Research Article

Research on a Target Detection Method for Alternanthera Philoxeroides in the Rice Seedling Stage Based on Single-Shot Multibox Detector

Xiangwu Deng [ID],[1] **Zhuwen Liu,**[1] **Kunsong Gong,**[1] **Song Liang,**[1] **Guangjun Qiu** [ID],[2,3] **and Long Qi** [ID][4]

[1]*College of Electronic Information Engineering, Guangdong University of Petrochemical Technology, Maoming 525000, China*
[2]*Institute of Facility Agriculture of Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China*
[3]*Life Science and Technology School, Lingnan Normal University, Zhanjiang 524048, China*
[4]*College of Engineering, South China Agricultural University, Guangzhou 510000, China*

Correspondence should be addressed to Xiangwu Deng; dengxiangwu123456@163.com

Alternanthera philoxeroides, an invasive alien malignant weed, competes with rice for water, fertilizer, light, and growth space before seedling closure stages, which commonly stresses the growth of rice. Chemical herbicides are mainly used to control weeds. However, excessive use of chemical herbicides could lead to serious environmental pollution. With the rapid development of artificial intelligence and deep-learning techniques, it is possible to reduce the use of chemical herbicides by spraying herbicides on only precise regions of weeds in rice fields. To improve the accuracy of the model in identifying weed regions, in this study the performance among the you only look once (YOLO) series and single-shot multibox detector (SSD) one-stage target detection models, i.e., YOLOv3, YOLOv4-tiny, YOLOv5-s, and SSD target detection networks, is discussed. The experimental results showed that the SSD-based target detection model for *Alternanthera philoxeroides* was better than the YOLO series due to its higher recall, mAP (mean average precision), and $F1$ values, which reached 0.874, 0.942, and 0.881, respectively. Meanwhile, the target detection model based on SSD performed better than the YOLO series when dealing with mutual occlusion images between seedlings and *Alternanthera philoxeroides*. In conclusion, in this study a high-accuracy method is provided for detecting precise regions of *Alternanthera philoxeroides* by constructing a model based on SSD, contributing to the reduction of environmental pollution.

1. Introduction

Weeds mainly compete with rice seedlings before the closure stages for fertilizer, light, growth space, and other resources [1]. Consequently, weeds could affect the effective tillering and growth of rice seedlings, and could even reduce the yield by up to 45% in the later stage [2]. *Alternanthera philoxeroides* [3] is a malignant weed in paddy fields that invaded foreign plants worldwide and have a great impact on habitats and indigenous species. The growth competition ability of *Alternanthera* philoxeroides is stronger than that of rice. As the weed population density increases, the height of rice plants decreases. The presence of *Alternanthera philoxeroides* seriously hinders the growth and development of rice, resulting in decreases in rice spike length and effective spike number, and a significant reduction in rice yield [4]. Due to its strong asexual propagation ability, *Alternanthera philoxeroides* cannot be effectively removed through mechanical weeding. To date, spraying chemical herbicides is the main method for controlling *Alternanthera philoxeroides* [5]. The chemical weeding method has the advantages of high efficiency and can save considerable time and manpower. However, chemical weeding includes continuous and full spraying. Thus, crops cannot be distinguished from weeds,

and herbicides will be evenly distributed in the spray area, which is undesirable. This process not only pollutes the environment but also causes certain damage to rice seedlings when indiscriminate herbicide spraying is adopted in paddy fields. Therefore, chemical herbicides should be sprayed effectively in areas with weeds. With the rapid development of artificial intelligence and deep learning [6], it is possible to spray chemical herbicides accurately according to the regional position of weeds in rice fields [7, 8].

Due to the strong feature learning ability of deep learning, such as convolutional neural network (CNNs) extracting features through convolution, AlexNet [9], and VGGNet [10], the target detection field has been rapidly developed and applied in the agricultural field [11]. The RCNN series includes target detection algorithms such as the first-stage target detection algorithm and the second-stage target detection algorithm. First, the RCNN target detection method [12] was proposed based on the selective search algorithm. Although the accuracy of this method has been greatly improved compared with that of traditional manual feature detection, RCNN has shortcomings such as slow training and large-disk space requirements. To avoid the shortcomings of RCNN, fast RCNN directly uses the softmax function to replace the traditional SVM classifier [13]. This method extracts the features of the proposed region on the convolution feature map of the last layer, avoids the problem of repeated convolution calculation in RCNN, and improves the speed and accuracy. To avoid the large amount of time consumption required by the selective search method in Fast RCNN [14], Faster RCNN proposes a method to quickly generate fewer candidate frames with high quality and accuracy, that is, region proposed network (RPN) [15]. Faster RCNN was applied to weed identification in cotton fields under complex field backgrounds [16, 17] and achieved good results. Faster RCNN was deployed to the detection algorithm of weed targets with a weeding robot [18], and it achieved good results on lawn weeds [19].

The YOLO [20] and SSD series are first-stage target detection algorithms. SSD network can effectively learn the RPN idea of Fast RCNN. By referring to YOLO-related processing methods, it could predict multiple bounding boxes and corresponding categories at the same time in the detection process, and finally, generated detection results through the nonmaximum suppression method. SSD could use multiresolution feature maps for simultaneous detection in multiple degrees, and each feature map independently predict the target category and frame offset; thus, realizing a more accurate real-time monitoring framework.

The SSD target detection algorithm adopts a design idea based on regression. By adding multiple convolution layers to the basic VGGNet network, and regressing the categories and boundary boxes of multiple regions from multiple convolution feature maps, this method can better balance the accuracy and efficiency of target detection. Three datasets [21] were constructed with different resolutions for weeds in cotton fields, and the YOLOv3 model was used to optimize different datasets. The experimental results showed that the models could meet the production needs. YOLOv4 was used to detect weeds at the seedling stage of 3–5 leaves of corn, 52 pixels × 52 pixels of corn seedlings, and 13 pixels × 13 pixels of weeds could be detected [22]. The method of YOLO V5s was used to apply target detection before fruit thinning, which is significant to achieve early production forecast and automatic fruit thinning [23].

Due to the small row spacing and plant spacing of rice seedlings at the seedling stage, the gap between the seedlings returning to green after transplanting [24, 25], provided time and space for the propagation of *Alternanthera philoxeroides*. With the gradual tillering of seedlings and the growth of weeds, seedlings and weeds inevitably block each other in paddy fields. These factors pose a great challenge to target detection based on the target frame. Therefore, in this paper the target detection method of *Alternanthera philoxeroides* in the seedling stage of rice fields is studied. The problem of shelter is very common in the agricultural field. Seedlings and weeds in the rice seedling stage are often intertwined. However, when rice seedlings and weeds are seriously blocked, the detection algorithm faces a large challenge, and it is an urgent problem to be solved in the current field of rice field weed detection. In this paper, taking weeds in the rice seedling stage as the research object, aiming at images with complex backgrounds in natural environments an end-to-end weed target detection model is proposed to improve the recognition rate of small targets of weeds in the rice seedling stage in images with complex backgrounds.

2. Materials and Methods

2.1. Collection of Sample Images in Paddy Fields. The sample images of *Alternanthera philoxeroides* in rice fields were collected 15 days after rice transplanting. The rice seedlings were not sealed when the images were collected. A total of 210 images were collected and unified to the model input size by size transformation for further processing. The four model input dimensions are shown in Table 1. Each image contained at least one target regional of *Alternanthera philoxeroides*. Figure 1 showed a sample image of rice seedlings and two regions of *Alternanthera philoxeroides* that were manually labeled with a red wire frame.

2.2. Process of Detecting Target Regions. The research flow of the target detection method of *Alternanthera philoxeroides*, the weed in the rice seedling stage, based on the SSD model, is given as follows (Figure2):

(1) RGB sample images of *Alternanthera philoxeroides* in a natural paddy field environment were obtained.

(2) The artificial target frame containing *Alternanthera philoxeroides* was marked on each image to form the corresponding target label; 80% of all RGB sample images of *Alternanthera philoxeroides* and their corresponding labels were randomly selected as training samples, and the remaining 20% of the sample images were used as testing samples.

(3) Based on the SSD model, a sketch image target detection model of weedy lotus seeds in the rice seedling stage was constructed.

TABLE 1: The model detection performance for *Alternanthera philoxeroides*.

Model	Input dimension	Network	mAP	F1	Precision	Recall
YOLOv3	416 × 416	DarkNet-53	0.661	0.392	**0.979**	0.246
YOLOv4-tiny	416 × 416	CSPDarkNet	0.791	0.701	0.909	0.571
YOLOv5-s	640 × 640	Focus + CSPDarkNet	0.939	0.880	0.908	0.851
SSD	300 × 300	VGGNet	**0.942**	**0.881**	0.884	**0.874**

The highest result was bold-facing.

FIGURE 1: A sample image containing rice seedlings and *Alternanthera philoxeroides* in a paddy field.
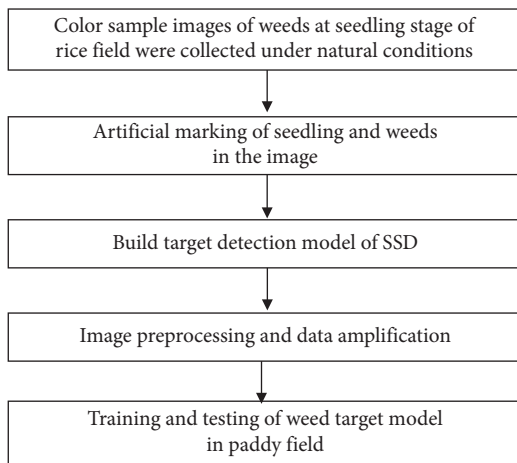

FIGURE 2: Flow chart of the target detection method for *Alternanthera philoxeroides*.

(4) The model was trained with the corresponding target frame label sample images formed by the RGB sample images of the *Alternanthera philoxeroides* weed and the artificial target frame label information. The SSD model was used to detect the position of *Alternanthera philoxeroides*, and the position of *Alternanthera philoxeroides* in the rice field was output to realize weed position detection in the rice field seedling stage. Then, we tested and analyzed each target detection model based on the test set.

2.3. Target Detection Model Based on SSD. The SSD algorithm improves the model detection ability for small objects. At present, multiscale feature detection represented by a feature pyramid has become a basic method to improve the model detection ability for small targets. The feature pyramid model first appeared in the feature pyramid networks (FPN) detection model. This method extracts feature images of different resolutions from the backbone network and then uses the upsampling method to improve the resolution of the deep low-resolution image and fuse the features with the shallow high-resolution feature layer through a lateral branch network. Finally, the detector is used to detect objects of different scales in a fused multiscale feature layer. This method extracts features of different scales from feature layers of different resolutions, which is conducive to the performance improvement of the detector. However, at the same time, the amount of calculation is increased and the reasoning performance is reduced. To improve the operation speed, SSD adopts a simple multiscale feature layer prediction and does not fuse different scales. Therefore, it has a fast reasoning speed and improves the model detection ability for small targets.

In this study, the model realizes the detection target through three steps. First, a convolution layer in a deep convolution neural network extracts the depth features of a weed image. Second, a region box extraction algorithm is used to locate the position of the weed region. Third, the extracted features are used to distinguish the category of weeds. In the SSD regression framework, which is shown in Figure 3, a pretrained VGG-16 convolutional neural network model was selected as the basic network and the target detection task was realized by modifying and fine tuning the VGG-16 network. Based on SSD regression, the position and category information of multiple targets could be obtained and combined with the idea of an RPN multiscale anchor frame, the real-time detection effect was realized under the premise of ensuring a certain level of accuracy. The convolution neural network was used as the backbone network, and an auxiliary structure were added to the network to generate a multiscale feature map detection method. Convolution layers were added to the end of the truncated basic network, and the size of these convolution layers was gradually decreased to obtain the predicted multiscale detection value. The detected convolution model was different for each feature layer. Using the multiscale regional features of each position of a whole image for regression not only had the characteristics of fast detection speed but could also greatly improve the accuracy of region box prediction.
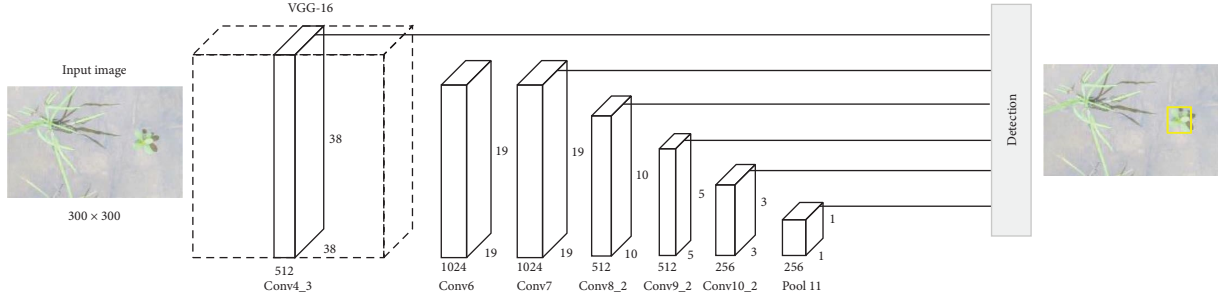
FIGURE 3: Algorithm flow chart of SSD.

SSD processes are performed on multiple feature maps, and SSD directly performs classification and region box regression using a scoring mechanism. As shown in Figure 4, an input image with small set of 8×8 or 4×4 default boxes with different aspect ratios were evaluated at each position on the feature map and the SSD processed in the network performs convolution. For each default box, the shape offset and confidence for all object categories were predicted. During training, these default boxes were first matched to the real labeled area boxes. The SSD method is based on a feedforward convolutional neural network, which generates a set of region boxes with a fixed size and score of object categories and finally achieves detection by utilizing a nonmaximum suppression (NMS) process.

2.3.1. Priori Bounding Box. A priori bounding box refers to a frame with a fixed size and position that is preset before the image is processed to cover the possible objects at a given position. The preset of a priori frame inevitably deviates from the real object, so the priori frame will be fine-tuned through the regression method to approach the edge of the real object as much as possible. The concept of a priori frame was first proposed in YOLOv1, was greatly improved by SSD and, became a classic example of a detection algorithm. The priori frame includes two parts the center and size of the priori frame. The center of the priori frame can be divided into $N \times N$ grids. The center of the grid is the center of the priori frame, which is also referred to as the anchor point. Due to the unique feature pyramid structure of SSD, the size of the feature map for detection is [38, 19, 10, 5, 3, 1], which is actually equivalent to dividing the input image into grids with sizes of $(38 \times 38, 19 \times 19, 10 \times 10, 5 \times 5, 3 \times 3, 1 \times 1)$. The calculation formula of the grid center coordinate is as follows:

$$(x, y) = \left[\frac{i + 0.5}{|f_k|}, \frac{j + 0.5}{|f_k|} \right], \quad i, j \in \left[0, |f_k| \right], k \in [1, m], \tag{1}$$

where i and j are the coordinates of the corresponding points of the feature map, f_k is the size of the feature map, and m is the number of effective feature layers.

An anchor point corresponds to 4 to 6 prior boxes of different sizes and proportions. The concept of the scaling factor is introduced here. The scaling factor refers to the ratio of the size of the box to the size of the original image. For example, if the scaling factor of the first layer is manually set

to 0.1, the size of the frame of this layer in the original image is 300×0.1, that is, 30 pixels. SSD uses the scaling factor recurrence formula to calculate the scaling factor of each layer. The recurrence formula is as follows:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1} (k - 1), k \in [1, m], \tag{2}$$

where Smin = 0.2, Smax = 0.9, and $m = 6$. According to the zoom factor and the size of the input image, the maximum and minimum sizes of the priori box in each effective feature layer can be calculated.

With each anchor point as the center, 4 to 6 boxes are preset according to different length and width sizes and proportions, which is a priori box. There are two squares, the size of the smaller square prior box is Sk and the side length of the larger square prior box is $\sqrt{s_k \cdot s_{k+1}}$. The side lengths of the other two rectangular boxes are $1/\sqrt{\text{aspect\_ratio} \cdot s_k}$, $\sqrt{\text{aspect\_ratio} \cdot s_k}$, $\sqrt{\text{aspect\_ratio} \cdot s_k}$, and $1/\sqrt{\text{aspect\_ratio} \cdot s_k}$, where aspect_ratio is a manually preset prior frames aspect ratio. A priori frame with different sizes and aspect ratios can be obtained.

2.3.2. Nonmaximum Suppression. NMS is a postprocessing method of the SSD algorithm when applied to target detection. When performing target detection, we need to obtain a priori frame of the detected target. However, there may be multiple different small targets in the same detection image, so multiple different priori frames will be generated in the image. The NMS method is added to the end of the SSD algorithm. The purpose is to select a candidate box that is closest to the real labeled prior box that can also filter out the candidate box with a low intersection score of the real labeled box and give up the redundant priori box that has a low intersection score. Finally, the priori box that is closest to the real label is selected.

2.4. Evaluation Performance of Target Detection Models. Generally, the performance of a target detection model can be evaluated in terms of classification accuracy, precision, recall rate, precision and recall (PR) curve, average precision (AP), mAP, etc.

The four basic evaluation indicators are TPs (true positives), which means that positive samples are allocated and the correct samples are allocated, representing correctly classified positive samples. TNs (true negatives) refer to the
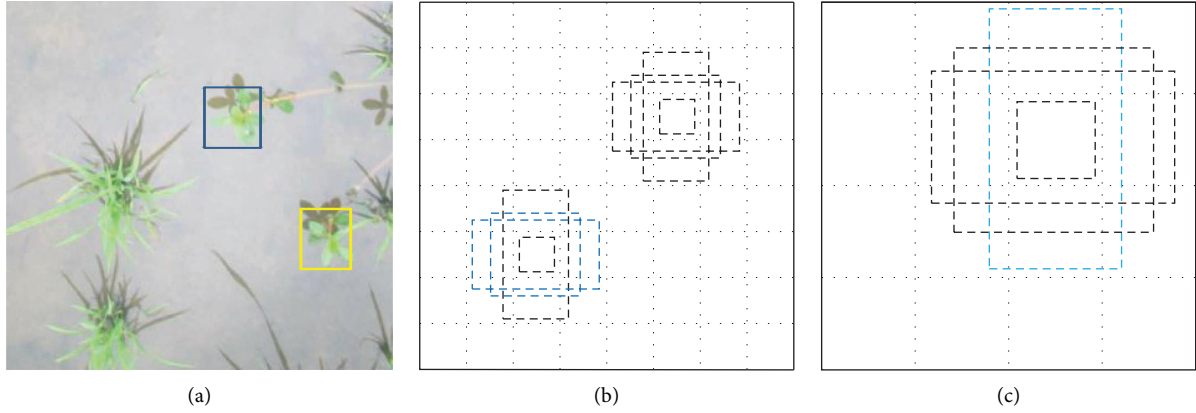
FIGURE 4: The detection process of SSD: (a) target block diagram, (b) characteristic diagram of 8×8, and (c) characteristic diagram of 4×4.

negative samples assigned and correct samples assigned representing correctly classified negative samples. FPs (false-positives) refer to samples that are incorrectly allocated as positive samples and represents negative samples that are misclassified. FNs (false-negative) refer to samples that are incorrectly assigned as negative samples and represents positive samples that are misclassified.

2.4.1. Recall. The specific meaning of recall is as follows: the proportion of the number of correct and positive classes detected by the classifier to the number of all positive classes. It evaluates the ability of the classifier to detect positive samples.

The calculation formula of recall is as follows:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

2.4.2. Precision. The accuracy rate is the probability of correct detection among all detected targets. Precision and accuracy are different. Accuracy is for all samples, and precision is only for some of the samples detected (including false detection). The calculation formula is as follows (4):

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{4}$$

2.4.3. Accuracy. Accuracy is the proportion of correct predictions in all predictions.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{5}$$

2.4.4. Intersection over Union (IoU). To evaluate the positioning accuracy of the model, it is necessary to calculate the IoU, which is the degree of overlap between the real label box and the network prediction box. The smaller the IoU is, the farther the distance between the network prediction box and the real annotation box. Only when the IoU value between the real box and the prediction box is greater than the set threshold value is the prediction box considered a true positive, and when the IoU value between the real box and the prediction box is less than the set threshold value, it is considered a false-positive. The IoU calculates the ratio of the intersection and union of the "predicted border" and "real border," as shown in Figure 5.

2.4.5. Average Precision (AP). Precision and recall is a pair of contradictory opposite indicators. Generally, when the precision is high, the recall is low; when the recall rate is high, the precision rate is low. Therefore, only in some simple tasks can the recall rate and precision rate be high. Therefore, the AP is proposed to measure the performance of the model more comprehensively. Before looking at the AP, let us take a look at the precision-recall (PR) curve; that is, the horizontal axis is recall and the vertical axis is precision. The AP represents the average value of the detector in each recall case, which corresponds to the area under the PR curve (AUC).

From a discrete point of view, the AP can be expressed as follows:

$$\text{AP} = \frac{\sum P_{ri}}{\sum r}, \tag{6}$$

where P_{ri} represents the value of p corresponding to $r - i$ on the PR curve, and $\sum r = 1$.

2.4.6. mAP. The AP is an identifier for one category, and the mAP is the average of the AP from the dimension of categories, so it can evaluate the performance of multiple classifiers. The size of the map must be in the interval of [0, 1], and the larger the map is, the better. This index is also the most critical index in the target detection algorithm.

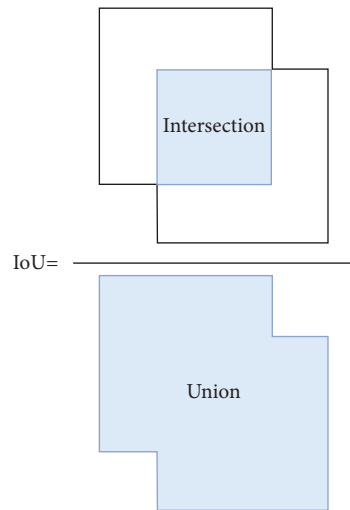IoU= $\dfrac{\text{Intersection}}{\text{Union}}$

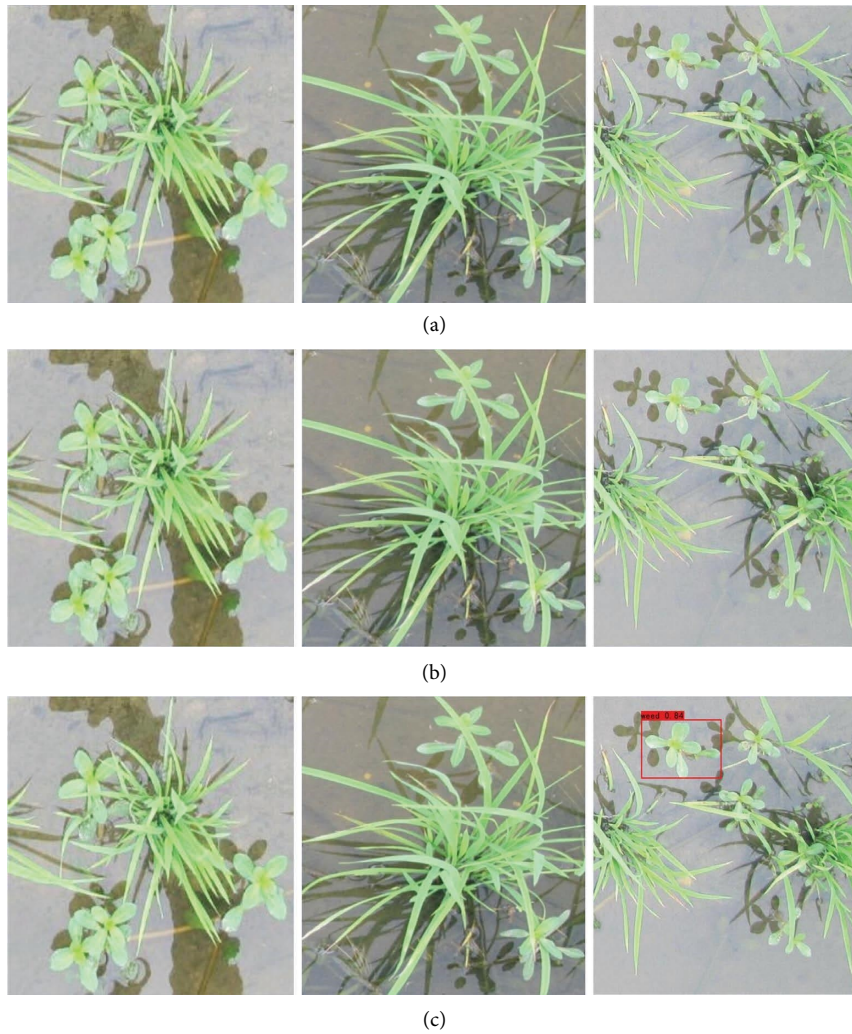Figure 5: IOU calculation.

(a)

(b)

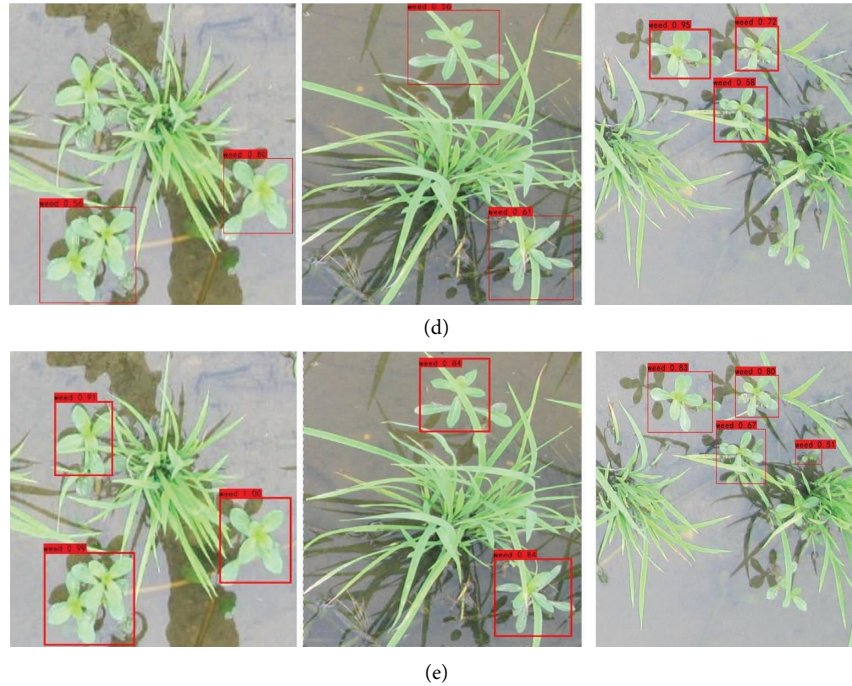
(c)

Figure 6: Continued.

(d)

(e)

FIGURE 6: Detection results of different models under occlusion: (a) original images, (b) YOLOv3, (c) YOLOv4-tiny, (d) YOLOv5-s, and (e) SSD.

$$mAP = \frac{AP}{num\_classes}. \tag{7}$$

3. Results and Discussion

3.1. Development Environment. The model training environment used a Windows 10 64-bit operating system. Anaconda was used for virtual environment construction, the Python compiler was PyCharm and the deep learning framework was PyTorch1.2.0. The dataset annotation used the wizard annotation assistant to record the BBox information of weed samples, and the GPU was a GTX 1080 Ti with 11 GB video memory.

3.2. Target Detection Results. Table 1 summarizes the model performance of the one-stage target detection model YOLO series (YOLOv3, YOLOv4 tin, and YOLOv5-s) and SSD. The precision value of the target detection model based on the YOLO series was higher than that of the SSD model, but the recall, mAP, and F1 values of the target detection model based on SSD were higher than those of the YOLO series.

Due to the small row spacing and plant spacing of rice seedlings, there would inevitably be mutual shielding between seedlings and *Alternanthera philoxeroides*. In view of the above situation, as shown in Figure 6, the detection effect of the target detection model of *Alternanthera philoxeroides* based on SSD was better than that of the YOLO series. As shown in Figure 6(a), the YOLOv3 model did not detect any weed targets, as shown in Figure 6(b), YOLOv4-tiny only detected one weed target, as shown in

Figure 6(c), and YOLOv5-s lost one weed target. As shown in Figure 6(d), the SSD model could identify all weed targets.

3.3. Target Detection Analysis. As shown in Figure 7, in the comparison of PR curves of different models, the abscissa is recall and the ordinate is precision. The area at the bottom left of the PR diagram represents the effect of the model on the dataset. The results showed that the effect of the SSD model with the dataset of *Alternanthera philoxeroides* was higher than that of the YOLOv3, YOLOv4-tiny, and YOLOv5-s models.

In the comparison of the recall curves of different models in Figure 8, the abscissa is the threshold value of the reserved prediction frame of the algorithm model and the ordinate is recall. The curve represents the ratio of the number of correct frames predicted by the algorithm model to the number of all real frames when the threshold was fixed and the threshold value was 0.5. When the threshold was larger, the model could still maintain a high recall value, which proved that the algorithm model performed well with the dataset. The results showed that the SSD model had a better recall curve than the YOLOv3, YOLOv4-tiny, and YOLOv5-s models.

In the comparison of the F1 value curves of different models in Figure 9, the abscissa is the threshold value of the reserved prediction frame of the algorithm model and the ordinate is F1. To discuss the performance of different algorithms, the concept of the F1 value was proposed on the basis of precision and recall to evaluate precision and recall as a whole. When the curve of the F1 value contained more areas, the detection effect of the model with the dataset was
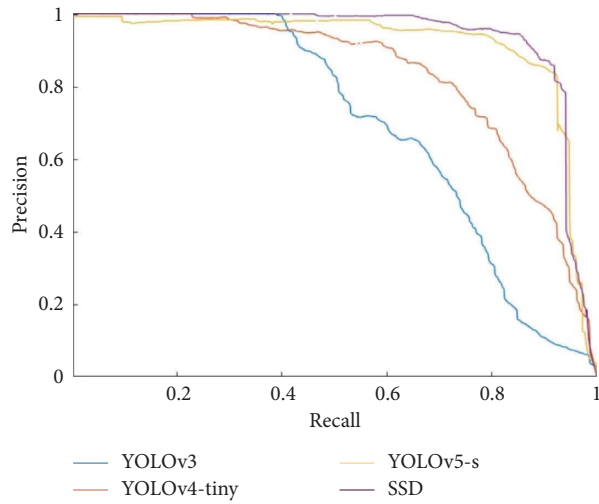
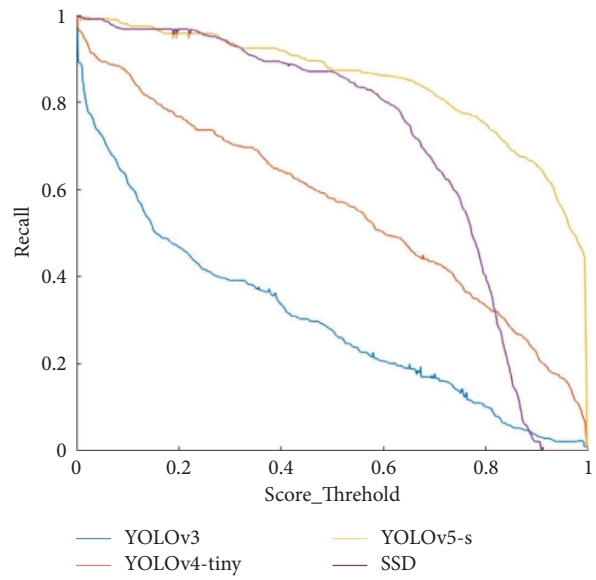FIGURE 7: Precision and recall curve with different models.

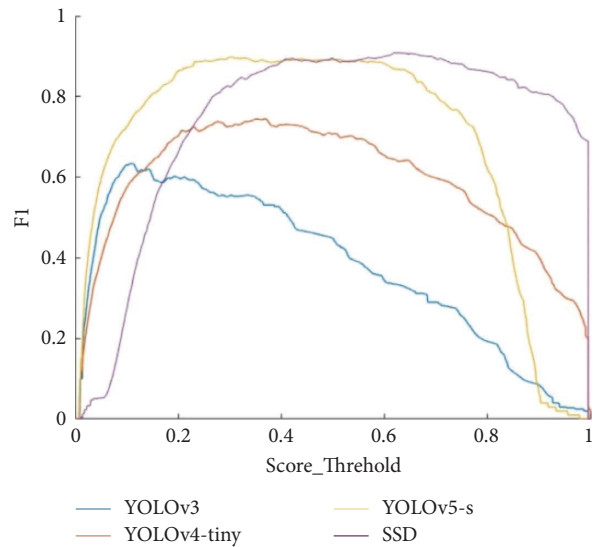FIGURE 8: Recall curve with different models.

FIGURE 9: *F*1 value curve with different models.

better. The results showed that the evaluation effect of the SSD model on the *F*1 value curve of the *Alternanthera philoxeroides* dataset was better than that of the YOLOv3, YOLOv4-tiny, and YOLOv5-s models.

4. Conclusions

In this study, a target detection method for *Alternanthera philoxeroides* in paddy fields was successfully developed by constructing a model based on an SSD network, and the position information of the *Alternanthera philoxeroides* target was obtained by combining the idea of the RPN multiscale anchor frame. Under the premise of ensuring certain accuracy, the real-time detection effect was realized. While comparing the performance of models constructed with the target detection networks of YOLOv3, YOLOv4 tin, YOLOv5-s, and SSD, the values of recall, map, and F1 of the target detection model of *Alternanthera philoxeroides* based on SSD were higher than other models, reaching 0.874 0.942 and 0.881, respectively. For the mutual occlusion images between seedlings and *Alternanthera philoxeroides*, the performance effect of the target detection model based on SSD was also better than that of YOLOv3, YOLOv4-tiny, and YOLOv5-s. It could scrupulously be concluded that the proposed method might somehow contribute to the reduction of environmental pollution by only spraying chemical herbicides within the precise region of *Alternanthera* philoxeroides. However, the shortcomings of the research are as follows: first, the sample of *Alternanthera philoxeroides* in paddy field is not rich enough and second, the model is not embedded in the terminal device for experiment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request. The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] M. Curt, P. Mauri, M. Sanz et al., "The ability of the Arundo donax crop to compete with weeds in central Spain over two growing cycles," *Industrial Crops and Products*, vol. 108, pp. 86–94, 2017.

[2] H. Jensen, L. B. Jacobsen, S. M. Pedersen, E. B. Tavella, S. M. Pedersen, and E. Tavella, "Socioeconomic impact of widespread adoption of precision farming and controlled traffic systems in Denmark," *Precision Agriculture*, vol. 13, no. 6, pp. 661–67777, 2012.

[3] H. Yan, F. Lei, Y. Zhao, F. Li, D. Wu, and C. Zhu, "Prediction of the spatial distribution of Alternanthera philoxeroides in China based on ArcGIS and MaxEn," *Global Ecology and Conservation*, vol. 21, Article ID e00856, 2020.

[4] L. Q. Yu, Y. Fujii, Y. J. Zhou et al., "Response of Response of Exotic Invasive Weed Alternanthera philoxeroides to Environmental Factors and Its Competition with Ricexotic invasive weed Alternanthera philoxeroides to environmental factors and its competition with rice," *Rice Science*, vol. 14, no. 1, pp. 49–55, 2007.

[5] Y. Liu, B. Liu, X. F. Wang, and C. H. Guan, "Problems and solutions for chemical weed control in China," *Agrochemicals*, vol. 44, no. 7, pp. 289–293, 2005.

[6] A. S. M. M. Hasan, F. Sohel, D. Diepeveen et al., "A survey of deep learning techniques for weed detection from images," *Computers and Electronics in Agriculture*, vol. 184, Article ID 106067, 2021.

[7] C. Liu, H. Lin, Y. Li, L. Gong, and Z. Miao, "Analysis on status and development trend of intelligent control Technology for agricultural equipment," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 51, no. 01, pp. 1–18, 2020.

[8] D. Fan, Q. Yao, B. Yang, Y. Zhou, Z. Guan, and J. Tang, "Progress in research on intelligentization of field weed recognition and weed control Technology," *Scientia Agricultura Sinica*, vol. 43, no. 09, pp. 1823–1833, 2010.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[10] K. E. Hengjin, D. A. N. Chen, L. I. Xiaoli, T. A. N. G. Yunbo, S. H. A. H. Tejal, and R. A. N. J. A. N. Rajiv, "Towards brain big data classification: epileptic EEG identification with a lightweight VGGNet on Global MIC," *IEEE Access*, vol. 2, no. 99, p. 1, 2018.

[11] N. Rai, Y. Zhang, B. G. Ram et al., "Sreekala Bajwa, Xin Sun. Applications of deep learning in precision weed management: Applications of deep learning in precision weed management: A review review," *Computers and Electronics in Agriculture*, vol. 206, Article ID 107698, 2023.

[12] R. Girshick, J. Donahue, Trevor Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.

[13] X. Li, C. Fu, X. Li, Z. Wang, X. Li, and Z. Wang, "Improved Improved Faster R-CNN for Multi-Scale Object Detectionaster RCNN for multi-scale object detection," *Journal of Computer-Aided Design and Computer Graphics*, vol. 31, no. 7, pp. 1095–1101, 2019.

[14] J. R. R. Uijlings, K. E. A. Van de Sande, T. Gevers et al., "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[15] C. Cao, B. Wang, W. Zhang et al., "An Improved Faster RCNN for Small Object Detection," *IEEE Access*, vol. 7, pp. 106838–106846, 2019.

[16] M. Peng, J. Xia, and H. Peng, "Efficient recognition of cotton and weed in field based on Faster RCNN by integrating FPN," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 35, no. 20, pp. 202–209, 2019.

[17] X. Fan, J. Zhou, Y. Xu, K. Li, and D. Wen, "Identification and localization of weeds based on optimized in faster RCNN cotton seedling stage," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 52, no. 05, pp. 26–34, 2021.

[18] L E. Montoya-Cavero, R. Díaz de León Torres, A. Gómez-Espinosa et al., "Vision systems for harvesting robots: Vision systems for harvesting robots: Produce detection and local-izationroduce detection and localization," *Computers and Electronics in Agriculture*, vol. 192, Article ID 106562, 2022.

[19] C. Li, S. Lu, S. Yuan, and Z. Wang, "Weed identification algorithm of weeding robot based on Faster RCNN," *Journal of Chinese Agricultural Mechanization*, vol. 40, no. 12, pp. 171–176, 2019.

[20] J. Chen, H. Wang, H. Zhang et al., "Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion," *Computers and Electronics in Agriculture*, vol. 202, Article ID 107412, 2022.

[21] J. Xue, J. Dai, Q. Zhao, G. Zhang, M. Cui, and N. Jiang, "Cotton field weed detection based on low-altitude drone image and YOLOv3," *Journal of Shihezi University*, vol. 37, no. 01, pp. 21–27, 2019.

[22] L. Quan, F. Xia, W. Jiang et al., "Research on recognition of maize seedlings and weeds in maize field based on Convolutional Neural Network YOLO v4 convolutional neural network," *Journal of Northeast Agricultural University*, vol. 52, no. 07, pp. 89–98, 2021.

[23] D. Wang, D. He, Wang, and D. He, "Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning," *Biosystems Engineering*, vol. 210, pp. 271–281, 2021.

[24] X. Deng, M. Xu, L.. Qi, G. Sun, and L. Song, "Weeds recognition at seedling stage in paddy fields based on deep feature," *Journal of Agricultural Mechanization Research*, vol. 43, no. 8, pp. 27–30, 2021.

[25] X. Deng, M. Xu, L. Qi, G. Sun, L. Song, and J. Jing, "Recognition of weeds at seedling stage in paddy fields using convolutional neural network and transfer learning," *Journal of Agricultural Mechanization Research*, vol. 43, no. 10, pp. 167–171, 2021.