

## Article

# FCB-YOLOv8s-Seg: A Malignant Weed Instance Segmentation Model for Targeted Spraying in Soybean Fields

Zishang Yang, Lele Wang, Chenxu Li and He Li \*

College of Mechanical and Electrical Engineering, Henan Agriculture University, Zhengzhou 450002, China; zsyang@henau.edu.cn (Z.Y.); wanglele@henau.edu.cn (L.W.); lichenxu@stu.henau.edu.cn (C.L.)

\* Correspondence: chungbuk@163.com

**Abstract:** Effective management of malignant weeds is critical to soybean growth. This study focuses on addressing the critical challenges of targeted spraying operations for malignant weeds such as *Cirsium setosum*, which severely threaten soybean yield in soybean fields. Specifically, this research aims to tackle key issues in plant protection operations, including the precise identification of weeds, the lightweight deployment of segmentation models, real-time requirements for spraying operations, and the generalization ability of models in diverse field environments. To address these challenges, this study proposes an improved weed instance segmentation model based on YOLOv8s-Seg, named FCB-YOLOv8s-Seg, for targeted spraying operations in soybean fields. The FCB-YOLOv8s-Seg model incorporates a lightweight backbone network to accelerate computations and reduce model size, with optimized Squeeze-and-Excitation Networks (SENet) and Bidirectional Feature Pyramid Network (BiFPN) modules integrated into the neck network to enhance weed recognition accuracy. Data collected from real soybean field scenes were used for model training and testing. The results of ablation experiments revealed that the FCB-YOLOv8s-Seg model achieved a mean average precision of 95.18% for bounding box prediction and 96.63% for segmentation, marking an increase of 5.08% and 7.43% over the original YOLOv8s-Seg model. While maintaining a balanced model scale, the object detection and segmentation accuracy of this model surpass other existing classic models such as YOLOv5s-Seg, Mask-RCNN, and YOLACT. The detection results in different scenes show that the FCB-YOLOv8s-Seg model performs well in fine-grained feature segmentation in complex scenes. Compared with several existing classical models, the FCB-YOLOv8s-Seg model demonstrates better performance. Additionally, field tests on plots with varying weed densities and operational speeds indicated an average segmentation rate of 91.30%, which is 6.38% higher than the original model. The proposed algorithm shows higher accuracy and performance in practical field instance segmentation tasks and is expected to provide strong technical support for promoting targeted spray operations.

**Keywords:** targeted spraying; weed recognition; instance segmentation; YOLOv8-Seg; FasterNet



**Citation:** Yang, Z.; Wang, L.; Li, C.; Li, H. FCB-YOLOv8s-Seg: A Malignant Weed Instance Segmentation Model for Targeted Spraying in Soybean Fields. *Agriculture* **2024**, *14*, 2357.

<https://doi.org/10.3390/agriculture14122357>

Academic Editor: Jiaqiang Zheng

Received: 17 November 2024

Revised: 19 December 2024

Accepted: 20 December 2024

Published: 21 December 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

Soybeans are an important source of oil and grain and play an essential role in poultry feed, international trade, etc. However, various malignant weeds in soybean fields threaten the crop's yield and quality [1]. These weeds compete with soybeans for essential resources such as water, light, nutrients, and space, especially during the early and middle growth stages [2]. Consequently, effective management of malignant weeds is essential. It is necessary to ensure a high yield and high quality of crops, and it is a global concern that demands ongoing attention and action [3,4].

Common weeding methods include biological weeding, manual weeding, chemical weeding, mechanical weeding, and intelligent weeding. Biological weeding involves introducing non-native organisms to suppress weed growth, but this approach may pose risks to the local ecosystem [5]. Manual weeding has a high weed removal rate and minimal crop damage, but it is unsuitable for large-area farmland. Chemical weeding, which uses

spraying machinery for large-scale applications, is highly efficient. However, this method cannot precisely target the location of weeds, leading to a large degree of agricultural resource waste. Additionally, excessive chemical dispersal into the air and water can cause environmental pollution and pose health risks to humans and animals [6]. Traditional mechanical weeding primarily relies on equipment such as weeders and cultivators to physically remove or cut weeds. Although effective before crop emergence, its low level of automation often leads to accidental damage to crops post-emergence, causing irreversible harm. Targeted weeding technologies, which focus on specific weed locations, have become a key area of research in precision agriculture and the era of Agriculture 4.0 [7]. Intelligent weeding methods (such as targeted spraying, row spraying, and laser weeding) offer precise weed targeting, minimizing crop harm and reducing environmental chemical pollution. To meet the growing demands of smart agriculture, agricultural production must adopt more efficient and sustainable practices [8]. Notably, targeted spraying technology puts a higher demand on accuracy and speed for weed location identification. Therefore, developing a weed recognition algorithm with high accuracy, lightweight design, and fast recognition speed is of great research significance to improve the intelligence level of agricultural machinery and promote the development of precision agriculture.

For the problem of field weed recognition, scholars have carried out extensive related research work using computer vision technology [9,10]. Traditional intelligent weed recognition methods mainly rely on image processing and machine learning algorithms to extract texture, shape, color, and spectrum features. Tufail et al. [11] implemented a support vector machine (SVM) classifier using the texture, shape, and color features of tobacco plants, with a classification accuracy of 96%. Agarwal et al. [12] proposed a machine learning weed classification system based on the fusion of shape and texture features. They studied the performance of SVM, K-nearest neighbors (KNN), multi-layer perceptron (MLP), and Naive Bayes classifiers through 10-fold cross-validation. Results indicated that the SVM classifier combined with shape, shape curvature, and texture features achieved an overall accuracy of 99.33%. These methods demonstrate the potential of machine learning for weed detection in controlled environments. However, this approach has certain limitations in real-world scenarios. These methods rely heavily on manually designed feature engineering, which makes them less robust to changes in environmental conditions [10]. In addition, traditional machine learning methods such as SVM are mostly used for image classification and are difficult to directly use for image instance segmentation. Therefore, these characteristics limit their application in precision pesticide application scenarios.

In contrast, modern deep learning methods have gained attention due to their ability to learn features directly from data. These methods can automatically extract high-level semantic and spatial features, making them highly effective for weed detection in diverse field conditions. As a result, researchers have improved the accuracy and robustness of weed identification through single-stage or two-stage neural network models [13]. Fan et al. [14] utilized an optimized Faster R-CNN (Region-Based Convolutional Networks) model to recognize cotton seedlings and seven common weeds, achieving a mean average precision (mAP) of 94.21%. Similarly, Zou et al. [15] employed an improved U-Net model for weed segmentation, which achieved an intersection-over-union ratio of 92.91%. Yang et al. [16] proposed a corn weed recognition model called SE-VGG16. Experimental results showed that the average accuracy of the SE-VGG16 model was 99.67%, which was 1.92 percentage points higher than the original VGG16 model. Sun et al. [17] employed ResNet101 as the feature extraction network of the Faster R-CNN model, and its average recognition accuracy on the VOC (Visual Object Classes) format dataset was as high as 80.89%. Wang et al. [18] proposed a dual attention network that utilizes branch attention blocks in the encoding stage and spatial attention blocks in the decoding stage to bridge the gap between high-level and low-level features. This model outperformed ExFuse, DeepLabv3+, and PSPNet on a weed segmentation dataset. Sodjinou et al. [19] used the U-Net and K-means subtractive algorithm to apply the semantic segmentation of crops and weeds. Kim et al. [20] proposed a multi-task semantic segmentation convolutional

neural network (MTS-CNN) for detecting crops and weeds. The correlation between crop and weed categories was enhanced by adding crops, weeds, and the loss of both in one-stage training, and the method was verified experimentally on three open databases to have higher segmentation accuracy than the existing technology. Xu et al. [21] proposed a method combining visible color index and instance segmentation based on encoder–decoder architecture, which improved the detection and segmentation accuracy of densely distributed weeds and soybean crops.

The deep learning models mentioned above primarily aim to enhance the accuracy of crop and weed recognition. However, their application is limited on devices with restricted storage resources and computing power due to the large number of model parameters, high complexity, and slow recognition speed. With the rapid advancement of deep learning technology, the superior performance of single-stage YOLO (You Only Look Once) series models is becoming increasingly evident [22–25]. Ahmed et al. [26] designed a model based on YOLOv3 and Darknet-53 for detecting common weeds in corn and soybean fields, achieving precise target localization and boundary box output. Zhu et al. [27] proposed an improved YOLOx weed detection model, which integrates deep networks with a lightweight attention mechanism, enabling effective identification of different weed types in maize seedling fields. Fan et al. [28] developed a lightweight network based on YOLOv5 (YOLO-WDNet) for detecting cotton and weeds, achieving a 9.1% improvement in mAP<sub>0.5</sub> and a 57.14% reduction in inference time. Wang et al. [29] designed a weed target detection model based on YOLOv5s for addressing weed management in straw-covered corn fields, facilitating targeted spraying operations. Additionally, Rai et al. [30] explored a weed detection method using an optimized YOLOv7-tiny architecture integrated with edge computing technology to identify agricultural weeds in images and videos, although field operation testing has not yet been conducted. These studies demonstrate the high feasibility and fast detection potential of using single-stage deep learning models for weed detection in agricultural fields. Building on this foundation, this study aims to further explore the application of YOLOv8 as a benchmark model in targeted spraying object segmentation tasks. As one of the most advanced object detection algorithms available, YOLOv8 offers significant performance improvements over CNN (convolutional neural network) models such as YOLOv3, YOLOx, YOLOv5, and YOLOv7, along with advantages in real-time applications.

In targeted spraying research, the author designed a real-time targeted spraying system [31] and conducted a precise detection study on dense soybean seedlings using airborne edge devices [32]. The above research bases provide adequate support for deploying ground-based targeted spraying devices and algorithms for this study. Unlike previous work, this study employs a YOLO-based instance segmentation model to detect the malignant weed. In future targeted spraying applications, it will be necessary to use instance segmentation algorithms to detect weeds rather than target detection and semantic segmentation algorithms. The target detection algorithm can only provide bounding box information, and it is difficult to locate the precise contours of the weeds accurately. Although semantic segmentation can classify each pixel in the image, it cannot distinguish different individuals of the same species. In contrast, instance segmentation can simultaneously complete pixel-level classification and individual-level differentiation, ensuring the accurate identification of each weed. This precision is crucial for targeted spraying. Accurately identifying the contours and locations of weeds can significantly reduce herbicide waste, minimize crop damage, and enhance operational efficiency and environmental sustainability in precision agriculture.

This study utilizes the YOLOv8s-Seg model as the benchmark for the instance segmentation improvement model. By introducing a lightweight network and optimizing the model structure, an improved model for weed segmentation is proposed. The model is deployed and tested on an embedded platform to verify its feasibility in a resource-constrained environment, aiming to provide support for the promotion of targeted spraying technology in the future. Specifically, the contributions of this paper are as follows:

- (1) A lightweight weed instance segmentation detection model named FCB-YOLOv8s-Seg was developed by incorporating improved FasterNet, C2fSE, and BiFPN modules.
- (2) The FCB-YOLOv8s-Seg model was trained on self-collected soybean field weed images, demonstrating a high ability to discern subtle difference between weeds.
- (3) The comprehensive performance analysis and comparative experiments of the FCB-YOLOv8s-Seg model show that it outperforms existing baseline models in target detection and instance segmentation accuracy, while exhibiting better generalization ability and stability.
- (4) The FCB-YOLOv8s-Seg model was deployed on ground-based targeted spraying pesticide vehicles and tested in real soybean field environments, validating the method's effective segmentation performance and stable operational capability.

## 2. Materials and Methods

### 2.1. Weed Image Acquisition and Dataset Preprocessing

#### 2.1.1. Weed Image Acquisition

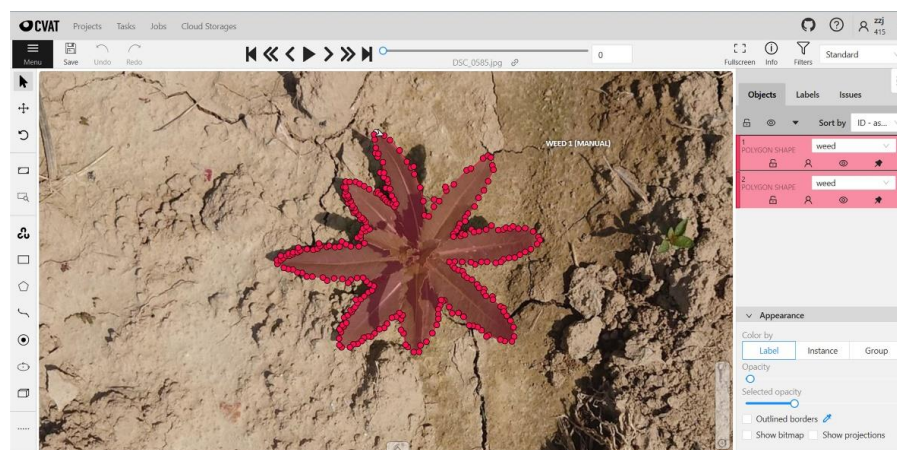
The weed object studied in this study is *Cirsium setosum* (*Cirsium arvense* var. *integrifolium*). It is a perennial and common malignant weed characterized by its well-developed stolons, strong herbicide resistance, and difficulty to control within dryland crops such as soybeans, corn, and wheat. The experimental images were collected in August 2023 in the soybean experimental field of the Changyuan Branch of the Henan Agricultural Science Academy in Changyuan City, Henan Province (35.428° N, 114.289° E). The acquisition device used in this study was a smartphone with a main camera of 50 megapixels and a lens with an aperture of  $f/1.9$  which supported optical image stabilization and autofocus. In this study, the resolution of the acquired image was set to  $4096 \times 3072$  pixels, the exposure parameters were set to automatic mode, and the image was saved in JPG format. The image acquisition method was a handheld acquisition device, which was orthogonal to the ground to collect weed images. To ensure the diversity of the dataset, image acquisition covered different time periods of multiple days. *Cirsium setosum* images were taken under complex conditions of lighting and background. After shooting, the quality of the *Cirsium setosum* images was assessed by removing severely blurry images, images without weeds, or images with incomplete weed boundary information. Finally, 500 samples of high-quality *Cirsium setosum* images were selected as the base dataset. Figure 1 shows some sample *Cirsium setosum* images.



**Figure 1.** Images collected from the *Cirsium setosum* dataset.

### 2.1.2. Production of the Base Dataset

The image annotation tool CVAT (Computer Vision Annotation Tool, Version 2.7.6) was used to annotate the dataset during data processing. Figure 2 illustrates the specific annotation process. The polygon tool was used to draw the outline of the weeds in the uploaded image with the mouse, which was then marked as “weed”. After completing the annotation, the annotation file was exported and converted into the COCO (Common Objects in Context) data format. This labeling file contains the coordinates, shape, size, and other details of all *Cirsium setosum*s in each image.



**Figure 2.** The CAVT annotation tool.

### 2.1.3. Data Division and Augmentation

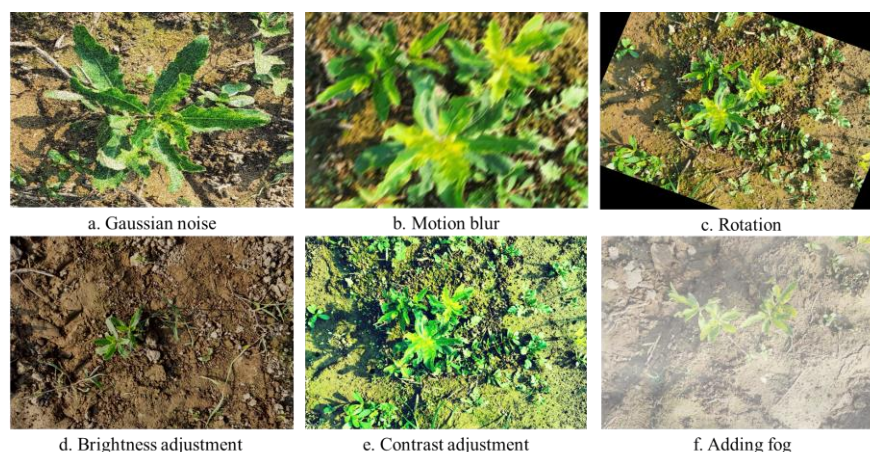
If the *Cirsium setosum* images in the base dataset are used directly for model training, underfitting is risky. Moreover, the operation state of the camera mounted on the intelligent machinery is shaky or bumpy. Therefore, the images in the basic dataset are relatively uniform, and the model’s generalization ability is weak. To improve the model’s generalization and enhance its ability to predict *Cirsium setosum*s under varying input images, it was necessary to carry out data augmentation processing on the 500 *Cirsium setosum* images in the base dataset mentioned above.

Before data augmentation, this study selected 50 images as a validation set and 100 images as a test set. These two sets of images were not processed and were used to evaluate the model’s performance. The remaining 350 images were used for training with image augmentation. The selected augmentation methods included Gaussian noise, motion blur, rotation, brightness adjustment, contrast adjustment, and adding fog. Figure 3 shows the effect of image augmentation. After augmentation, the *Cirsium setosum* dataset was expanded to 2250 images.

## 2.2. YOLOv8s-Seg Network Model

The YOLO series algorithms are single-stage and real-time object detection algorithms widely used in precision agriculture, smart orchards, and ecological unmanned farms [33,34]. The YOLOv8 series models are designed by the Ultralytics team based on the YOLOv5 series models [35]. Therefore, the YOLOv8-Seg (ultralytics-8.1.9) network follows the YOLOv5 network architecture. The official website provides five different specifications of the network models: YOLOv8n-Seg, YOLOv8s-Seg, YOLOv8m-Seg, YOLOv8l-Seg, and YOLOv8x-Seg. Considering the evaluation indicators (such as the mAP and model size), this study chose YOLOv8s-Seg as the experimental model. The YOLOv8s-Seg model consists of the backbone, neck, and prediction network. To improve performance, several optimizations have been made. First, the C2f module replaces the C3 module in the YOLOv5 backbone. The C2f module enhances feature extraction by using cross-layer connections and split operations, which improve gradient flow and enable more robust feature extraction, particularly in complex scenes such as weed detection. To maintain

a lightweight model, the C2f module reduces computation by adjusting the number of channels at different scales. The neck network continues to use the PANet structure for multi-scale feature fusion. However, the YOLOv8s-Seg removes convolution operations in the upsampling stage, which reduces the number of layers and parameters, resulting in faster computation and improved real-time performance. The prediction network draws on YOLACT, removing the objectness branch from the YOLOv5 model and retaining only the decoupled structure with classification and regression branches (Decoupled-Head). The prediction head consists of detection and segmentation branches, which are used to perform the detection and segmentation tasks, respectively. The detection branch generates class, bounding box, and mask coefficients, while the segmentation branch generates prototype masks. Additionally, the YOLOv8s-Seg model employs an anchor-free method to reduce the number of predicted boxes and enhance the effect of non-maximum suppression. After the non-maximal suppression module, the prediction mask coefficients and prototype masks for each instance are combined by matrix multiplication. The final instance is constructed through crop and threshold operations. For loss function calculation, the YOLOv8s-Seg model uses the Task Aligned Assigner positive sample assignment strategy and adopts CIOU (Complete Intersection Over Union) loss, BCE (Binary Cross Entropy) loss, and distribution focus loss to improve accuracy in both detection and segmentation tasks.



**Figure 3.** Demonstration of image augmentation effect.

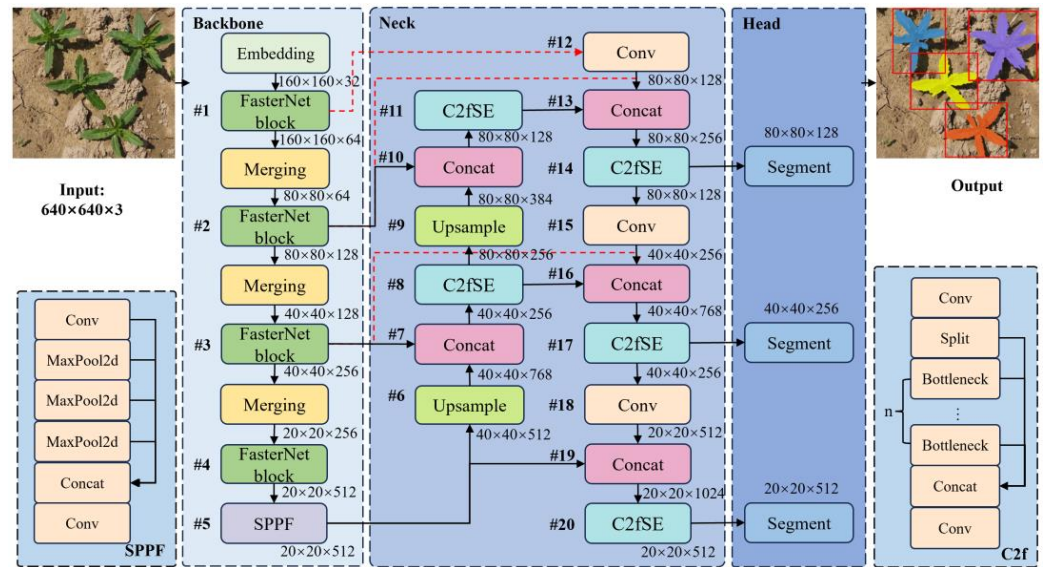
### 2.3. Improvement Scheme of the YOLOv8s-Seg Model

#### 2.3.1. Improvement of the Backbone Network Based on FasterNet

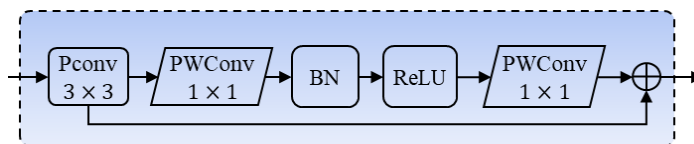
The YOLOv8s-Seg model uses the CSPDarknet module as the backbone feature extraction network. Although its cross-stage partial connection method reduces the amount of computation to a certain extent, the complex network structure and more convolutional layers limit its application on edge devices with limited computing resources. To solve this problem, many researchers have tried to use lightweight networks (such as Ghost-Net, MobileNet, and ShuffleNet) for feature extraction. These methods reduce the model complexity by introducing technologies such as deep separable convolution, Ghost convolution, and group convolution. However, these technologies usually require additional data processing operations (such as connection, shuffle, pooling, etc.). The frequent memory access of these operations will significantly affect the real-time performance of the detection model and increase the actual running time and computational overhead. To this end, this study chose to replace the backbone network of YOLOv8s-Seg with FasterNet. FasterNet has become the preferred solution for lightweight networks through its simplified network structure, efficient feature extraction mechanism, and optimized design for multiple hardware platforms. Its compact structure and small number of parameters significantly reduce the computational complexity. At the same time, it reduces memory access by optimizing convolution operations, greatly improving the real-time performance and operation efficiency of the model. Compared with other lightweight networks, FasterNet not only

reduces redundant data processing operations, but also can run efficiently on hardware platforms such as CPU, ARM, and GPU, making full use of computing resources, and further enhancing its applicability on edge devices.

As shown in Figure 4, the backbone network of the FCB-YOLOv8s-Seg model mainly contains an embedding layer, four fusion layers, three FasterNet block modules, and a Spatial Pyramid Pooling Fusion (SPPF) module. Among them, the embedding layer comprises a conventional  $4 \times 4$  with a step size of 4 for spatial downsampling. The fusion layers consist of a traditional  $2 \times 2$  with a step size of 2 for channel number expansion. As the main feature extraction module, the FasterNet block module shown in Figure 5 comprises one partial convolution layer (PConv) and two point-wise convolution layers (PWConv). This residual block improves detection accuracy and speed while reducing memory access and computational redundancy. In addition, the two PWConv layers are interspersed with batch normalization (BN) and rectified linear units (ReLU). This setting of normalization and activation layers can achieve lower latency and maintain functional diversity. Specifically, the role of the BN layer is to accelerate the training speed and improve accuracy. The role of the ReLU activation function is to speed up the model training speed and avoid the gradient disappearance problem. The SPPF module has not been changed in the FCB-YOLOv8s-Seg model structure.



**Figure 4.** The structure of the FCB-YOLOv8s-Seg model. Note: #1, . . . , #20 represent the names of the network output of layer 1st, . . . , 20th; Conv is composed of Convolution2d, BatchNorm2d, and SiLU.



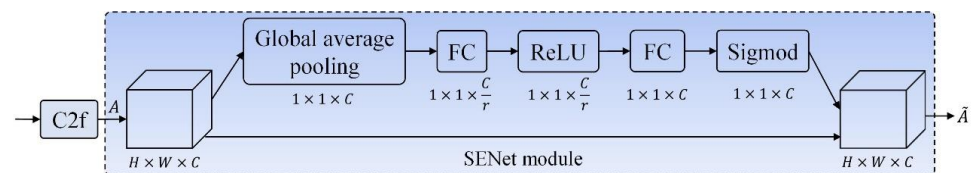
**Figure 5.** Structural components of the FasterNet block.

### 2.3.2. C2fSE Attention Module

Traditional convolutional neural networks typically use convolution and pooling layers to extract image features. However, this structure does not explicitly model the relationships between feature channels, which can result in some channels not contributing as effectively to the model’s performance. To address this issue, the SENet channel attention mechanism introduces two key operations, squeeze and excitation, allowing the model to adaptively learn and focus on the importance of each channel for the task. According to the task needs, the channel contributions in the feature map are weighted and adjusted,

enhancing the model's ability to discriminate features and prediction performance [36]. The feature map remains consistent before and after passing through the SENet module. Therefore, the SENet channel attention mechanism can be added anywhere in the model structure. In this study, the SENet attention mechanism is integrated with the C2f module to form the C2fSE module, which is placed in the 8th, 11th, 14th, 17th, and 20th layers, as shown in Figure 4. Introducing and optimizing the attention mechanism helps improve the *Cirsium setosum* segmentation accuracy of the FCB-YOLOv8s-Seg model.

Figure 6 illustrates the components of the C2fSE module structure. Assume the feature map output from the C2f module is  $A$ , with dimensions  $H \times W \times C$ , where  $H$ ,  $W$ , and  $C$  represent the image's height, width, and number of channels, respectively. The feature map  $A$  undergoes three main steps: squeeze, excitation, and scale in the SENet module. First, the feature map  $A$  is squeezed. The spatial dimension information of each channel is compressed into a single value through global average pooling to obtain a global description of each channel. Next is the excitation operation, where two fully connected layers and a nonlinear activation function are used to learn, generate, and normalize the importance weights of each channel, capturing the relationships between channels. In the final scaling process, the weights generated from the excitation step are multiplied with the original feature map to obtain the output feature map. This operation can adjust the contribution of each channel, highlight important channels, and suppress unimportant ones.



**Figure 6.** The C2fSE module structure diagram.

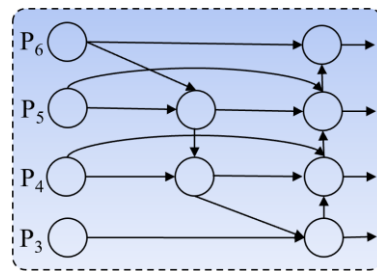
### 2.3.3. Introduction and Optimization of the BiFPN Module

The YOLOv8s-Seg model primarily relies on the PANet module for feature fusion. Although this module can merge features from different scales, it is ineffective in adjusting and optimizing feature map weights. This limitation can reduce the model's performance, particularly in complex detection and segmentation tasks where the balance between different types of information is crucial. To address this issue, this study introduces the BiFPN, which is an improvement of the Feature Pyramid Network (FPN) [37]. The BiFPN realizes top-down and bottom-up feature fusion by introducing bidirectional connections between different feature layers. In the bottom-up path, the low-level features are gradually transferred to the high-level, enhancing the ability of high-level features to capture local details. In the top-down path, the high-level features are gradually transferred to the low-level, strengthening the semantic expression ability of the low-level features. In addition, BiFPN introduces a learnable weight mechanism, which allows the network to adaptively adjust the weights of feature maps at different levels, optimize the trade-off between spatial and semantic information, and thus more effectively extract features and enhance the performance of the model. Figure 7 illustrates the BiFPN module structure, where P3-P6 represents the feature layers from low to high.

In this study, two main tasks were performed when introducing and optimizing the BiFPN module: (1) Adjusting the neck network structure of the YOLOv8s-Seg model. Specifically, the output position of the  $80 \times 80 \times 128$  detection head was changed. The Conv layer, Concat layer, and C2fSE layer were added after the C2fSE module in the 11th layer to enhance the capability of extracting rich semantic features in the deep network. (2) Constructing the base structure of the BiFPN module. Specifically, the Concat layers of the neck network's 12th, 13th, and 16th layers are cross-layer fused with the output feature maps of the backbone network's 1st, 2nd, and 3rd layers (as shown by the red dashed lines in Figure 4). By introducing this structure, the BiFPN module enables each feature



map to be weighted appropriately, improving the fusion of features at different levels and enhancing overall feature extraction.



**Figure 7.** The BiFPN module structure diagram.

## 2.4. Model Training and Evaluation Indicators

### 2.4.1. Model Training Environment

In deep learning networks' training and prediction processes, a large quantity of convolutional and matrix computations often require GPU parallel computing to accelerate model training and inference. This study conducted all model training and testing processes on the same server platform. The PyTorch deep learning framework and the NVIDIA GeForce RTX 3060 GPU were employed on an Ubuntu server to implement parallel computing through CUDA. The CUDNN acceleration library was also integrated into the deep learning framework to accelerate model training and inference further. The specific experimental environment configuration is shown in Table 1.

**Table 1.** Model training and prediction environment configuration.

Running Environment	Version
CPU	Intel Core i5-12600KF
GPU	NVIDIA GeForce RTX 3060 12GB
Server environment	Ubuntu 20.04.3 LTS
Deep learning framework	Pytorch 3.8.18
CUDA	11.0
Python version	3.11

### 2.4.2. Model Training and Testing Parameter Settings

In this study, the YOLOv8-related series models were built using the PyTorch library. The input image size was  $640 \times 640$ . Appropriate training parameters were selected through model pre-training. Specifically, the stochastic gradient descent optimizer was used in the training process, and the training duration was 300 epochs. The initial learning rate was set to 0.01, and the momentum parameter was set to 0.937. To prevent overfitting, the weight decay was set to 0.0005. The batch sizes were set to 16 and 1 during training and testing, respectively. Additionally, the FCB-YOLOv8s-Seg model turns off the mosaic enhancement operation in the last ten epochs to improve the model's prediction accuracy [38]. When testing the models' frame rate (FPS), 12 tests were performed after the server warm-up. The longest and shortest inference times were discarded to eliminate the hardware fluctuation effects, and the average of the remaining 10 test results was calculated as the final FPS result value. The intersection-over-union (IOU) threshold for non-maximum suppression was set to 0.5, and the confidence threshold was set to 0.5.

### 2.4.3. Model Evaluation Indicators

After completing the training of the *Cirsium setosum* instance segmentation model, it is necessary to evaluate the performance of the trained model quantitatively. Common instance segmentation evaluation indicators include precision (P), recall (R), average precision (AP), and mean average precision (mAP). To clarify the interpretation of these indicators,

relevant parameters are defined as follows: true positive (TP) represents correct positive predictions, false positive (FP) indicates incorrect positive predictions, false negative (FN) refers to missed positive cases, and true negative (TN) denotes correct negative predictions. The calculation formulas of each indicator used in this study are shown in (1)–(5).

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (4)$$

$$IOU = \frac{A \cap B}{A \cup B} \quad (5)$$

In the formulas,  $i$  represents the  $i$ -th class,  $n$  represents the total number of classes.  $A$  represents the ground truth bounding box, and  $B$  represents the predicted bounding box.

In practical applications, only when the precision and recall of a model are both at a high level can the model be considered effective. However, the relationship between the two is often negatively correlated. If the precision is improved, the number of samples that are actually predicted to be true will decrease. The total number of positive samples remains unchanged, resulting in a decrease in recall. Similarly, if the recall is increased, the precision will decrease accordingly. The AP index is introduced to balance precision and recall and help choose the best trade-off between the two. AP represents the precision of the model for a specific class, defined as the area under the P–R (precision–recall) curve. The mAP indicates the average precision across all classes, with higher mAP values signifying better model performance.

In the instance segmentation model of this study, the AP value is equivalent to the mAP value. The suffix “\_obj” is used for object detection box predictions, and “\_mask” for segmentation predictions. These two indicators are used to measure the model’s performance in target detection and instance segmentation tasks. The prediction box mAP focuses on the accuracy of the target location, while the segmentation mAP focuses on the accuracy of the target shape. By calculating and comprehensively comparing these two mAP indicators separately, the instance segmentation performance of the model can be comprehensively evaluated.

### 3. Results and Discussion

#### 3.1. Ablation Experiments

This study conducted ablation experiments to validate the effectiveness of different improvement schemes for model detection and segmentation performance. Each improvement method was added sequentially to verify its necessity. All experiments were conducted under consistent conditions to ensure fair and impartial comparisons. The results of relevant indicators are shown in Table 2.

As can be seen from Table 2, with the continuous recursion of the model improvement scheme (from YOLOv8s-Seg to FCB-YOLOv8s-Seg), the mAP indicators of the prediction box and the segmentation prediction box show an overall trend of gradual improvement. First, after replacing the backbone network in the YOLOv8s-Seg model with FasterNet, the prediction box and segmentation results were improved by 0.8% and 1.7%, respectively, compared with the original model. The overall trend is increased P, decreased R, increased mAP, faster inference speed, and reduced model size. The decrease in recall indicates an increase in missed detections (undetected *Cirsium setosums*). The lightweight design of

FasterNet may enhance the detection ability of high-confidence weed targets, making the model more inclined to output higher-confidence results and filtering out some lower-confidence candidate boxes. At the same time, during feature extraction, more attention may be paid to weed targets with strong salience, while some smaller, complex, or edge targets are ignored. Because some weeds are missed, the overall recall becomes lower. However, it is worth noting that moderate misses may be acceptable in actual agricultural applications, especially when the model can significantly reduce false positives (identifying non-weeds as weeds), thereby reducing unnecessary processing work. Adding the C2fSE module to the neck network helps to extract rich semantic features in the deep network, so that the detection box and segmentation results increase by 2% and 0.9%, respectively, compared with the YOLOv8-Seg+FasterNet model. Based on this, the FCB-YOLOv8s-Seg model after introducing BiFPN has a mAP of 95.18% (prediction box) and 96.63% (segmentation), which are 5.08% and 7.43% higher than the original model, respectively. At the same time, the inference FPS increased by 7.7%, and the model size decreased by 1.8%. It can be seen from the vertical column data in the table that although the improvement scheme still needs to be further improved in terms of reasoning speed and model lightweightness, it nevertheless plays a significant positive role in the recognition accuracy of *Cirsium setosum* in soybean fields.

**Table 2.** Results of the ablation experiment.

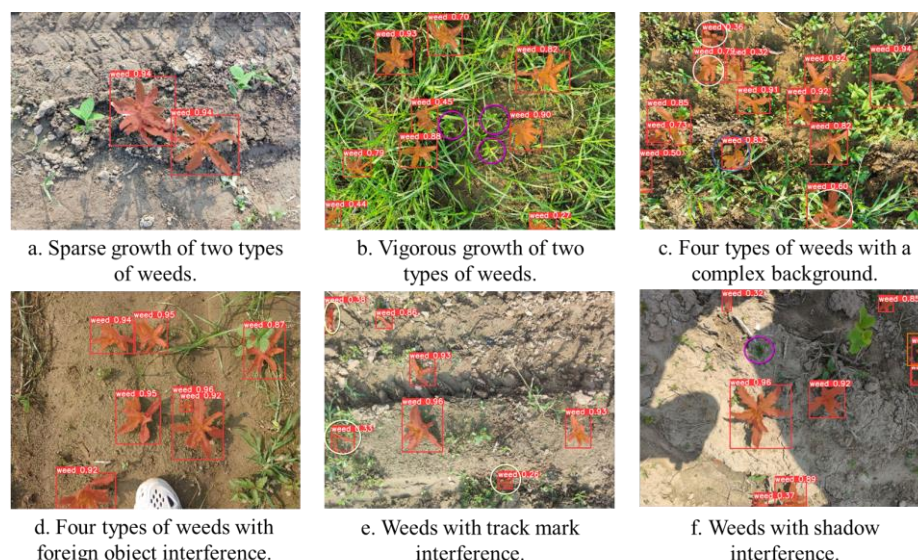
Models	P_Obj (%)	R_Obj (%)	mAP_obj (%)	P_Mask (%)	R_Mask (%)	mAP_Mask (%)	FPS	Size (MB)
YOLOv8s-Seg	82.2	88.0	90.1	81.7	87.5	89.2	75.47	22.7
+FasterNet	89.3 <sup>+7.1</sup>	80.2 <sup>-7.8</sup>	90.9 <sup>+0.8</sup>	89.0 <sup>+7.3</sup>	80.4 <sup>-7.1</sup>	90.9 <sup>+1.7</sup>	86.2 <sup>+10.74</sup>	17.9 <sup>-4.8</sup>
+FasterNet+C2fSE	90.9 <sup>+1.6</sup>	85.9 <sup>+5.7</sup>	92.9 <sup>+2</sup>	89.9 <sup>+0.9</sup>	85.9 <sup>+5.5</sup>	91.8 <sup>+0.9</sup>	81.3 <sup>-4.91</sup>	18.1 <sup>+0.2</sup>
+FasterNet+C2fSE+BiFPN	91.95 <sup>+1.05</sup>	86.4 <sup>+0.5</sup>	95.18 <sup>+2.28</sup>	91.3 <sup>+1.4</sup>	89.95 <sup>+4.05</sup>	96.63 <sup>+4.83</sup>	81.3 <sup>+0</sup>	22.3 <sup>+4.2</sup>

Note: (1) The upper right corner value of the number in this table represents the change value of each indicator compared with the previous model, where “+” represents an increase and “-” represents a decrease. (2) The bold part is the optimal improvement scheme finally selected in this study. (3) FCB-YOLOv8s-Seg is the abbreviation of YOLOv8s-Seg+FasterNet+C2fSE+BiFPN.

### 3.2. Segmentation Effect of the FCB-YOLOv8s-Seg Model in Different Scenes

In the early stages of soybean seedlings, various weeds thrive. Their growth rate far exceeds that of the soybean seedlings. In this study, images of mixed growth of *Cirsium setosum* and soybean seedlings in various scenes were collected to test the performance of the FCB-YOLOv8s-Seg model. Figure 8 shows the *Cirsium setosum* segmentation effect in soybean fields under different complexity scenes.

Figure 8a shows a scene where soybean seedlings and two types of weeds coexist with sparse growth. The background in this scene is relatively simple, and the soybean seedlings are in the cotyledon stage. The FCB-YOLOv8s-Seg model can completely segment *Cirsium setosum*, demonstrating its effectiveness in a simple background. In the scene shown in Figure 8b, the two mixed weeds compete intensely with the soybean seedlings for survival space. Although most of the *Cirsium setosum* were thoroughly segmented, three missed detections were due to severe occlusion by other types of weeds (as shown in the purple circles in Figure 8b). For this scene, it is recommended that other types of weeds be controlled through targeted drug control first. Then, the FCB-YOLOv8s-Seg model can be used to recognize *Cirsium setosum* to improve detection accuracy. Figure 8c depicts a soybean seedling scene with four mixed weeds under a complex background. In this scene, the FCB-YOLOv8s-Seg model successfully segmented nine *Cirsium setosum* plants. Still, three other types of weeds (as shown in the white circle in Figure 8c) and one bean seedling (as shown in the blue circle in Figure 8c) were also misdetected. Although these misdetections will increase the workload for targeted operations, they also suggest the potential of the FCB-YOLOv8s-Seg model to detect other types of weeds. Future work should optimize the model for multi-weed segmentation and identification tasks.



**Figure 8.** Segmentation effect of the FCB-YOLOv8s-Seg model in different scenes.

Compared with the three scenes mentioned above, the scenes in Figure 8d–f involve added artificial interference. In Figure 8d, there is interference from a shoe. The FCB-YOLOv8s-Seg model successfully segmented seven *Cirsium setosum*. There is interference from agricultural machinery vehicles in Figure 8e. The FCB-YOLOv8s-Seg model successfully segmented four *Cirsium setosum* in this scene. Still, there were also three places where other types of weeds were misdetected (as shown in the white circle in Figure 8e), showing the challenges of the FCB-YOLOv8s-Seg model in dealing with vehicle interference. Under the shadow interference in Figure 8f, the FCB-YOLOv8s-Seg model successfully segmented eight *Cirsium setosum* and missed one *Cirsium setosum* due to its small shape (as shown in the purple circle in Figure 8f). Additionally, the right side of Figure 8f shows incomplete segmentation due to the partial exposure of the target (as shown in the orange rectangular box in Figure 8f). Further improvements should focus on small target detection and camera angle adjustment. The FCB-YOLOv8s-Seg model performs well in complex scenes and fine-grained feature segmentation, demonstrating higher applicability in practical instance segmentation tasks. The FCB-YOLOv8s-Seg model will provide strong technical support for future in-vehicle deployment.

### 3.3. Comparison of Performance Parameters with Other Models

To further explore whether the FCB-YOLOv8s-Seg segmentation model has performance advantages in *Cirsium setosum* recognition and segmentation, this study compared it with classic segmentation models such as Mask-RCNN, YOLACT, and YOLOv5s-Seg. The same training and validation sets were used for all models. Table 3 presents the performance test results of each model.

**Table 3.** The performance test results of different models.

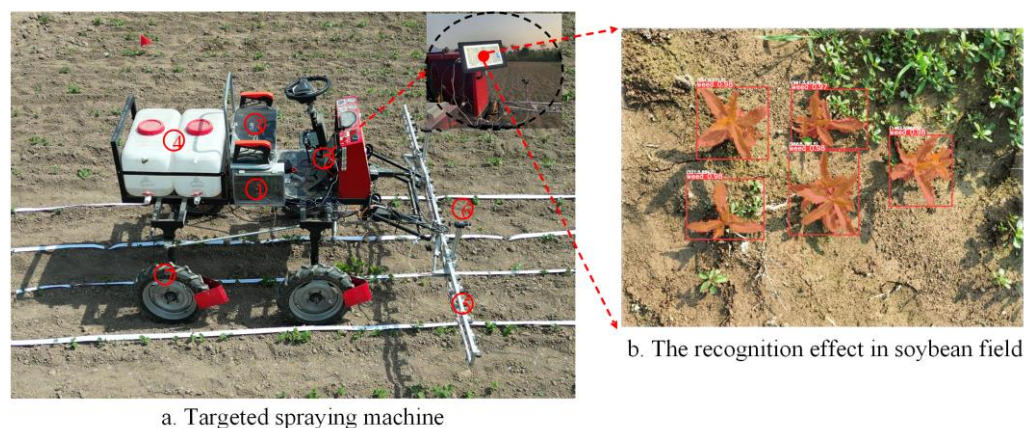
Models	mAP_Obj (%)	mAP_Mask (%)	Size (MB)
YOLOv5s-Seg	88.80	87.65	14.5
MaskRCNN (ResNet 50)	72.19	60.85	335
YOLACT	82.30	84.50	117
YOLOv8s-Seg	90.10	89.20	22.7
FCB-YOLOv8s-Seg	95.18	96.63	22.3

The FCB-YOLOv8s-Seg model significantly outperforms other models’ target detection accuracy, notably surpassing the Mask R-CNN model by 23 percentage points. Regarding segmentation accuracy, the FCB-YOLOv8s-Seg model also maintains a clear lead, with

mAP\_mask values exceeding those of the other models by 8.98%, 35.78%, and 12.13%, respectively. In comparison, although the YOLOv5s-Seg model has the smallest size, its accuracy is lower than that of the FCB-YOLOv8s-Seg model. The Mask RCNN and YOLACT models are too large to be deployed in environments with strict computing resources and memory restrictions. Overall, the FCB-YOLOv8s-Seg model exhibits the best performance in target detection and segmentation accuracy while maintaining a relatively small model size. The above evaluation indicator results thoroughly verify its strong feasibility and high performance advantage in the *Cirsium setosum* segmentation task.

### 3.4. Field Test and Result Analysis

To explore whether the FCB-YOLOv8s-Seg model improves the targeted spraying operation in soybean fields, this study deployed the YOLOv8s-Seg model before and after the improvement on the ground target pesticide application vehicle for testing. The data collection device used was a GoPro HERO11 Black action camera with a primary camera resolution of 27 megapixels. In the experimental field shown in Figure 9a, three 20 m × 3 m strip areas were designated as sampling areas. In each sampling area, testing was conducted at 2 km/h, 3 km/h, and 4 km/h. The number of *Cirsium setosum* in each strip area and the number of successfully segmented *Cirsium setosum* on the human-machine interface were recorded to calculate each area's *Cirsium setosum* recognition segmentation rate. Each treatment level was repeated three times, resulting in 27 experiments. The average number of successful segmentations from the three experiments was calculated, and the experimental results are shown in Table 4. Figure 9b illustrates the recognition effect of *Cirsium setosum* in soybean fields under stable conditions.



**Figure 9.** Soybean field targeted spraying test site and recognition results. Note: 1. edge control device; 2. human-machine interface; 3. cab; 4. pesticide box; 5. spraying mechanism; 6. camera; 7. rack.

From the statistical results in Table 4, it can be seen that there are obvious differences in *Cirsium setosum* density between different plots. As *Cirsium setosum* density increases, the average segmentation rate of the two algorithms shows a declining trend. The plot with lower *Cirsium setosum* density exhibits the best segmentation effect, with the highest average segmentation rate of 94.16%. This may be because when the number of *Cirsium setosum* is small, the algorithm has a lighter burden and can segment the *Cirsium setosum* more accurately. Conversely, the plot with higher *Cirsium setosum* density shows poorer segmentation performance, with the lowest average segmentation rate at 83.8%. The increased *Cirsium setosum* density likely adds to the algorithm's processing burden and leads to more false negatives and positives, adversely affecting segmentation accuracy.

**Table 4.** Statistical results of field segmentation tests in soybean fields.

Area Number	Speed (km/s)	Number of <i>Cirsium arvense</i>	YOLOv8s-Seg		FCB-YOLOv8s-Seg	
			Segmentation Count	Average Segmentation Rate (%)	Segmentation Count	Average Segmentation Rate (%)
1	2	97	87	86.60	95	94.16
	3		85		91	
	4		80		88	
2	2	142	130	83.80	136	89.67
	3		119		127	
	4		108		119	
3	2	106	96	84.91	102	90.88
	3		91		97	
	4		83		90	
Total		345	293.00	84.93	315	91.30

In the same plot, the performance of both models generally decreases as the operational speed increases from 2 km/s to 4 km/s. This decline may be attributed to two factors: first, the increase in speed may cause rapid changes in light conditions or increased camera shake, which affects image quality. Second, higher speeds reduce the time available for the algorithm to process each frame, leading to lower segmentation effectiveness. Therefore, reducing the forward speed in segmenting *Cirsium setosum*s in soybean fields helps improve the average segmentation rate of *Cirsium setosum*s. However, in practical applications, balancing segmentation accuracy and operational efficiency requires optimizing the forward speed based on different plot conditions or introducing more stable image capture and processing technologies to mitigate the negative effects of increased speed. Notably, the FCB-YOLOv8s-Seg model demonstrates greater stability and higher segmentation rates under varying speeds.

Comprehensive analysis shows that the total number of thorn grass in the three plots is 345, and the FCB-YOLOv8s-Seg algorithm successfully segmented 315 plants. The average segmentation rate reached 91.3%, 6.38 percentage points higher than the YOLOv8s-Seg algorithm. These data once again verify the effectiveness of the improved scheme. Since the actual field conditions are not flat, the average segmentation rate in actual operation (91.3%) is 5.06 percentage points lower than the segmentation results under stable camera conditions (96.63%). This difference is reasonable. To shorten the gap, the next step is to optimize the lighting conditions, upgrade the camera configurations, or research the image recovery algorithms under the shaking state to enhance the segmentation effect under high-speed operation.

While the FCB-YOLOv8s-Seg model has demonstrated significant improvements in segmentation accuracy and real-time performance, certain challenges persist. Future research will focus on addressing the limitations in detecting small and occluded targets by exploring advanced attention mechanisms and multi-scale feature enhancement techniques. Additionally, research will aim to improve model adaptability to varying environmental conditions, such as changes in lighting, background complexity, and crop growth stages, ensuring robustness across diverse real-world scenarios. Furthermore, efforts will be directed towards optimizing the model for high-density weed distribution to maintain a balance between precision and recall in complex agricultural settings. At the same time, this study primarily focused on *Cirsium setosum* in soybean fields, which presents a limitation in terms of generalizability. To expand the applicability of the model, future work will involve extending its use to other crops and weed types by creating a more diverse dataset that captures the variability in weed morphology and environmental conditions. However, challenges such as differences in weed and crop characteristics, background complexity, and the need for additional annotation resources must be addressed to ensure robust model performance across multiple scenarios. These studies aim to further enhance the

model's practicality and reliability, ultimately contributing to the development of intelligent agricultural applications on a broader scale.

#### 4. Conclusions

This study proposes an FCB-YOLOv8s-Seg model for instance segmentation of *Cirsium* in soybean fields. The YOLOv8s-Seg model's backbone was replaced with the lightweight FasterNet network, and the neck network was enhanced with the optimized C2fSE attention mechanism and BiFPN module. These modifications significantly improved the model's inference speed and feature extraction capabilities. Through a series of experimental verifications, the following conclusions were drawn:

- (1) Ablation experiments show that the FCB-YOLOv8s-Seg model is significantly better than the original model in terms of mAP evaluation indicators, achieving 95.18% and 96.63% mAP in bounding box detection and segmentation, respectively, which were 5.08% and 7.43% higher than the original model, respectively. At the same time, the inference FPS increased by 7.7%, and the model size decreased by 1.8%. The model's target detection and segmentation accuracy performance surpass some existing models while maintaining a balanced model size.
- (2) Image detection results in different scenes show that the FCB-YOLOv8s-Seg model performs well in complex scenes and fine-grained feature segmentation.
- (3) Comparative analysis with classic segmentation algorithms, such as Mask-RCNN, YOLACT, and YOLOv5s-Seg, demonstrates the feasibility and performance advantages of the FCB-YOLOv8s-Seg model in *Cirsium setosum* segmentation tasks.
- (4) In the field targeted spraying operation test, the average segmentation rate of the FCB-YOLOv8s-Seg model reached 91.3%, which was 6.38% better than the original model, further proving its practical reliability in smart agriculture applications.

In summary, the proposed FCB-YOLOv8s-Seg segmentation model offers greater practicality and reliability in real-world applications, providing solid technical support for the development of intelligent agricultural operations. Despite these significant achievements, challenges remain in adequately addressing the variability and complexity of targeted spraying scenarios in the field. Future work will focus on optimizing model adaptability and further improving detection capabilities in dynamic and complex agricultural environments.

**Author Contributions:** Conceptualization, Z.Y. and L.W.; methodology and software, Z.Y. and L.W.; investigation, C.L.; data curation, C.L.; writing—original draft preparation, Z.Y.; writing—review and editing, L.W.; supervision and project administration, H.L.; funding acquisition, Z.Y. and H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Science and Technology R&D Plan Joint Fund of Henan Province (30602873) and the China Agriculture Research System of MOF and MARA (CARS-04-PS28).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to possible further research.

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

#### References

1. Darbyshire, M.; Coutts, S.; Bosilj, P.; Sklar, E.; Parsons, S. Review of weed recognition: A global agriculture perspective. *Comput. Electron. Agric.* **2024**, *227*, 109499. [[CrossRef](#)]
2. García-Navarrete, O.L.; Correa-Guimaraes, A.; Navas-Gracia, L.M. Application of Convolutional Neural Networks in Weed Detection and Identification: A Systematic Review. *Agriculture* **2024**, *14*, 568. [[CrossRef](#)]
3. Grün, E.; Alves, A.F.; da Silva, A.L.; Zanon, A.J.; Corrêa, A.R.; Leichtweis, E.M.; Neto, R.C.A.; da Rosa Ulguim, A. How Do Off-Season Cover Crops Affect Soybean Weed Communities? *Agriculture* **2024**, *14*, 1509. [[CrossRef](#)]
4. Sharma, G.; Shrestha, S.; Kunwar, S.; Tseng, T.-M. Crop diversification for improved weed management: A review. *Agriculture* **2021**, *11*, 461. [[CrossRef](#)]

5. Gaskin, J. Recent contributions of molecular population genetic and phylogenetic studies to classic biological control of weeds. *BioControl* **2024**, *69*, 353–360. [[CrossRef](#)]
6. 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](#)]
7. 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. *Netw. Comput. Appl.* **2021**, *187*, 103107. [[CrossRef](#)]
8. Andreasen, C.; Scholle, K.; Saberi, M. Laser weeding with small autonomous vehicles: Friends or foes? *Front. Agron.* **2022**, *4*, 841086. [[CrossRef](#)]
9. Jin, X.; Che, J.; Chen, Y. Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access* **2021**, *9*, 10940–10950. [[CrossRef](#)]
10. Khan, A.; Ilyas, T.; Umraiz, M.; Mannan, Z.I.; Kim, H. Ced-net: Crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. *Electronics* **2020**, *9*, 1602. [[CrossRef](#)]
11. Tufail, M.; Iqbal, J.; Tiwana, M.I.; Alam, M.S.; Khan, Z.A.; Khan, M.T. Identification of tobacco crop based on machine learning for a precision agricultural sprayer. *IEEE Access* **2021**, *9*, 23814–23825. [[CrossRef](#)]
12. Agarwal, D. A machine learning framework for the identification of crops and weeds based on shape curvature and texture properties. *Int. J. Inf. Technol.* **2024**, *16*, 1261–1274. [[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. Fan, X.; Zhou, J.; Xu, Y.; Li, K.; Wen, D. Identification and Localization of Weeds Based on Optimized Faster R-CNN in Cotton Seedling Stage. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 26–34. [[CrossRef](#)]
15. Zou, K.; Chen, X.; Wang, Y.; Zhang, C.; Zhang, F. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. *Comput. Electron. Agric.* **2021**, *187*, 106242. [[CrossRef](#)]
16. 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](#)]
17. Sun, Z.; Zhang, C.; Ge, L.; Zhang, M.; Li, W.; Tan, Y. Image Detection Method for Broccoli Seedlings in Field Based on Faster R-CNN. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 216–221. [[CrossRef](#)]
18. Wang, H.; Song, H.; Wu, H.; Zhang, Z.; Deng, S.; Feng, X.; Chen, Y. Multilayer feature fusion and attention-based network for crops and weeds segmentation. *J. Plant Dis. Prot.* **2022**, *129*, 1475–1489. [[CrossRef](#)]
19. Sodjinou, S.G.; Mohammadi, V.; Mahama, A.T.S.; Gouton, P. A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. *Inf. Process. Agric.* **2022**, *9*, 355–364. [[CrossRef](#)]
20. Kim, Y.H.; Park, K.R. MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds. *Comput. Electron. Agric.* **2022**, *199*, 107146. [[CrossRef](#)]
21. Xu, B.; Fan, J.; Chao, J.; Arsenijevic, N.; Werle, R.; Zhang, Z. Instance segmentation method for weed detection using UAV imagery in soybean fields. *Comput. Electron. Agric.* **2023**, *211*, 107994. [[CrossRef](#)]
22. Chen, C.; Zheng, Z.; Xu, T.; Guo, S.; Feng, S.; Yao, W.; Lan, Y. Yolo-based uav technology: A review of the research and its applications. *Drones* **2023**, *7*, 190. [[CrossRef](#)]
23. Jiang, P.; Qi, A.; Zhong, J.; Luo, Y.; Hu, W.; Shi, Y.; Liu, T. Field cabbage detection and positioning system based on improved YOLOv8n. *Plant Methods* **2024**, *20*, 96. [[CrossRef](#)]
24. Jing, X.; Wang, Y.; Li, D.; Pan, W. Melon ripeness detection by an improved object detection algorithm for resource constrained environments. *Plant Methods* **2024**, *20*, 127. [[CrossRef](#)]
25. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
26. 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](#)]
27. 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](#)]
28. Fan, X.; Sun, T.; Chai, X.; Zhou, J. YOLO-WDNet: A lightweight and accurate model for weeds detection in cotton field. *Comput. Electron. Agric.* **2024**, *225*, 109317. [[CrossRef](#)]
29. Wang, X.; Wang, Q.; Qiao, Y.; Zhang, X.; Lu, C.; Wang, C. Precision Weed Management for Straw-Mulched Maize Field: Advanced Weed Detection and Targeted Spraying Based on Enhanced YOLO v5s. *Agriculture* **2024**, *14*, 2134. [[CrossRef](#)]
30. Rai, N.; Zhang, Y.; Villamil, M.; Howatt, K.; Ostlie, M.; Sun, X. Agricultural weed identification in images and videos by integrating optimized deep learning architecture on an edge computing technology. *Comput. Electron. Agric.* **2024**, *216*, 108442. [[CrossRef](#)]
31. Li, H.; Guo, C.; Yang, Z.; Chai, J.; Shi, Y.; Liu, J.; Zhang, K.; Liu, D.; Xu, Y. Design of field real-time target spraying system based on improved yolov5. *Front. Plant Sci.* **2022**, *13*, 1072631. [[CrossRef](#)]
32. Yang, Z.; Liu, J.; Wang, L.; Shi, Y.; Cui, G.; Ding, L.; Li, H. Fast and Precise Detection of Dense Soybean Seedlings Images Based on Airborne Edge Device. *Agriculture* **2024**, *14*, 208. [[CrossRef](#)]



33. Wang, L.; Zhao, Y.; Xiong, Z.; Wang, S.; Li, Y.; Lan, Y. Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model. *Front. Plant Sci.* **2022**, *13*, 965425. [[CrossRef](#)] [[PubMed](#)]
34. Terven, J.; Córdova-Esparza, D.; Romero-González, J. 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](#)]
35. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 1 January 2024).
36. Fang, M.; Liang, X.; Fu, F.; Song, Y.; Shao, Z. Attention mechanism based semi-supervised multi-gain image fusion. *Symmetry* **2020**, *12*, 451. [[CrossRef](#)]
37. Tan, M.; Pang, R.; Le, Q. 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.
38. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430. [[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.