

## Article

# Improved YOLOv8-Seg Based on Multiscale Feature Fusion and Deformable Convolution for Weed Precision Segmentation

Zhuxi Lyu , Anjiang Lu \* and Yinglong Ma

College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China; zhuxilv@gmail.com (Z.L.)

\* Correspondence: ajlu@gzu.edu.cn

**Abstract:** Laser-targeted weeding methods further enhance the sustainable development of green agriculture, with one key technology being the improvement of weed localization accuracy. Here, we propose an improved YOLOv8 instance segmentation based on bidirectional feature fusion and deformable convolution (BFFDC-YOLOv8-seg) to address the challenges of insufficient weed localization accuracy in complex environments with resource-limited laser weeding devices. Initially, by training on extensive datasets of plant images, the most appropriate model scale and training weights are determined, facilitating the development of a lightweight network. Subsequently, the introduction of the Bidirectional Feature Pyramid Network (BiFPN) during feature fusion effectively prevents the omission of weeds. Lastly, the use of Dynamic Snake Convolution (DSConv) to replace some convolutional kernels enhances flexibility, benefiting the segmentation of weeds with elongated stems and irregular edges. Experimental results indicate that the BFFDC-YOLOv8-seg model achieves a 4.9% increase in precision, an 8.1% increase in recall rate, and a 2.8% increase in mAP50 value to 98.8% on a vegetable weed dataset compared to the original model. It also shows improved mAP50 over other typical segmentation models such as Mask R-CNN, YOLOv5-seg, and YOLOv7-seg by 10.8%, 13.4%, and 1.8%, respectively. Furthermore, the model achieves a detection speed of 24.8 FPS on the Jetson Orin nano standalone device, with a model size of 6.8 MB that balances between size and accuracy. The model meets the requirements for real-time precise weed segmentation, and is suitable for complex vegetable field environments and resource-limited laser weeding devices.

**Keywords:** YOLOv8; weed segmentation; bi-directional feature pyramid network; dynamic snake convolution; laser weeding



**Citation:** Lyu, Z.; Lu, A.; Ma, Y. Improved YOLOv8-Seg Based on Multiscale Feature Fusion and Deformable Convolution for Weed Precision Segmentation. *Appl. Sci.* **2024**, *14*, 5002. <https://doi.org/10.3390/app14125002>

Received: 21 April 2024

Revised: 2 June 2024

Accepted: 5 June 2024

Published: 7 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The issue of weeds inevitably impacts crop yields of agricultural production. In China, there are 1430 species (varieties) of agricultural weeds, leading to a reduction in grain production of up to 60 million tons and economic losses of up to USD 30 billion. Current main weeding control methods include biological weeding, chemical weeding, traditional mechanical weeding, and targeted weeding [1–3]. Biological weeding involves introducing foreign organisms to suppress weed growth, but these organisms can threaten local ecosystems [4]. Extensive application of herbicides plays a crucial role in controlling weeds and boosting agricultural efficiency, yet, this extensive application can induce resistance in weeds and significantly affect soil, biota, and human health [5]. Traditional mechanical weeding is effective pre-emergence, but post-emergence inaccuracy can irreversibly damage crops. Targeted weeding technology, which focuses on the specific locations of weeds, has gradually become a focal point in the era of precision agriculture and Agriculture 4.0 [6].

Current targeted weeding technologies can be categorized into targeted herbicide spraying and laser-targeted burning. With the rapid advancement of computer vision, deep learning convolutional neural networks have been widely implemented in targeted

herbicide spraying [7–10]. To enhance the accuracy of targeted weeding, researchers have optimized DeepLabv3, YOLOv7, the Deep Residual Convolutional Neural Network (DRCN), and the YOLOv4 backbone network, achieving detection accuracies of 91.53%, 94.96%, 97.3%, and 98.52%, respectively [11–14]. Through precise targeted spraying, these enhancements have effectively reduced the amount of herbicides used. However, the use of herbicides cannot be completely eliminated, which contradicts the principles of sustainable green agriculture [15]. Laser-targeted weeding, as a branch of targeted weeding, has become feasible with the maturation of compact laser devices and the enhanced transferability of computer vision models. This technology allows for the precise control of lasers via high-speed visual computing on devices with limited computational power [10,16–18]. Therefore, laser-targeted weeding methods are poised to become a more efficient and environmentally beneficial solution for weed management [19]. In earlier research, a color differentiation algorithm was used to extract plants from the soil background, followed by size differentiation to distinguish between crops and weeds [20]. This method proved to be inefficient and less effective in complex agricultural fields. Zhu et al. [21,22] designed a neural-network-based blue laser weeding robot, incorporating a lightweight attention module into the YOLOx-Darknet architecture. Experimental tests showed a weed detection accuracy of 88.94% with a seedling damage rate of 4.53%. However, there is still significant room for improvement in the model's precision. Fatima et al. [23] compared the YOLOv5 and SSD-ResNet50 networks using datasets of three crops and four weeds, with localization via bounding boxes. Tests demonstrated that the YOLOv5-trained model achieved 88% detection accuracy and 27 FPS when transferred to the Nvidia Xavier AGX standalone device. In laser weeding operations, enhancing the model's accuracy is a crucial strategy to minimize seedling damage rates and improve production efficiency.

The application of laser weeding has been predominantly restricted to bounding box localization (object detection) of weeds. Previous researchers have endeavored to enhance object detection algorithms such as ResNet and YOLO series. Their goal was to calculate the central position of weeds based on the rectangular coordinates obtained from object detection, thereby guiding the laser to accurately target the weeds, meeting practical application requirements. However, object detection methods have consistently fallen short in precisely delineating the edges of weed stems and leaves. Rakhmatulin et al.'s [9] experimental results from laser weeding indicate that the optimal time for weed removal is when the weed stem diameter is less than 2 mm during the 3–4 leaf stage, necessitating precise targeting of the weed stem and leaves with lasers for effective removal. This introduces new requirements for control processing. Laser weeding equipment requires an understanding of the complete plant shape to accurately locate the boundaries of weed stems and leaves, addressing the issue of bounding box localization in target detection algorithms, which fails to control the laser for precise cutting. Thus, this paper first applies image segmentation technology to the field of laser weeding and proposes an improved YOLOv8 instance segmentation algorithm based on bidirectional feature fusion and deformable convolution (BFFDC-YOLOv8-seg). This improved instance segmentation algorithm supports precise localization of weeds during the laser weeding process, effectively reducing crop damage rates. In order to conduct this study, the research consists of the following tasks:

1. The dataset images are processed by adding Gaussian noise and adjusting the color space to enhance the model's generalization and robustness towards weed edge features in agricultural environments.
2. By obtaining model scales compatible with laser weeding equipment and developing pretrained weights more suited for agricultural settings, the accuracy and speed of model training are enhanced.
3. It introduces the Bidirectional Feature Pyramid Network (BiFPN) [24], an efficient weighted bidirectional framework for cross-scale connections and fast normalized feature fusion method. This approach significantly enhances the network's ability to focus on small targets, effectively addressing the challenge of detecting inconspicuous features in complex backgrounds.

4. DSCConv is integrated to enhance the network's capability to segment irregular edges of plant stems and leaves, enabling accurate weed segmentation.

## 2. Materials and Methods

### 2.1. Image Collection and Dataset Construction

#### 2.1.1. Data Collection and Annotation

The dataset in this paper consists of images captured on site in Guiyang, Guizhou Province, China. It includes eleven types of weeds: *Amaranthus blitum*, *Cirsium arvense*, *Senna tora*, *Portulaca oleracea*, *Digitaria sanguinalis*, *Ipomoea nil*, *Euphorbia heyneana*, *Cyperus rotundus*, *Mollugo stricta*, *Platostoma palustre*, and *Eleusine indica*, as shown in Figure 1. Each weed type has 120 images, all in JPEG format with a resolution of  $640 \times 640$  pixels. Images were manually annotated using LabelMe to acquire data labels in JSON format, which were then converted into TXT files containing multiple coordinate points.



**Figure 1.** Images of eleven types of weeds.

#### 2.1.2. Dataset Augmentation and Construction

In the original dataset, images of each weed type were divided into training, validation, and test sets in a 7:2:1 ratio, with 84 images for training, 24 for validation, and 12 for testing. Deep learning models depend heavily on the image features within a dataset. Rich image data improve model accuracy and help prevent overfitting. However, noise in the images and errors in annotations can decrease the accuracy of the model. To compensate for insufficient data and prevent network overfitting, this paper enhances the dataset by applying  $90^\circ$  clockwise and counterclockwise rotations, adding 0.1% Gaussian noise, and varying  $\pm 25\%$  saturation and  $\pm 15\%$  brightness. These methods expanded the number of training images from 924 to 2772, as shown in Table 1.

**Table 1.** Number of images before and after data augmentation.

	Training Images	Validation Images	Test Images	Total Images
Before Augmentation	924	264	132	1320
After Augmentation	2772	264	132	3168







### 2.2.2. Structure of the BFFDC-YOLOv8-Seg Network

To enhance the accuracy of weed segmentation in complex, cluttered agricultural fields with overlapping plants, while ensuring model simplicity and real-time segmentation, this paper introduces the new BFFDC-YOLOv8-seg instance segmentation network. As shown in Figure 3, the network reconstructs the Concat module in the original Neck structure, replacing the existing FPN and PAN with BiFPN for innovative multiscale feature fusion, effectively enhancing the network’s ability to detect small targets. The introduction of DSConv to replace some of the  $3 \times 3$  convolutions in the original Backbone, integrating features extracted by  $3 \times 3$  convolutions with DSConv, increases the flexibility of the convolution kernels, thus improving the accuracy of segmentation for irregular edges of plant stems and leaves. The BFFDC-YOLOv8-seg network achieves precise segmentation of highly similar, cluttered, and overlapping weeds in complex field backgrounds, with features that ensure simplicity and real-time performance, making it feasible for operation on standalone devices.

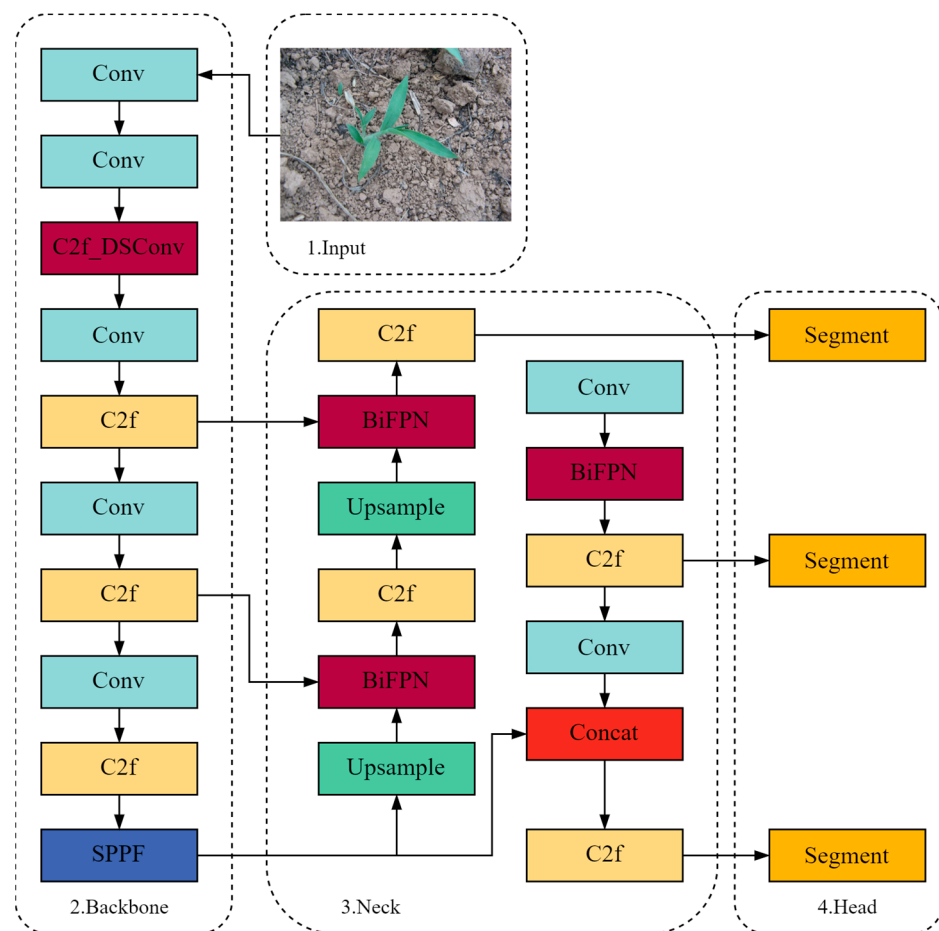


Figure 3. The BFFDC-YOLOv8-seg network structure diagram.

- **Appropriate weight documents and scales**

Ultralytics officially provides five different scales of networks (N/S/M/L/X) and corresponding initial weight files on Github to cater to various application scenarios. These training weight files, essential for assisting training, enhance the accuracy and speed of training, containing model parameters such as weights and biases for each layer. However, the official weight files, trained on the COCO2017 dataset, lack the capability to perceive vegetable field environments. To better acclimate the model to vegetable fields and achieve improved training outcomes, this paper utilizes a public plant instance segmentation dataset with over 5000 images. By iterating 200 times on the original YOLOv8(N/S/M/L/X)-seg

networks, this study obtains training weights adapted to vegetable field environments, serving as optimized training weights for the model.

Table 2 clearly shows that under the same training batches, the S model has significantly lower inference speed compared to the N model, with only a 1.3% improvement in accuracy and a substantial increase in model size. The M/L/X models, compared to the N model, show a notable increase in size and a significant decrease in inference speed, with a maximum of only 1.9% improvement in accuracy. With no significant gains in detection accuracy, the larger models require substantial storage resources and higher processing power, making them unsuitable for resource-limited laser weeding devices. Therefore, this paper chooses to optimize the N-scale model.

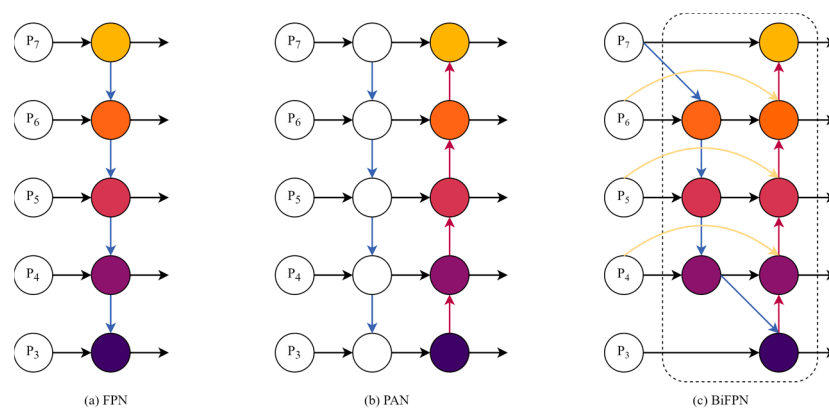
**Table 2.** Results of training multiscale models.

Scale	Depth	Width	mAP50	FPS	Size (MB)
N	0.33	0.25	0.859	277.7	6.8
S	0.33	0.50	0.872	147.0	23.9
M	0.67	0.75	0.875	33.2	54.9
L	1.00	1.00	0.876	5.4	92.3
X	1.00	1.25	0.878	1.2	548

• **Multiscale feature fusion**

The Concat module in the Neck section, which includes both FPN and PAN, plays a critical role in the fusion of image information. As depicted in Figure 4a and described by Equation (1), traditional FPN networks use input features  $P_i^{in} = (P_3^{in} \dots P_7^{in})$  with 3–7 layers, where  $P_i^{in}$  denotes the resolution level  $in/2^i$  of the input image. Features are aggregated from top to bottom, and *resize* usually involves upsampling or downsampling for resolution matching, while *Conv* is typically used for feature processing. This results in feature fusion being limited by a unidirectional flow of information, which is ineffective in extracting features of small weed targets in agricultural settings with high similarity and indistinct color features. To enhance the detection capabilities of small targets in agricultural environment, this paper introduces a Bidirectional Feature Pyramid Network (BiFPN).

$$\begin{aligned}
 P_7^{out} &= Conv(P_7^{in}) \\
 P_6^{out} &= Conv(P_6^{in} + resize(P_7^{out})) \\
 &\dots \\
 P_3^{out} &= Conv(P_3^{in} + resize(P_4^{out}))
 \end{aligned}
 \tag{1}$$



**Figure 4.** Topology diagram of the FPN [26], PAN [27], and BiFPN [24] networks.

BiFPN represents an efficient bidirectional framework for cross-scale connections and fast normalization of feature fusion. From the network topology (Figure 4b,c), it can be seen that BiFPN modifies the multiscale connections within the PAN architecture: Initially,

it removes network nodes that have a single input feature edge without fusion, creating a simplified bidirectional network. Subsequently, when the original input and output nodes are at the same level, an additional pathway is added between them to enable more feature fusion without significantly increasing computational costs; Lastly, unlike PAN, which features only one top-down and one bottom-up pathway, each bidirectional (top-down and bottom-up) pathway is implemented as a feature network layer, repeated multiple times to achieve advanced feature fusion.

Different input features have varying resolutions, and compared to the high resolution of crops, the smaller resolution of weed inputs leads to a significant imbalance in the network's output contributions. Traditional methods treat all input features equally without distinction, which is not ideal in practical applications. Tan et al. [24] assigned varying weights to input features, significantly enhancing the network's performance in detecting small objects. Therefore, we propose adding an additional weight to each input feature, enabling the network to learn the significance of each feature and preventing it from overlooking the small-scale features of weeds. Based on this concept, BiFPN uses fast normalized fusion, as described in Equation (2), an efficient and stable weighted fusion mechanism that applies an *ReLU* activation function after each  $w_i$  to ensure that  $w_i \geq 0$  and  $\varepsilon = 0.001$  are small, keeping each normalized weight between 0 and 1. Fast normalized fusion, similar in learning behavior and accuracy to Equation (3) (Softmax-based fusion), omits the *Softmax* operation, allowing BiFPN to run 30% faster on GPUs. To further enhance efficiency, we use depthwise separable convolutions for feature fusion and add batch normalization and activation after each convolution.

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i \quad (2)$$

$$O = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} \cdot I_i \quad (3)$$

The original Concat module is restructured, using BiFPN to replace the traditional FPN and PAN for a novel feature fusion approach, thereby assigning higher weights to small object features, enhancing the focus on small targets.

- **Deformable convolution**

In laser weeding operations, targeting the critical tissue parts of weeds with laser beams is essential for effective weed eradication; imprecise targeting can increase the accidental injury rate to crop seedlings. The original YOLOv8-seg network relies on the detection accuracy of bounding boxes within the Backbone, but square bounding boxes are not sensitive to the local information of irregular targets. Therefore, to enhance the network's perception of the irregular edges of weeds, deformable convolutions (DCNs) [29] are considered for integration into the Backbone, allowing some  $3 \times 3$  convolutional kernels to adjust their shapes to fit the irregular structures of weeds, while maintaining the stability of the convolutional structure and reducing deviation. Given that Dynamic Snake Convolution (DSConv) [30] performs well in segmenting tubular structures, adapting to slender and twisted local structural features to enhance geometric structure perception, this paper introduces DSConv, constructing convolutional kernels with strong perception of irregular curves.

This section elucidates the application of DSConv in extracting irregular local features of weed stems and leaf edges, assuming standard 2D convolutional coordinates  $K$ , with the center coordinate as  $K_i = (x_i, y_i)$ . The original  $3 \times 3$  convolutional kernel is  $K$ , represented by

$$K = \{(x-1, y-1), (x-1, y), \dots, (x+1, y+1)\} \quad (4)$$

By introducing deformation offsets  $\Delta$ , the convolutional kernel becomes more flexible, focusing on the irregular edges of tubular weed stems and leaves. Figure 5 linearizes the standard kernel in both axial directions, expanding it into a kernel of size 9. Taking the



X axial direction as an example, each grid position in  $K$  is denoted as  $K_{i\pm c} = (x_{i\pm c}, y_{i\pm c})$ , where  $c = 0, 1, 2, 3, 4$  represents the horizontal distance from the center grid. The selection of each grid position  $K_{i\pm c}$  in kernel  $K$  is a cumulative process. Starting from the central position  $K_i$ , the position away from the center grid depends on the position of the previous grid:  $K_{i\pm 1}$  increases by an offset  $\Delta = \{\delta | \delta \in [-1, 1]\}$  relative to  $K_i$  [30]. Therefore, the offsets need to be accumulated to ensure that the kernel conforms to a linear structural form.

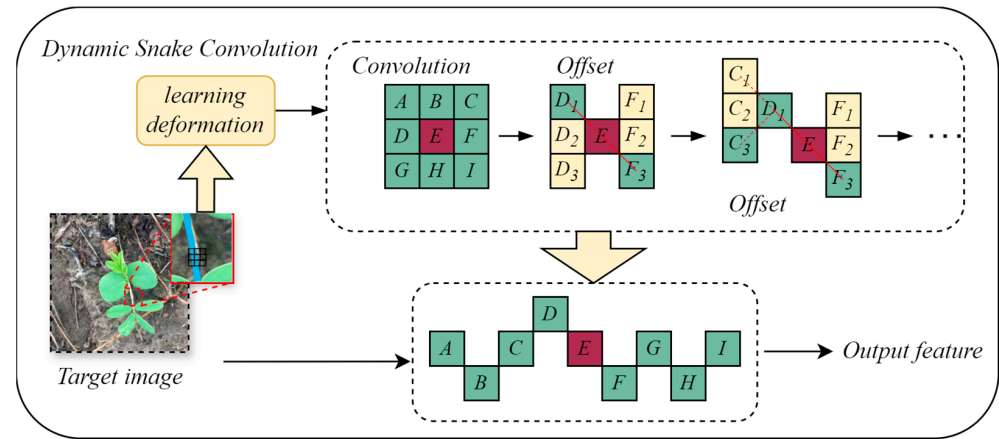


Figure 5. An example of DSConv convolution kernel changes [30].

The change in the X-axis direction is

$$K_{i\pm c} = \begin{cases} (x_{i\pm c}, y_{i\pm c}) = (x_i + c, y_i + \sum_{i=1}^{i+c} \Delta y) \\ (x_{i\pm c}, y_{i\pm c}) = (x_i - c, y_i + \sum_{i=1}^{i-c} \Delta y) \end{cases} \quad (5)$$

The change in the Y-axis direction is

$$K_{j\pm c} = \begin{cases} (x_{j\pm c}, y_{j\pm c}) = (x_j + \sum_{j=1}^{j+c} \Delta x, y_j + c) \\ (x_{j\pm c}, y_{j\pm c}) = (x_j + \sum_{j=1}^{j-c} \Delta x, y_j - c) \end{cases} \quad (6)$$

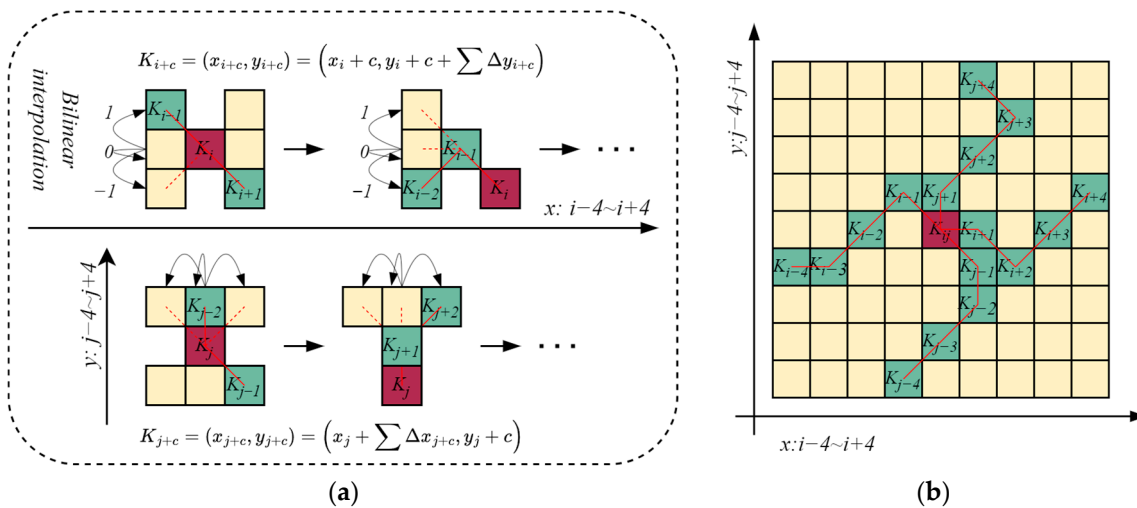
Since the offset  $\Delta$  is typically a decimal, while coordinates are usually in integer form, bilinear interpolation is employed, expressed as

$$K = \sum_{k'} B(K', K) \cdot K' \quad (7)$$

Here,  $K$  represents the decimal positions in Equations (5) and (6).  $K'$  enumerates all integer spatial positions.  $B$  is a bilinear interpolation kernel, which can be decomposed into two one-dimensional kernels:

$$B(K, K') = b(K_x, K'_x) \cdot b(K_y, K'_y) \quad (8)$$

As shown in Figure 6, the changes in the two-dimensional (X and Y axes) setup enable the dynamic serpentine convolutional kernels described in this paper to cover a  $9 \times 9$  perceptual field during their deformation, better adapting to elongated tubular structures and enhancing the perception of critical features. The use of  $3 \times 3$  convolutional kernels to perform the functions of  $9 \times 9$  kernels allow for greater flexibility in the model's kernels while keeping scale increases minimal.



**Figure 6.** (a) The method for calculating the coordinates of DSCConv. (b) The receptive field of DSCConv [30].

### 2.3. Model Training and Outputs

This study’s model training was conducted on a Windows 10 operating system. The computer was equipped with a 13th Gen Intel Core i5-13600KF CPU @ 5.1 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX 4060 8 GB GPU. Under the PyTorch deep learning framework, a multiclass segmentation task neural network training model was constructed, with the main software versions as shown in Table 3. The training configuration included an input image size of  $640 \times 640$ , batch size of 16, momentum of 0.937, initial learning rate of 0.0001, and 50 iterations.

**Table 3.** Training and testing environment.

Configuration	Allocation
CUDA version	11.3
Python version	3.8
PyTorch version	1.12

The trained model was output in onnx format and tested on an embedded edge computing device (NVIDIA Jetson Orin nano 4 GB) running Ubuntu 20.04 OS, featuring an arm64 CPU (512 CUDA cores) with TensorRT acceleration.

### 2.4. Model Evaluation Criteria

This paper evaluates the model’s detection accuracy and image segmentation capabilities through metrics such as precision, recall, and mean average precision (mAP). Frames per second (FPS) measures the inference speed on hardware, i.e., the number of images processed per second by the device.

Precision (P) and Recall (R) are calculated using a confusion matrix, which includes True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Precision is the proportion of actual targets in the total predicted targets by the network, representing the classification accuracy of the network, whereas recall is the ratio of true targets correctly predicted by the network to the actual number of true targets. The corresponding formulas are as follows:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{9}$$

$$Recall = \frac{TruePositive}{True\ Positive + False\ Negative} \tag{10}$$

Intersection over union (*IoU*) represents the ratio of intersection to union in object detection, used to measure the degree of overlap between the model-generated boundaries and the original annotated boundaries. When *IoU* exceeds 0.5, it is considered that the object is detectable. If the true region is *A* and the annotated region is *B*, then

$$IoU = \frac{A \cap B}{A \cup B} \quad (11)$$

The average precision (*AP*) for a single category is determined by ranking the model's predictions based on their confidence scores and calculating the area under the precision–recall (*PR*) curve, as follows:

$$AP = \int_0^1 P(R)d(R) \quad (12)$$

Mean average precision (*mAP*) indicates the average precision across multiple categories, with *mAP*<sub>50</sub> representing the *mAP* at a 50% *IoU* threshold. *mAP*<sub>50–95</sub> is a stricter evaluation metric, calculating the average of the *AP* at each *IoU* threshold from 50% to 95%, with increments of 0.05, allowing for a more accurate assessment of model performance at various *IoU* thresholds.

$$mAP_{50-95} = \frac{AP_{IoU=0.5+} + AP_{IoU=0.55} + AP_{IoU=0.6} + \dots + AP_{IoU=0.95}}{n} \quad (13)$$

### 3. Results

#### 3.1. Ablation Experiments and Model Training Details

##### 3.1.1. Ablation Experiments

This paper constructs ablation comparative experiments to verify the efficacy of optimization techniques, ensuring there is no competition or conflict among various methods, and that resource utilization is optimal to avoid inconsistencies and confusion in the model training and decision-making process. Two distinct optimization approaches were individually integrated into the original YOLOv8-seg network for comparative analysis of YOLOv8-seg, BiFPN YOLOv8-seg, DSCConv YOLOv8-seg, and BiDS-YOLOv8-seg networks using the publicly available coco128 dataset, training 50 batches, as shown in Table 4.

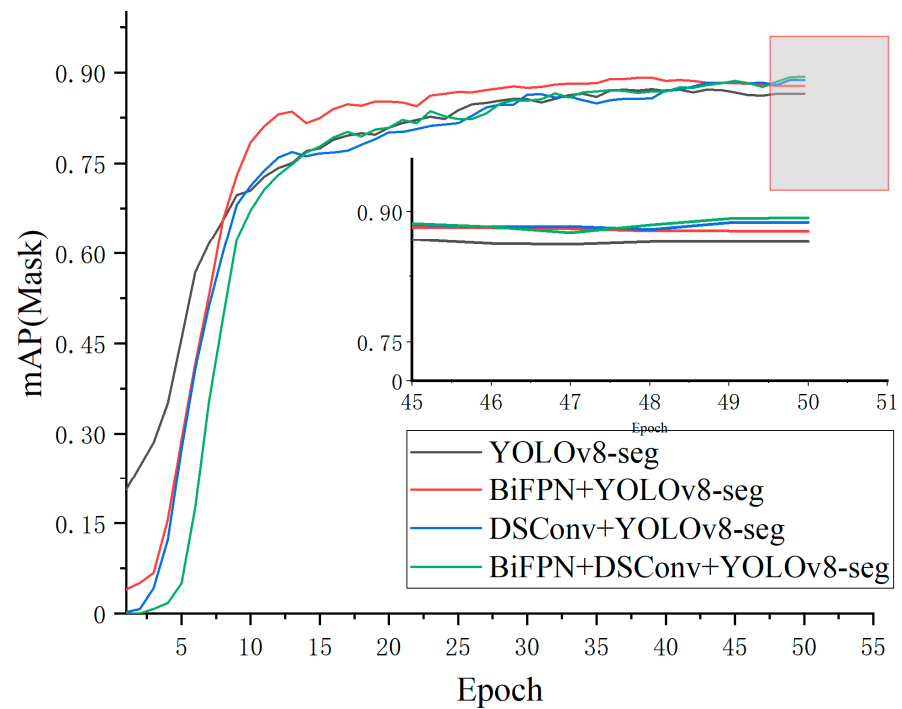
**Table 4.** Performance of YOLOv8-seg combined with various optimization techniques.

Network	Precision	Recall	mAP50	mAP50-95
YOLOv8-seg	0.904	0.811	0.875	0.637
BiFPN + YOLOv8-seg	0.914	0.836	0.889	0.641
DSCConv + YOLOv8-seg	0.912	0.811	0.887	0.636
BiFPN + DSCConv + TOLOv8-seg	0.917	0.835	0.893	0.640

The introduction of BiFPN enhanced the model's precision, recall, mAP50, and mAP50:95 to 91.4%, 83.6%, 88.9%, and 64.1%, respectively, showing improvements over the original model by 1%, 2.5%, 1.8%, and 0.4%. This indicates that BiFPN effectively enhances the model's ability to fuse features across different scales. The introduction of DSCConv resulted in an 8% increase in precision and a 1.6% increase in mAP50 compared to the original model, indicating that deformable convolutions perform well in segmenting object edges, beneficial for the segmentation of weed stems and irregular leaf surfaces. However, recall and mAP50:95 are similar to the original network, suggesting that DSCConv's feature recognition across multiple categories is somewhat limited, leading to poor balance, and, thus, should not be used in isolation.

The mAP50 curves for different models across training batches are shown in Figure 7. The BiDS-YOLOv8-seg model, formed by the fusion of BiFPN, DSCConv, and YOLOv8-seg, exhibits the best detection performance with significant improvements in all metrics, achieving a precision of 91.7%, recall rate of 83.5%, and mAP50 of 89.3%.



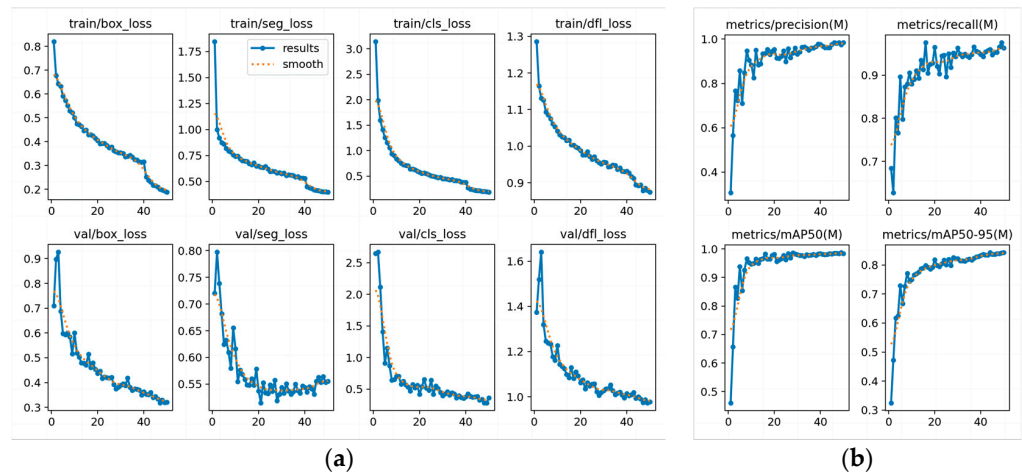


**Figure 7.** The mAP50 curves of various optimized networks throughout the iteration process.

### 3.1.2. Training Results for the BFFDC-YOLOv8-Seg

In the present study, we trained the BFFDC-YOLOv8-seg on a vegetable field weed dataset, with Figure 8 detailing various performance metrics during the training and validation processes. During model training, we observed that the Box Loss (box\_loss) rapidly decreased from a high initial value to below 0.2 and converged, demonstrating high precision in detecting weed boundaries without overfitting. The Segmentation Loss (seg\_loss) also showed a significant decreasing and converging trend, validating the model's effectiveness in segmentation tasks. The decrease in Classification Loss (cls\_loss) indicated a gradual enhancement in the model's reliability for object classification. The reduction in the specific loss metric, Direction/Flow Loss (dfl\_loss), reflected the model's learning capability in predicting object direction or flow. During model validation, all loss metrics showed a converging trend similar to training, with final stable values slightly higher than training results, indicating that the model also has good generalization capabilities on data outside the training set. Moreover, the Mask's precision (precision(M)) and recall (recall(M)) remained above 0.9, with the mean average precision at 50% IoU (mAP50) exceeding 0.9, and at a stricter 50–95% IoU (mAP50-95) also demonstrating performance above 0.8, further confirming the model's strong segmentation performance and robustness under rigorous evaluation standards.

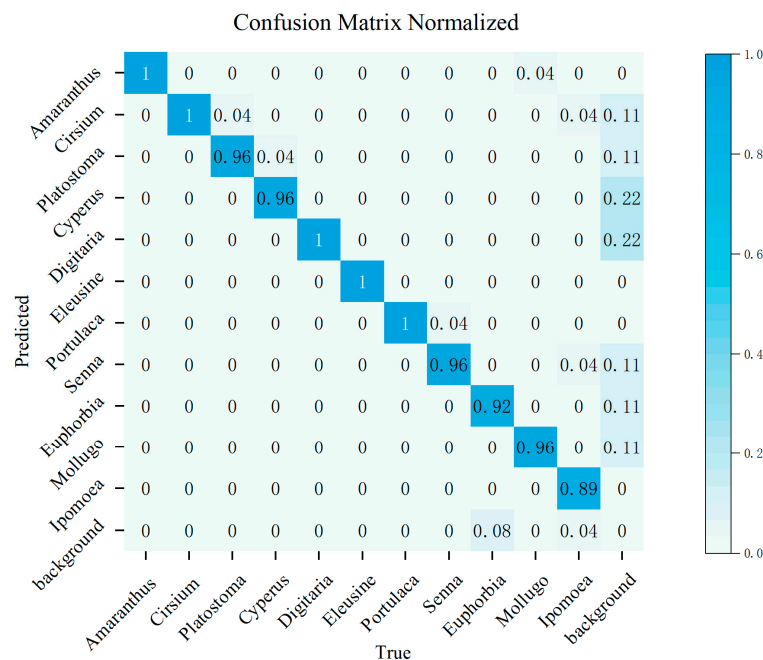
In summary, the final loss values are low and stable, a good balance is achieved between precision and recall, and the model performs excellently in segmenting small weed targets at different IoU thresholds, effectively learning weed characteristics without overfitting, achieving excellent training results.



**Figure 8.** Performance evaluation of the BFFDC-YOLOv8-seg model during training and validation. (a) The upper part of the figure displays four loss metrics on the training set, including Box Loss (box\_loss), Segmentation Loss (seg\_loss), Classification Loss (cls\_loss), and Direction/Flow Loss (dfl\_loss), shown as actual values (blue dots) and smooth curves (orange dashed lines). The lower part corresponds to the variation of the same loss metrics on the validation set. (b) The figure sequentially displays the curves for precision, recall, mAP50, and mAP50-95 of the segmentation Mask.

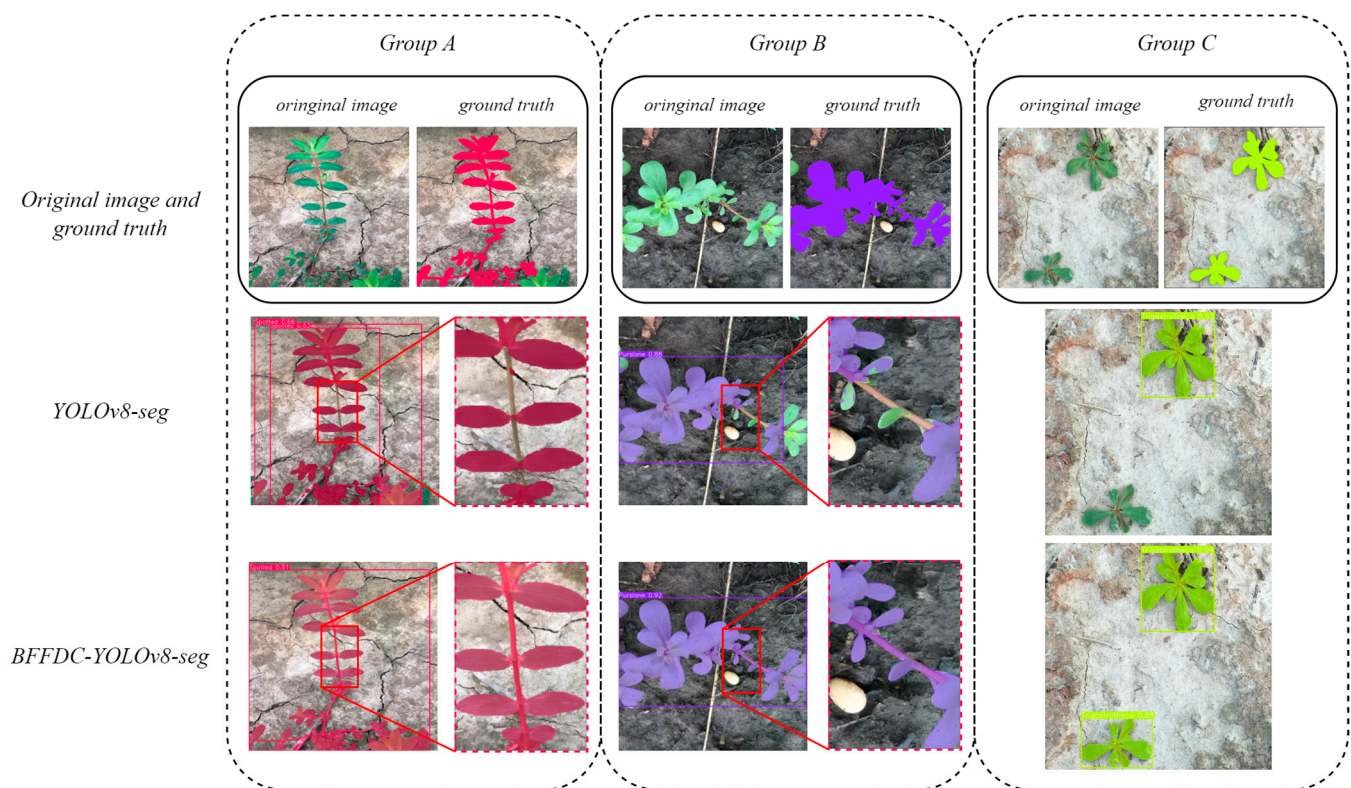
### 3.1.3. BFFDC-YOLOv8-Seg Detection and Segmentation Effect

This study tested the BFFDC-YOLOv8-seg model using reserved test set images to verify its actual detection and segmentation effects. A normalized confusion matrix for each weed category was created, as shown in Figure 9, and results were assessed based on the classification accuracy of the weeds. The matrix results indicate that the model performs uniformly in detection, achieving 100% accuracy for *Amaranthus*, *Cirsium*, *Digitaria*, *Eleusine*, and *Portulaca*, 96% for *Platostoma*, *Cyperus*, *Senna*, and *Mollugo*, 92% for *Euphorbia*, and 89% for *Ipomoea*. During testing, the model exhibited numerous errors in detecting *Euphorbia* and *Ipomoea*, missing some *Euphorbia* instances and misidentifying some *Ipomoea* as *Cirsium* or *Senna*. This issue could be mitigated by collecting more data to enhance the model’s ability to distinguish similar physical features such as the shape, texture, and color of leaves.



**Figure 9.** Normalized confusion matrix for detection accuracy of various weeds.

By comparing masks, the study found that the original model (YOLOv8-seg) failed to adequately segment weed stems and edges and missed small targets, as shown in Figure 10. The analysis suggests that the original network's insufficient convolution flexibility and limited ability to perceive irregular edges resulted in the omission of stems and edges; inadequate feature fusion also led to the nondetection of small targets. The weighted bidirectional multiscale feature fusion in BFFDC-YOLOv8-seg model enhances the network's ability to perceive small targets; DSCConv, with its more flexible convolution kernels, has a stronger ability to detect stems and irregular edges, improving edge segmentation accuracy. Experiments demonstrate that in complex agricultural environments, our model outperforms the original in extracting small target features and segmenting irregular edges. Although our model demonstrates superior segmentation capabilities for irregular small targets, some gaps caused by overlapping leaves are still present within the mask. These gaps can lead to ineffective laser burning, thereby reducing overall work efficiency. Consequently, there is still room for improvement in edge segmentation.



**Figure 10.** Groups A, B, and C demonstrate the model's ability to segment slender stems of plants, edges of weed leaves, and detect small objects, respectively.

### 3.2. Comparison of the Performance with the Other Segmentation Models

To validate the efficiency and advantages of the BFFDC-YOLOv8-seg weed segmentation model in this study, we compared it with mainstream segmentation networks such as Mask-RCNN [31], YOLOv5-seg [32], and YOLOv7-seg [33]. The same dataset and training parameters were used for training the models, and tests were conducted on the test set. The complete experiment assessed the accuracy (P), recall (R), mean average precision (mAP), inference speed (FPS), and model size for each segmentation network, as shown in Table 5.



**Table 5.** Comparison results of different segmentation model performances.

Model	Precision	Recall	mAP50	mAP50-95	FPS	Size (MB)
Mask RCNN	0.895	0.876	0.88	0.682	34	228
YOLOv5-seg	0.701	0.781	0.854	0.593	227	4.2
YOLOv7-seg	0.917	0.95	0.975	0.749	18.3	76.4
YOLOv8-seg	0.926	0.894	0.96	0.776	270	6.8
Ours	0.975	0.975	0.988	0.842	101	6.8

From a comprehensive comparison of tests, it is evident that the BFFDC-YOLOv8-seg model proposed in this paper exceeds the original YOLOv8-seg model by 2.8% in mAP50 for segmenting small weed targets, and surpasses the Mask-RCNN, YOLOv5-seg, and YOLOv7-seg models by 10.8%, 13.4%, and 1.3% respectively, showing the highest precision in detecting and segmenting weeds in complex agricultural environments. In terms of real-time detection capabilities, Mask-RCNN, with its complex model and high-density computations, and YOLOv7-seg's large-scale network require high computational costs, resulting in inference speeds of only 34 and 18.3 FPS, respectively. After increasing the sampling scale, this paper's network has a reduced inference speed compared to YOLOv8-seg's 270 FPS, but at 101 FPS, it fully meets the real-time detection needs of slow-moving laser weeding device, while maintaining optimal segmentation accuracy. Regarding model size, the network retains its lightweight characteristics, making the 6.8 MB model more suitable for deployment on resource-limited automatic laser weeders compared to the Mask-RCNN and YOLOv7-seg models.

### 3.3. Testing on Standalone Devices

To ensure that the BFFDC-YOLOv8-seg model adapts well to standalone devices, this paper ports the trained model to the standalone embedded device Jetson Orin nano (4 GB) for compatibility testing. The Nvidia Jetson, as an embedded AI computing platform, significantly reduces the computational cost of deep learning models, facilitating broader application of compact laser weeding devices. This paper also uses precision, recall, mAP50, mAP59-95, and FPS metrics to evaluate the model's performance on standalone devices. As shown in Table 6, the model achieves a 95.8% mAP50 on the Jetson Orin nano and processes images at 24.8 FPS, suitable for real-time weed segmentation.

**Table 6.** Testing results on Jetson Orin nano.

Box				Mask				FPS
Precision	Recall	mAP50	mAP50-95	Precision	Recall	mAP50	mAP50-95	
0.974	0.924	0.958	0.916	0.974	0.924	0.958	0.817	24.8

## 4. Discussion

In research on automated laser weeding, many scholars focus on object detection algorithms, neglecting the coarseness of bounding box localization, which can lead to accidental damage to seedlings and incomplete eradication of weeds. Conversely, with significant improvements in the efficiency and portability of segmentation algorithms, instance segmentation techniques are more advantageous for locating weeds in complex environments. Yue et al. [34] reported an mAP50 of 92.2% using an improved YOLOv8-seg model for segmenting tomato diseases at different stages, demonstrating significant potential of the YOLOv8 series in segmentation tasks.

Feature extraction of small and irregular objects has always been a hot topic in computer vision, with multiscale feature fusion and dynamic deformable convolutions being the most common in addressing segmentation of such objects. In agricultural production, there are many applications for the segmentation of small targets, such as determining the extent of infection in plants by segmenting and locating plant pests and diseases. There-

fore, precise segmentation of irregular small targets in complex backgrounds remains a significant challenge in agricultural applications.

The present study demonstrates that the BFFDC-YOLOv8-seg effectively enhances weed segmentation capabilities in complex agricultural environments, making it suitable for small, cost-effective, automated laser weeding devices and capable of efficiently detecting and segmenting weeds in vegetable fields. Although this research has made some progress, many key technical issues still need to be addressed to make laser-targeted weeding equipment more aligned with practical needs: a more comprehensive dataset of field weeds needs to be established to increase the diversity of feature collection and further enhance the model's accuracy and generalizability; the model's detection precision needs further optimization and its complexity reduced to meet the real-time demands and resource constraints of practical applications; image coordinates obtained from the model's segmentation need to be integrated with a positioning algorithm for precise target localization; a mismatch between the speed of the laser source and detection efficiency leads to reduced efficiency, necessitating a more efficient laser source. Although this research has achieved certain results, further optimization of the model's detection precision and reduction in complexity remains a focus. We hope that our efforts will lead to widespread adoption of laser weeding devices equipped with the BFFDC-YOLOv8-seg model in agricultural production, promoting sustainable green agriculture.

## 5. Conclusions

Addressing the challenges of inaccurate weed targeting and high crop damage rates associated with traditional laser weeding equipment, the present study is the first to apply image segmentation technology to the field of laser weeding, using instance segmentation to precisely guide the laser targeting. We propose a BFFDC-YOLOv8-seg weed segmentation model that integrates dynamic snake convolution and a weighted bidirectional feature pyramid network to precisely locate the boundaries of weed stems and leaves, addressing the issues of inaccurate weed targeting and high crop damage rates in traditional laser weeding equipment. In the original YOLOv8 network, weighted multiscale feature fusion is implemented to enhance the network's ability to perceive small objects; DSCConv replaces some  $3 \times 3$  convolutional kernels, making the existing convolutional layers more flexible and better adapted to the irregular edges of weed stems and leaves. Experimental results demonstrate that the BFFDC-YOLOv8-seg model achieves an accuracy of 97.5%, a recall of 97.5%, an mAP50 of 98.8%, and an mAP50-95 of 84.2% on the test set. Compared to the current mainstream models such as Mask-RCNN, YOLOv5-seg, YOLOv7-seg, and YOLOv8-seg, this model achieves better mAP50 values, with improvements of 10.8%, 13.4%, 1.3%, and 2.8%, respectively. Additionally, the model size is only 6.8 MB, and in stability tests on the Jetson Orin nano independent device, the actual scenes captured by the camera can be processed at a detection speed of 24.8 FPS with an mAP50 of 95.8%. The proposed model delivers precise real-time weed segmentation, making it suitable for small, cost-effective automatic laser weeding devices.

**Author Contributions:** Conceptualization, Z.L. and A.L.; methodology, Z.L. and A.L.; software, Z.L.; validation, Z.L. and Y.M.; formal analysis, Z.L. and A.L.; investigation, Z.L.; resources, Y.M.; data curation, Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, Z.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Undergraduate Training Programs for Innovation and Entrepreneurship of Guizhou University (gzusc (2023)015).

**Data Availability Statement:** The weed dataset from agricultural fields used in this study contains sensitive regional information. In accordance with relevant laws and regulations, this dataset is not publicly accessible to prevent potential environmental risks and misuse. Therefore, the dataset will not be stored in any public databases nor shared via other public means. Researchers interested in accessing these data for verifying the validity of the research findings may contact the corresponding author of this paper. Access may be considered within a properly regulated and confidential frame-

work, provided that sufficient justification and compliance with confidentiality requirements are met. The publicly available plant instance segmentation dataset mentioned in Section 2.2.2 and the COCO128 dataset discussed in Section 3.1.1 of this paper can be accessed and downloaded via the Kaggle platform through the following link: <https://www.kaggle.com/datasets/leozhuxi/weed-seg>.

**Acknowledgments:** The author would like to extend sincere gratitude to the College of Big Data and Information Engineering for providing the necessary laboratory facilities and equipment that were crucial for conducting this research. Special thanks are also owed to my supervisor, whose guidance was invaluable throughout the research process. Additionally, I wish to express my deep appreciation to my family, whose unwavering support and encouragement have been a constant source of strength. Their belief in my abilities has been a great motivation.

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

## References

1. Wozniak, A. Mechanical and chemical weeding effects on the weed structure in durum wheat. *Ital. J. Agron.* **2020**, *15*, 102–108. [[CrossRef](#)]
2. Panta, S.; Schwarzländer, M.; Weyl, P.S.R.; Hinz, H.L.; Winston, R.L.; Eigenbrode, S.D.; Harmon, B.L.; Bacher, S.; Paynter, Q. Traits of insect herbivores and target weeds associated with greater biological weed control establishment and impact. *BioControl* **2024**, *69*, 221–236. [[CrossRef](#)]
3. Gao, W.-T.; Su, W.-H. Weed Management Methods for Herbaceous Field Crops: A Review. *Agronomy* **2024**, *14*, 486. [[CrossRef](#)]
4. Gaskin, J. Recent contributions of molecular population genetic and phylogenetic studies to classic biological control of weeds. *BioControl* **2023**, *69*, 353–360. [[CrossRef](#)]
5. Gamble, A.V.; Price, A.J. The intersection of integrated pest management and soil quality in the resistant weed era. *Ital. J. Agron.* **2021**, *16*, 1875. [[CrossRef](#)]
6. Raj, M.; Gupta, S.; Chamola, V.; Elhence, A.; Garg, T.; Atiquzzaman, M.; Niyato, D. A survey on the role of Internet of Things for adopting and promoting Agriculture 4.0. *J. Netw. Comput. Appl.* **2021**, *187*, 103107. [[CrossRef](#)]
7. Kaya, A.; Keceli, A.S.; Catal, C.; Yalic, H.Y.; Temucin, H.; Tekinerdogan, B. Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* **2019**, *158*, 20–29. [[CrossRef](#)]
8. Nevavuori, P.; Narra, N.; Lipping, T. Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* **2019**, *163*, 104859. [[CrossRef](#)]
9. Hasan, A.S.M.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G.K. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [[CrossRef](#)]
10. Coleman, G.R.Y.; Bender, A.; Hu, K.; Sharpe, S.M.; Schumann, A.W.; Wang, Z.; Bagavathiannan, M.V.; Boyd, N.S.; Walsh, M.J. Weed detection to weed recognition: Reviewing 50 years of research to identify constraints and opportunities for large-scale cropping systems. *Weed Technol.* **2022**, *36*, 741–757. [[CrossRef](#)]
11. Yu, H.; Che, M.; Yu, H.; Zhang, J. Development of Weed Detection Method in Soybean Fields Utilizing Improved DeepLabv3+ Platform. *Agronomy* **2022**, *12*, 2889. [[CrossRef](#)]
12. Li, J.; Zhang, W.; Zhou, H.; Yu, C.; Li, Q. Weed detection in soybean fields using improved YOLOv7 and evaluating herbicide reduction efficacy. *Front. Plant Sci.* **2023**, *14*, 1284338. [[CrossRef](#)]
13. Babu, V.S.; Ram, N.V. Deep Residual CNN with Contrast Limited Adaptive Histogram Equalization for Weed Detection in Soybean Crops. *Trait. Du Signal* **2022**, *39*, 717–722. [[CrossRef](#)]
14. Zhao, J.; Tian, G.; Qiu, C.; Gu, B.; Zheng, K.; Liu, Q. Weed Detection in Potato Fields Based on Improved YOLOv4: Optimal Speed and Accuracy of Weed Detection in Potato Fields. *Electronics* **2022**, *11*, 3709. [[CrossRef](#)]
15. Liu, L.; Liu, K. Can digital technology promote sustainable agriculture? Empirical evidence from urban China. *Cogent Food Agric.* **2023**, *9*, 2282234. [[CrossRef](#)]
16. Rakhmatulin, I.; Andreasen, C. A Concept of a Compact and Inexpensive Device for Controlling Weeds with Laser Beams. *Agronomy* **2020**, *10*, 1616. [[CrossRef](#)]
17. Wang, M.; Leal-Naranjo, J.-A.; Ceccarelli, M.; Blackmore, S. A Novel Two-Degree-of-Freedom Gimbal for Dynamic Laser Weeding: Design, Analysis, and Experimentation. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 5016–5026. [[CrossRef](#)]
18. Mwitwa, C.; Rains, G.C.; Prostko, E. Evaluation of Inference Performance of Deep Learning Models for Real-Time Weed Detection in an Embedded Computer. *Sensors* **2024**, *24*, 514. [[CrossRef](#)]
19. Mwitwa, C.; Rains, G.C.; Prostko, E. Evaluation of Diode Laser Treatments to Manage Weeds in Row Crops. *Agronomy* **2022**, *12*, 2681. [[CrossRef](#)]
20. Xiong, Y.; Ge, Y.; Liang, Y.; Blackmore, S. Development of a prototype robot and fast path-planning algorithm for static laser weeding. *Comput. Electron. Agric.* **2017**, *142*, 494–503. [[CrossRef](#)]
21. Zhu, H.; Zhang, Y.; Mu, D.; Bai, L.; Zhuang, H.; Li, H. YOLOX-based blue laser weeding robot in corn field. *Front. Plant Sci.* **2022**, *13*, 1017803. [[CrossRef](#)]
22. Zhu, H.; Zhang, Y.; Mu, D.; Bai, L.; Wu, X.; Zhuang, H.; Li, H. Research on improved YOLOx weed detection based on lightweight attention module. *Crop Prot.* **2024**, *177*, 106563. [[CrossRef](#)]



23. Fatima, H.S.; ul Hassan, I.; Hasan, S.; Khurram, M.; Stricker, D.; Afzal, M.Z. Formation of a Lightweight, Deep Learning-Based Weed Detection System for a Commercial Autonomous Laser Weeding Robot. *Appl. Sci.* **2023**, *13*, 3997. [[CrossRef](#)]
24. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 19–20 June 2020; pp. 10778–10787.
25. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [[CrossRef](#)]
26. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
27. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
28. Bolya, D.; Zhou, C.; Xiao, F.; Lee, Y.J. YOLACT: Real-Time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9156–9165.
29. Meng, Y.; Men, H.; Prasanna, V. Accelerating Deformable Convolution Networks. In Proceedings of the 2022 IEEE 30th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), New York, NY, USA, 15–18 May 2022; p. 1.
30. Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; Yang, G. Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 6047–6056.
31. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
33. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
34. Yue, X.; Qi, K.; Na, X.; Zhang, Y.; Liu, Y.; Liu, C. Improved YOLOv8-Seg Network for Instance Segmentation of Healthy and Diseased Tomato Plants in the Growth Stage. *Agriculture* **2023**, *13*, 1643. [[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.