electronics

Article

RVDR-YOLOv8: A Weed Target Detection Model Based on Improved YOLOv8

Yuanming Ding *, Chen Jiang, Lin Song, Fei Liu and Yunrui Tao

Communication and Network Key Laboratory, Dalian University, Dalian 116622, China; jiangchen1412@163.com (C.J.); songlin@dlu.edu.cn (L.S.); liufei20210529@163.com (F.L.); tyr9573648@163.com (Y.T.)

* Correspondence: dingyuanming@dlu.edu.cn

Abstract: Currently, weed control robots that can accurately identify weeds and carry out removal work are gradually replacing traditional chemical weed control techniques. However, the computational and storage resources of the core processing equipment of weeding robots are limited. Aiming at the current problems of high computation and the high number of model parameters in weeding robots, this paper proposes a lightweight weed target detection model based on the improved YOLOv8 (You Only Look Once Version 8), called RVDR-YOLOv8 (Reversible Column Dilation-wise Residual). First, the backbone network is reconstructed based on RevCol (Reversible Column Networks). The unique reversible columnar structure of the new backbone network not only reduces the computational volume but also improves the model generalisation ability. Second, the C2fDWR module is designed using Dilation-wise Residual and integrated with the reconstructed backbone network, which improves the adaptive ability of the new backbone network RVDR and enhances the model's recognition accuracy for occluded targets. Again, GSConv is introduced at the neck end instead of traditional convolution to reduce the complexity of computation and network structure while ensuring the model recognition accuracy. Finally, InnerMPDIoU is designed by combining MPDIoU with InnerIoU to improve the prediction accuracy of the model. The experimental results show that the computational complexity of the new model is reduced by 35.8%, the number of parameters is reduced by 35.4% and the model size is reduced by 30.2%, while the mAP$_{50}$ and mAP$_{50\text{-}95}$ values are improved by 1.7% and 1.1%, respectively, compared to YOLOv8. The overall performance of the new model is improved compared to models such as Faster R-CNN, SSD and RetinaNet. The new model proposed in this paper can achieve the accurate identification of weeds in farmland under the condition of limited hardware resources, which provides theoretical and technical support for the effective control of weeds in farmland.

Keywords: deep learning; weed identification; YOLOv8; lightweight model

check for updates

Citation: Ding, Y.; Jiang, C.; Song, L.; Liu, F.; Tao, Y. RVDR-YOLOv8: A Weed Target Detection Model Based on Improved YOLOv8. *Electronics* **2024**, *13*, 2182. https://doi.org/10.3390/electronics13112182

Academic Editors: Haibin Wu, Aili Wang and Yuji Iwahori

Received: 20 April 2024
Revised: 16 May 2024
Accepted: 23 May 2024
Published: 3 June 2024

1. Introduction

Weeds pose a serious threat to crop production [1]. It is estimated that weed causes crop yield losses of up to 43% worldwide every year. [2]. The traditional method of weed management is spraying herbicides [3]. The use of chemical herbicides is important to protect crop health and increase yields [4]. However, the standard in agriculture is to spray herbicides extensively. Therefore, even if there are no weeds in the ground, herbicides should be applied evenly. This behaviour not only adversely affects the environment but also causes economic losses to the farming operation [5,6]. Therefore, the ability to accurately identify and spray weeds by automated weed control systems is crucial for maintaining and enhancing global food productivity [4].

In traditional image processing techniques, analysing and extracting morphological and textural features of weed species is widely used to identify weeds in crops [7]. However, the process of extracting important features takes a long time and is susceptible to bias. In order to improve the generalisation of the model, machine learning techniques such as support vector machines have been used to train computers for automatic weed recognition [8]. In [9], the authors proposed a texture-based classifier for segmenting weeds in major crops by considering the combination of wavelet features in neural networks. In [10], a machine vision approach for weed identification using support vector machines was proposed. However, traditional machine learning algorithms are time-consuming and prone to bias in extracting key features [11].

Deep learning techniques are superior in feature extraction using convolutional neural networks [12]. For this reason, deep-learning-based object detection techniques have been widely used in object recognition [13]. In [14], the authors combined a ResNeXt feature extraction network with a Faster R-CNN [15] model to obtain good recognition results on crop seedling and weed image datasets. In [16], the authors replaced the VGG network portion in the original SSD [17] network with ResNet [18]. The improved SSD model achieved an average detection accuracy of 89.7% for surface defects in solid wood. In [19], the authors predicted wheat ears under different conditions based on migration learning and RetinaNet [20]. The experiments proved that RetinaNet achieved high recognition performance and recognition speed. From the above research results, the recognition method based on deep learning can well overcome the shortcomings of traditional recognition methods.

In recent years, YOLO-based deep learning methods have been widely used in object recognition research [21]. This method can effectively detect small targets in complex scenes and have higher detection speeds compared with other methods [22,23]. In [24], the authors conducted experiments on RetinaNet, SSD and YOLOv3 [25] for real-time pill recognition and verified the effectiveness of the YOLOv3 algorithm. In [26], the authors constructed a new backbone based on YOLOv4. By introducing a multi-branch structure and combining methods such as dilation convolution, the new model improved the AP value of small target weeds by 15.1% and the mAP by 4.2%. In [27], the authors constructed a new model, YOLO-CBAM, by introducing the attention mechanism into YOLOv5. The mAP of the new model was improved by 2.55% compared to YOLOv5, and its detection speed and effectiveness could meet the requirements of real-time weed monitoring in the field. In [28], the authors constructed a weed detection model called YOLOv7-FWeed based on YOLOv7. The new model improved the accuracy of weed identification using an F-ReLU and MaxPool multi-head self-attention module. The results show that the method outperforms YOLOv7 in many aspects.

Ultralytics introduced YOLOv8 in 2023, proving its superiority as the state-of-the-art version of YOLO by comparing it with previous YOLO series models [29]. However, since the core processing device of a weeding robot has limited computational and storage resources, it is of great practical significance to investigate a weeding recognition method that can satisfy both requirements of high accuracy and lightweight [30]. In order to solve the above problems, this paper designs a new method based on YOLOv8. The new method can not only effectively identify multiple types of weeds, but also effectively reduce the hardware overhead of the model. The main work of the paper can be summarised as follows:

1. Based on RevColNet, the backbone network of YOLOv8 is reconfigured to reduce the computational complexity and the number of parameters of the model, while improving the feature extraction capability of the model.
2. The C2fDWR module is designed based on the Dilation-wise Residual of the DWRSeg model and integrated into the RevCol backbone to form a new backbone RVDR, which improves the model recognition capability and makes up for the model's shortcomings for small target detection.
3. GSConv and VoVGSCSPC are used instead of the traditional convolution module and CSP module at the neck end of the model. This improved method reduces the size of the model while ensuring its performance.

4. The original network loss function is optimised using InnerMPDIoU Loss to provide a more accurate loss metric.

2. Materials and Methods

2.1. YOLOv8 Model

Among the various existing detection methods, YOLO is widely used for the recognition of various types of objects due to its fast and accurate features. YOLOv8 is the most advanced model of the YOLO series at present, which is built on the basis of YOLOv5, with some new features added to further improve the accuracy and speed of recognition. The architecture of the YOLOv8 model includes the input end, the backbone module, the neck module and the head modules, which can be classified into five types according to their depth and width: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x.

On the input side, adaptive scaling of the image, input dimension adjustment, mosaic data enhancement and other functions are realised; the backbone module is mainly composed of the CBS (Context-Based Spatial Attention) module, the C2f (Faster Implementation of CSP Bottleneck with 2 convolutions) module and the SPPF (Spatial Pyramid Pooling Fast) module. The CBS module contains the convolution, the batch normalisation and the SiLU activation function. The use of the CBS module can speed up the convergence of the model and prevent phenomena such as the disappearance of the gradient from occurring. The C2f module is based on the C3 module; it is obtained by adding the Split operation and jumping layer connection, which can make the gradient flow of the model richer under the premise of ensuring the model is lightweight. The SPPF module can realise the fusion of multi-scale features. The neck module uses a combination of FPN and PAN to make the feature fusion more adequate. The head module uses the current mainstream decoupled head structure. This structure separates detection and classification and determines the positive and negative samples based on the scores obtained by weighting the scores of classification and regression. This structure effectively improves the performance of the model. The structure diagram of the YOLOv8 model is shown in Figure 1 below.

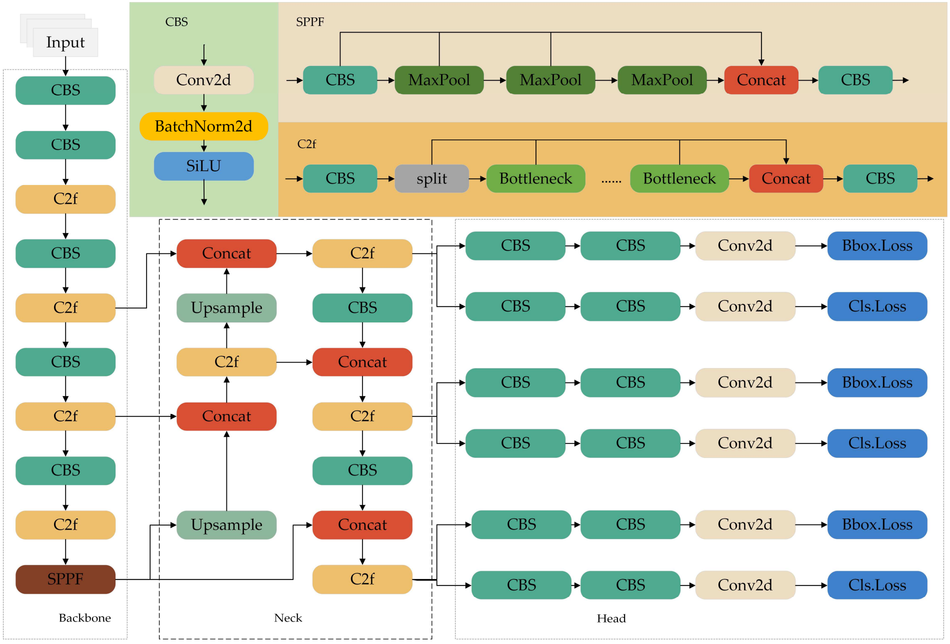

Figure 1. YOLOv8 network structure.

2.2. Improved YOLOv8 Model

The improved model structure is shown in Figure 2 below. Firstly, the model backbone network is reconstructed using RevColNet (Reversible Column Network), which effectively reduces the complexity of the model. Secondly, the designed C2fDWR module is incorporated into the backbone network to improve the recognition ability of the model. Again, GSConv is introduced at the neck end to effectively reduce the number of parameters of the model while maintaining the detection accuracy. Finally, the original loss function is replaced using InnerMPDIoU to improve the model generalisation ability. Through the above improvements, not only the hardware overhead of the model is effectively reduced, but also the prediction accuracy of the model is improved.

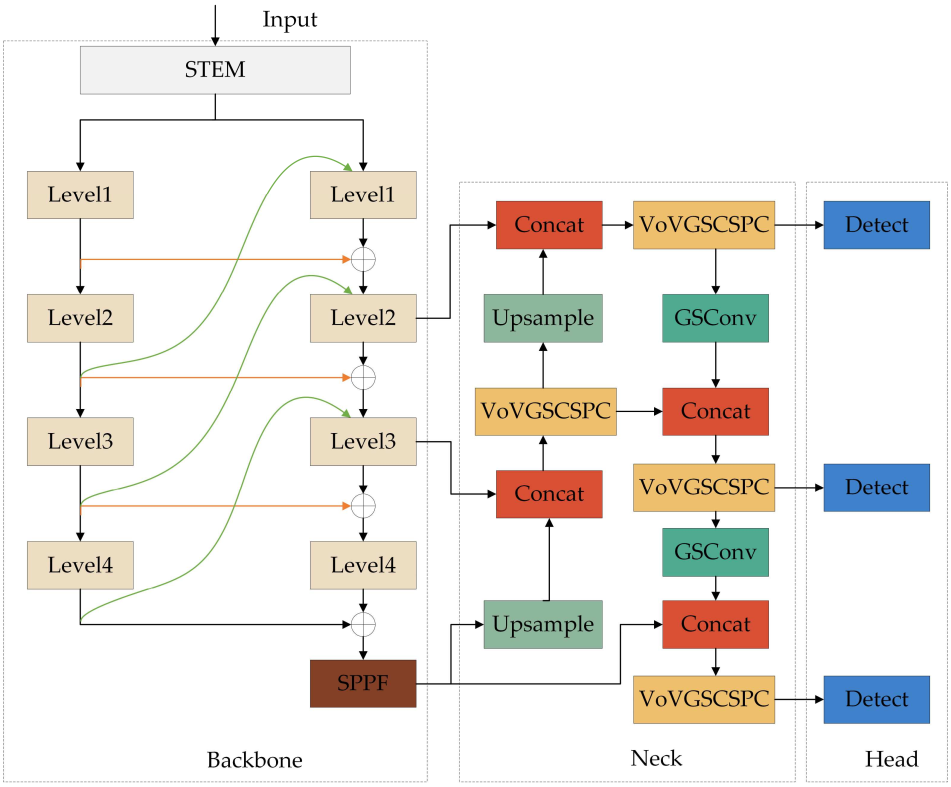

Figure 2. Improved model structure diagram.

2.2.1. Reconfiguration of the Backbone Network RevCol

The current YOLO series model backbone uses a top–down structure. This structure is prone to losing the information embedded in the image when extracting features, which in turn leads to the degradation of model performance. In order to solve this problem, this paper proposes a reconstructed backbone network based on the Reversible Connected Multi-Column Networks [31]. RevCol breaks through the information transfer mode of traditional straight-through networks and adopts a multi-input design, where the starting point of each column contains low-level information, and the semantic information is extracted from it through a compressed image channel. Adopting a reversible connection design between columns makes the network reversible, ensures that the data are transmitted without loss between columns and adds supervision at the end of each column to limit the feature extraction in each column. The macrostructure of RevCol is shown in Figure 3.

The input image is first segmented into a number of non-overlapping regions, which are then processed in each of the four hierarchical modules and finally combined with the inputs of the reversible operations to obtain the final result.

The microstructure of RevCol is shown below in Figure 4, where each level in Figure 4a performs feature extraction by downsampling and ConvNeXt, and Figure 4b demonstrates the design of reversible connections taken between columns, from which it can be seen that there are two inputs to each level, one from the previous hierarchy in the same column and the other from the previous column in the next hierarchy. The equations for the two inputs are shown in Equations (1) and (2).

$$X_t = F_t(X_{t-1}, X_{t-m+1}) + \gamma X_{t-m}, \tag{1}$$

$$X_{t-m} = \gamma^{-1}[X_t - F_t(X_{t-1}, X_{t-m+1})], \tag{2}$$

where X_t is the level t feature, $F_t(.)$ is the activation function, γ is an invertible operation and γ^{-1} is its inverse. Each column is composed of m feature maps within a group.

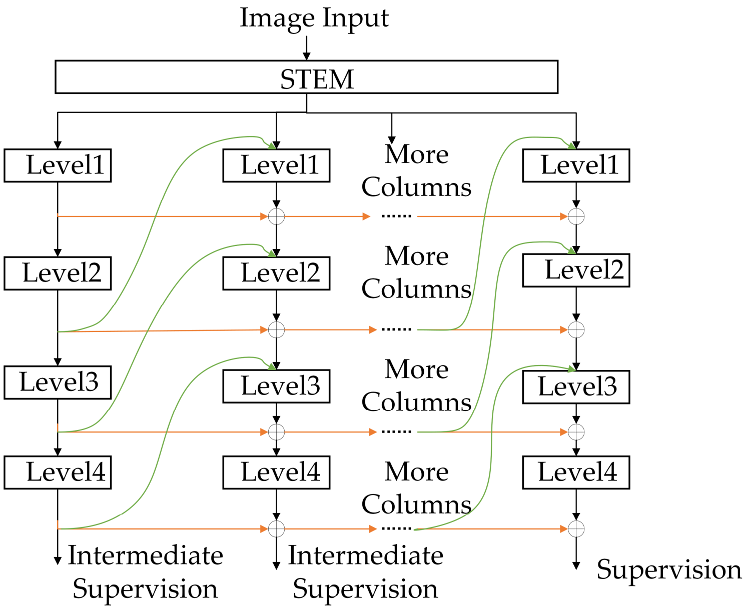

Figure 3. The macrostructure of RevCol.

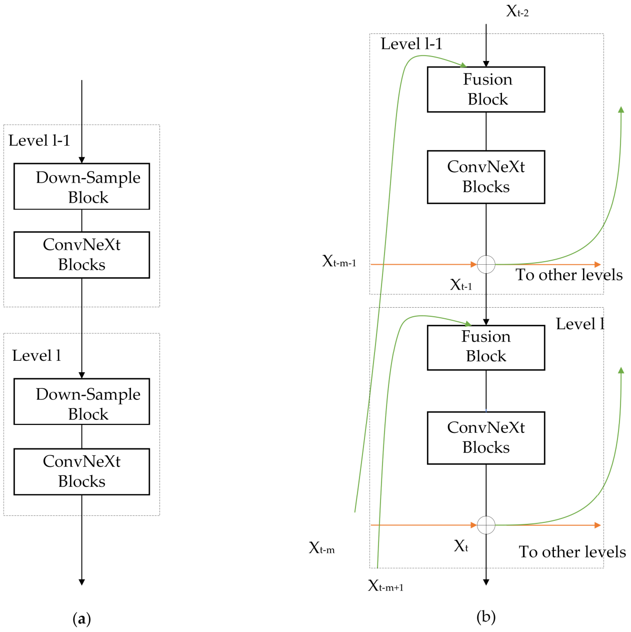

Figure 4. (**a**) Firsts column level; (**b**) second and subsequent middle columns.

In order to avoid the backbone network being too complex, leading to a rise in the complexity and parameter count of the model, this paper sets the number of columns of RevCol to 2. At the same time, the operations in Fusion Block are reconstructed, and for the high-level semantic information, the downsampling is performed by convolution, batch normalisation and activation function operations. For low-level semantic information, convolution and upsampling are used for the operations, while the ConvNeXt module is replaced in level with the C2f module in YOLOv8.

2.2.2. Expandable Residual Attention Module DWR (Dilation-Wise Residual)

The traditional YOLO series model has certain deficiencies in small-target detection due to its multi-scale nature, so this paper introduces a fusion-expandable residual attention module [32]. This module is mainly applied to deep networks, and the multi-branching structure meets the network's needs for different sizes of receptive fields, and its structure is shown in Figure 5. For each branch of the input feature map, a 3 × 3 kernel normalised convolution operation is performed, and then the batch normalisation layer and ReLU layer are combined for feature extraction. Since each output channel contains several small spatial regions that need to be refined, the final output result is composed of these regions. Then, on this basis, the semantic information is extracted by using the deep 3 × 3 convolution, and then the semantic residuals are obtained using the BN layer, followed by concatenation operations on the branches, and then all the feature maps are fused using the point-by-point convolution method to obtain the final residuals. Finally, the final residuals are merged on the input feature maps to construct a more complete feature representation. In addition, features extracted with small receptive fields are relatively important, so the number of hollow depth convolution channels with the lowest null rate is set to c, and the number of other channels to c/2.

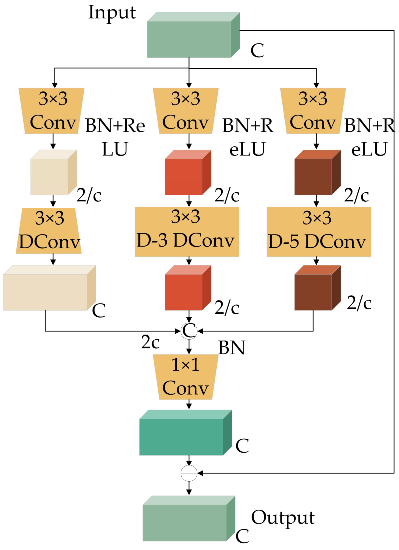

Figure 5. The structure of DWR.

Based on the DWR module, this paper further designs the C2fDWR module, whose structure is shown in Figure 6 below.

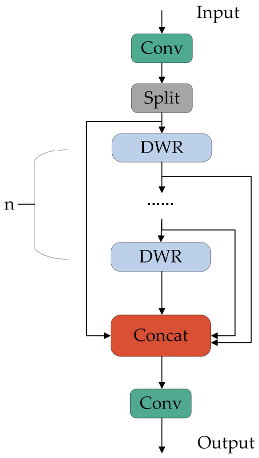

Figure 6. The structure of C2FDWR.

In order to make up for the deficiency of the model in small-target identification, the designed C2fDWR module replaces the C2f module of the reconfigured backbone RevCol, thus forming a new backbone network RVDR, whose structure is shown in Figure 7 below. The STEM module can divide the input image into several non-overlapping blocks.

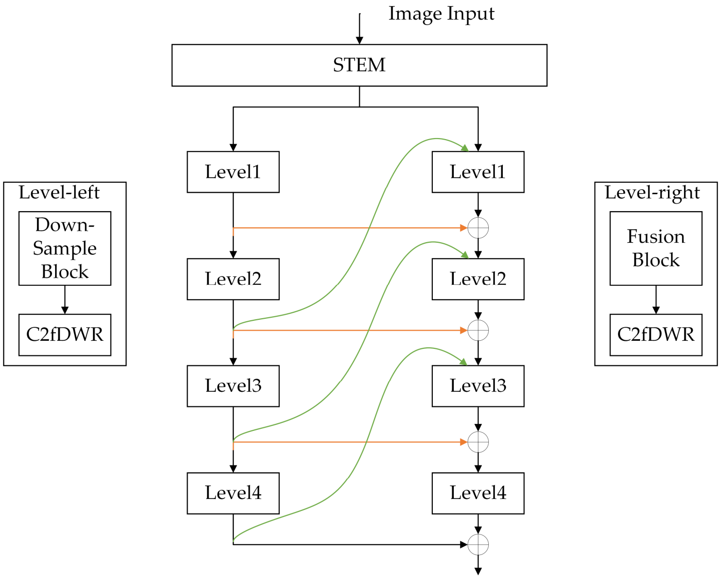

Figure 7. The new backbone network structure.

2.2.3. GSConv

To better accommodate weeding machine equipment with limited computational and storage resources, the model needs to be designed for lightweighting. This design not only reduces the computational cost, but also speeds up the model detection. And replacing the traditional convolution with depth-separable convolution is a good method of lightweighting. Compared with traditional convolution, depth-separable convolution can effectively reduce the computational amount by layering the feature layer of the input channel, but it also causes the loss of information between the channels. To address this problem, this paper introduces the GSConv module based on depth-separable convolution [33], whose main structure is shown in Figure 8 below. The number of input channels is C1, and the number of output channels is C2. Firstly, after a standard convolution to make the number of channels into C2/2, then through the depth-separable convolution to obtain the same number of channels, finally, the two results will be subjected to joining and mixing operations. Using GSConv can keep the information of multiple channels and enhance the feature expression ability of the image while reducing the amount of operations.

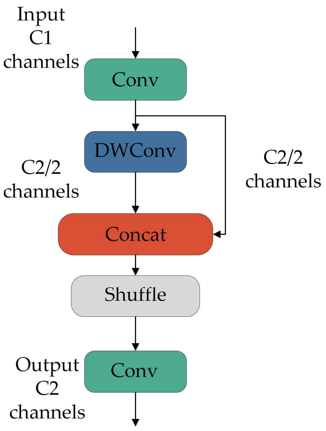

Figure 8. The structure of GSConv.

In this paper, GSConv is introduced to the neck layer to reduce the number of parameters and the computational complexity of the neck module. And the VoVGSCSPC based on GSConv and GSbottleneck is used to replace the CSP module of the original model to further improve the performance of YOLOv8n. GSbottleneck and VoVGSCSPC are structured as shown in Figure 9 below.

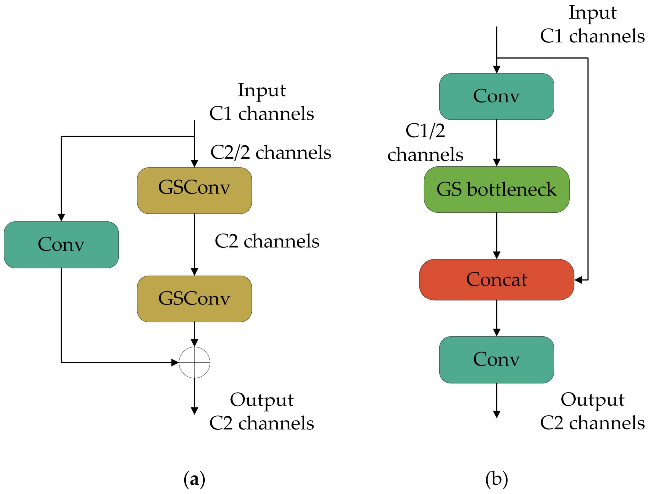

Figure 9. Improved module based on GSConv: (**a**) GSbottleneck module; (**b**) VOVGSCSPC module.

2.2.4. Based on the Auxiliary Border InnerMPDIoU

The loss function used in the YOLOv8 model for bounding box regression is CIoU Loss. CIoU is based on DIoU, and the aspect ratio of the bounding box is added to the loss function so as to improve the accuracy of regression. But most of the BBR loss functions represented by CIoU may have the same value under different prediction results, which reduces the convergence speed and accuracy of bounding box regression. Therefore, in this paper, a novel loss function based on the minimum point distance, MPDIoU, is introduced as a loss function to improve the bounding box regression of the YOLOv8 model. MPDIoU compares the similarity between the predicted bounding box and the actual labelled bounding box during the bounding box regression process by directly calculating the distance of the key points between the predicted box and the real box [34]. The formula for MPDIoU is as follows:

$$LMPDIoU = 1 - MPDIoU, \tag{3}$$

$$MPDIoU = IoU - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2}, \tag{4}$$

$$d_1^2 = \left(x_1^B - x_1^A\right)^2 + \left(y_1^B - y_1^A\right)^2, \tag{5}$$

$$d_2^2 = \left(x_2^B - x_2^A\right)^2 + \left(y_2^B - y_2^A\right)^2, \tag{6}$$

$$IoU = \frac{inter}{union}, \tag{7}$$

where A and B are the prediction frame and the real frame, respectively, w and h denote the width and height of the input image, respectively, (x_1^A, y_1^A) and (x_2^A, y_2^A) denote the coordinates of the upper-left point of A and the coordinates of the lower-right point of A, respectively, inter represents the intersection of the Target Box and Anchor Box, union represents the union of the Target Box and Anchor Box, and (x_1^B, y_1^B) and (x_2^B, y_2^B) stand for the coordinates of the upper-left point of B and the coordinates of the lower-right point of B, respectively.

Most of the existing IoU improvement methods use the addition of new loss items. This approach ignores the limitations that the loss term itself has. In this paper, InnerIoU is introduced into MPDIoU, and a new loss function InnerMPDIoU is proposed as the model bounding box regression loss function.

InnerIoU is a new improvement method. It differs from the traditional improvement algorithm by analysing the use of different scales of auxiliary borders in the regression to calculate the loss, so as to speed up the speed of border regression. InnerIoU introduces a scale factor ratio to control the size of the auxiliary borders to calculate the loss, and the general value of the ratio is within [0.5,1.5]. When the ratio is larger than 1, it will produce an auxiliary border that is larger than the real border, thus increasing the effective range of regression, which is helpful for the regression of low IoU samples. When the ratio is less than 1, an auxiliary margin computational loss smaller than the true margin is incurred, making the absolute value of the regression gradient larger than the absolute value of the actual margin IoU gradient. Therefore, a high IoU sample regression is favoured at this time [35]. The InnerMPDIoU calculation process is shown below.

$$IoU^{inner} = \frac{inter}{union}, \tag{8}$$

$$L_{InnerMPDIoU} = L_{MPDIoU} + IoU - IoU^{inner}, \tag{9}$$

where inter represents the intersection of the InnerTarget Box and InnerAnchor Box, and union represents the union of InnerTarget Box and InnerAnchor Box. The formula of inter and union is as follows:

$$inter = \left[min(b_r^{gt}, b_r) - max(b_l^{gt}, b_l)\right] \times \left[(min(b_b^{gt}, b_b) - max(b_t^{gt}, b_t)\right], \tag{10}$$

$$union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter, \tag{11}$$

where b_l, b_r, b_t and b_b represent the positions of the four vertices of the InnerAnchor Box, b_l^{gt}, b_r^{gt}, b_t^{gt} and b_b^{gt} represent the positions of the four vertices of the InnerTarget Box, ratio represents the scale factor, w^{gt} and h^{gt} represent the width and height of the InnerTarget Box, and w and h represent the width and height of the InnerAnchor Box, respectively. The formulae for these variables are as follows:

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, \tag{12}$$

$$b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2}, \tag{13}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, \tag{14}$$

$$b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2}, \tag{15}$$

$$b_l = x_c - \frac{w \times ratio}{2}, \tag{16}$$

$$b_r = x_c + \frac{w \times ratio}{2}, \tag{17}$$

$$b_t = y_c - \frac{h \times ratio}{2}, \tag{18}$$

$$b_b = y_c + \frac{h \times ratio}{2}, \tag{19}$$

where $\left(x_c^{gt}, y_c^{gt}\right)$ represents the centre point in the InnerTarget Box and (x_c, y_c) represents the centre point in the InnerAnchor Box.

3. Experiments

3.1. Dataset Production

This experiment is based on the publicly available dataset Weed25 [36], with farmland and grassland as the research object. The dataset contains a total of 14,035 images and 25 weed categories. The 25 categories of weeds not only belong to 14 families, but also contain the growth period of each type of weed, which is more diverse than the existing dataset, and it is very suitable for applying to the training of weed detection models. The dataset also includes deteriorating conditions, such as plants overlapping with weeds, which makes detection difficult.

Due to the differences in the number of images of each weed in Weed25, we selected 12 of them with a similar number of images of different types of weeds as training samples, and sequentially allocated the dataset of each category into the training set, validation set and test set in the ratio of 8:1:1. The specific data distribution is shown in Figure 10.

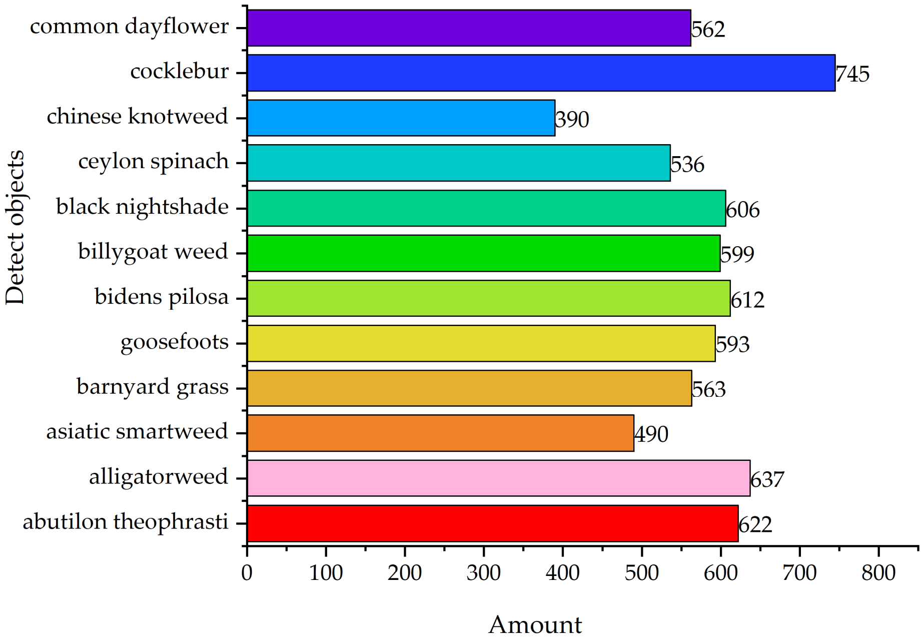

Figure 10. The number of each detected object in the dataset.

3.2. Experimental Configuration

This experiment uses Ubuntu 20.02.4 LTS as the operating system, the GPU is the NVIDIA GeForce RTX 2080Ti graphics card with 10 GB of video memory (ASUS, Taipei, China), the Anaconda3 development training virtual environment is used, the code environment is Python 3.9, PyTorch1.12.1 and Cuda is 11.3. The experiment uses the hyperparameter configuration: an initial learning rate of 0.01, an epoch of 300, a batch size of 32, a momentum setting of 0.937, a weight loss factor of 0.0005 and an optimiser of SGD. Images were taken at a height and angle of approximately 30–50 cm and 60–90°, respectively, with a digital camera (Nikon D5300 SLR, Osaka, Japan) or a smartphone (Huawei Enjoy 9S, Chongqing, China).

3.3. Model Evaluation Indicators

In this paper, the main evaluation indexes of the model are accuracy rate, recall rate and mean average precision. The accuracy rate is the proportion of the number of samples predicted by the detector to be positive classes that are really positive classes, the recall rate is the proportion of the number of samples predicted by the detector to be positive classes that are really positive classes, and mAP refers to the mean average precision, which indicates the average precision of the detector's detection results for all classes. The formulas for the definition of accuracy and recall are as follows.

$$P = \frac{TP}{TP + FP}, \tag{20}$$

$$R = \frac{TP}{TP + FN}, \tag{21}$$

where TP means both the prediction and actual are true; FP means the prediction is true and the actual is false; and FN means the prediction is false and the actual is true.

AP represents how well the model performs on each category. mAP is the average of all APs, which measures the performance of the model on the overall dataset. mAP is defined by the formula shown below. mAP_{50} represents the total category average accuracy when IoU is equal to 0.5, and $mAP_{50:95}$ represents the average accuracy of different IoU values (from 0.5 to 0.95 in steps of 0.05).

$$mAP = \frac{1}{n}\sum_{i=1}^{n}\int_{0}^{1} P(R)dR \tag{22}$$

In addition, the number of model parameters, computational complexity and model file size were included in the model evaluation metrics in order to reflect the advantages over other models.

4. Results and Discussion

4.1. Ablation Experiments

In order to prove the effectiveness of the improved method, this study conducted ablation experiments on the Weed25 dataset, and the results of the experiments are shown in Table 1 below.

Table 1. Improved point ablation experiment.

Models	mAP_{50} (%)	$mAP_{50\text{-}95}$ (%)	FLOPs (G)	Params (M)	Size (MB)	P (%)	R (%)
YOLOv8	92.1	62.3	8.1	3.00	6.3	90.4	88.5
+RevCol	92.8	62.9	6.3	2.27	4.9	91.6	87.9
+Slimneck	92.8	62.9	7.1	2.70	5.7	90.6	87.6
+RevCol+Slimneck	92.7	62.4	5.3	1.97	4.4	90.4	86.2
+RevCol+C2fDWR+Slimneck	92.9	63.2	5.2	1.94	4.4	91.5	86.3
+RevCol+C2fDWR+Slimneck+InnerMPDIoU	93.8	63.4	5.2	1.94	4.4	92.9	88.3

The analysis of the table shows that after incorporating RevCol for the YOLOv8 model, the number of parameters of the model decreased by 24.2%, the model size and GFLOPS decreased by 22.2%, and after incorporating Slimneck for the YOLOv8 model, the number of parameters of the model decreased by 10%, the GFLOPS decreased by 12.3% and the model size decreased by 9.5%. This proves the effectiveness of both improvement strategies. After combining and integrating these two improvements into the network, although there was a slight decrease in $mAP_{50\text{-}95}$ values, the model's parameter count, computational complexity and model size were further reduced. And after incorporating C2fDWR on this basis, not only the parameters and computational complexity of the model were reduced, but the mAP$_{50\text{-}95}$ value was also improved from 62.4% to 63.4%, and finally after incorporating InnerMPDIoU, the mAP$_{50}$ value and accuracy were improved while ensuring that the other parameters were basically unchanged. Finally, compared with YOLOv8, the improved model proposed in this study reduced the computational complexity by 35.8%, the number of parameters by 35.4% and the model size by 30.2%, while the mAP$_{50}$ and mAP$_{50\text{-}95}$ values were improved to 93.8% and 63.4%, respectively, the precision value was improved to 92.9% and the performance of the improved model outperformed the original YOLOv8 model in several aspects.

The results of the ablation experiments based on other innovations for the ratio in InnerMPDIoU are shown in Table 2 below.

Table 2. Ratio ablation experiment.

Models	Ratio	mAP$_{50}$(%)	mAP$_{50\text{-}95}$(%)	P(%)	R(%)
RVDR-YOLOv8	0.71	92.4	62.8	89.8	88.1
RVDR-YOLOv8	0.81	92.6	62.9	88.8	88.1
RVDR-YOLOv8	0.91	92.6	63.3	91.6	86.1
RVDR-YOLOv8	1	92.9	63.3	91.8	87.1
RVDR-YOLOv8	1.15	93.1	63.0	89.8	87.5
RVDR-YOLOv8	1.17	93.2	63.3	90.1	89.0
RVDR-YOLOv8	1.25	93.8	63.4	92.9	88.3
RVDR-YOLOv8	1.27	93.1	63.4	89.9	89.3
RVDR-YOLOv8	1.35	93.0	63.3	89.0	89.1

When the ratio > 1, the obtained margins are larger than the actual margins, which is favourable for the regression of low IoU samples. The experimental data show that the model performs better at a ratio > 1 than at a ratio < 1, and the experiment works best overall when the ratio = 1.25. The exact value of ratio needs to be adjusted according to the dataset.

4.2. Comparison Experiments

In order to compare the models in terms of performance, we chose Faster R-CNN, SSD, YOLOv3, YOLOv4tiny, YOLOv7tiny, YOLOv7l, EfficientDet and RetinaNet for the comparative experiments, and chose the mAP$_{50\text{-}95}$ values, the mAP$_{50}$ values, the number of model parameters and the computational complexity as the comparison metrics. The results of the comparison experiments are shown in Table 3 below.

Table 3. The results of the experiments for comparison with other models.

Models	mAP$_{50}$(%)	mAP$_{50\text{-}95}$(%)	FLOPs(G)	Params(M)
Faster R-CNN	89.6	55.8	369.9	136.9
SSD	88.2	52.6	278.0	25.1
YOLOv3	84.5	51.6	155.4	61.6
YOLOv4tiny	74.9	38.7	16.2	5.9
YOLOv5n	92.2	61.8	7.1	2.5
YOLOv7tiny	90.7	57.6	13.2	6.0
YOLOv7l	91.4	58.9	105.3	37.3
EfficientDet	87.2	53.6	11.6	6.6
RetinaNet	90.3	59.0	126.7	19.9
YOLOv8n	92.1	62.3	8.1	3.0
RVDR-YOLOv8	93.8	63.4	5.2	1.9

As can be seen from Table 3, the improved model RVDR-YOLOv8 not only has the smallest computational complexity and number of parameters, but also has a great improvement in detection ability, which is the best performance among many models. Based on the above conclusions, the RVDR-YOLOv8 model proposed in this paper, although not as fast as YOLOv8 in detection speed, still achieves real-time detection, and thus outperforms other models in the weed detection task in an integrated manner.

4.3. Experimental Analysis

In order to demonstrate the enhancement effect of the proposed model compared to the original model, comparative experiments are conducted in this paper. Table 4 shows the AP values for each category derived from the original model and the improved model tested on the dataset, and through the table it can be seen that the improved model has improved the AP values of most of the categories to different degrees, indicating that the improved method improves the model's detection accuracy of the target. Values where there has been an improvement are shown in red and deteriorating values are shown in green.

Table 4. The YOLOv8n and our model for each category of AP.

Detect Objects	**AP(%)**	
	YOLOv8n	**Ours (RVDR-YOLOv8)**
Abutilon theophrasti	74.7	76.2
Alligatorweed	57.6	58.8
Asiatic smartweed	33.4	32.6
Barnyard grass	60.1	61.2
Goosefoots	69.9	72.2
Bidens pilosa	72.6	73.0
Billygoat weed	65.9	65.1
Black nightshade	53.9	54.8
Ceylon spinach	77.9	80.5
Chinese knotweed	36.3	40.5
Cocklebur	74.6	75.2
Common dayflower	70.7	70.6

In order to visualise the feature extraction capability of the network, we use the Grad-GAM technique to plot the heatmap of YOLOv8 and RVDR-YOLOv8 for target recognition attention, as shown in Figure 11 below. By analysing the heatmap, it can be seen that the improved model exhibits higher intensity in recognising the target region, thus suggesting that the improved model has a stronger capability compared to the baseline model in extracting the features and exploiting the features.

Figure 12 shows the confusion matrix of the improved model, where the horizontal axis represents the true values and the vertical axis represents the predicted values, by analysing the picture it can be seen that most of the true values correspond to the predicted values, which shows that the model has a good performance.

Figure 13 shows the prediction results of YOLOv8 and RVDR-YOLOv8, and the comparison shows that the improved model not only achieves accurate target localisation, but also has some improvement for the confidence level.

Figure 14 shows the comparison graph between the original model and the improved model in terms of the number of parameters, model size and GFLOPS, from which it can be seen that the improved model has a greater degree of reduction in all the three evaluation metrics, which demonstrates that the improved model is lighter compared to the original model and is more suitable for application to weeding robots with limited computational and storage resources.

Figure 15 shows the comparison of the improved model RVDR-YOLOv8 with Faster R-CNN, SSD and other models in terms of the number of parameters. The comparison results show that the improved model has a greater degree of reduction in the number of parameters compared to the other models.

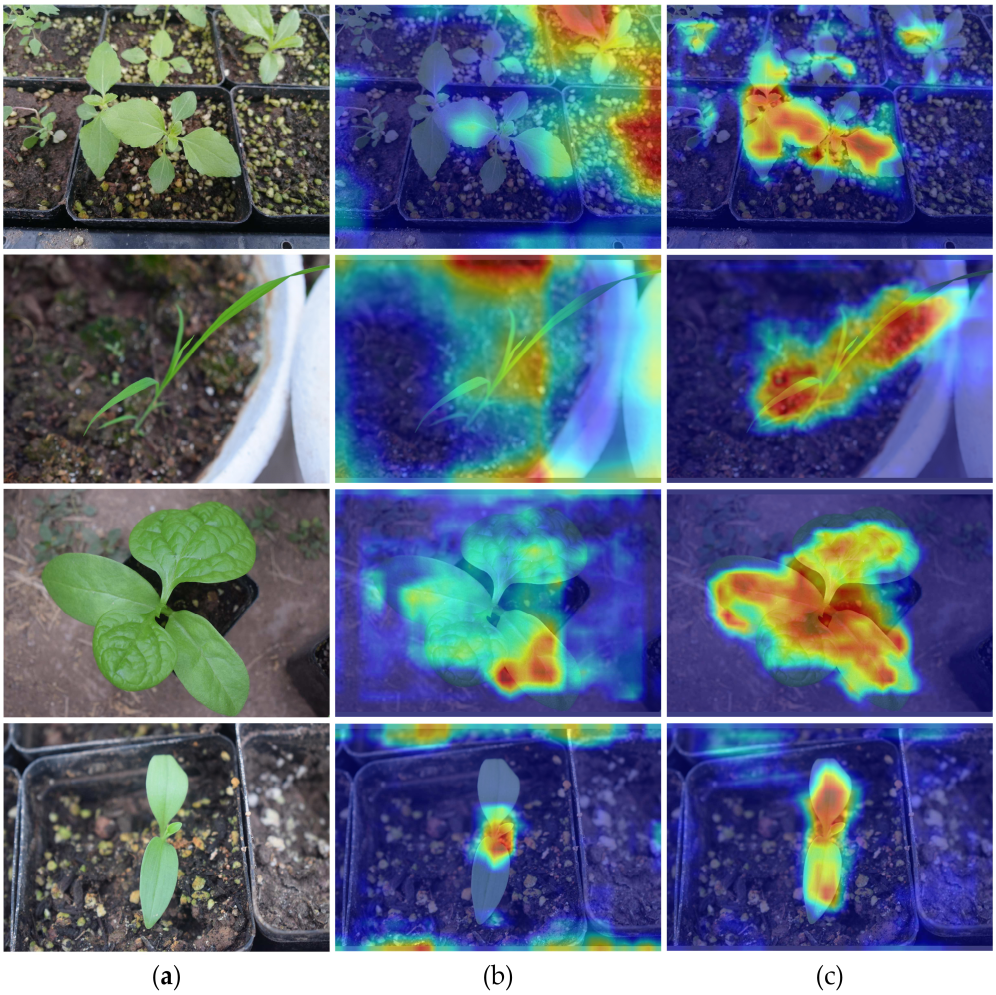

Figure 11. (**a**) Input image; (**b**) YOLOv8 Grad-CAM figure; (**c**) RVDR-YOLOv8 Grad-CAM figure.

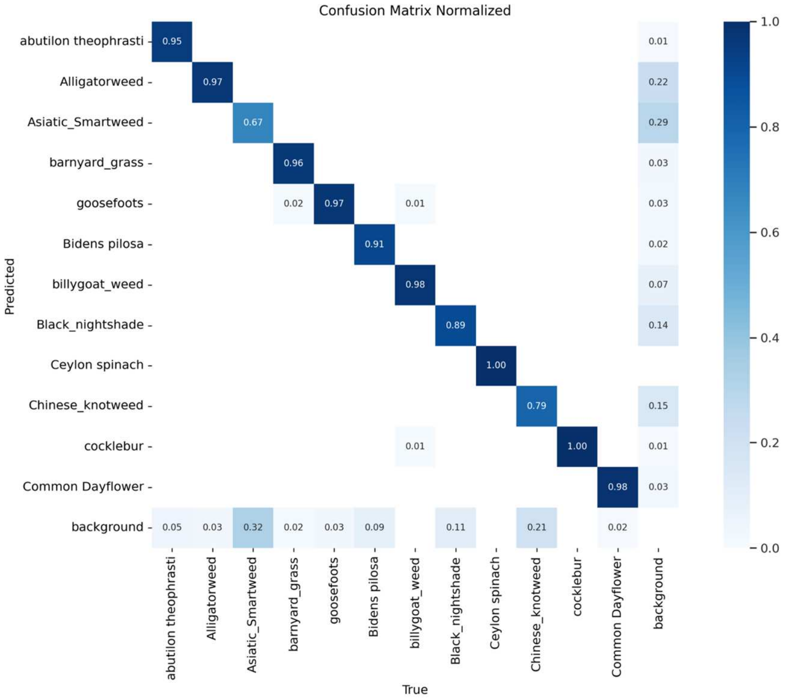

Figure 12. Confusion matrix graph of the improved YOLOv8 algorithm.

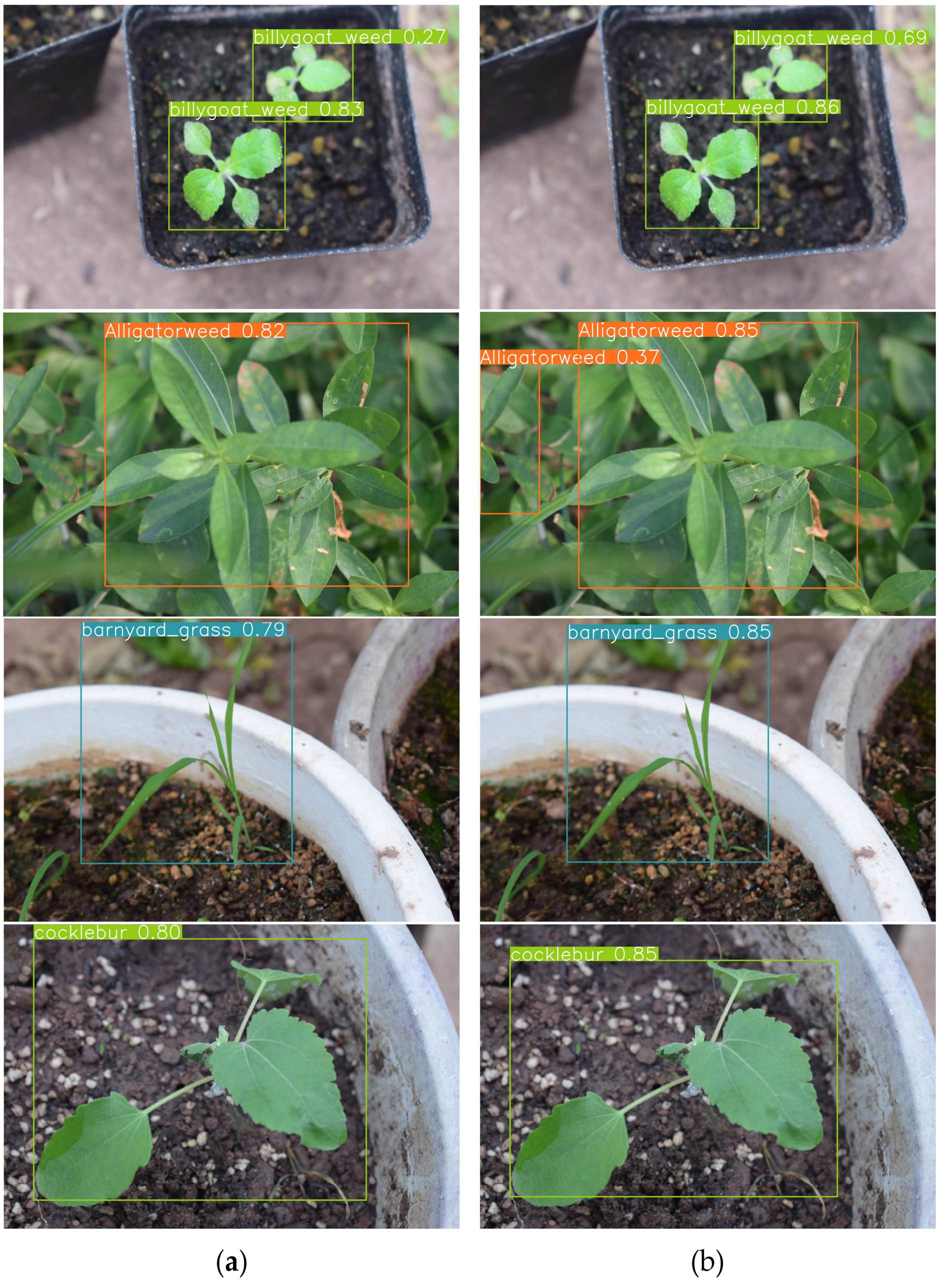

Figure 13. Comparison of detection results: (**a**) detection effect of YOLOv8 model; (**b**) detection effect of RVDR-YOLOv8 model.

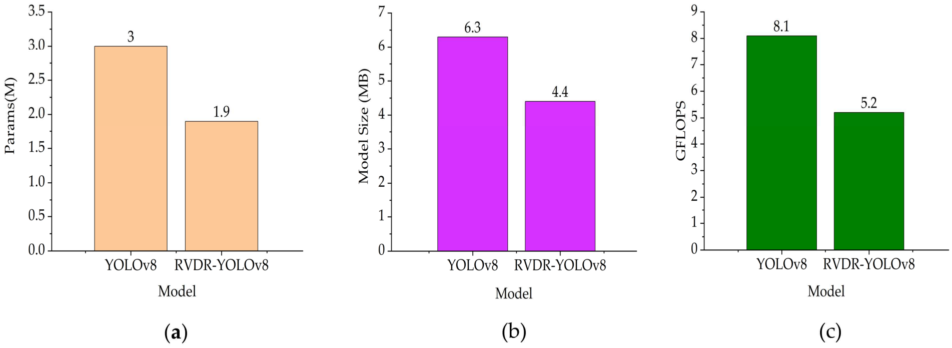

Figure 14. Comparison of YOLOv8 and improved model results: (**a**) parameter comparison results; (**b**) model size comparison results; (**c**) GFLOPS comparison results.

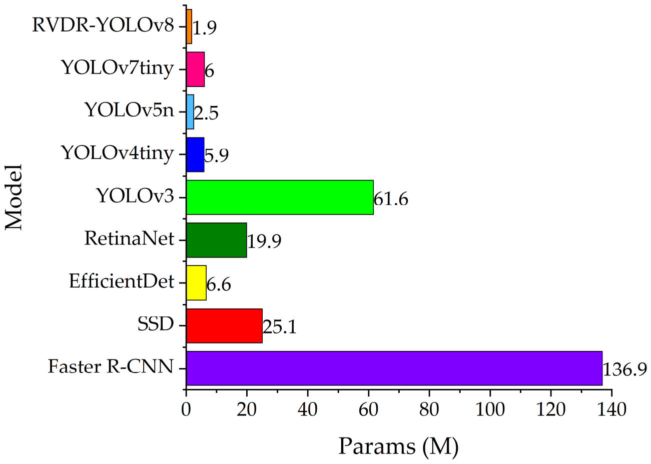

Figure 15. Results of the improved model RVDR-YOLOv8 versus other models in terms of the number of parameters.

5. Conclusions

The accurate identification of weeds is very important for implementing field weed control. In this paper, a weed target detection model based on improved YOLOv8 is proposed, which not only can accurately detect the target, but also makes the model more lightweight. Specifically, firstly, the computational complexity of the model and the number of network parameters are effectively reduced by reconfiguring the backbone network RVDR. Subsequently, the GSConv and VoVGSCSPC modules are used at the neck end to reduce the size of the model. Finally, the original loss function is optimised by InnerMPDIoU Loss to improve the model detection accuracy. Compared to the original model, the computational complexity is reduced by 35.8%, the number of parameters is reduced by 35.36%, the model size is reduced by 30.15%, the mAP_{50} value is improved to 93.8% and the $mAP_{50\text{-}95}$ value is improved to 63.4%, which makes the proposed improved model well suited to be used in weed control robots with limited computational and storage resources.

Although the improved model is able to achieve real-time detection, the FPS value is decreased compared to the original model. Follow-up work will aim to improve the model detection accuracy while reducing the model inference time.

Author Contributions: Conceptualization, Y.D. and C.J.; methodology, C.J. and L.S.; software, C.J.; validation, C.J. and F.L.; investigation, Y.D. and Y.T.; writing—original draft preparation, C.J.; writing—review and editing, C.J. and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

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

References

1. Zhang, Y.; Wang, M.; Zhao, D.; Liu, C.; Liu, Z. Early weed identification based on deep learning: A review. *Smart Agric. Technol.* **2023**, *3*, 100123. [CrossRef]
2. Chen, D.; Lu, Y.; Li, Z.; Young, S. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Comput. Electron. Agric.* **2022**, *198*, 107091. [CrossRef]
3. Li, Y.; Al-Sarayreh, M.; Irie, K.; Hackell, D.; Bourdot, G.; Reis, M.M.; Ghamkhar, K. Identification of weeds based on hyperspectral imaging and machine learning. *Front. Plant Sci.* **2021**, *11*, 611622. [CrossRef]
4. Elstone, L.; How, K.Y.; Brodie, S.; Ghazali, M.Z.; Heath, W.P.; Grieve, B. High speed crop and weed identification in lettuce fields for precision weeding. *Sensors* **2020**, *20*, 455. [CrossRef]
5. Zhao, X.; Wang, X.; Li, C.; Fu, H.; Yang, S.; Zhai, C. Cabbage and weed identification based on machine learning and target spraying system design. *Front. Plant Sci.* **2022**, *13*, 924973. [CrossRef]

6. Etienne, A.; Ahmad, A.; Aggarwal, V.; Saraswat, D. Deep learning-based object detection system for identifying weeds using uas imagery. *Remote Sens.* **2021**, *13*, 5182. [CrossRef]
7. Rai, N.; Zhang, Y.; Ram, B.G.; Schumacher, L.; Yellavajjala, R.K.; Bajwa, S.; Sun, X. Applications of deep learning in precision weed management: A review. *Comput. Electron. Agric.* **2023**, *206*, 107698. [CrossRef]
8. Ahmad, A.; Saraswat, D.; Aggarwal, V.; Etienne, A.; Hancock, B. Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. *Comput. Electron. Agric.* **2021**, *184*, 106081. [CrossRef]
9. Bakhshipour, A.; Jafari, A.; Nassiri, S.M.; Zare, D. Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst. Eng.* **2017**, *157*, 1–12. [CrossRef]
10. Tellaeche, A.; Pajares, G.; Burgos-Artizzu, X.P.; Ribeiro, A. A computer vision approach for weeds identification through Support Vector Machines. *Appl. Soft Comput.* **2011**, *11*, 908–915. [CrossRef]
11. Yu, H.; Che, M.; Yu, H.; Ma, Y. Research on weed identification in soybean fields based on the lightweight segmentation model DCSAnet. *Front. Plant Sci.* **2023**, *14*, 1268218. [CrossRef]
12. Yang, L.; Xu, S.; Yu, X.; Long, H.; Zhang, H.; Zhu, Y. A new model based on improved VGG16 for corn weed identification. *Front. Plant Sci.* **2023**, *14*, 1205151. [CrossRef]
13. Zhang, J.; Gong, J.; Zhang, Y.; Mostafa, K.; Yuan, G. Weed identification in maize fields based on improved Swin-Unet. *Agronomy* **2023**, *13*, 1846. [CrossRef]
14. Mu, Y.; Feng, R.; Ni, R.; Li, J.; Luo, T.; Liu, T.; Li, X.; Gong, H.; Guo, Y.; Sun, Y.; et al. A faster R-CNN-based model for the identification of weed seedling. *Agronomy* **2022**, *12*, 2867. [CrossRef]
15. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
16. Yang, Y.; Wang, H.; Jiang, D.; Hu, Z. Surface detection of solid wood defects based on SSD improved with ResNet. *Forests* **2021**, *12*, 1419. [CrossRef]
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Cham, Switzerland, 2016; pp. 21–37.
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
19. Li, J.; Li, C.; Fei, S.; Ma, C.; Chen, W.; Ding, F.; Wang, Y.; Li, Y.; Shi, J.; Xiao, Z. Wheat ear recognition based on RetinaNet and transfer learning. *Sensors* **2021**, *21*, 4845. [CrossRef]
20. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *arXiv* **2017**, arXiv:1708.02002.
21. Zhai, X.; Huang, Z.; Li, T.; Liu, H.; Wang, S. YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection. *Electronics* **2023**, *12*, 3664. [CrossRef]
22. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef]
23. Elmessery, W.M.; Gutiérrez, J.; El-Wahhab, G.G.A.; Elkhaiat, I.A.; El-Soaly, I.S.; Alhag, S.K.; Al-Shuraym, L.A.; Akela, M.A.; Moghanm, F.S.; Abdelshafie, M.F. YOLO-based model for automatic detection of broiler pathological phenomena through visual and thermal images in intensive poultry houses. *Agriculture* **2023**, *13*, 1527. [CrossRef]
24. Tan, L.; Huangfu, T.; Wu, L.; Chen, W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 324. [CrossRef]
25. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
26. Wu, H.; Wang, Y.; Zhao, P.; Qian, M. Small-target weed-detection model based on YOLO-V4 with improved backbone and neck structures. *Precis. Agric.* **2023**, *24*, 2149–2170. [CrossRef]
27. Wang, Q.; Cheng, M.; Huang, S.; Cai, Z.; Zhang, J.; Yuan, H. 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.* **2022**, *199*, 107194. [CrossRef]
28. Li, J.; Zhang, W.; Zhou, H.; Yu, C.; Li, Q. Weed detection in soybean fields using improved YOLOv7 and evaluating herbicide reduction efficacy. *Front. Plant Sci.* **2024**, *14*, 1284338. [CrossRef]
29. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **2023**, *11*, 677. [CrossRef]
30. Li, Y.; Fan, Q.; Huang, H.; Han, Z.; Gu, Q. A modified YOLOv8 detection network for UAV aerial image recognition. *Drones* **2023**, *7*, 304. [CrossRef]
31. Cai, Y.; Zhou, Y.; Han, Q.; Sun, J.; Kong, X.; Li, J.; Zhang, X. Reversible column networks. *arXiv* **2022**, arXiv:2212.11696.
32. Wei, H.; Liu, X.; Xu, S.; Dai, Z.; Dai, Y.; Xu, X. DWRSeg: Rethinking Efficient Acquisition of Multi-scale Contextual Information for Real-time Semantic Segmentation. *arXiv* **2022**, arXiv:2212.01173.
33. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv* **2022**, arXiv:2206.02424.
34. Siliang, M.; Yong, X. Mpdiou: A loss for efficient and accurate bounding box regression. *arXiv* **2023**, arXiv:2307.07662.

35. Zhang, H.; Xu, C.; Zhang, S. Inner-iou: More effective intersection over union loss with auxiliary bounding box. *arXiv* **2023**, arXiv:2311.02877.
36. Wang, P.; Tang, Y.; Luo, F.; Wang, L.; Li, C.; Niu, Q.; Li, H. Weed25: A deep learning dataset for weed identification. *Front. Plant Sci.* **2022**, *13*, 1053329. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.