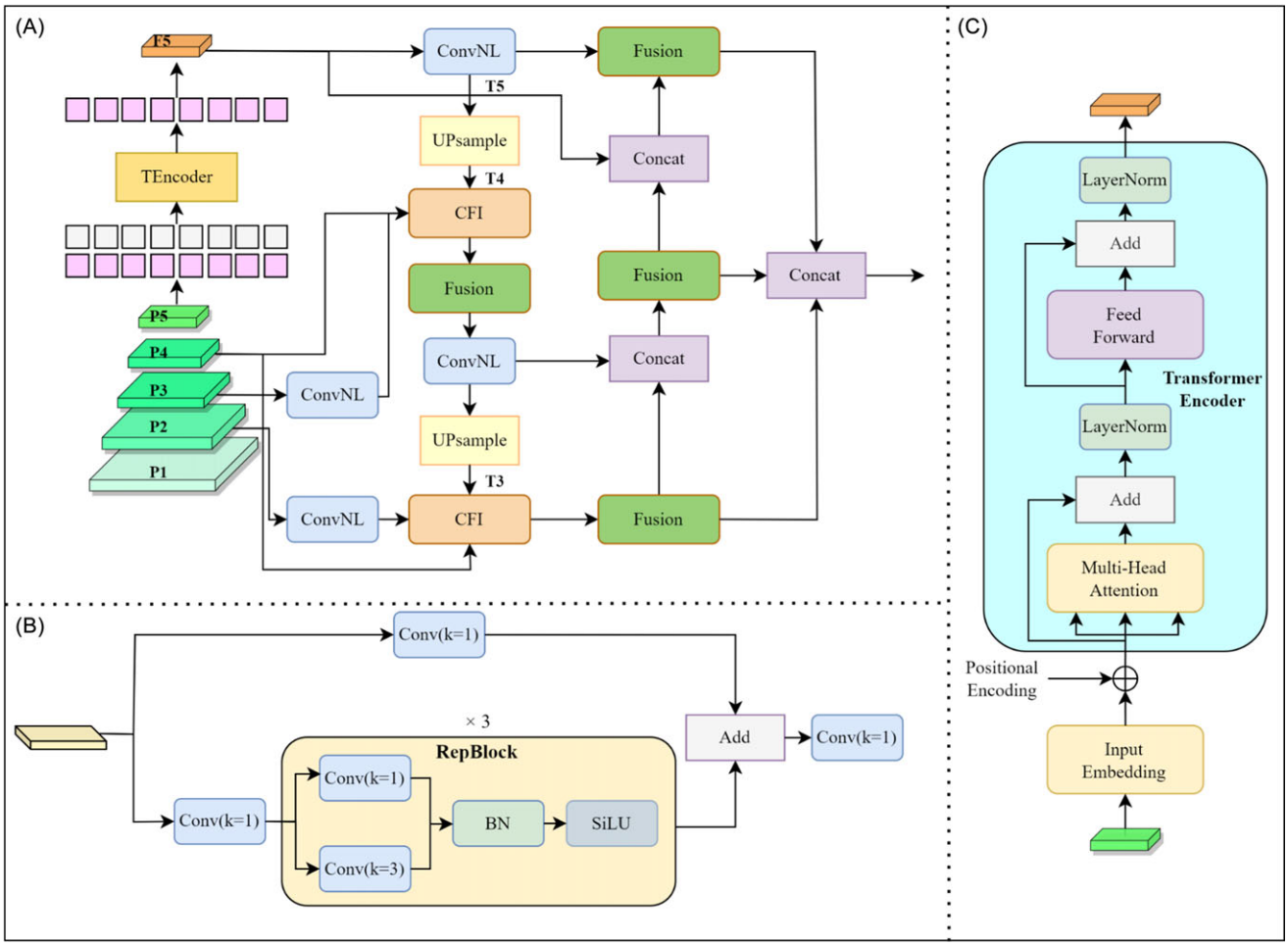


Figure 7. The structure of feature complement fusion module (FCFM) and its components: (A) the structure of FCFM, with P1–P5 representing different levels of features; (B) the structure of Fusion module; and (C) the structure of transform encoder (TEncoder) module. CFI, complementary feature integration;TEncoder, transform encoder; BN, batch normalization.

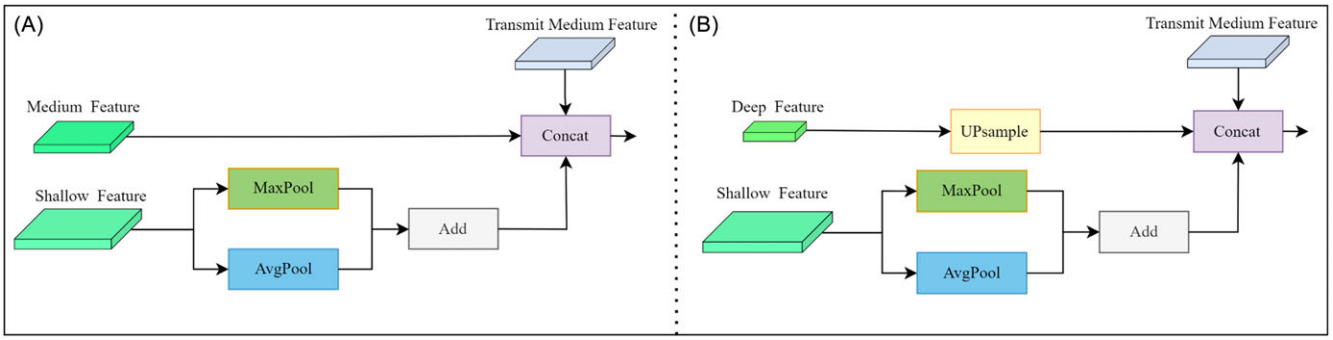


Figure 8. The structure of two types of complementary feature integration (CFI) modules: (A) the structure of CFI-A, and (B) the structure of CFI-B.

the loss contribution for negative samples ($q = 0$). Positive samples with a large q value will have a larger loss contribution; thus, the model allow s focus on high-quality weed samples during loss training, improving the detection accuracy of small-target weeds.

Results and Discussion

Comparison of Backbone Networks

The comparison results of WeedDETR using different backbone networks are shown in Table 3. The RepCBNet collaboratively mines weed edges and texture details in images through a two-branch structure consisting of a deep feature extraction branch and a shallow gradient retention branch. Through this synergistic design, the model comprehensively captures spatial and semantic information of small-target weeds, thereby effectively improving detection accuracy. The $AP0.5_{weed}$ of the RepCBNet is improved by 1.8%, 1.5%, 3.0%, 3.2%, and 5.0% compared with ResNet-34 (He et al. 2016), MobileNetv3-L (Howard et al. 2019), Swin Transformer-Tiny (Liu et al. 2021), HGNetv2-L (the backbone of RT-DETR) (Lv et al. 2023), and ConvNeXtV2-Atto (Woo et al. 2023), respectively.

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

Table 3. Comparison of WeedDETR with different backbone networks.

Backbone	mAP0.5(AP0.5$_{Cabbage}$/AP0.5$_{Weed}$)[a]	mAP0.5:0.95[b]	Parameters		FLOPs[c]		FPS[d]
HGNetv2-L	0.815(0.917/0.712)	0.538	31.99	million	103.40	giga	64.03
ResNet-34	0.821(0.916/0.726)	0.541	30.03	million	88.40	giga	69.09
MobileNetv3-L	0.818(0.907/0.729)	0.531	11.71	million	26.70	giga	81.95
ConvNeXtV2-Atto	0.789(0.884/0.694)	0.504	17.75	million	45.50	giga	65.47
Swin Transformer-Tiny	0.807(0.900/0.714)	0.528	36.31	million	97.00	giga	46.67
RepCBNet	0.827(0.910/0.744)	0.552	19.87	million	56.90	giga	76.58

[a]mAP0.5, mean average precision at 0.5 intersection over union threshold; AP0.5$_{Cabbage}$, average precision at 0.5 intersection over union threshold for Chinese cabbage categories; AP0.5$_{Weed}$, average precision at 0.5 intersection over union threshold for weed categories.
[b]mAP0.5:0.95, the mean average precision computed across intersection over union thresholds from 0.5 to 0.95 with 0.05 intervals.
[c]FLOPs, floating-point operations.
[d]FPS, frames per second.

Table 4. Comparison of WeedDETR with different complementary feature integration (CFI) modules.

Type[a]	mAP0.5(AP0.5$_{Cabbage}$/AP0.5$_{Weed}$)[b]	mAP0.5:0.95[c]	Parameters		FLOPs[d]	
Baseline	0.815(0.917/0.712)	0.538	31.99	million	103.40	giga
CFI-B	0.821(0.920/0.723)	0.541	22.94	million	70.10	giga
CFI-A	0.829(0.918/0.739)	0.551	22.96	million	70.50	giga

[a]CFI-A, mode A of the CFI module; CFI-B, mode B of the CFI module.
[b]mAP0.5, mean average precision at 0.5 intersection over union threshold; AP0.5$_{Cabbage}$, average precision at 0.5 intersection over union threshold for Chinese cabbage categories; AP0.5$_{Weed}$, average precision at 0.5 intersection over union threshold for weed categories.
[c]mAP0.5:0.95, the mean average precision computed across intersection over union thresholds from 0.5 to 0.95 with 0.05 intervals.
[d]FLOPs, floating-point operations.

Table 5. Comparison of the results between focal loss (FL) and varifocal loss (VFL).

FL	VFL	mAP0.5(AP0.5$_{Cabbage}$/AP0.5$_{Weed}$)[a]	mAP0.5:0.95[b]	Parameters		FLOPs[c]	
		0.815(0.917/0.712)	0.538	31.99	million	103.4	giga
√		0.817(0.917/0.718)	0.539	31.99	million	103.4	giga
	√	0.822(0.919/0.724)	0.542	31.99	million	103.4	giga

[a]mAP0.5, mean average precision at 0.5 intersection over union threshold; AP0.5$_{Cabbage}$, average precision at 0.5 intersection over union threshold for Chinese cabbage categories; AP0.5$_{Weed}$, average precision at 0.5 intersection over union threshold for weed categories.
[b]mAP0.5:0.95, the mean average precision computed across intersection over union thresholds from 0.5 to 0.95 with 0.05 intervals.
[c]FLOPs, floating-point operations.

The PadConv block in RepCBNet extends the context-awareness range through padding operations, which enhance the detailed discrimination of weeds while maintaining parameter efficiency. The number of parameters of the RepCBNet is 54.7%, 62.1%, and 66.2% compared with Swin Transformer-Tiny, HGNetv2-L, and ResNet-34, respectively, achieving the optimal detection performance with a lighter architecture. Although MobileNetv3-L and ConvNeXtV2-Atto have fewer parameters, their lightweight design sacrifices some feature extraction capability, resulting in inadequate ability to detect small-target weeds. These results show that using the RepCBNet as the backbone can effectively extract fine-grained information, improve detection accuracy, and achieve a balance between computation and accuracy.

Effectiveness of the CFI Module

We compared the detection performance of the model when using different types of CFI modules, and the results are shown in Table 4. Compared with the model without the CFI module, the model increased AP0.5$_{Weed}$ by 1.1% and 2.7% with the CFI-B and the CFI-A, respectively. The results indicate that the CFI structure effectively fuses shallow features with richer detail information, achieving full integration of shallow and deep features. Compared with the CFI-B, the CFI-A is able to detect small-target weeds more accurately while using similar computational effort. This phenomenon may be due to the fact that direct access to the underlying information, rather than using deeper information for upsampling, is more conducive to feature complementarity and avoiding feature confusion (Wang et al. 2023). The experimental results demonstrate that the CFI-A module mitigates the problem of small-target weed information being ignored in the deep network through feature complementary fusion, hence its use in composing the FCFM.

Comparison of Loss Functions

The loss function only affects the computation of losses during model training, as it does not increase the parameters and the FLOPs. The experimental results are shown in Table 5, where the AP0.5$_{Weed}$ increased by 0.6% and 1.2% after using FL and VFL in training, respectively. VFL is more capable of focusing on hard to detect weed samples than FL, thus effectively improving model detection performance (Du and Jiao 2022; Peng et al. 2022).

Ablation Experiment

Three improvements improve the detection performance to varying degrees, as shown by the ablation experiment results in Table 6. The RepCBNet reduces the number of parameters while acquiring feature representations at a finer granularity. The FCFM

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

Table 6. Results of ablation experiment.

RepCBNet	FCFM[a]	VFL[b]	mAP0.5(AP0.5$_{Cabbage}$/AP0.5$_{Weed}$)[c]	mAP0.5:0.95[d]	Parameters		FLOPs[e]	
			0.815(0.917/0.712)	0.538	31.98	million	103.40	giga
√			0.827(0.910/0.744)	0.552	19.87	million	56.90	giga
	√		0.829(0.918/0.739)	0.551	22.96	million	70.50	giga
		√	0.822(0.919/0.724)	0.542	31.99	million	103.40	giga
√	√		0.830(0.916/0.743)	0.553	19.92	million	58.20	giga
√		√	0.833(0.916/0.749)	0.560	19.87	million	56.90	giga
	√	√	0.831(0.919/0.742)	0.548	22.96	million	70.50	giga
√	√	√	0.839(0.920/0.757)	0.558	19.92	million	58.20	giga

[a]FCFM, feature complement fusion module.
[b]VFL, varifocal loss.
[c]mAP0.5, mean average precision at 0.5 intersection over union threshold; AP0.5$_{Cabbage}$, average precision at 0.5 intersection over union threshold for Chinese cabbage categories; AP0.5$_{Weed}$, average precision at 0.5 intersection over union threshold for weed categories.
[d]mAP0.5:0.95, the mean average precision computed across intersection over union thresholds from 0.5 to 0.95 with 0.05 intervals.
[e]FLOPs, floating-point operations.

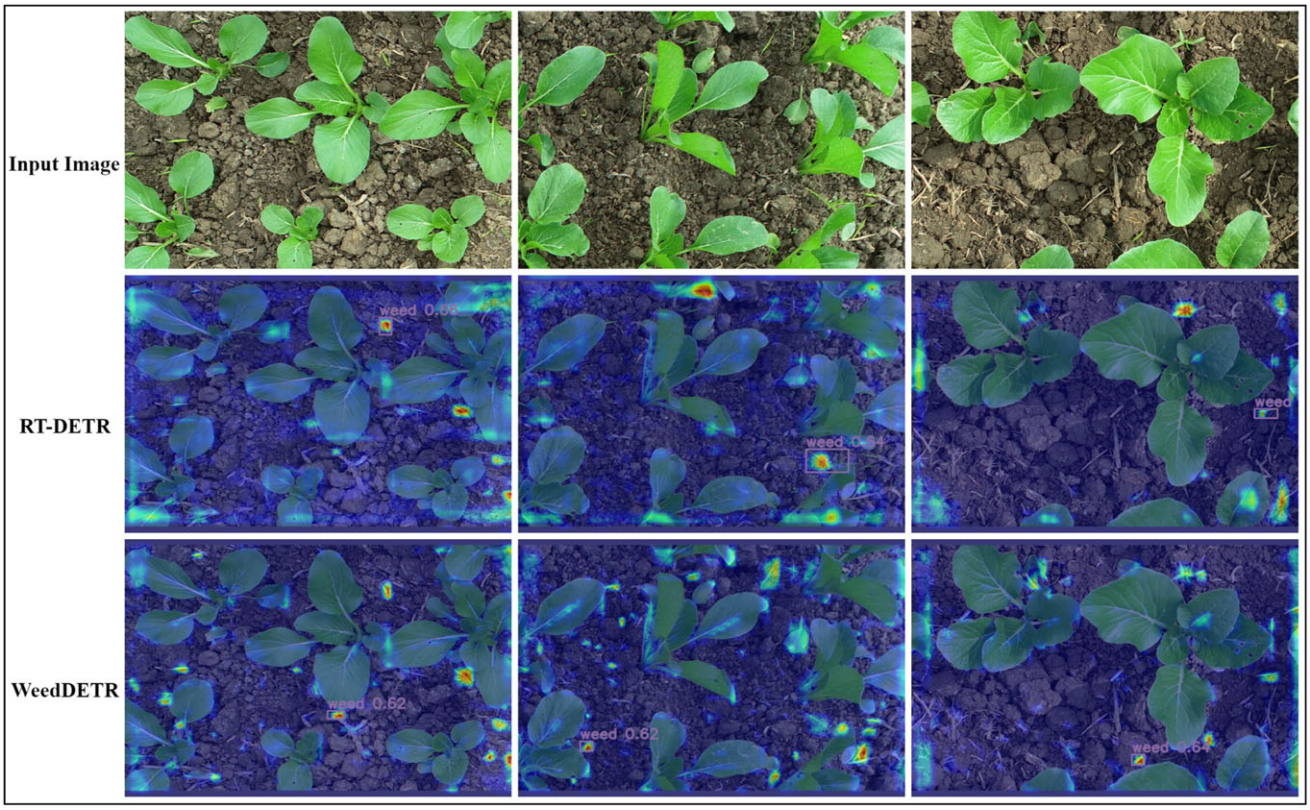


Figure 9. Heat map comparison of RT-DETR and WeedDETR. The darker red areas in the heat maps indicate the areas of the feature maps that the models focus on.

module incorporates multi-layered low-level features, which effectively improves the accuracy of weed detection, resulting in a 2.7% improvement in the AP0.5$_{Weed}$. By introducing VFL in the loss calculation, the loss weights of complex samples are increased, and the weed detection accuracy is improved. The FCFM provides rich weed features as discriminative guidance for VFL during sample reweighting, while VFL compels the model to prioritize learning the critical spatial features of difficult samples captured by FCFM. Their synergistic interaction achieves a 3.0% improvement in AP0.5$_{Weed}$. The experimental results showed that WeedDETR effectively improved the accuracy of small-target weed detection by 2.4% for mAP0.5 and 4.5% for AP0.5$_{Weed}$ compared with the RT-DETR.

The heat map for WeedDETR and RT-DETR is shown in Figure 9. The darker red areas in the heat maps indicate the areas of the feature maps that the models focus on. RT-DETR has insufficient perception of small-target weeds and is prone to miss them, while WeedDETR is able to focus more comprehensively on small-target weeds and has better weed detection performance.

Re-parameterization Experiment

The re-parameterization operation is applied only during inference, merging training-stage multi-branch structures into a single-branch equivalent to eliminate computational redundancy while preserving the original training model architecture (Zhang and Wan 2024). As shown in Table 7, the operation reduces 16.9% of the parameters and 16.8% of the FLOPs in the inference process, which improves the efficiency while maintaining detection

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

Table 7. Parameters and floating-point operations (FLOPs) change during training and inference.

State	mAP0.5(AP0.5$_{Cabbage}$/AP0.5$_{Weed}$)[a]	mAP0.5:0.95[b]	Parameters		FLOPs	
training	0.839(0.920/0.757)	0.558	23.99	million	70	giga
inference	0.839(0.920/0.757)	0.558	19.92	million	58.2	giga

[a]mAP0.5, mean average precision at 0.5 intersection over union threshold; AP0.5$_{Cabbage}$, average precision at 0.5 intersection over union threshold for Chinese cabbage categories; AP0.5$_{Weed}$, average precision at 0.5 intersection over union threshold for weed categories.
[b]mAP0.5:0.95, the mean average precision computed across intersection over union thresholds from 0.5 to 0.95 with 0.05 intervals.

Table 8. Comparison of detection results of different detection models.

Models[a]	mAP0.5(AP0.5$_{Cabbage}$/AP0.5$_{Weed}$)[b]	mAP0.5:0.95[c]	Parameters		FLOPs[d]		FPS[e]
YOLOv3-SPP	0.801(0.911/0.690)	0.524	104.71	million	283.10	giga	55.56
YOLOv5-L	0.820(0.917/0.722)	0.542	53.13	million	134.70	giga	66.38
YOLOv6-M	0.805(0.916/0.694)	0.526	51.98	million	161.10	giga	67.52
YOLOv8-L	0.818(0.916/0.721)	0.547	43.61	million	164.80	giga	75.24
Faster R-CNN	0.484(0.846/0.123)	0.317	136.71	million	401.71	giga	34.88
SSD	0.619(0.811/0.493)	0.349	23.75	million	273.61	giga	64.53
RetinaNet	0.497(0.784/0.209)	0.372	36.35	million	163.85	giga	44.51
WeedDETR	0.839(0.920/0.757)	0.558	19.92	million	58.20	giga	76.28

[a]YOLO, You Only Look Once; Faster R-CNN, faster regions with convolutional neural networks; SSD, single-shot detector.
[b]mAP0.5, mean average precision at 0.5 intersection over union threshold; AP0.5$_{Cabbage}$, average precision at 0.5 intersection over union threshold for Chinese cabbage categories; AP0.5$_{Weed}$, average precision at 0.5 intersection over union threshold for weed categories.
[c]mAP0.5:0.95, the mean average precision computed across intersection over union thresholds from 0.5 to 0.95 with 0.05 intervals.
[d]FLOPs, floating-point operations.
[e]FPS, frames per second.

accuracy. The efficiency–accuracy decoupling optimization strategy of re-parameterization enhances the model's computational efficiency, thereby facilitating its deployment on agricultural edge devices with limited memory and computational resources.

Comparison of Results with Other Detections

The performance of WeedDETR was comprehensively compared with state-of-the-art detection models, including Faster R-CNN (Ren et al. 2017), SSD (Liu et al. 2016), RetinaNet (Lin et al. 2017b), and YOLO series models represented by YOLOv3-SPP (Redmon and Farhadi 2018), YOLOv5-L (Jocher 2020), YOLOv6-M (Li et al. 2022), and YOLOv8-L (Ultralytics 2023), all of which were trained from scratch, with the results shown in Table 8. Faster R-CNN and RetinaNet weed detection performed ineffectively, while the YOLO models achieved better detection results. Compared with YOLOv5-L, the best performer in the YOLO series, WeedDETR has a 1.9% improvement in mAP0.5, a 3.5% improvement in AP0.5$_{Weed}$, and a 1.6% improvement in mAP0.5:0.95. WeedDETR achieves dual efficiency in parameters and computational complexity with 19.92 M parameters and 58.20 G FLOPs, while attaining the highest real-time detection speed of 76.28 FPS among comparative models.

The PR curves for the four models with the highest detection accuracy are illustrated in Figure 10. The PR curve of WeedDETR comprises a larger closed region compared with YOLOv5-L, YOLOv6-M, and YOLOv8-L, which indicates that the proposed model exhibits higher detection accuracy.

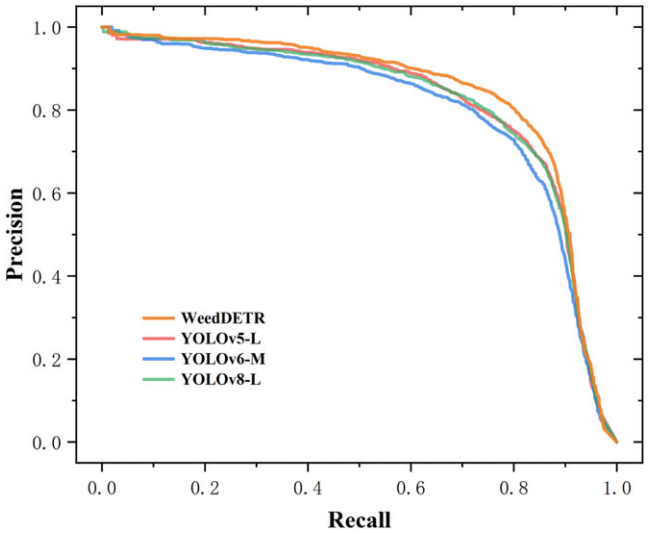


Figure 10. Comparison of precision-recall (PR) curves.

Visualization of Prediction Results with Other Mainstream Detections

A comparison of detection results among the four most accurate models is presented in Figures 11 and 12. The detection results of the model under three complex backgrounds (shadow occlusion, rice straw occlusion, and waterbody interference) are presented in Figure 11. All models accurately detected Chinese cabbage in both shadow-obscured and straw-obscured backgrounds. But in the waterbody interference background, YOLOv5-L and YOLOv6-M showed false detection of marginal Chinese cabbage leaves, as shown in Figure 11C. As illustrated in Figure 11A and 11B, WeedDETR more accurately captures weeds that are shaded or obscured than other models, demonstrating its robustness of detection in complex environments. RepCBNet accurately extracts information about the differences between weeds and background, allowing WeedDETR to efficiently detect weeds obscured by shadows or straw.

All models accurately detected Chinese cabbage as an obvious target, but for small-target weeds, there was partial weed miss-detection in all models except WeedDETR, as shown in Figure 12. The limited multilevel utilization of shallow features in YOLO

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

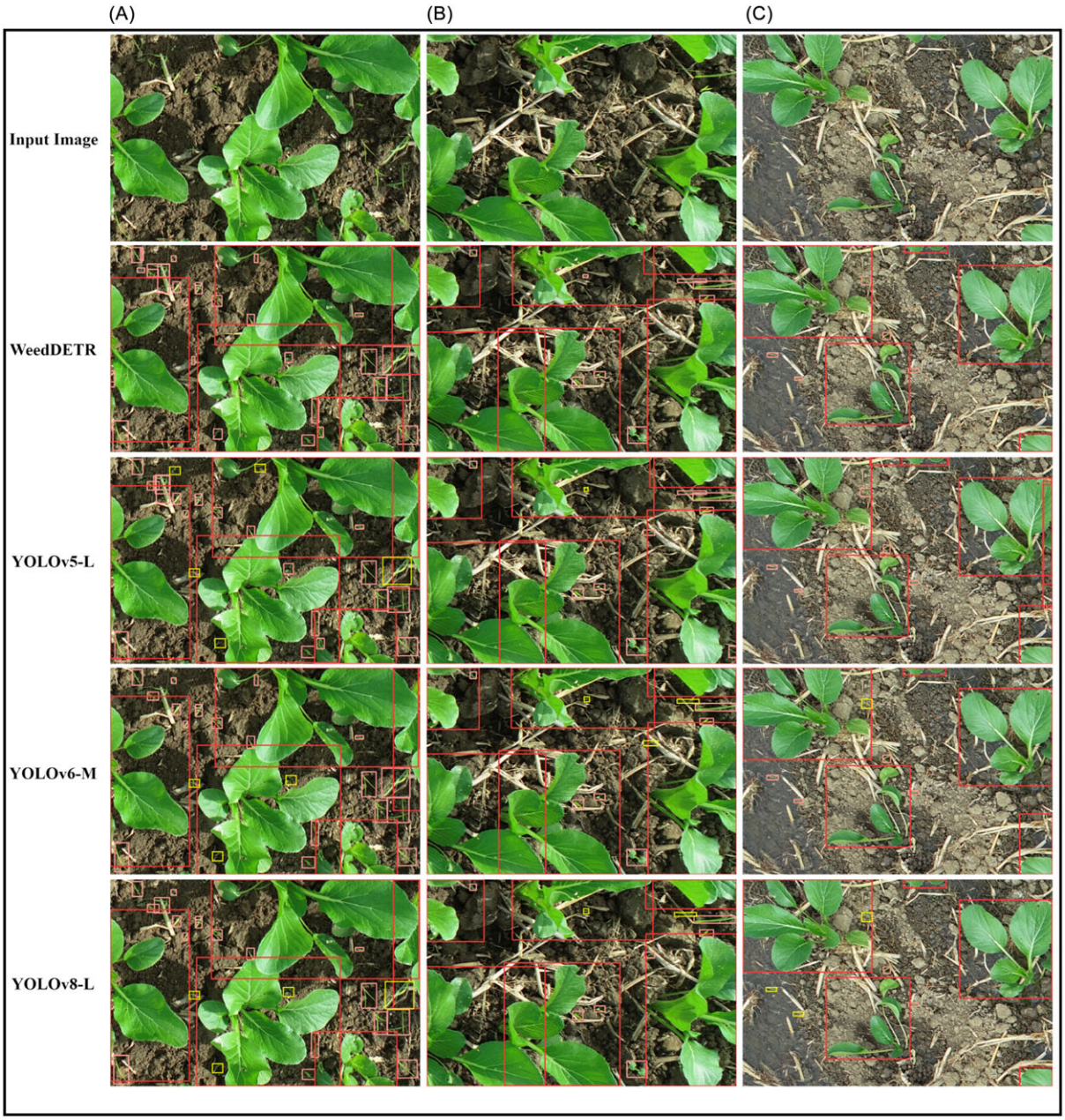


Figure 11. Comparison of detection results by different models in complex background. The three scenarios are: (A), shadow occlusion, (B) rice straw occlusion, and (C) waterbody interference. Red boxes represent detected Chinese cabbage, brown boxes represent detected weeds, and yellow boxes represent missed weeds.

series models might lead to progressive degradation of small-target weed representations during deep network propagation, potentially contributing to suboptimal weed detection performance, particularly under dense small-target weeds scenarios, as shown in Figure 12B (Zhang 2023; Zhang et al. 2024). To address this phenomenon, WeedDETR effectively mitigates the loss of small-target features and achieves enhanced weed detection accuracy through the collaboration of shallow feature information supplementation and cross-scale global semantic feature fusion.

Comparative results show that WeedDETR exhibits better performance in accurately detecting small-target weeds in complex backgrounds. Based on the conducted analysis, the proposed model is able to accurately detect small-target weeds in UAV-captured images, effectively mitigating the phenomenon of underdetection of small-target weeds and accelerating the inference speed through re-parameterization convolution. Compared with other detection models, WeedDETR detects weeds with higher accuracy and faster inference, which can meet the field deployment requirements of UAVs for weed detection applications. Additionally, we have developed a weed imagery detection system built upon WeedDETR, showcasing its ability to detect weeds in high-resolution drone-captured images, with implementation details provided in the Supplementary Material.

Conclusions

To address the lack of current UAV-based weed detection datasets and the limited performance of weed detection models in weed

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

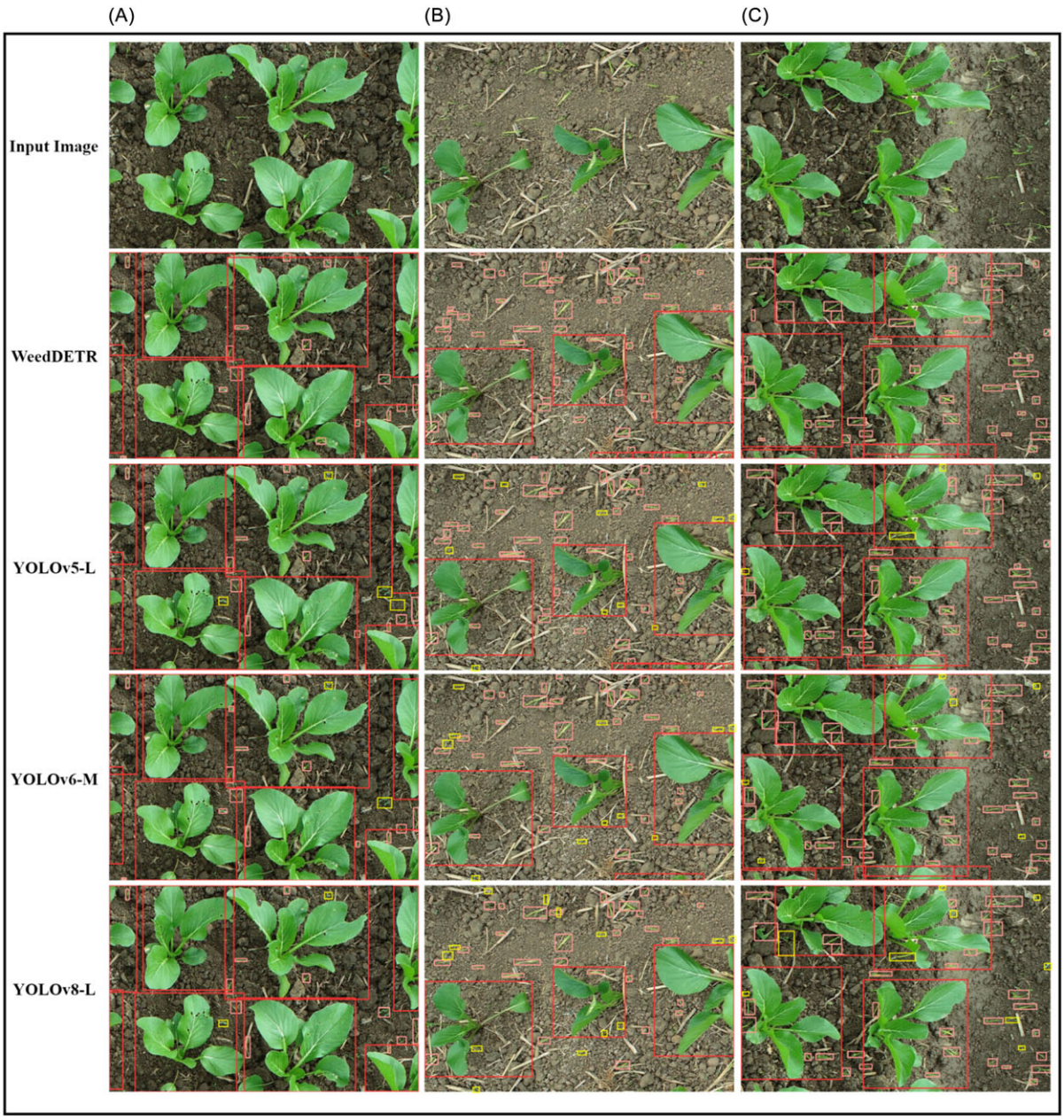


Figure 12. Comparison of detection results by different models in small-target weeds scenarios. Red boxes represent detected Chinese cabbage, brown boxes represent detected weeds, and yellow boxes represent missed weeds.

detection, we constructed a high-quality field UAV weed detection dataset and proposed the WeedDETR based on the characteristics of small-target weeds. The WeedDETR achieved 73.9% and 91.8% AP0.5 in the weed and Chinese cabbage categories with 76.28 FPS, outperforming the existing state-of-the-art detection models. In addition, the proposed model establishes a highly reliable algorithmic foundation for intelligent weeding equipment. The weed density heat maps generated by the model can further guide variable spraying systems to achieve weed-targeted precision spraying, thereby reducing herbicide usage (Xu et al. 2025). Furthermore, the model can be extended to dynamic monitoring of herbicide-resistant weeds by analyzing spatial dispersion patterns of specific weed populations through continuous multi-season data, thereby providing data-driven support for optimizing crop rotation systems and herbicide rotation strategies (Vasileiou et al. 2024).

While WeedDETR demonstrated robust performance on the GZWeed dataset, its generalizability is limited by the single-crop scenario of the current dataset with unsegmented weed classes. Moreover, the PyTorch-based weights of WeedDETR can be further converted to lightweight inference frameworks such as TensorFlow Lite and ONNX (Open Neural Network Exchange) to enhance computational efficiency in agricultural terminals. In our next work, we will construct a multi–crop field weed dataset leveraging UAV platforms and systematically evaluate the model's generalization capability for cross-crop weed detection. Besides, we aim to improve the inference efficiency of the model by compressing model parameters and computational overhead through knowledge distillation and model pruning, thereby advancing the implementation of SSWM.

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/wsc.2025.10035

Data availability statement. The dataset with annotation is accessible to the public and thoroughly documented on GitHub: https://github.com/sxyang4399/GZWeed.

Funding statement. This study was supported by the National Key Research and Development Program of China (2021YFE0113700), the National Natural Science Foundation of China (32360705; 31960555), the Guizhou Provincial Science and Technology Program (CXTD[2025]041; GCC[2023]070; HZJD[2022]001), and the Program for Introducing Talents to Chinese Universities (111 Program; D20023).

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

References

Bochkovskiy A, Wang C-Y, Liao H-YM (2020) YOLOv4: optimal speed and accuracy of object detection. arXiv database 2004.10934. https://arxiv.org/abs/2004.10934

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S (2020) End-to-end object detection with transformers. Pages 213–229 *in* Proceedings from the European Conference on Computer Vision. Glasgow: Springer

Dang F, Chen D, Lu Y, Li Z (2023) YOLOWeeds: a novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. Comput Electron Agric 205:107655

Ding X, Zhang X, Ma N, Han J, Ding G, Sun J (2021) RepVGG: making VGG-style ConvNets great again. Pages 13733–13742 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN: Institute of Electrical and Electronics Engineers

Du F-J, Jiao S-J (2022) Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection. Sensors 22:3537

Gallo I, Rehman AU, Dehkordi RH, Landro N, La Grassa R, Boschetti M (2023) Deep object detection of crop weeds: performance of YOLOv7 on a real case dataset from UAV images. Remote Sens 15:539

Gašparović M, Zrinjski M, Barković Đ, Radočaj D (2020) An automatic method for weed mapping in oat fields based on UAV imagery. Comput Electron Agric 173:105385

Gerhards R, Andújar Sanchez D, Hamouz P, Peteinatos GG, Christensen S, Fernandez-Quintanilla C (2022) Advances in site-specific weed management in agriculture—a review. Weed Res 62:123–133

Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. Pages 580–587 *in* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: Institute of Electrical and Electronics Engineers

He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. Pages 2961–2969 *in* Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: Institute of Electrical and Electronics Engineers

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Pages 770–778 *in* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: Institute of Electrical and Electronics Engineers

Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H (2019) Searching for MobileNetV3. Pages 1314–1324 *in* Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: Institute of Electrical and Electronics Engineers

Hu K, Wang Z, Coleman G, Bender A, Yao T, Zeng S, Song D, Schumann A, Walsh M (2023) Deep learning techniques for in-crop weed recognition in large-scale grain production systems: a review. Precis Agric 25:1–29

Islam N, Rashid MM, Wibowo S, Xu C-Y, Morshed A, Wasimi SA, Moore S, Rahman SM (2021) Early weed detection using image processing and machine learning techniques in an Australian chilli farm. Agriculture 11:387

Jocher G (2020) YOLOv5 by Ultralytics. https://github.com/ultralytics/yolov5. Accessed: May 31, 2025

Khan S, Tufail M, Khan MT, Khan ZA, Anwar S (2021) Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. Precis Agric 22:1711–1727

Li C, Li L, Jiang H, Weng K, Geng Y, Li L, Ke Z, Li Q, Cheng M, Nie W, Li Y, Zhang B, Liang Y, Zhou L, Xu X, *et al.* (2022) YOLOv6: a single-stage object detection framework for industrial applications. arXiv database 2209.02976. https://arxiv.org/abs/2209.02976

Li Y, Tang Y, Liu Y, Zheng D (2023) Semi-supervised counting of grape berries in the field based on density mutual exclusion. Plant Phenomics 5:0115

Lin J, Chen X, Cai J, Pan R, Cernava T, Migheli Q, Zhang X, Qin Y (2023) Looking from shallow to deep: hierarchical complementary networks for large scale pest identification. Comput Electron Agric 214:108342

Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017a) Feature pyramid networks for object detection. Pages 2117–2125 *in* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: Institute of Electrical and Electronics Engineers

Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2017b) Focal loss for dense object detection. Pages 2980–2988 *in* Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: Institute of Electrical and Electronics Engineers

Liu S, Qi L, Qin H, Shi J, Jia J (2018) Path aggregation network for instance segmentation. Pages 8759–8768 *in* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: Institute of Electrical and Electronics Engineers

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. Pages 21–37 *in* Proceedings of the European Conference on Computer Vision. Amsterdam: Springer

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: hierarchical vision transformer using shifted windows. Pages 10012–10022 *in* Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: Institute of Electrical and Electronics Engineers

Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. arXiv database 1711.05101. https://arxiv.org/abs/1711.05101

Lv W, Zhao Y, Xu S, Wei J, Wang G, Cui C, Du Y, Dang Q, Liu Y (2023) DETRs beat YOLOs on real-time object detection. Pages 16965–16974 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: Institute of Electrical and Electronics Engineers

Miho H, Pagnotta G, Hitaj D, De Gaspari F, Mancini LV, Koubouris G, Godino G, Hakan M, Diez CM (2024) OliVaR: improving olive variety recognition using deep neural networks. Comput Electron Agric 216:108530

Peña JM, Torres-Sánchez J, De Castro AI, Kelly M, López-Granados F (2013) Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PLoS ONE 8:e77151

Peng H, Li Z, Zhou Z, Shao Y (2022) Weed detection in paddy field using an improved RetinaNet network. Comput Electron Agric 199:107179

Rai N, Zhang Y, Ram BG, Schumacher L, Yellavajjala RK, Bajwa S, Sun X (2023) Applications of deep learning in precision weed management: a review. Comput Electron Agric 206:107698

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You Only Look Once: unified, real-time object detection. Pages 779–788 *in* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: Institute of Electrical and Electronics Engineers

Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. arXiv database 1804.02767. https://arxiv.org/abs/1804.02767

Reedha R, Dericquebourg E, Canals R, Hafiane A (2022) Transformer neural network for weed and crop classification of high resolution UAV images. Remote Sens 14:592

Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149

Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S (2019) Generalized intersection over union: a metric and a loss for bounding box regression. Pages 658–666 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA: Institute of Electrical and Electronics Engineers

Roboflow (2025) Roboflow: Computer Vision Tools for Developers and Enterprises. https://roboflow.com. Accessed: May 31, 2025

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 128:336–359

Ultralytics (2023) Ultralytics YOLOv8. Version 8.0.0. https://github.com/ultralytics/ultralytics. Accessed: May 31, 2025

Valente J, Hiremath S, Ariza-Sentís M, Doldersum M, Kooistra L (2022) Mapping of *Rumex obtusifolius* in nature conservation areas using very high resolution UAV imagery and deep learning. Int J Appl Earth Obs Geoinf 112:102864

Vasileiou M, Kyrgiakos LS, Kleisiari C, Kleftodimos G, Vlontzos G, Belhouchette H, Pardalos PM (2024) Transforming weed management in sustainable agriculture with artificial intelligence: a systematic literature review towards weed identification and deep learning. Crop Prot 176:106522

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. Pages 6000–6010 *in* Proceedings of the Conference on Neural Information Processing Systems. Long Beach, CA: Neural Information Processing Systems Foundation

Veeranampalayam Sivakumar AN, Li J, Scott S, Psota E, J. Jhala A, Luck JD, Shi Y (2020) Comparison of object detection and patch-based classification deep learning models on mid- to late-season weed detection in UAV imagery. Remote Sens 12:2136

Wang C, He W, Nie Y, Guo J, Liu C, Han K, Wang Y (2023) Gold-YOLO: efficient object detector via gather-and-distribute mechanism. Pages 51094–51112 *in* Advances in Neural Information Processing Systems. New Orleans: Neural Information Processing Systems Foundation

Wang Q, Cheng M, Huang S, Cai Z, Zhang J, Yuan H (2022) A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed *Solanum rostratum* Dunal seedlings. Comput Electron Agric 199:107194

Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon IS, Xie S (2023) ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders. Pages 16133–16142 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: Institute of Electrical and Electronics Engineers

Xu H, Li T, Hou X, Wu H, Shi G, Li Y, Zhang G (2025) Key technologies and research progress of intelligent weeding robots. Weed Sci 73:e25

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

Zhang H, Wang Y, Dayoub F, Sünderhauf N (2021) VarifocalNet: an IoU-aware dense object detector. Pages 8514–8523 *in* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN: Institute of Electrical and Electronics Engineers

Zhang L, Wan Y (2024) Partial convolutional reparameterization network for lightweight image super-resolution. J Real-Time Image Proc 21:187

Zhang Y, Ye M, Zhu G, Liu Y, Guo P, Yan J (2024) FFCA-YOLO for small object detection in remote sensing images. IEEE Trans Geosci Remote Sens 62:1–15

Zhang Z (2023) Drone-YOLO: an efficient neural network method for target detection in drone images. Drones 7:526

Zhu X, Su W, Lu L, Li B, Wang X, Dai J (2021) Deformable DETR: deformable transformers for end-to-end object detection. arXiv database 2010.04159. https://arxiv.org/abs/2010.04159

https://doi.org/10.1017/wsc.2025.10035 Published online by Cambridge University Press