sustainability

MDPI

Article

An Improved U-Net Model Based on Multi-Scale Input and Attention Mechanism: Application for Recognition of Chinese Cabbage and Weed

Zhongyang Ma [1], Gang Wang [1,*], Jurong Yao [2], Dongyan Huang [1], Hewen Tan [1], Honglei Jia [1] and Zhaobo Zou [3]

[1] College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China; zyma20@mails.jlu.edu.cn (Z.M.); huangdy@jlu.edu.cn (D.H.); hwtan20@mails.jlu.edu.cn (H.T.); jiahl@vip.163.com (H.J.)
[2] First Administrative Department of People's Canal, Sichuan Dujiangyan Water Conservancy Development Center, Chengdu 611000, China; 10340043@henu.edu.cn
[3] Changchun Zhongda Tractor Manufacturing Co., Ltd., Changchun 130062, China; zzb1967918@163.com
* Correspondence: gw611004@jlu.edu.cn

Abstract: The accurate spraying of herbicides and intelligent mechanical weeding operations are the main ways to reduce the use of chemical pesticides in fields and achieve sustainable agricultural development, and an important prerequisite for achieving these is to identify field crops and weeds accurately and quickly. To this end, a semantic segmentation model based on an improved U-Net is proposed in this paper to address the issue of efficient and accurate identification of vegetable crops and weeds. First, the simplified visual group geometry 16 (VGG16) network is used as the coding network of the improved model, and then, the input images are continuously and naturally down-sampled using the average pooling layer to create feature maps of various sizes, and these feature maps are laterally integrated from the network into the coding network of the improved model. Then, the number of convolutional layers of the decoding network of the model is cut and the efficient channel attention (ECA) is introduced before the feature fusion of the decoding network, so that the feature maps from the jump connection in the encoding network and the up-sampled feature maps in the decoding network pass through the ECA module together before feature fusion. Finally, the study uses the obtained Chinese cabbage and weed images as a dataset to compare the improved model with the original U-Net model and the current commonly used semantic segmentation models PSPNet and DeepLab V3+. The results show that the mean intersection over union and mean pixel accuracy of the improved model increased in comparison to the original U-Net model by 1.41 and 0.72 percentage points, respectively, to 88.96% and 93.05%, and the processing time of a single image increased by 9.36 percentage points to 64.85 ms. In addition, the improved model in this paper has a more accurate segmentation effect on weeds that are close to and overlap with crops compared to the other three comparison models, which is a necessary condition for accurate spraying and accurate weeding. As a result, the improved model in this paper can offer strong technical support for the development of intelligent spraying robots and intelligent weeding robots.

Keywords: image identification; semantic segmentation; U-Net; ECA; Chinese cabbage; weed

Citation: Ma, Z.; Wang, G.; Yao, J.; Huang, D.; Tan, H.; Jia, H.; Zou, Z. An Improved U-Net Model Based on Multi-Scale Input and Attention Mechanism: Application for Recognition of Chinese Cabbage and Weed. *Sustainability* **2023**, *15*, 5764. https://doi.org/10.3390/su15075764

Academic Editor: Gwanggil Jeon

Received: 15 February 2023
Revised: 13 March 2023
Accepted: 21 March 2023
Published: 26 March 2023

1. Introduction

Weeds growing in fields not only raise the risk of agricultural diseases [1] but also compete with crops for sunlight, water, fertilizers, and other nutrients, which negatively impacts crop growth and yield [2,3]. As a result, timely and effective weed removal has historically been a key area of study. Currently, chemical and mechanical weeding are the two main methods used to manage weeds. Chemical weeding relies heavily on spraying herbicides evenly across the field regardless of the presence of weeds, which not only results in the excessive use of chemical pesticides and brings a series of environmental

pollution problems but also leaves pesticide residues on the surface of crops, endangering human health [4]. While inter-row weeding makes up the majority of mechanical weeding, inter-plant weeding still has to be improved due to the difficulty of control and the high degree of intelligence required [5]. Therefore, to reduce the use of biochemical pesticides and achieve the green and sustainable development of agriculture, it is essential to realize the accurate spraying of chemical herbicides and the intelligent operation of weeding mechanisms, and accurate and quick identification of crops and weeds is a prerequisite for realizing these technologies [6,7].

Numerous machine learning algorithms have been widely used in crop and weed recognition studies, such as the artificial neural network (ANN) [8,9], Naive Bayes model, AdaBoost algorithm [10], decision tree, K-means [11,12], fuzzy K-means [13,14], and support vector machine (SVM) [15,16], among others. However, these methods typically extract features, such as color, morphology, texture, spectrum, and spatial distribution, of crops and weeds in the form of vectors. They then choose the appropriate classifiers for detection and recognition, which are easily affected by light, shading, and noise, and can only achieve good results in specific environments with low robustness, making it difficult to meet the recognition needs of pesticide precision spraying and intelligent weeding operations. The convolutional neural network (CNN) was initially used by AlexNet [17] in 2012 to solve the classification problem of a large-scale image dataset, and the results were outstanding. Since then, CNN has become a research hotspot in image processing, and numerous outstanding algorithms have been developed using CNN as their foundation. In 2014, Simonyan and Zisserman [18] proposed VGG based on CNN, demonstrating that the network performance could be enhanced by increasing the network depth. The name VGG is derived from the acronym of the Visual Geometry Group at the University of Oxford. Since then, CNN has continued to develop deeper, and deeper network models such as GoogLeNet [19] and ResNet [20] have emerged. These algorithms have also been used to investigate the identification of crops and weeds [21–23]. The concept of transfer learning is using knowledge or patterns acquired on one task or domain to another related task or domain [24]. In deep learning, there are some similarities between various challenges. The over-fitting issue can be effectively reduced and the model convergence speed can be increased by transferring the model parameters trained for one job to another. Transfer learning has also been used more in the area of crop and weed identification [25–27].

Target detection and semantic segmentation are most widely used in the recognition of crops and weeds. Target detection can be considered as a combination of two tasks, target localization and classification, i.e., locating the position of an object in an image and identifying the class to which the object belongs [28]. To quickly and accurately recognize different types of weeds in peanut fields, Zhang et al. [29] proposed the EM-YOLOv4-Tiny weed identification model based on the YOLOv4-Tiny target detection model, multiscale detection, and the attention mechanism; Kang et al. [30] proposed a weed detection method in sugar beet based on SSD model based on multi-scale fusion module and feature improvement; Partel et al. [31] proposed three improved target detection models based on the YOLOv3 model and implemented a precision intelligent sprayer for weed recognition with sunflower and weed and pepper and weed as datasets; Peng et al. [32] compared two target detection networks, Faster R-CNN and YOLOv3, and structurally optimized the Faster R-CNN network by introducing a feature pyramid network in the RPN network to generate target candidate frames to achieve the efficient identification of cotton field weeds in complex backgrounds. The first semantic segmentation model, fully convolutional networks (FCN), which segmented images by end-to-end training of convolutional neural networks, was suggested by Long et al. in 2015 [33]. The segmentation accuracy has been significantly improved compared with traditional methods. Semantic segmentation understands images at the pixel level and identifies and classifies each pixel based on the semantic information contained in the image [34], which can identify object locations and boundaries in images more accurately than target detection at the image level and labeled frames. Lottes et al. [35] achieved the semantic segmentation of sugar

beet and weeds using FCN with sequence information; Ma et al. [36] proposed a SegNet semantic segmentation model based on FCN and achieved high classification accuracy in the segmentation of rice seedlings and weeds. Kamath et al. [37] studied semantic segmentation models, such as PSPNet and SegNet, for the recognition of rice crops and weeds, and all obtained good results with over 90% accuracy.

U-Net [38], as a classical variant of the first semantic segmentation model FCN, is named after its overall structure of "U" shape and was originally proposed for medical image segmentation. Compared with FCN, U-Net uses dimensional splicing for feature fusion in the jump connection part, which can retain more feature information and has higher segmentation accuracy, and U-Net outperforms other coding–decoding structure networks for both small target segmentation tasks and small sample datasets. The VGG network uses small convolutional kernels repeatedly to deepen the network. Despite having a straightforward structure, it performs exceptionally well at picture recognition. VGG16 is a typical structure in the VGG network and is frequently used as a feature extraction network for U-Net since it is well-suited for classification and localization tasks. Yu et al. [39] investigated the potential of the U-Net model to segment maize tassel, and the results showed that the segmentation accuracy of the U-Net model with VGG16 as the feature extraction network for tassels at the all the tasseling stages was better than that of U-Net model with MobileNet; Sugirtha et al. [40] also confirmed that U-Net with the VGG16 encoder shows better performance than the ResNet-50 encoder when segmenting urban streets. In order to accomplish the reliable detection of navigation lines in different growth periods of potato, Yang et al. [41] presented a fitting approach of feature midpoint modification and replaced the original U-Net's feature extraction structure with VGG16; Zou et al. [42] proposed an image-enhancement method based on the random synthesis of "foreground" and "background", and reduced the number of convolutional layers in the U-Net model network, achieving the semantic segmentation of field weed images. Qian et al. [43] also used VGG16 to replace the encoder in the original U-Net network and added a repeated criss-cross attention to the U-Net network's skip connection; the experiments showed that the segmentation accuracy indexes of the improved U-Net network were higher than those of other comparative algorithms. In order to achieve the online quality detection of machine-harvested soybean, Jin et al. [44] employed U-Net as the basic network structure, combined with the VGG16 network, and added the convolutional block attention module (CBAM) after the feature maps were extracted in the encoder. Zou et al. [45] pre-trained a decoding network using image segmentation tasks on similar datasets and effectively segmented field wheat crops and weeds based on an improved U-Net model.

Although the aforementioned studies also produced promising findings, most of them increased the depth and width of the network to improve the detection accuracy of the model without considering the number of model parameters, model size, and recognition speed, which are crucial for reducing resource consumption and achieving real-time recognition effects in constrained hardware environments [46]. Therefore, to remedy the above deficiencies, this paper proposes a semantic segmentation model based on the U-Net network. The main contributions of this paper are the following:

(1) A dataset of Chinese cabbage crops and weeds at seedling stage was created;
(2) To accomplish the effective, precise, and quick detection of Chinese cabbage crops and weeds, the U-Net model was enhanced by the lateral integration of multi-scale feature maps and the addition of the efficient channel attention (ECA);
(3) The revised U-Net model put forth in this study can operate in a lower hardware environment configuration than the original U-Net, which reduces memory costs and conserves resources. Additionally, the upgraded model's picture-processing speed is quicker than the original U-Net, better meeting the demands of smart agriculture for the real-time detection of crop and weed;
(4) The proposed model has a more precise segmentation effect on weeds near and overlapping with crops, which can offer a strong technical foundation for the growth of precision agriculture.

Overall, this study proposes a semantic segmentation model that can accurately identify weeds and Chinese cabbage crops, which can offer technical assistance for attaining agricultural sustainable development. The remainder of the article is organized as follows. The dataset needed for model training is included in Section 2 along with detailed explanations of the individual strategies used to enhance the U-Net model. The results of this study are presented and discussed in Section 3. Section 4 summarizes the research results of this paper, points out the limitations of this study, and outlines the future research directions.

2. Materials and Methods

2.1. Image Acquisition

The images needed for the study were taken between 16 August and 18 August 2022 in the Zhanlin Green Agricultural Picking Garden, Changchun City, Jilin Province, China (125°12′33″ E, 43°59′27″ N). The image acquisition site is located in the hinterland of the Northeast China Plain, which has a temperate monsoon climate, where the Chinese cabbage was planted in seedbeds, and transplanted at 4–6 leaves with plant spacing of 40–45 cm and row spacing of 55–60 cm. When the images were taken, the Chinese cabbage was in the seedling stage, 7–10 days after transplanting. The acquisition equipment is shown in Figure 1, and the RGB industrial camera (the camera is produced by Sichuan Weixin Vision Technology Co., Ltd., China, and the specific specifications of the camera are shown in Table 1) was mounted vertically on the mobile trolley with a height of 65 cm above the ground and an imaging area of 65 × 110 cm; the body of the mobile trolley and tires were excluded from the imaging area. The field image was continuously collected when the mobile trolley was moving, and a total of 345 pictures were gathered.

Figure 1. Field image acquisition location and equipment.

Table 1. Performance parameters of RGB industrial camera.

Name	Technical Specifications
Model	SY011HD-V1
Sensor type	Cmos
Sensor size	1/2.7″ inches
Maximum resolution	1920 × 1080 pixels
Signal-to-noise ratio	62 dB
Pixel size	3 μm × 3 μm
Maximum frequency	60 fps
Operating temperature	−20 °C~70 °C
Operating humidity	15~85%

2.2. Image Annotation and Data Enhancement

As seen in Figure 2, this study manually annotated the collected photographs using the image annotation program Labelme (version 3.16.7, relying on Anaconda software for implementation) to create corresponding label files for the sections of the images that

represented the cabbage crops and weeds. As illustrated in Figure 3, this study uses four data enhancement techniques to enlarge the original image dataset to improve the dataset's variety and avoid the model being overfitted owing to a lack of data, specifically as follows: (a) Gaussian noise—the image is augmented with random Gaussian noise with mean value of 0 and variance of 0.05; (b) Random rotation—the image is rotated at random to the left or right, with a maximum angle of 30 degrees for each direction; (c) Random cropping—the cropped image area is 0.7 times the original image, and the cropped image is enlarged to the same size as the original image; (d) Random flip—the image is randomly flipped either horizontally, vertically, or diagonally, with a probability of 1/3 for each flip. Each data improvement method has a probability of 0.5 to be activated, and finally, the original dataset was expanded to four times its original size, with 1104 sheets as the training set and 276 sheets as the test set.

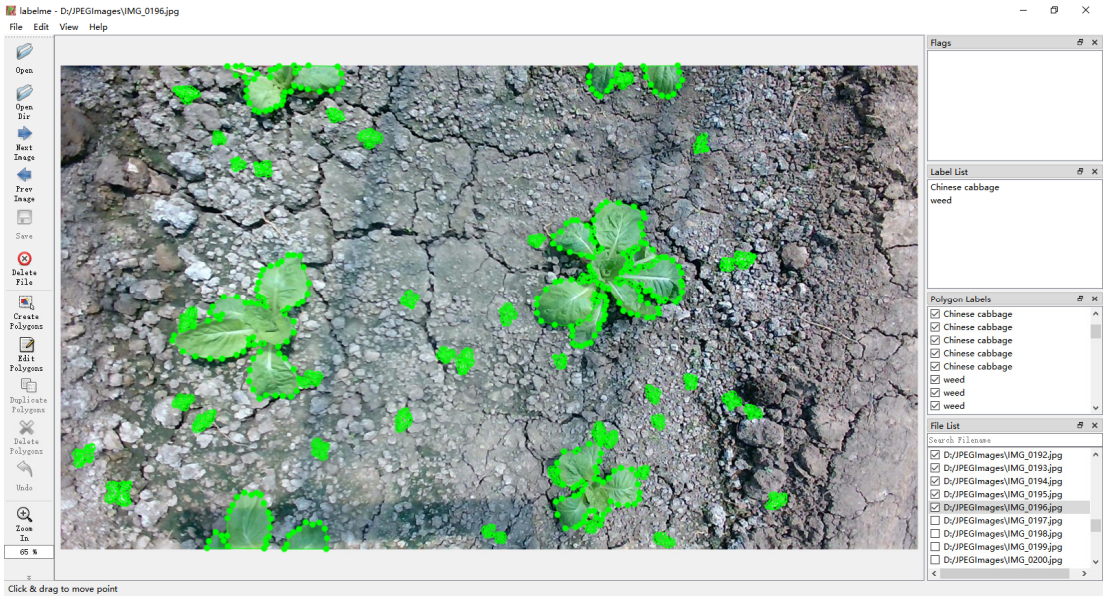

Figure 2. Labelme labeling interface.

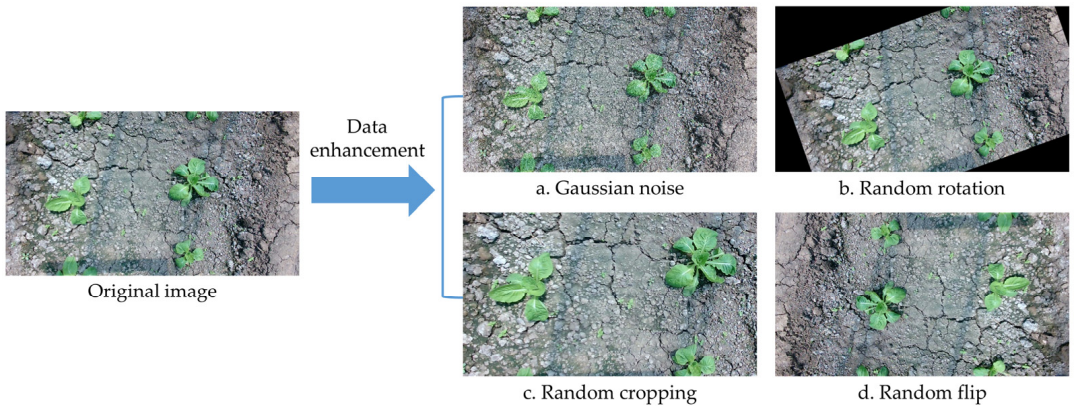

Figure 3. Data enhancement.

2.3. Construction of Semantic Segmentation Model

2.3.1. Multi-Scale Feature Map Input

To make the model retain more details of the feature maps, this paper employs average pooling to continuously down-sample the input images to generate feature maps of various sizes, and input the above feature maps from the network side based on the U-Net network. More specifically, the input RGB three-channel images are continuously pooled using a

pooling kernel of size $f = 2$ and step size $s = 2$, and the pooling principle is shown in Figure 4a. The feature map's height and breadth will decrease to half of their original size after each pooling, as shown in Figure 4a, but the number of channels will remain the same. In this study, the size of the second layer feature map is 256 × 256, three repetitions of averaging pooling are completed, and the size of the final feature map is 64 × 64, as shown in Figure 4b, which constitutes the multi-scale input feature map of this study.

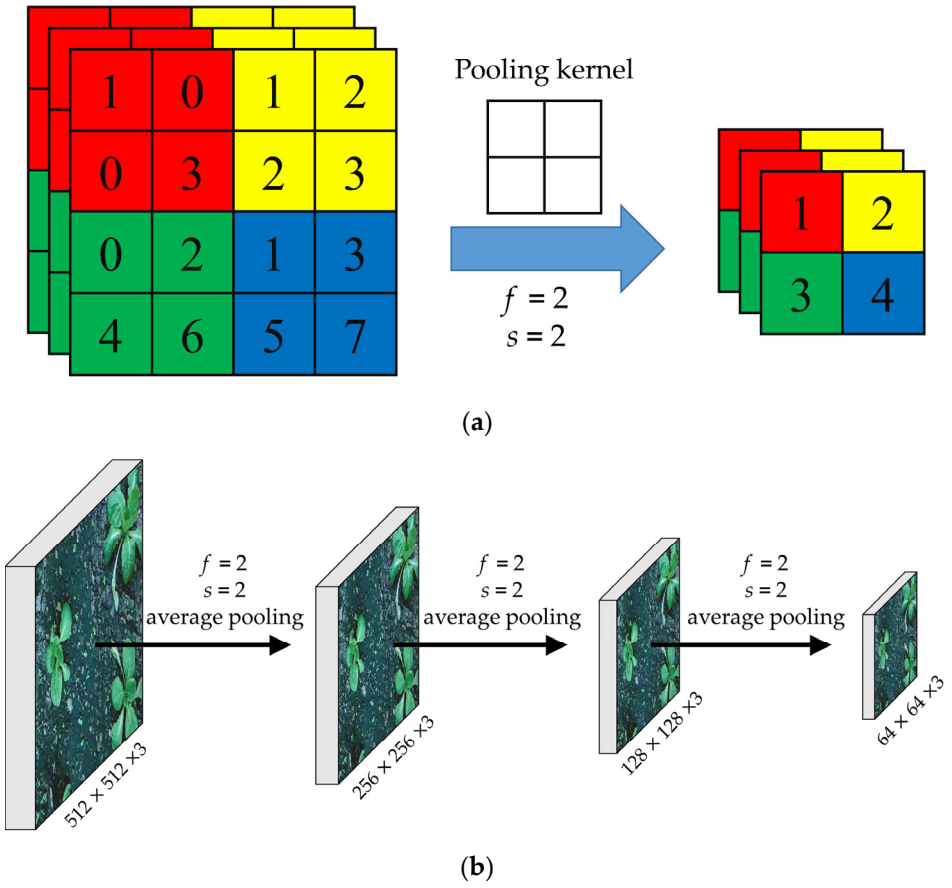

Figure 4. Principle of multi-scale feature map generation: (**a**) Average pooling principle; (**b**) Multi-scale feature map.

2.3.2. Attention Mechanism

This study introduces the efficient channel attention (ECA) [47] mechanism before the feature fusion of the U-Net network, and its precise structure is illustrated in Figure 5. The model can become more concentrated on the extraction of target features according to this mechanism. For the input feature map $\chi \in R^{W \times H \times C}$, the ECA module first aggregates the spatial information of each channel through global average pooling (GAP) to obtain a global description feature of $1 \times 1 \times C$, and then uses a one-dimensional convolution with a kernel size of k to determine the weights of each channel. Finally, the resulting channel weights are multiplied with the corresponding elements of the input feature map to obtain the final output feature map $\widetilde{\chi} \in R^{W \times H \times C}$, which uses one-dimensional convolutional cross-channel interaction instead of fully connected layers to effectively reduce the computational effort and complexity of the model. Furthermore, the width, height, and number of channels of the feature map are left unchanged after the ECA module.

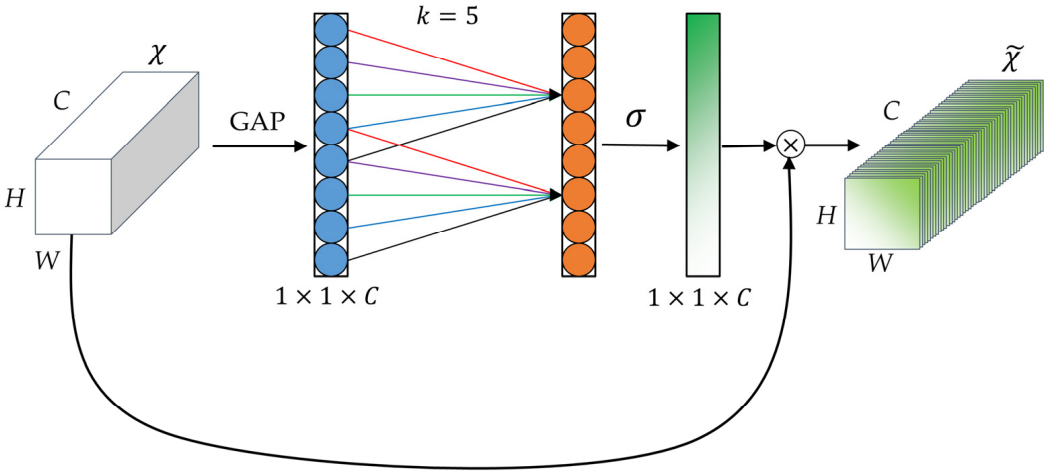

Figure 5. Structure diagram of the ECA.

2.3.3. Overall Structure of the Model

This paper streamlines the 16-layer VGG16 network before introducing it to the network to decrease the number of network parameters and increase the efficiency of network operation. First, the three fully connected layers of the VGG16 network that take up a significant portion of the network's parameters are eliminated, followed by a reduction in the number of convolutional layers of the VGG16 network. As shown in Figure 6, finally, the VGG16 network's layers are reduced to six, with the final convolutional layer's channel count increasing from 512 to 1024, and it is then incorporated into the coding network of the improved model. At the same time, the input feature maps in this study are down-sampled using both average pooling and maximum pooling, and the three feature maps of various sizes that were generated by down-sampling from the 2×2 average pooling layer are input from the network laterally, and feature-fused with the feature maps produced by down-sampling from the 2×2 maximum pooling layer in a dimensional splicing manner. The improved model coding network is as follows. The input RGB image has three channels and the size of 512×512. Initially, it adjusts the number of channels to 64 through two 3×3 convolution layers and extracts the valid information it contains. Next, it changes the image size to 256×256 through 2×2 maximum pooling layers and performs feature fusion with a feature map of the same size from the lateral direction. After the fusion, the image size remains unchanged and the number of channels increases to 67; then, the number of channels is adjusted to 128 by 3×3 convolutional layers again, and the image size is changed to 128×128 by 2×2 maximum pooling layers again, and feature fusion is continued with the same-sized feature map from the lateral direction, and so forth. Finally, the size of the feature map at the end of the coding network of this model is 32×32 and the number of channels is 1024.

The model employs four ECA modules in the decoding network, which is as follows. The decoding network takes the feature map produced at the end of the coding network as the input image, which is first up-sampled through the 2×2 up-sampling layer, increasing the image size to 64×64 and keeping the number of channels constant, and then, feature fusing with the feature map of the same size from the jump connection after passing through the ECA module together. Following the fusion, the image size is maintained, while the number of channels increases to 1536. Next, 3×3 convolutional layers are applied to further adjust the number of channels, and 2×2 up-sampling layers are applied to further increase the image's size. This process is repeated until the image size is changed to 512×512 and the number of channels is 64. Lastly, 1×1 convolution is used to adjust the number of channels of the final feature map produced by the decoding network to the number of categories. Each pixel of the image is then classified, and the number of categories in this study is 3.

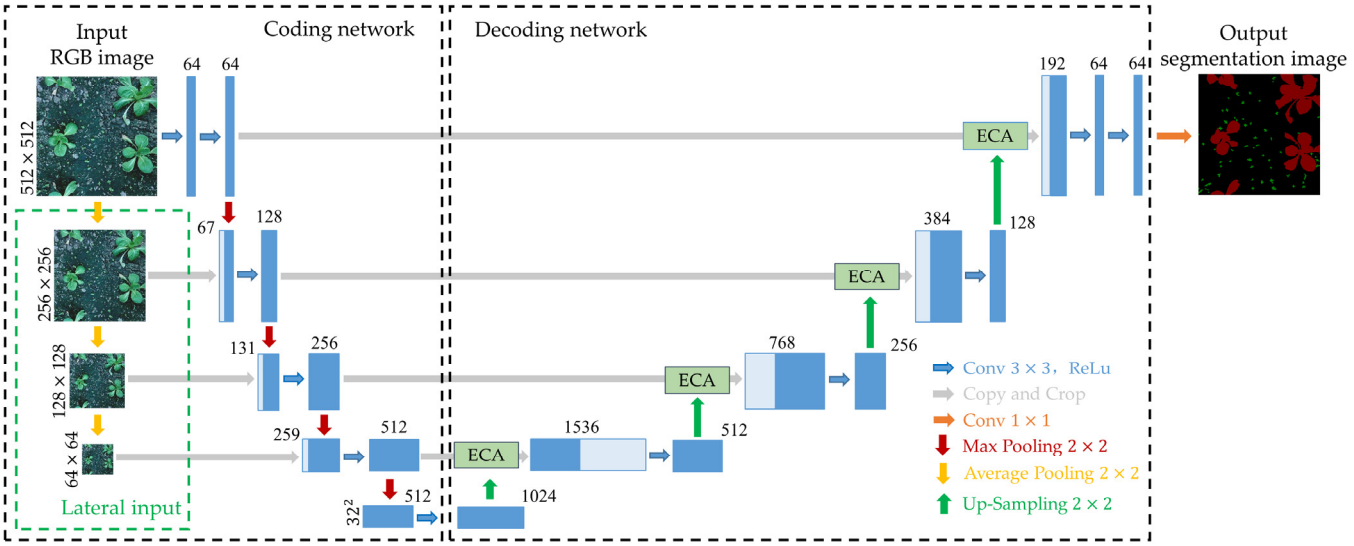

Figure 6. An improved U-Net semantic segmentation model based on multi-scale input and attention mechanism.

2.4. Model Training Environment and Performance Evaluation

The deep learning framework TensorFlow was used for model training and testing. The computing hardware environment is as follows: AMD Ryzen 7 5800X 8-Core Processor, 3.80 GHz Main Frequency, 16 GB RAM; NVIDIA GeForce RTX 3060 Graphics Processor, 12 GB Video Memory. The operating system is Windows 10, together with CUDA 11.3, cuDNN 8.2.1, Python 3.7, and TensorFlow 2.5. The model's starting learning rate is 1×10^{-4}, its learning rate momentum is 0.9, the batch size is set to 4, the size of the input image is set to 512 × 512, and the number of iterations is 300. The "Adam" optimizer is used to optimize the network, which can constantly correct the learning rate to prevent the model from local fitting during the training process.

In this study, the model's performance is assessed in four areas: segmentation accuracy, model parametric number, model size, and segmentation speed. The segmentation speed is measured as the average time the model takes to process a single image. The average of the sum of the intersection and merge ratios between each category's true and projected labels is known as the mean intersection over union (MIOU). Mean pixel accuracy (MPA) is the average of the sum of the percentage of correct predicted pixel values for each category over the total pixel values. Therefore, the segmentation accuracy is indicated by MIOU and MPA, which are computed as follows:

$$MIOU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{FN+FP+TP} \quad (1)$$

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

where k denotes the total number of categories excluding the background category. In this study, we need to distinguish between crops and weeds in addition to the background; therefore, $k = 2$; TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

3. Results and Discussion

3.1. Ablation Experiment

An ablation experiment was carried out in this study to examine the contribution of the VGG16+Cutting, multi-scale input, and ECA module to enhance U-Net. The results

are displayed in Table 2. The VGG16+Cutting means employing the simplified VGG16 as the coding network of U-Net and cutting the number of the convolutional layers of the decoding network of U-Net, and the details of this approach can be seen in Figure 6.

Table 2. Comparison of the effect of adding each module.

Model	VGG16 + Cutting	Multi-Scale Input	ECA	MIOU	Single Image Time Consuming/ms	Model Parameters
U-Net	×	×	×	87.55	71.55	31,379,075
Optimization 1	√	×	×	86.42	57.70	15,745,923
Optimization 2	×	√	×	87.96	72.81	31,403,267
Optimization 3	×	×	√	89.18	76.63	31,379,113

The addition of the VGG16+Cutting module, as can be observed in Table 2, reduces the MIOU of the model by 1.13% but also decreases the model parameters by 49.82% and the single-image time consumption by 13.85 milliseconds. In order to make the model lighter and better suited for real-time detection, we believe that a minor loss of accuracy is worthwhile. The MIOU of the model is increased with the addition of the multi-scale input module by 0.41%, but only at the expense of an increase in single-image time consumption of 1.26 milliseconds and an increase of 0.08% in model parameters. This is due to the fact that the multi-scale input module can increase the input image's number of channels to retain more information, whereas the number of channels of the image in this study was only increased briefly during the image feature fusion to avoid the model parametric number surge, and then, the number of channels was immediately restored to the original U-Net network with 3×3 convolutional layers. Contrarily, although refs. [35,46,48] also enhanced the model by boosting the number of image channels to achieve better segmentation, these enhancements were made by directly fusing RGB and NIR images to create a four-channel image input into the network, and this method would significantly increase the number of model parameters.

Furthermore, the MIOU of the U-Net model increased by 1.63 percentage points when the ECA module was included, showing that the ECA module can significantly improve the model's segmentation accuracy. The attention gate (AG) module, squeeze and excitation (SE) module, and convolutional block attention module (CBAM) were added to the U-Net model by John et al. [49], Yu et al. [50], and Jin et al. [44], respectively. Although the addition of these attention mechanism modules improves the segmentation accuracy of the model, it also introduces many new model parameters and increases the complexity of the network. In contrast, the ECA module used in this work is a lightweight module, and it can be seen in Table 2 that the number of model parameters is only slightly increased after the ECA module is added.

3.2. Comparison of the Overall Accuracy of the Model

The change curves of the mean intersection over union and mean pixel accuracy on the training set of the improved model in this paper and the original U-Net model as well as the current widely used semantic segmentation models PSPNet and DeepLab V3+ [51] are shown in Figure 7. The computational results are shown in Table 3, and the improved model in this paper is MSECA-Unet.

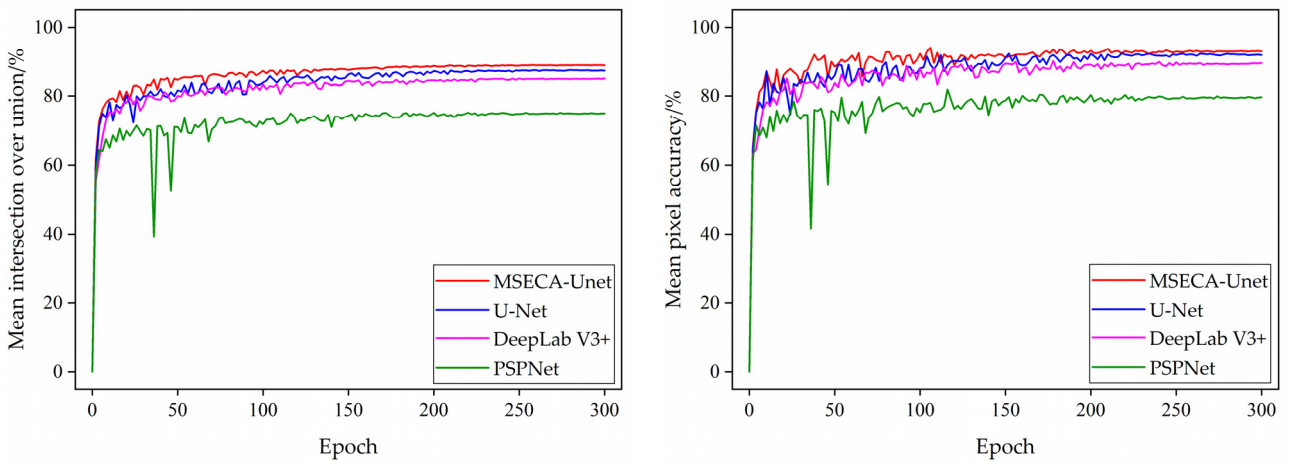

Figure 7. Comparison of segmentation accuracy of each model in the training set: (**a**) Variation curve of mean intersection over union; (**b**) Variation curve of mean pixel accuracy.

Table 3. Comparison of model segmentation accuracy.

Model	Mean Intersection over Union/%	Mean Pixel Accuracy/%
PSPNet	74.90	79.60
DeepLabV3+	85.00	89.49
U-Net	87.38	91.95
MSECA-Unet	88.95	93.02

In contrast to PSPNet, DeepLab V3+ and the original U-Net model, the MIOU and MPA of the MSECA-Unet model on the training set are higher, as shown in Figure 7. Additionally, as can be seen in Figure 7a, the improved MSECA-Unet model converged after 130 iterations, stabilized near the highest value earlier, and did so significantly more quickly than the other three comparison models. This is because, in this paper, the ECA module, which can successfully prevent the activation of irrelevant information and noise in the network, is introduced before the fusion of features in the U-Net network, so that it only fuses the feature information that requires attention, which decreases the time loss in feature fusion, and hence, quickens the model's convergence, which is consistent with the conclusions reached by Zhang et al. [29] when introducing the ECA module into the YOLOv4-Tiny network, and by Zhao et al. [52] when introducing the ECA module into DenseNet network.

As shown in Table 3, the improved MSECA-Unet model's MIOU is 88.95% and the MPA is 93.02% on the training set, which is higher than the 87.38% and 91.95% of the original U-Net model, and also higher than the corresponding indexes of the other two commonly used semantic segmentation models, which indicates that the improved MSECA-Unet network in this paper significantly improves the model's segmentation accuracy, and the MSECA-Unet model has a better segmentation effect on the Chinese cabbage and weed training set compared with the U-Net, PSPNet, and DeepLab V3+ models.

The MSECA-Unet model, as well as the U-Net, PSPNet, and DeepLab V3+ models, are also assessed on the test set in this work. The prediction results are displayed in Table 4, whereas Table 5 shows the number of model parameters, model size, and prediction speed; Table 6 shows the model accuracy, precision, and F1-score.

Table 4. Comparison of model prediction results.

Model	Intersection over Union/%				Pixel Accuracy/%			
	Background	Weed	Crop	MIOU	Background	Weed	Crop	MPA
PSPNet	98.31	38.67	87.81	74.93	99.26	45.89	93.85	79.67
DeepLabV3+	99.02	63.43	92.71	85.06	99.58	73.25	95.98	89.60
U-Net	99.16	69.87	93.62	87.55	99.58	80.58	96.84	92.33
MSECA-Unet	99.24	73.62	94.02	88.96	99.64	82.58	96.92	93.05

Table 5. Comparison of model parameters, model size, and single-image time consumption.

Model	Model Parameters	Model Size/MB	Single-Image Time Consumption/ms
PSPNet	4.91×10^7	178.85	67.48
DeepLab V3+	4.13×10^7	158.42	76.32
U-Net	3.14×10^7	119.77	71.55
MSECA-Unet	1.58×10^7	60.27	64.85

Table 6. Comparison of model accuracy, precision, and F1-score.

Model	Accuracy/%	Precision/%	F1-Score/%
PSPNet	93.84	87.76	83.52
DeepLab V3+	97.19	92.82	91.18
U-Net	97.84	93.39	92.86
MSECA-Unet	98.24	94.56	93.80

As can be seen in Table 4, the intersections over union and pixel accuracy of all models for weed segmentation are much lower than their corresponding metrics for background and crop segmentation, which is due to the high density and small area of weeds in the dataset collected in this study, which possess greater segmentation difficulty compared to background and crop with large areas and small numbers. In addition, Table 4 shows that for background, weeds, and crops, the proposed MSECA-Unet model in this paper produced the best results in terms of the intersection over union and pixel accuracy with 99.24%, 73.62%, 94.02%, and 99.64%, 82.58%, and 93.05%, respectively, as opposed to the original U-Net model with 99.16%, 69.87%, 93.62%, and 99.58%, 80.58%, 96.84%, which are increased by 0.08%, 3.75%, 0.40% and 0.06%, 2.00% and 0.08%, respectively. Thus, it can be seen, in addition to having a higher intersection over union and pixel accuracy than the original U-Net model for all categories in this study, the MSECA-Unet model also significantly increased these metrics for weeds, the hardest category to segment, which strongly supports the efficacy of the improvements made in this paper.

In Tables 4 and 5, we can see that the MIOU of the original U-Net model is 87.55% and the MPA is 92.33%, while the MIOU of the MSECA-Unet model proposed in this paper is 88.96% and the MPA is 93.05%, which are improved by 1.41 and 0.72 percentage points, respectively. This is due to the fact that the original U-Net model down-samples the feature map four times in order to obtain deeper feature information, which causes the network to lose a lot of detailed information that cannot be recovered by the subsequent up-sampling operation, and affects the segmentation accuracy of the network. While this study incorporates the multi-scale feature map produced by average pooling into the network, which effectively addresses the aforementioned information loss issue and boosts the model's segmentation accuracy. Meanwhile, the original U-Net model uses jump connections to combine the spatial data from the up-sampled paths with the spatial data from the down-sampled paths. However, this brings many redundant underlying features and noise, which affect the segmentation accuracy and speed of the network. In this paper, the ECA module is introduced before the network feature fusion. Increasing the

target feature weight and reducing the weight of the useless or small-effect features make the model focus more on the target feature extraction and improve the model's feature extraction efficiency and accuracy.

Additionally, the proposed MSECA-Unet model has 1.58×10^7 model parameters and a model size of 60.27 MB, which are both 49.68% less than the original U-Net model's 3.14×10^7 and 119.77 MB. Moreover, the proposed MSECA-Unet model's single-image time consumption is 64.85 ms, which is 9.36% faster than the original U-Net model's 71.55 ms. This indicates that the proposed MSECA-Unet model has a faster segmentation speed than the original U-Net model, and that it is more capable of meeting the requirements of real-time crop and weed detection. This is because the model coding network is simplified according to the simple features of cabbage and weed in the images. The simplified coding network can maintain the same image feature extraction capability while consuming fewer computational resources. Meanwhile, this study also simplifies the model decoding network by reducing the number of convolutional layers, which is due to the fact that the images in this study are not complex and the decoding network does not need more abstract features. The reduction in the number of convolutional layers of the coding and decoding networks causes a decrease in the number of model parameters and model size, and speeds up the segmentation of the model. In contrast, refs. [39,40,43] directly use the VGG16 network as the encoding network for U-Net without simplifying VGG16, which also achieves better segmentation results but increases the width and depth of the network and requires a more optimal environment configuration to run the model. Chen et al. [53] achieved the accurate segmentation of grains, branches, and straws in hybrid rice grain images by improving the U-Net model, but the improvement they made was still to make the model extract richer semantic information by increasing the depth of the model. The advancements made in this work, however, strive to obtain the largest gain effect with the fewest possible factors. The ECA module introduced before the network feature fusion is a lightweight module, which has fewer parameters, and also when integrating the multi-scale feature maps into the backbone feature extraction network, it is chosen to integrate from the lateral direction, which effectively avoids the significant growth of the network parameters. Due to these advancements, the model can have fewer model parameters and a smaller model size while still preserving the segmentation effect, and the smaller number of model parameters and model sizes allow the model to run in a relatively low hardware environment configuration, reducing memory costs and saving resource consumption.

In addition, the MSECA-Unet model proposed in this paper also significantly outperforms the current semantic segmentation models DeepLab V3+ and PSPNet. The MSECA-Unet model's MIOU and MPA are improved by 3.90% and 3.45%, respectively, over DeepLab V3+, while the number of model parameters, model size, and single-picture time consumption are decreased by 61.74%, 61.96%, and 15.03%, respectively. In comparison to PSPNet, the MIOU and MPA of the MSECA-Unet model are increased by 14.03% and 13.38%, and the number of model parameters, model size, and single-image time consumption are reduced by 67.82%, 66.30%, and 3.90%, respectively. In summary, the segmentation speed (single-image time consumption) of the proposed MSECA-Unet model is significantly faster than the other three semantic segmentation models, and its segmentation accuracy (MIOU and MPA) is also significantly improved with a significant reduction in the number of model parameters and model size, indicating that the proposed model is more suitable for application in the recognition of Chinese cabbage crops and weeds.

As can be seen in Table 6, the MSECA-Unet model proposed in this paper has the best accuracy, precision, and F1-score compared with U-Net, DeepLab V3+ and PSPNet. The accuracy, precision, and F1-score of the MSECA-Unet model each increased by 0.4%, 1.17%, and 0.94%, respectively, when compared to U-Net. In order to make the model more lightweight and improve the segmentation speed of the model, references [42,45] decreased the U-Net model's convolutional layer count in a manner similar to this study. Despite the fact that the segmentation speed of the improved model for farmland weeds was considerably increased, the reduction in a significant number of model parameters

resulted in a decrease in model precision, and later, other improvements of the model were unable to make up for this loss. The MSECA-Unet model's accuracy, precision, and F1-score increased in comparison to DeepLab V3+ by 1.05%, 1.74%, and 2.62%, respectively; in comparison to PSPNet, they increased by 4.4, 6.8, and 10.28 percentage points, respectively.

3.3. Comparison of Model Segmentation Effects

Randomly selected images in the test set are used as sample images to obtain their segmentation effects on each model. In order to observe the segmentation effect more clearly, the original image was fused with the predicted label image after reducing the transparency and the segmentation effect of each model was displayed in Figure 8. To facilitate the observation of the differences in segmentation effects between different models, certain regions in the segmentation effect map were locally enlarged and the weeds in the locally enlarged map were numbered in the labelled image, as shown in Figure 9. The Chinese cabbage crop is presented in red in the figure, and the weed is presented in green.

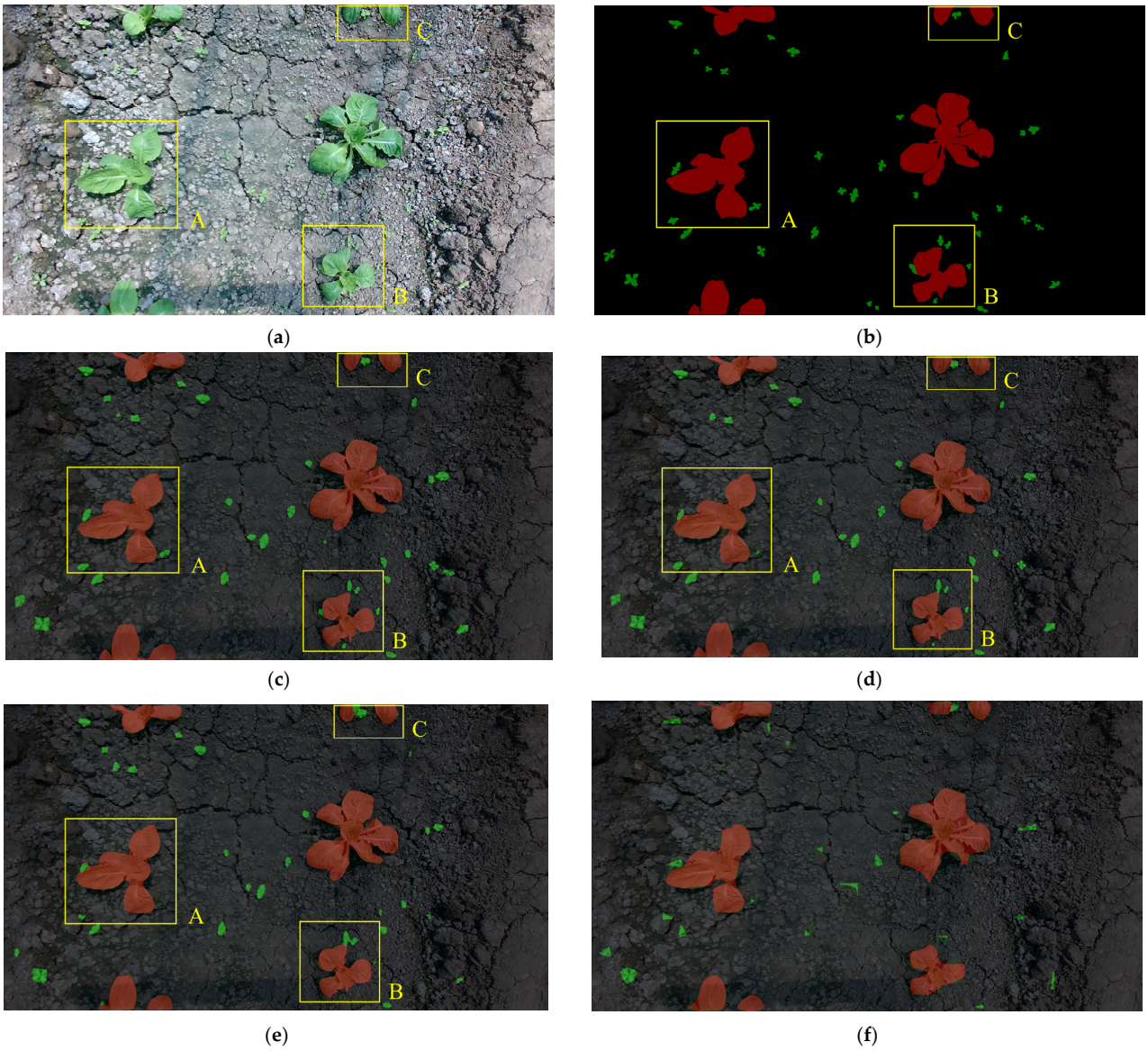

Figure 8. Segmentation results of different models: (**a**) Original image; (**b**) Image label; (**c**) Segmentation result of MSECA-Unet; (**d**) Segmentation result of U-Net; (**e**) Segmentation result of DeepLab V3+; (**f**) Segmentation result of PSPNet. In the figure, the A, B and C are the areas to be enlarged in order to facilitate comparison of the segmentation effect of each model.

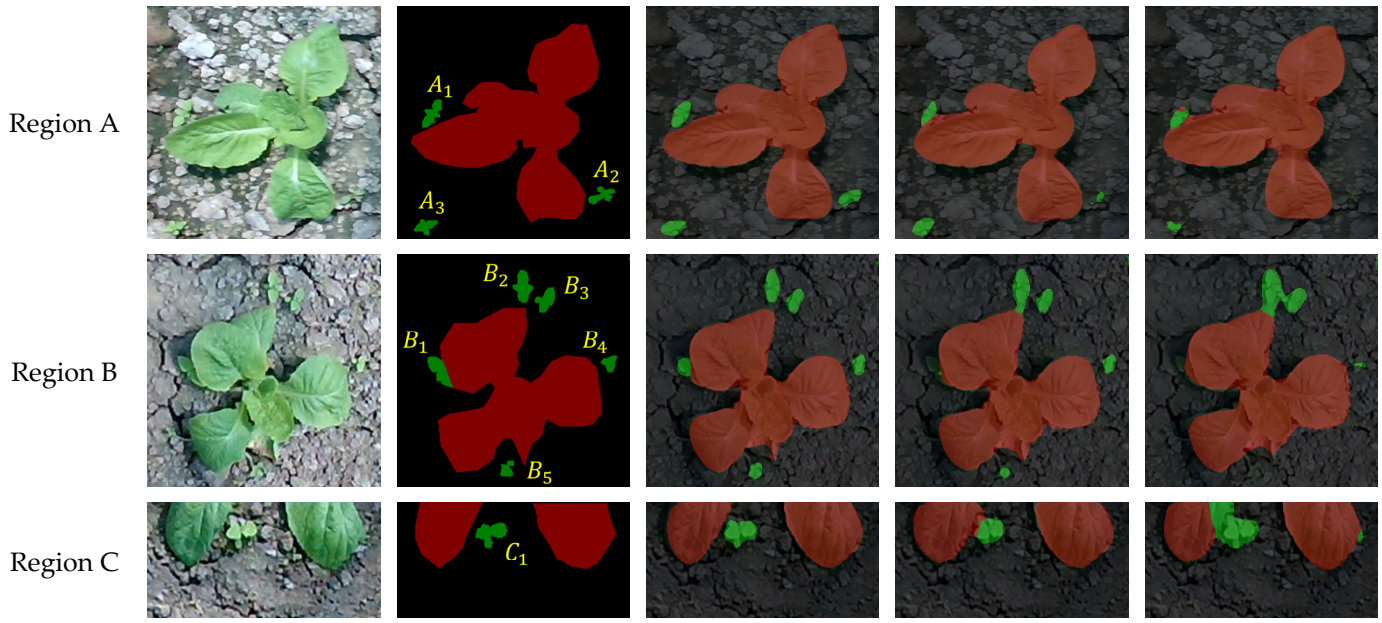

Figure 9. Comparison of segmentation results for regions A, B, and C of different models: (**a**) Original image; (**b**) Image label; (**c**) MSECA-Unet model; (**d**) U-Net model; (**e**) DeepLab V3+ model. In the figure, A_1–A_3, B_1–B_5 and C_1 are the numbers of each weed in order to facilitate comparison of the segmentation effect of each model for each weed.

According to Figure 8, the MSECA-Unet model that was suggested in this study has the optimal segmentation effect and its segmentation effect is most similar to the labeled picture. In contrast, the segmentation effect of the PSPNet model is the least satisfying. In Figure 8f, it is obvious that the segmentation area of the Chinese cabbage crop by the PSPNet model has deviated seriously from the original area of the image, and the mis-segmentation and under-segmentation of weeds in the image are serious, making it impossible to correctly identify weeds.

The MSECA-Unet model has the best segmentation impact on weeds A_2, A_3, B_4, and B_5, according to the images of regions A, B, and C in Figure 9, while the DeepLab V3+ model has the lowest segmentation effect, segmenting weeds A_3 and B_4 partially, and failing to segment weeds A_2 and B_5. Additionally, for weeds A_1, B_2, and C_1, which are close to the crop, the MSECA-Unet model can accurately segment the gap between them and the crop, while the U-Net and DeepLab V3+ models have mis-segmentation issues when segmenting weeds A_1, B_2, and C_1, which incorrectly segment the background as crop or weed, causing the crop and weed prediction labels to be mixed together directly without segmenting the gaps between them. Moreover, the DeepLab V3+ model had the worst segmentation effect, which not only mixed weed B_2 with the crop, but also mixed weed B_3 at the same time. Additionally, the MSECA-Unet model had the best segmentation of weed B_1, which overlapped with the crop. While the U-Net and DeepLab V3+ models under-segmented weed B_1 severely, the U-Net model only segmented a very tiny region, and the DeepLab V3+ model did not segment it at all.

In summary, compared with U-Net and DeepLab V3+ models, the MSECA-Unet model has the best performance, which can not only accurately segment the weeds overlapping with crops, but also has the most accurate segmentation effect on the gap between crops and weeds, and the accurate segmentation of weeds close to crops and overlapping with crops is an important prerequisite for accurate spraying and accurate weed control.

4. Conclusions

To solve the problem of the efficient and accurate identification of vegetables and weeds in the field, and to realize the accurate spraying of herbicides and intelligent weeding operations, a semantic segmentation model, MSECA-Unet, based on an improved U-Net is proposed in this paper, which improves its segmentation accuracy and achieves efficient, accurate, and quick identification of Chinese cabbage crops and weeds by laterally integrating multi-scale inputs and introducing the efficient channel attention (ECA) mechanism with a substantial reduction in the number of model parameters and model size.

The suggested MSECA-Unet model outperformed the currently popular semantic segmentation models PSPNet and DeepLab V3+, as well as the original U-Net model with MIOU and MPA values of 88.96% and 93.05%, respectively, on the dataset for Chinese cabbage and weed. They improved by 1.41 and 0.72 percentage points, respectively, in comparison to the MIOU and MPA of the original U-Net model, which supported the efficacy of the model in this work. Moreover, the proposed MSECA-Unet model has a model parameter number of 1.58×10^7 and a model size of 60.27 MB, both of which are decreased by 49.68 percentage points when compared to the original U-Net model. This suggests that the model can operate in a lower hardware environment configuration, reducing memory costs and saving resource consumption. The MSECA-Unet model better satisfies the requirements of real-time crop and weed detection, consuming 64.85 ms for a single image, which is 9.36 percentage points less than the original U-Net model. This further confirms the usefulness of the model in this study.

Finally, by comparing the segmentation effects of the test set images on various models, it can be seen that the proposed MSECA-Unet model has more accurate segmentation effects on weeds close to and overlapping with the crop than the other three models, which is a necessary prerequisite for accurate spraying and accurate weeding. As a result, the proposed MSECA-Unet model can provide strong technical support for the development of intelligent spraying robots and intelligent weeding robots.

The MSECA-Unet model proposed in this paper is lightweight, and the fast recognition speed is its advantage, but it also has some limitations. This model only identifies weeds, but does not classify the species of them. Therefore, the model is unable to select the corresponding herbicide according to the type of weed when guiding the intelligent spraying robot to spray accurately. In addition, the model is poorly adaptable and needs to be retrained on a new dataset when used in other crop fields, and cannot be applied to multiple crops at the same time. Therefore, for future work, we will consider the further classification of weed species and expand the dataset for model training to include more crop species and weeds in the dataset to develop a more adaptable model that can be adapted to different crops and weeds.

Author Contributions: Conceptualization, Z.M. and G.W.; methodology, Z.M. and G.W.; software, Z.M., J.Y. and D.H.; validation, J.Y., D.H. and H.T.; formal analysis, H.T. and H.J.; investigation, Z.Z.; resources, H.J. and Z.Z.; data curation, D.H. and H.T.; writing—original draft preparation, Z.M. and G.W.; writing—review and editing, Z.M. and G.W.; visualization, Z.M., J.Y. and D.H.; supervision, H.J. and Z.Z.; project administration, G.W. and H.J.; funding acquisition, G.W. and H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Research Project of the Jilin Provincial Education Department (grant number: JJKH20221022KJ) and the Science and Technology Development Project of Jilin Province (grant number: 20220203081SF, 20220508113RC, 20210202019NC).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The location for the image acquisition was kindly provided by Jilin Province Zhanlin Green Agriculture Technology Co., Ltd.

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

References

1. Zhao, A. Analysis of the danger of weeds in agricultural fields and their classification. *Agric. Technol.* **2013**, *33*, 140.
2. Hamuda, E.; Glavin, M.; Jones, E. A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* **2016**, *125*, 184–199. [CrossRef]
3. Wang, A.; Zhang, W.; Wei, X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* **2019**, *158*, 226–240. [CrossRef]
4. Qi, Y.; Li, J.; Yan, B.; Deng, Z.; Fu, G. Impact of herbicides on wild plant diversity in agro-ecosystems: A review. *Biodivers. Sci.* **2016**, *24*, 228–236. [CrossRef]
5. Chen, Z.; Zhang, C.; Li, N.; Sun, Z.; Li, W.; Zhang, B. Study review and analysis of high performance intra-row weeding robot. *Trans. CSAE* **2015**, *31*, 1–8. [CrossRef]
6. Xing, Q.; Ding, S.; Xue, X.; Cui, L.; Le, F.; Li, Y. Research on the development status of intelligent field weeding robot. *J. Chin. Agric. Mech.* **2022**, *43*, 173–181.
7. Ma, X.; Qi, L.; Liang, B.; Tan, Z.; Zuo, Y. Present status and prospects of mechanical weeding equipment and technology in paddy field. *Trans. CSAE* **2011**, *27*, 162–168. [CrossRef]
8. Bakhshipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.* **2018**, *145*, 153–160. [CrossRef]
9. Shah, T.M.; Nasika, D.P.B.; Otterpohl, R. Plant and weed identifier robot as an agroecological tool using artificial neural networks for image identification. *Agriculture* **2021**, *11*, 222. [CrossRef]
10. Xu, K.; Li, H.; Cao, W.; Zhu, Y.; Chen, R.; Ni, J. Recognition of weeds in wheat fields based on the fusion of RGB images and depth images. *IEEE. Access* **2020**, *8*, 110362–110370. [CrossRef]
11. Tang, J.; Wang, D.; Zhang, Z.; He, L.; Xin, J.; Xu, Y. Weed identification based on K-means feature learning combined with convolutional neural network. *Comput. Electron. Agric.* **2017**, *135*, 63–70. [CrossRef]
12. Tang, J.; Zhang, Z.; Wang, D.; Xin, J.; He, L. Research on weeds identification based on K-means feature learning. *Soft Comput.* **2018**, *22*, 7649–7658. [CrossRef]
13. Tellaeche, A.; Burgos-Artizzu, X.P.; Pajares, G.; Ribeiro, A. On combining support vector machines and fuzzy K-means in vision-based precision agriculture. *Int. J. Comput. Inf. Eng.* **2007**, *1*, 844–849.
14. Yang, S.; Hou, M.; Li, S. Three-Dimensional Point Cloud Semantic Segmentation for Cultural Heritage: A Comprehensive Review. *Remote Sens.* **2023**, *15*, 548. [CrossRef]
15. Wang, C.; Li, Z. Weed recognition using SVM model with fusion height and monocular image features. *Trans. CSAE* **2016**, *32*, 165–174.
16. Zheng, Y.; Zhu, Q.; Huang, M.; Guo, Y.; Qin, J. Maize and weed classification using color indices with support vector data description in outdoor fields. *Comput. Electron. Agric.* **2017**, *141*, 215–222. [CrossRef]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
18. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
19. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Rabinovich, A. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842.
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
21. De Camargo, T.; Schirrmann, M.; Landwehr, N.; Dammer, K.-H.; Pflflanz, M. Optimized Deep Learning Model as a Basis for Fast UAV Mapping of Weed Species in Winter Wheat Crops. *Remote Sens.* **2021**, *13*, 1704. [CrossRef]
22. Teimouri, N.; Dyrmann, M.; Nielsen, P.; Mathiassen, S.; Somerville, G.; Jørgensen, R. Weed Growth Stage Estimator Using Deep Convolutional Neural Networks. *Sensors* **2018**, *18*, 1580. [CrossRef] [PubMed]
23. Dos Santos Ferreira, A.; Freitas, D.M.; Da Silva, G.G.; Pistori, H.; Folhes, M.T. Weed detection in soybean crops using ConvNets. *Comput. Electron. Agric.* **2017**, *143*, 314–324. [CrossRef]
24. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl. Based Syst.* **2015**, *80*, 14–23. [CrossRef]
25. Suh, H.K.; Ijsselmuiden, J.; Hofstee, J.W.; van Henten, E.J. Transfer learning for the classification of sugar beet and volunteer potato under fifield conditions. *Biosyst. Eng.* **2018**, *174*, 50–65. [CrossRef]
26. Bosilj, P.; Aptoula, E.; Duckett, T.; Cielniak, G. Transfer Learning between Crop Types for Semantic Segmentation of Crops versus Weeds in Precision Agriculture. *J. Field Robot* **2020**, *37*, 7–19. [CrossRef]
27. Naushad, R.; Kaur, T.; Ghaderpour, E. Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study. *Sensors* **2021**, *21*, 8083. [CrossRef]
28. Cao, J.; Li, Y.; Sun, H.; Xie, J.; Huang, K.; Pang, Y. A survey on deep learning based visual object detection. *J. Image Graph.* **2022**, *27*, 1697–1722.
29. Zhang, H.; Wang, Z.; Guo, Y.; Ma, Y.; Cao, W.; Chen, D.; Yang, S.; Gao, R. Weed Detection in Peanut Fields Based on Machine Vision. *Agriculture* **2022**, *12*, 1541. [CrossRef]

30. Kang, J.; Liu, G.; Guo, G. Weed detection based on multi-scale fusion module and feature enhancement. *Trans. CSAM* **2022**, *53*, 254–260.
31. Partel, V.; Kakarla, C.; Ampatzidis, Y. Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. *Comput. Electron. Agric.* **2019**, *157*, 339–350. [CrossRef]
32. Peng, M.; Xia, J.; Peng, H. Efficient recognition of cotton and weed in field based on Faster R-CNN by integrating FPN. *Trans. CSAE* **2019**, *35*, 202–209. [CrossRef]
33. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
34. Hui, X.; Zhu, Y.; Zhen, T.; Li, Z. Survey of image semantic segmentation methods based on deep neural network. *J. Front. Comput. Sci. Technol.* **2021**, *15*, 47–59.
35. Lottes, P.; Behley, J.; Milioto, A.; Stachniss, C. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2870–2877. [CrossRef]
36. Ma, X.; Deng, X.; Qi, L.; Jiang, Y.; Li, H.; Wang, Y.; Xing, X. Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PLoS ONE* **2019**, *14*, e0215676. [CrossRef] [PubMed]
37. Kamath, R.; Balachandra, M.; Vardhan, A.; Maheshwari, U. Classification of paddy crop and weeds using semantic segmentation. *Cogent Eng.* **2022**, *9*, 2018791. [CrossRef]
38. Olaf, R.; Philipp, F.; Thomas, B. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, Munich, Germany, 5–9 October 2015.
39. Yu, X.; Yin, D.; Nie, C.; Ming, B.; Xu, H.; Liu, Y.; Bai, Y.; Shao, M.; Cheng, M.; Liu, Y.; et al. Maize tassel area dynamic monitoring based on near-ground and UAV RGB images by U-Net model. *Comput. Electron. Agric.* **2022**, *203*, 107477. [CrossRef]
40. Sugirtha, T.; Sridevi, M. Semantic Segmentation using Modified U-Net for Autonomous Driving. In Proceedings of the 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 1–4 June 2022. [CrossRef]
41. Yang, R.; Zhai, Y.; Zhang, J.; Zhang, H.; Tian, G.; Zhang, J.; Huang, P.; Li, L. Potato Visual Navigation Line Detection Based on Deep Learning and Feature Midpoint Adaptation. *Agriculture* **2022**, *12*, 1363. [CrossRef]
42. 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]
43. Qian, C.; Liu, H.; Du, T.; Sun, S.; Liu, W.; Zhang, R. An improved U-Net network-based quantitative analysis of melon fruit phenotypic characteristics. *J. Food Meas. Charact.* **2022**, *16*, 4198–4207. [CrossRef]
44. Jin, C.; Liu, S.; Chen, M.; Yang, T.; Xu, J. Online quality detection of machine-harvested soybean based on improved U-Net network. *Trans. CSAE* **2022**, *38*, 70–80. [CrossRef]
45. Zou, K.; Liao, Q.; Zhang, F.; Che, X.; Zhang, C. A segmentation network for smart weed management in wheat fields. *Comput. Electron. Agric.* **2022**, *202*, 107303. [CrossRef]
46. Sun, J.; Tan, W.; Wu, X.; Shen, J.; Lu, B.; Dai, C. Real-time recognition of sugar beet and weeds in complex backgrounds using multi-channel depth-wise separable convolution model. *Trans. CSAE* **2019**, *35*, 184–190. [CrossRef]
47. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Montreal, QC, Canada, 13–19 June 2020.
48. Lottes, P.; Hörferlin, M.; Sander, S.; Müter, M.; Schulze, P.; Stachniss, L.C. An effective classification system for separating sugar beets and weeds for precision farming applications. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 5157–5163.
49. John, D.; Zhang, C. An attention-based U-Net for detecting deforestation within satellite sensor imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102685. [CrossRef]
50. Yu, H.; Men, Z.; Bi, C.; Liu, H. Research on Field Soybean Weed Identification Based on an Improved U-Net Model Combined With a Channel Attention Mechanism. *Front. Plant Sci.* **2022**, *13*, 1881. [CrossRef]
51. Zhao, X.; Yao, Q.; Zhao, J.; Jin, Z.; Feng, Y. Image semantic segmentation based on fully convolutional neural network. *Comput. Eng. Appl.* **2022**, *58*, 45–57.
52. Zhao, H.; Cao, Y.; Yue, Y.; Wang, H. Field weed recognition based on improved DenseNet. *Trans. CSAE* **2021**, *37*, 136–142. [CrossRef]
53. Chen, J.; Han, M.; Lian, Y.; Zhang, S. Segmentation of impurity rice grain images based on U-Net model. *Trans. CSAE* **2020**, *36*, 174–180. [CrossRef]