

## Article

# Weed Detection in Peanut Fields Based on Machine Vision

Hui Zhang, Zhi Wang, Yufeng Guo, Ye Ma, Wenkai Cao, Dexin Chen, Shangbin Yang and Rui Gao \*

College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China

\* Correspondence: ronnierrui@henau.edu.cn; Tel.: +86-13673626972

**Abstract:** The accurate identification of weeds in peanut fields can significantly reduce the use of herbicides in the weed control process. To address the identification difficulties caused by the cross-growth of peanuts and weeds and by the variety of weed species, this paper proposes a weed identification model named EM-YOLOv4-Tiny incorporating multiscale detection and attention mechanisms based on YOLOv4-Tiny. Firstly, an Efficient Channel Attention (ECA) module is added to the Feature Pyramid Network (FPN) of YOLOv4-Tiny to improve the recognition of small target weeds by using the detailed information of shallow features. Secondly, the soft Non-Maximum Suppression (soft-NMS) is used in the output prediction layer to filter the best prediction frames to avoid the problem of missed weed detection caused by overlapping anchor frames. Finally, the Complete Intersection over Union (CIoU) loss is used to replace the original Intersection over Union (IoU) loss so that the model can reach the convergence state faster. The experimental results show that the EM-YOLOv4-Tiny network is 28.7 M in size and takes 10.4 ms to detect a single image, which meets the requirement of real-time weed detection. Meanwhile, the mAP on the test dataset reached 94.54%, which is 6.83%, 4.78%, 6.76%, 4.84%, and 9.64% higher compared with YOLOv4-Tiny, YOLOv4, YOLOv5s, Swin-Transformer, and Faster-RCNN, respectively. The method has much reference value for solving the problem of fast and accurate weed identification in peanut fields.

**Keywords:** weed identification; YOLOv4-Tiny; attention mechanism; multiscale detection; precision agriculture



**Citation:** 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. <https://doi.org/10.3390/agriculture12101541>

Academic Editor: Michele Pisante

Received: 6 August 2022

Accepted: 22 September 2022

Published: 24 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Peanut is one of the leading oil crops in the world and is vital to global oil production. However, weed competition [1] is an essential factor restricting peanut production, reducing peanut production by 5–15% owing to annual grass damage. Research has shown that peanut production in farmlands with 20 weeds per square meter is 48.31% less than a no-weed control group. In addition, weeds facilitate the breeding and spread of diseases and insect pests, resulting in the frequent emergence of peanut diseases and insect pests [2]. The conventional weeding method of spraying pesticides incurs a significant amount of pesticide waste and causes irreversible pollution to the farmland. Owing to the development of precision agriculture [3], the investigation of site-specific weed management [4] for weed prevention and control has intensified gradually. An efficient detection and identification method for peanuts and weeds is necessary to achieve accurate weed control and management in the farmland.

Currently, many methods are proposed for weed detection, including remote sensing analysis [5], spectral identification [6], and machine vision identification [7]. The equipment required for remote sensing analysis and spectral identification methods is expensive and difficult to promote in agricultural production. The machine vision identification method has been widely used in weed identification because of its low cost and high portability. Bakhshish et al. [8] used Fourier descriptors and invariant moment features to form a shape feature set and implemented weed detection based on artificial neural networks. Rojas et al. [9] extracted the texture features of weeds using the gray-level co-occurrence matrix. They used principal component analysis to reduce the dimensionality of the

features and finally used a support vector machine algorithm to complete the classification. Although these methods achieve the identification of crops and weeds, they rely excessively on the manual design and selection of image features, are susceptible to environmental factors such as lighting, and have poor stability and low recognition accuracy.

The development of deep learning technology [10] has enabled convolutional neural networks to reveal deeper features in images, which possess stronger generalization ability than manually selected features. Gai et al. [11] proposed an improved YOLOv4 model for fast and accurate detection of cherry fruit in complex environments. Khan et al. [12] established a weed identification system for pea and strawberry fields based on an improved Faster-RCNN, whose maximum average accuracy for weed recognition was 94.73%. Sun et al. [13] used YOLOv3 to identify Chinese cabbages in a vegetable field. They employed image processing methods to tag plants around Chinese cabbages as weeds. To detect weeds in a carrot field, Ying et al. [14] incorporated deep separable convolutions and an inverted residual block structure into YOLOv4 and replaced its backbone network with MobileNetV3-Small, which improved the recognition speed of the model; however, the average recognition accuracy was only 86.62%. The studies mentioned above indicate that although deep learning can solve the problem of manual feature design in conventional image processing methods, the following issues remain: 1) although using a deep-seated network model for weed detection improves the recognition accuracy, the recognition speed cannot satisfy real-time requirements owing to its large volume; 2) improving the recognition speed by trimming the model network renders the model insensitive to smaller target recognition and reduces its recognition accuracy.

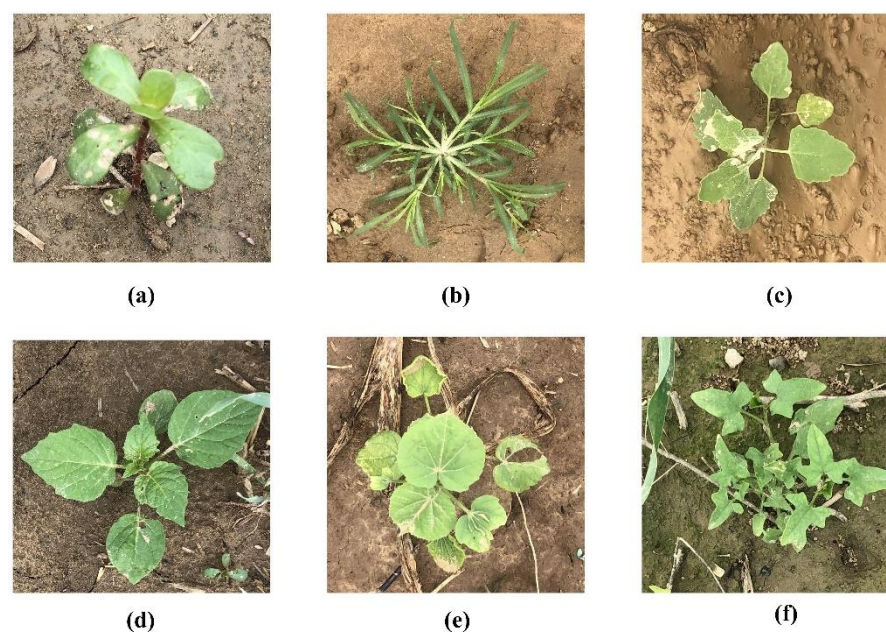
In this study, peanuts and six types of weeds were used as recognition objects, and a weed recognition model based on the improved YOLOv4-Tiny [15] was developed to address the issues above. First, based on YOLOv4-Tiny, CSPDarkNet53-Tiny [16] was used as the backbone network of the model to ensure real-time detection performance; next, a multiscale detection model was implemented by introducing the detailed information of shallow-layer features in an FPN [17] to improve the ability of smaller target recognition. In addition, an ECA [18] module was used to calibrate the effective feature layer to enhance key information pertaining to weeds in the image. Finally, the soft-NMS [19] function was used in the output prediction layer to replace the NMS [20] function to filter the prediction box.

## 2. Materials and Methods

### 2.1. Materials

#### 2.1.1. Data Acquisition

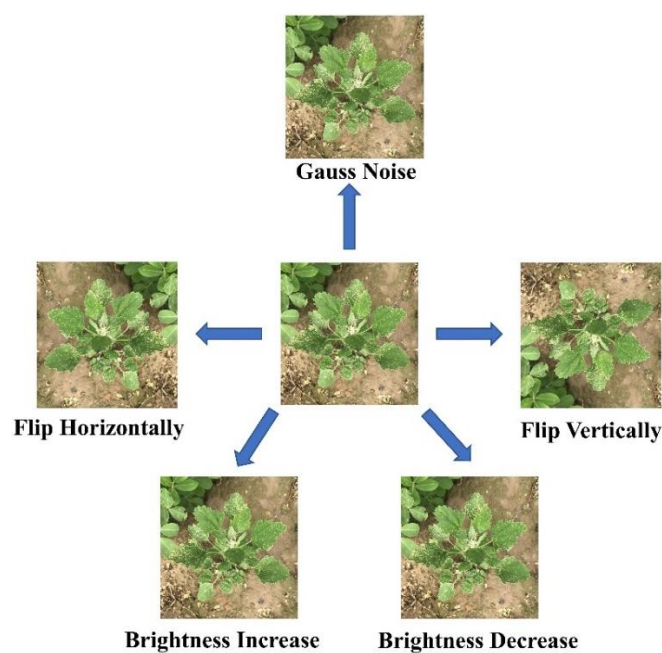
The weed images used in this study were obtained from peanut fields in more than 20 areas in Henan Province, China. A Fuji Finepix4500 camera was used to capture artificial images with a resolution of  $2017 \times 2155$  in JPG format; 855 images were obtained, including those of a single weed, sparsely distributed weeds, and overgrown weeds. The images were captured at 7:00, 13:00, and 17:00 via high-angle overhead shots from approximately 70 cm relative to the ground. Based on investigation and screening, the weed types selected were *Portulaca oleracea*, *Eleusine indica*, *Chenopodium album*, *Amaranth blitum*, *Abutilon theophrasti*, and *Calystegia*. No imbalance was indicated between any two types of weeds. The shape and color of the six weeds are shown in Figure 1.



**Figure 1.** Shape and color of six weeds. (a) *Portulaca oleracea*, (b) *Eleusine indica*, (c) *Chenopodium album*, (d) *Amaranth blitum*, (e) *Abutilon thophrasti*, (f) *Calystegia hederacea*.

### 2.1.2. Data Enhancement and Annotation

Overfitting in the training set caused by excessively small data sizes was prevented using the following methods: image horizontal and vertical flip, brightness increase and decrease (randomly increase or decrease the original brightness by 10%–20%), and Gaussian noise addition (variance  $\sigma = 0.05$ ) for random image enhancement [21]. Figure 2 shows an example of the effect of data enhancement. The data enhancement method was only used in the training set. The expanded dataset contained 3355 images. Information regarding weeds and peanuts in the image was annotated using the LabelImg software. The annotation format was Pascal VOC2007, and the file type was .xml. The dataset was categorized into training and test sets. The number of pictures in each dataset is shown in Table 1.



**Figure 2.** Data enhancement.

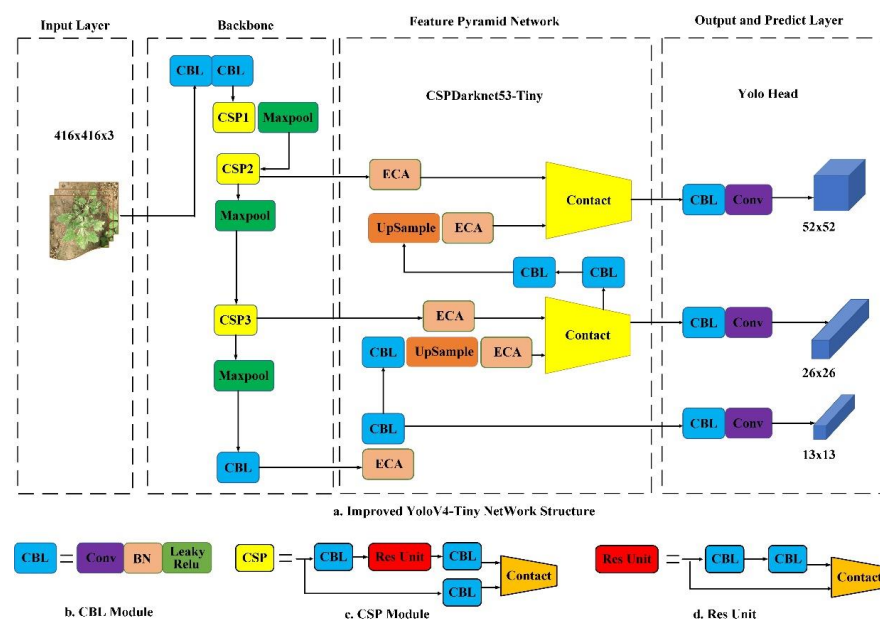
**Table 1.** Dataset after data enhancement.

Dataset	Train	Test	Total
Original Images	700	155	855
Flip Horizontally	500	0	500
Flip Vertically	500	0	500
Brightness Increase	500	0	500
Brightness Decrease	500	0	500
Gauss Noise	500	0	500
Total Number	3200	155	3355

## 2.2. Methods

### 2.2.1. EM-YOLOv4-Tiny Network

YOLOv4-Tiny comprises four components: an input layer, a backbone network, an FPN, and an output prediction layer. The images received were uniformly scaled to a size of  $416 \times 416$ . The features were extracted from CSPDarkNet53-Tiny and then sent to the FPN for feature fusion. The location and category information of the target was obtained in the output prediction layer. CSPDarkNet53-Tiny primarily comprises a CBL module and a cross-stage partial (CSP) module [22]. The CBL module comprises a convolutional layer, batch normalization, and a Leaky Relu [23] activation function in series. It is the smallest module in the overall network structure and is used for feature control splicing and sampling. The CSP module is an improved residual network structure that can segment the input feature map into two components: the main component stacks the residual, and the other is fused in series with the main component after some processing. CSPDarkNet53-Tiny contains three CSP modules: CSP1, CSP2, and CSP3. As the dimensions of the output feature map are reduced, the location information in the CSP module becomes increasingly vague, the detailed information becomes increasingly scarce, and the ability to detect smaller targets is gradually weakened. To solve these problems, a path connected to the CSP2 layer in the FPN was added, while the output characteristics of the CSP2 layer were fused with the upsampling results in the channel dimension to form an output focused on the detection of smaller targets. The EM-YOLOv4-Tiny network structure is shown in Figure 3.



**Figure 3.** EM-YOLOv4-Tiny network Structure, where Conv is convolution, BN is batch normalization, Leak Relu is activation function, Maxpool is maximum pooling, ResUnit is the residual unit, Upsample is upsampling, ECA is efficient channel attention module, Contact is the feature fusion method of adding channel numbers, Yolo Head is the prediction anchor, CBL is series fusion module of Conv, BN, and Leak Relu, and CSP is cross-stage partial module.



To further improve the detection accuracy, the ECA module was used repetitively to process the effective features in the FPN. The attention module suppressed the background information in the image and enhanced the key information through weight calibration [24]. Regarding the predicted output, the EM-YOLOv4-Tiny network yielded three outputs of different scales, namely  $13 \times 13$ ,  $26 \times 26$ , and  $52 \times 52$ .

### 2.2.2. ECA Attention Mechanisms

Multiscale prediction for hierarchical detection was utilized in this study to detect smaller targets. Although shallow features have smaller receptive fields, which can enable better detection of smaller targets, they result in considerable irrelevant noises, thus affecting the network's ability to assess the importance of information obtained from an image. By introducing the ECA attention module into the neck section of YOLOv4-Tiny, the weed features in the image could be further enhanced while irrelevant background weights were suppressed.

In the ECA network, the input features were first pooled globally, and a single numerical value was used to represent the characteristics of each channel. Next, a fast one-dimensional convolution [25] of size  $k$  was performed to assign weights for each channel to realize information exchange between channels. Finally, the weight proportion of each channel was generated using the sigmoid function [26], and features with channel attention were obtained by merging with the original input features. More details about the ECA network can be found in Appendix A.

### 2.2.3. Use of Complete Intersection over Union Loss

Owing to the scale invariance and non-negativity of the IoU [27], the latter is typically set as the bounding box loss function in conventional target detection networks. Specifically, IoU refers to calculating the ratio of the prediction box and the real box, which can better reflect the quality of the regression box. However, using IoU as the loss function still has some problems. On the one hand, when the positions of two bounding boxes do not intersect ( $\text{IoU} = 0$ ), the loss function will become non-differentiable. On the other hand, when the overlap rate of prediction frames is the same, IoU cannot accurately reflect the location information of both.

Therefore, the CIoU [28] was used in this study as the loss function for training. Additionally, the overlap degree and the distance between the prediction and real boxes were considered comprehensively, and the aspect ratio of the prediction box was added as a penalty term to stabilize the regression results. More details about CIoU loss can be found in Appendix B.

### 2.2.4. Soft-NMS Algorithm for Filtering Prediction Boxes

For the output and prediction of YOLOv4-Tiny, the NMS algorithm filters redundant prediction boxes around the target to be detected. The NMS algorithm deletes prediction boxes whose confidence is below the preset threshold, filters boxes that belong to the same category, and obtains the highest score in a specific area; hence, it effectively eliminates redundant bounding boxes. However, in cases involving dense weed growth or severe mutual occlusion between weeds and peanuts, the NMS algorithm deletes prediction boxes that belong to other targets, thus resulting in missed detections. To solve this issue, the soft-NMS instead of the original NMS was used in this study. When multiple prediction boxes appeared around a weed, their scores were multiplied by a weighting function to weaken those that overlapped with the box with the highest score. In this regard, the Gaussian [29] function was used as the weighting function, and the calculation is as follows:

$$\text{Score}_i = \text{Score}_i \cdot e^{\frac{-\text{IoU}(C_i, B)}{\sigma}} \quad (1)$$

where  $\text{Score}_i$  represents the score of the current box,  $C_i$  represents the current bounding box, and  $B$  represents the prediction box with the highest score. The greater the overlap

between the prediction box and the box with the highest score, the stronger the weakening ability of the weighting function and the lower the score assigned to it.

#### 2.2.5. Model Performance Evaluation Indices

In this study, indices typically used in multiclass target detection models, such as precision, recall rate, mean average precision (mAP), and F1 value, were used to evaluate the model performance.

Precision indicates the proportion of correct detections in all the prediction boxes, and Recall indicates the proportion of correctly detected label boxes in all label boxes.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

where TP represents the number of correctly detected weeds; FP represents the number of incorrectly detected weeds; and FN represents the number of missed detections of weeds.

AP represents the average precision of a class of detected objects, and mAP is the mean average value of AP for all classes.

$$AP = \int_1^0 \text{Precision d Recall}, \quad (4)$$

$$mAP = \frac{1}{N} \sum_1^N AP(k) \quad (5)$$

The F1 value can be regarded as a harmonic mean of Precision and Recall, as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

The evaluation indices selected in this study were calculated based on a threshold of 0.5. In the follow-up experiments, the mAP was used as the primary performance evaluation index of the model.

#### 2.2.6. Model Training

The software and hardware environment of model training and testing are shown in Table 2. In order to further improve the recognition accuracy of the model, this study used a transfer learning method to initialize the weights of the model. Before model training, the EM-YOLOv4-Tiny network was pretrained with the Pascal VOC dataset, and the weight file with the highest map in the training results was used as the pretraining weight to initialize the model. Meanwhile, the K-means [30] algorithm was used to cluster the anchor boxes in the dataset, and a total of 9 anchor boxes with different sizes were obtained: (19, 31), (56, 62), (90, 82), (103, 158), (149, 125), (175, 217), (250, 171), (241, 291), and (320, 335). This makes the true size of the anchor frame closer to the size of the weed to be detected. During training, the number of samples in each batch was set to 16, and the loading of the entire training set was considered an iteration. The adaptive moment estimation algorithm was used to optimize the model, the initial learning rate was set to 0.001, and the cosine annealing algorithm was employed for attenuation. After 150 iterations, the model converged.

**Table 2.** Training and test environment configuration table.

Configuration	Parameter
Operating System	Ubuntu 18.04.1 LTS
CPU	Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz
GPU	NVIDIA Tesla T4
Accelerate Environment	CUDA10.2 CuDNN7.6.5
Pytorch	1.2
Python	3.6.2

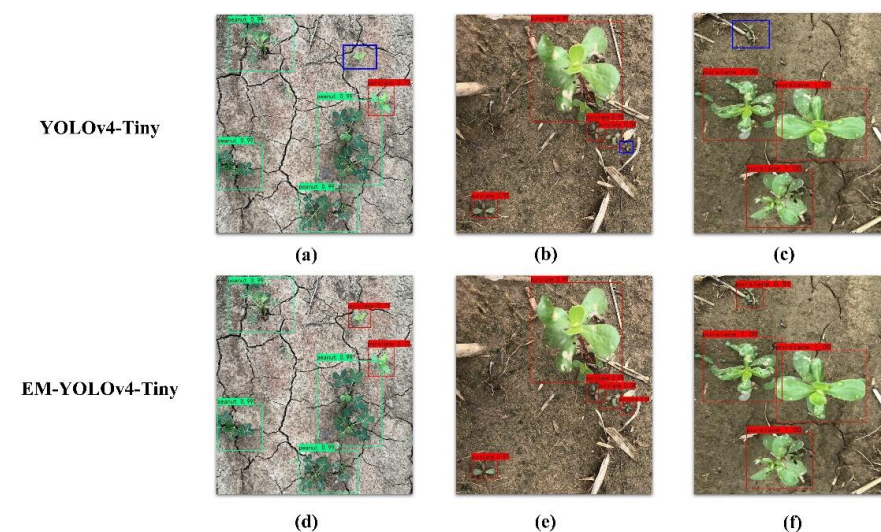
### 3. Results

#### 3.1. Performance Evaluation of EM-YOLOv4-Tiny

Based on the standard of the MS COCO dataset provided by Microsoft, weeds with a resolution lower than  $32 \times 32$  were defined as smaller targets. Several types of weeds exist in peanut fields, with some being smaller in morphological appearance than others. The standard YOLOv4-Tiny network tends to misdetect when identifying smaller targets. Based on the comparison results of EM-YOLOv4-Tiny and YOLOv4-Tiny using the same test set as shown in Table 3, the recognition precision rates of the EM-YOLOv4-Tiny for smaller targets and all targets were 89.65% and 94.54%, respectively, which surpassed the precision rates of the original network by 10.12% and 6.83%, respectively. The improved network combined the location and detailed information of the shallow-layer feature and improved the ability to identify smaller weeds via the addition of a channel attention mechanism, which suppresses the abundant noise in smaller receptive fields. The recognition performances before and after the network improvement are shown in Figure 4. The EM-YOLOv4-Tiny network included a new scale output in the neck section while the backbone network structure of the model remained unchanged, and the average inference time of each image increased to only 4.4 ms, indicating that the proposed network maintained a high inference speed while improving the recognition precision.

**Table 3.** Comparison of detection results of YOLOv4-Tiny and EM-YOLOv4-Tiny.

Models	mAP/%		Volume/MB	Time/ms
	Small Targets	All Targets		
YOLOv4-Tiny	79.53	87.71	22.4	6
EM-YOLOv4-Tiny	89.65	94.54	28.7	10.4



**Figure 4.** Comparison of detection results of YOLOv4-Tiny and EM-YOLOv4-Tiny, where (a–c) represent the recognition effect of the YOLOv4-Tiny model, and (d–f) represent the recognition effect of the EM-YOLOv4-Tiny model.

### 3.2. Performance Comparison of Improved Methods

To further demonstrate the effectiveness of the improved method in enhancing the model performance, different modules were benchmarked against the original YOLOv4-Tiny target detection network. The results are shown in Table 4.

**Table 4.** Influence of different improved modules on YOLOv4-Tiny network.

Method	Precision/%	Recall/%	mAP/%	F1/%	Time/ms
YOLOv4-Tiny	87.60	75.60	87.71	0.80	6.0
YOLOv4-Tiny + K-Means	91.80	74.80	88.90	0.82	6.0
YOLOv4-Tiny + K-Means+ Soft-NMS	88.16	84.91	90.37	0.86	6.0
YOLOv4-Tiny + K-Means+ Soft-NMS + scale3	95.40	82.90	93.72	0.89	9.0
YOLOv4-Tiny + K-Means+ Soft-NMS + scale3 + ECA(EM-YOLOv4-Tiny)	96.7	85.90	94.54	0.90	10.4

scale3 represents an improved strategy for employing multiscale detection in the network.

After obtaining the anchor box using the K-means clustering algorithm, the mAP and F1 values of the model were 1.2% and 2% higher than the original values, respectively, indicating a better match in size between the anchor box and the target to be detected. When using the soft-NMS algorithm to filter the prediction box, the recognition precision decreased. Still, the recall rate increased by approximately 10%, indicating the effectiveness of soft-NMS in improving missed detections. When a new functional layer was added to focus on detecting smaller targets, the detection time increased slightly, but the mAP and F1 values increased by approximately 3%. When the ECA attention mechanism was introduced into the network, the noise caused by shallow features was reduced, and Recall increased by 3%. In general, the proposed methods improved the weed detection performance of the network.

### 3.3. Performance Comparison of Different Attention Mechanisms

To further verify the advantages of the channel attention mechanism used in this study, under the same experimental conditions, the SE attention mechanism and CBAM attention machine were used as controls at the same location as the network. The experimental results are shown in Table 5.

**Table 5.** Performance comparison after using different attention modules.

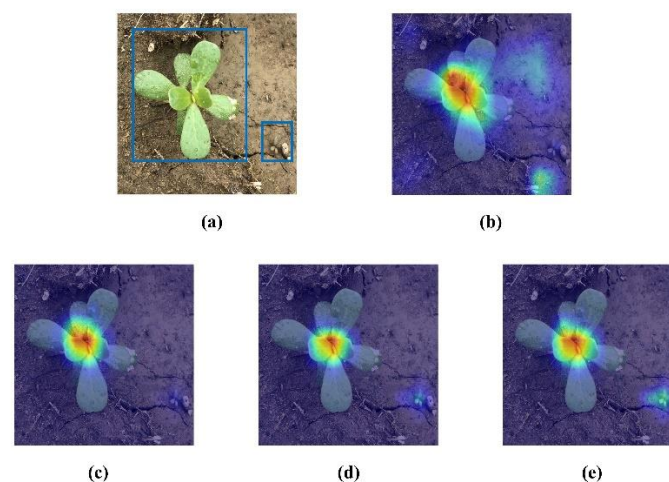
Method	Precision/%	Recall/%	mAP/%	F1/%	Time/ms
Base-SE	96.3	79.6	92.32	0.87	11
Base-CBAM	97.5	80.8	93.15	0.88	12
Base-ECA(EM-YOLOv4-Tiny)	96.7	85.9	94.54	0.90	10.4

Base represents the combined model obtained by using methods of K-Means, multiscale strategy, and soft-NMS, and its result can be found in Table 4.

Compared with the ECA attention network, the SE network uses a full connection to realize information exchange between channels, which increases the computational load and causes feature loss due to dimensionality reductions. The CBAM network is a convolutional block attention module that introduces location information in the channel dimension using the global maximum pool. However, it is limited to local range information instead of long-range dependent information. As shown in Table 5, after different attention mechanisms were added, each performance index improved compared with those of the original model. Among them, the ECA attention module outperformed the others; its mAP was higher than those of the other two attention modules by 2.22% and 1.39%, respectively, implying that the ECA network is more suitable for the model used in this study.



Similarly, in order to further explore the impact of the attention module on the weed detection model, the grad cam method was used in this study to visually analyze the features of the networks before and after adding the attention mechanism. From the detection results in Figure 5, we know that when the attention mechanism module is not added, the network will appear to pay attention to the background information when performing the detection. In contrast, the network incorporating the attention mechanism pays more attention to the information of the object to be detected through the recalibration of the weights. Comparing the feature visualization results of the three attention networks, the ECA network used in this study shades darker on the small target weeds in the images, indicating more attention to the information of small targets.



**Figure 5.** Visual heat map of attentional mechanisms, where (a) represents the original image; (b) represents the results of using the base model; (c) represents the results of using the base model and the SE attention mechanism; (d) represents the results of using the base model and the CBAM attention mechanism; (e) represents the results of using the base model and the ECA attention mechanism.

### 3.4. Comparison of Performance with Different Network Models

To verify the efficiency and practicability of the proposed model, several classical target detection models, such as YOLOv4, YOLOv5s, and the Faster-RCNN, were used to test the efficiency of weed detection. In the comparison experiments, strict control was exerted over the parameters. Specifically,  $416 \times 416$  images were used uniformly as the input to the training network, and identical training and test sets were used throughout the experiments. The results are shown in Table 6.

**Table 6.** Performance comparison results of multiple target detection networks.

Model	mAP/%	F1/%	Time/ms	Volume/MB	Parameter/ $\times 10^6$
Faster-RCNN	84.90	0.78	121	111.4	28.3
YOLOv4	89.76	0.80	25.2	234	64.0
YOLOv5s	87.78	0.86	15	27.1	7.1
Swin-Transformer	89.70	0.89	20.4	117.8	30.8
DETR	95.3	0.92	32.7	158.9	41
EM-YOLOv4-Tiny	94.54	0.90	10.4	27.8	6.8

As shown in Table 6, the average recognition accuracy of all types of networks for weeds exceeded 85%. The mAP of the EM-YOLOv4-Tiny network proposed herein was 94.54%, and its F1 value was 0.9, which is higher than those of the other four target detection networks. Because the test set contained a few smaller target weeds, the Faster-RCNN network did not construct an image pyramid and was insensitive to the detection of smaller targets, resulting in a low Recall and a mAP of only 87.71%. Compared with YOLOv4, the proposed network introduced multiscale detection and the attention mechanism based on

YOLOv4-Tiny, whose mAP and F1 were 4.78% and 10% higher than those of the YOLOv4 network, respectively. Moreover, the volume and number of parameters of the proposed model were much smaller than those of the original YOLOv4 network, indicating that the improved network preserved the merit of lightness. The lightweight YOLOv5s and EM-YOLOv4-Tiny exhibited similar model volumes and testing times; however, the mAP of YOLOv5s was only 87.78%, which was similar to that of the original YOLOv4-Tiny. Although the lightweight network had a simple structure, it was susceptible to overlooking occluded and smaller targets during detection.

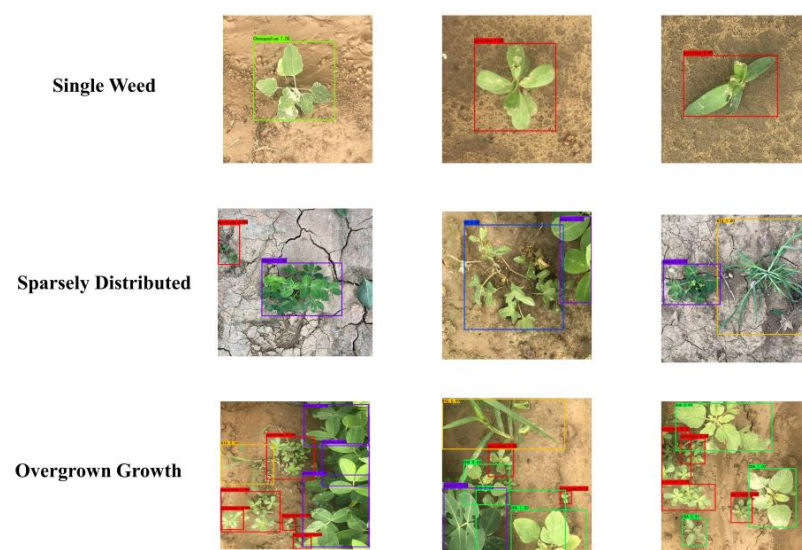
Transformer-based target detection networks like Swin-Transformer and DETR were also trained and tested on the dataset in this study. The recognition accuracy is generally better than that of the CNN-based network. Still, the size of the model and the slow detection speed is not conducive to the deployment and development of embedded devices. It is worth mentioning that the Transformer structure is on an unstoppable trend to overtake the CNN structure in the existing studies. In future research, this study will also consider incorporating the Transformer structure into EM-YOLOv4-Tiny, working to improve the accuracy of the model further.

### 3.5. Comparison of Performances under Different Scenarios

To evaluate the robustness of the model in different scenarios, three different datasets were prepared based on the different growth densities of peanuts and weeds: single weed, sparsely distributed weeds, and overgrown weeds. The test results obtained using the proposed network on the three datasets are shown in Table 7 and Figure 6. The experimental results show that the proposed model performed favorably in terms of weed detection under different growing conditions and accurately located peanuts and various weeds via boundary box regression. The average recognition accuracies of the three datasets mentioned above were 98.48%, 98.16%, and 94.3%, respectively, with a mean value of 96.98%. When the density of peanuts and weeds was high, the model accurately identified occluded weeds while demonstrating excellent recognition of small target weeds.

**Table 7.** Performance comparison results of models in different scenarios.

Scenarios	Precision/%	Recall/%	mAP/%	F1/%
Single Weed	94.67	96.03	98.48	0.95
Sparsely Distributed	95.97	93.21	98.16	0.94
Vigorous Growth	90.24	89.52	94.30	0.90
Mean	93.62	93.01	96.98	0.93



**Figure 6.** Effect of model recognition under different scenarios.

## 4. Discussion

### 4.1. Deep Learning for Weed Detection

In this study, the target detection technology based on deep learning was used to detect weeds in peanut fields and achieved good results. In similar weed detection work, many researchers [31,32] used unmanned aerial vehicles (UAVs) with intelligent sensors to detect weeds in the field. The UAV can cover a large area in a short time and generate a weed map of the field to guide the weeding device to the designated area for weeding. However, producing a weed map is very challenging due to the similarity of the crops and the weeds. In contrast, deep learning technology can automatically learn the discriminant characteristics between crops and weeds through a deep convolution neural network, which can better solve the problem of weed detection in a complex environment. Hussain [33] used the improved YOLOv3-Tiny network model to detect two kinds of weeds in the wild blueberry field, and the F1 values of the two kinds were 0.97 and 0.90, respectively. This also shows the great potential of the deep learning method in the field of weed detection. However, the actual agricultural production environment is often changeable and uncontrollable. The proposed method may also have certain limitations when the application scenario changes, such as a large increase in weed species and extreme weather. Although deep learning technology has a strong learning and adaptive ability, it must be combined with many other technologies to contribute to agricultural development.

### 4.2. Challenge of Small Target Detection

Small target detection has always been a research hotspot in the field of target detection. Multiscale detection and feature fusion are the most commonly used methods to solve the problem of small target detection. In this study, the idea of multiscale detection and the attention mechanism were introduced into YOLOv4-Tiny, which improved the recognition ability of the model for small target weeds. The multiscale feature learning method improves the sensitivity of the original network to small target detection by fusing the details of shallow features. The attention module recalibrates the input features with weights, which makes up for the defect that the receptive field of shallow features is small and easily produces noise. However, the existing feature fusion methods, such as concatenation, cannot fully take into account the feature information of the context, which also leads to the model missing or falsely detecting weeds on some small targets. In agricultural production, many application scenarios for small target detection will also exist. The pests are too small and mostly have protective colors, making pest detection a challenge in the pest control process. The accurate identification and positioning of small fruits and vegetables is also key to fruit and vegetable picking. Therefore, small target detection remains a more significant challenge in agriculture. Fortunately, the detection regarding small targets has been ongoing. Wei et al. [34] used a Path Aggregation Feature Pyramid Network (PAFPN) structure to fuse the multiscale features obtained by the Attention Mechanism Network to get high-level multiscale semantic features. The global feature fusion method, like PAFPN, is better than the local feature fusion method in small target detection. Therefore, in subsequent research we will consider adding appropriate feature fusion algorithms to our own networks to further improve the recognition ability of the model for small targets.

### 4.3. Limitations and Shortcomings

Although the network proposed in this study can better identify weeds in peanut fields, some noteworthy problems still need further research. First of all, the data in this study only include weeds in the peanut seedling stage, and the collected area is only in Henan Province, China. Future research will focus on collecting weed data in peanuts in other growth stages and will cover as many regions as possible. Secondly, although the network in this paper improves the recognition accuracy of the model compared with the original YOLOv4-Tiny network, it also increases the volume of the model to a certain extent. Zhang et al. [35] used the deep separation convolutional network to replace the

original convolutional network, which not only improved the accuracy of the model but also reduced the number of parameters and calculations of the network. In subsequent research, we plan to introduce this method into the network of this paper. Finally, the improvement strategy of the multiscale detection and the attention mechanism has been proved to be highly practical in this study. Still, other advanced research continues, such as on the Transformer [36], the Generative Adversarial Network [37], and so on, which have attracted more and more attention. It is worth further exploring the introduction of these technologies into our own network and improving the detection performance of the model.

## 5. Conclusions

To rapidly and accurately identify various types of weeds in peanut fields, a weed recognition method named EM-YOLOv4-Tiny was proposed. Based on YOLOv4-Tiny, multiscale detection and the attention mechanism were introduced, the CIOU was used as the loss function for training, and the soft-NMS method was used to screen the prediction box to improve the model performance in identifying small targets. The proposed model shows better recognition accuracy than Faster-RCNN, YOLOv5s, YOLOv4, and Swin-Transformer. In addition, the volume of the EM-YOLOv4-Tiny model was 28.7 M, and the single detection time was 10.9 ms, which rendered the model suitable for the embedded development of intelligent weeding robots.

In future work, this research will transplant the constructed model to a suitable embedded device for testing and select an intelligent spraying device to complete the precise weeding in the peanut field. In addition, the model will also be used in applications on smartphones so that farmers can better understand field information and make timely decisions.

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

**Funding:** This research was funded by the Key Research and Development program of Henan Province (No. 212102110028) and Henan Provincial Science and Technology Research and Development Plan Joint Fund (No. 22103810025).

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

**Data Availability Statement:** The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The author would like to thank all contributors to this study.

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

## Appendix A

The ECA network structure is shown in Figure A1. In the ECA network, a fast one-dimensional convolution with a convolution kernel  $k$  was performed to realize local cross-channel interactions, which reduced the computational workload and complexity of the entire connection layer. A positive interaction occurred between the channel dimension  $C$  and the convolution kernel size  $k$ , i.e., a larger  $C$  resulted in a larger  $k$ . The relationship between the two can be expressed as follows:

$$C = \varnothing(k) \quad (A1)$$

$C$  is typically measured in an exponential multiple of 2. Therefore, the relationship between the two can be more reasonably expressed as follows:

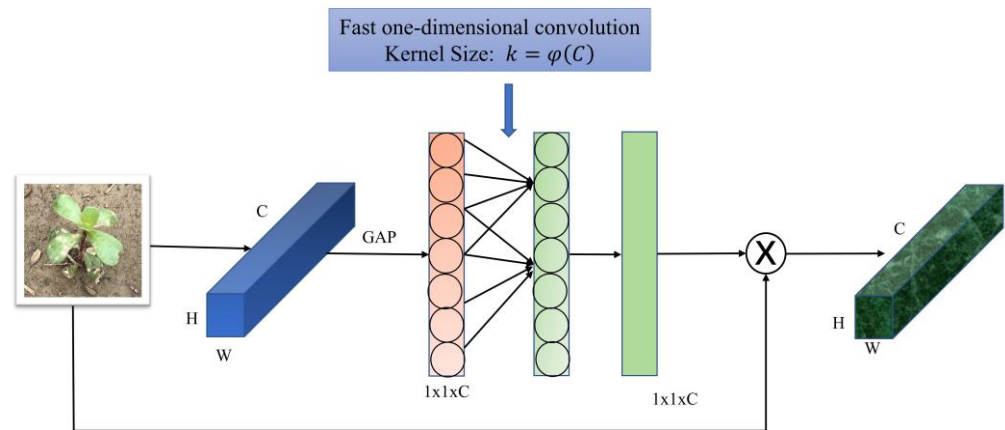
$$C = \varnothing(k) = 2^{(\gamma \times k - b)}, \quad (A2)$$



Here,

$$k = \varphi(C) = \left\lceil \frac{\log_2(C)}{r} + \frac{b}{\gamma} \right\rceil_{\text{odd}}, \quad (\text{A3})$$

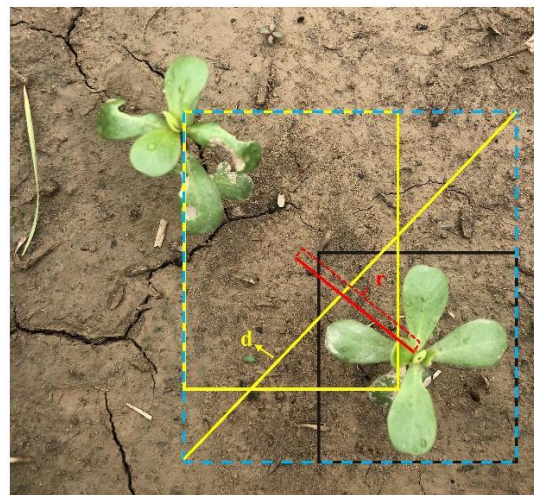
where  $|n|_{\text{odd}}$  represents the odd number closest to  $n$ , with  $\gamma$  and  $b$  being 2 and 1, respectively.



**Figure A1.** ECA network structure, where  $C$  is the channel dimension of the input data,  $H$  is the height of the input data, and  $W$  is the width of the input data. GAP denotes global average pooling, and  $k$  denotes the size of the convolution kernel using fast one-dimensional convolution.

## Appendix B

As shown in Figure A2, the CIoU bounding box regression loss function directly minimizes the normalized distance between the predicted box and the real target box, taking into account the overlapping area of the detection box as well as the distance from the center point of the detection box. The measurement parameter of the consistency of the aspect ratio between the detection frame and the real target frame is also added to make the model more inclined to optimize in the direction of the dense overlapping area.



**Figure A2.** CIoU diagram, where  $r$  represents the center point distance  $d$  of the two detection boxes, and  $d$  represents the distance between the diagonals of the smallest rectangle containing the two detection boxes.

The loss function of the CIoU is calculated as follows:

$$\text{CIoU}_{\text{Loss}} = 1 - \text{CIoU} = 1 - \text{IoU} + \frac{\rho^2(b, c)}{d^2} + av \quad (\text{A4})$$

where  $d$  represents the distance between the diagonals of the smallest rectangle containing the two boxes;  $b$  and  $c$  represent the coordinates of the central points of the real and prediction boxes, respectively;  $\rho^2(b, c)$  is the function for solving the Euclidean distance between the two mentioned points; and  $av$  is the penalty term for border scale.

The  $a$  in Equation (7) is the parameter used to balance the ratio, and  $v$  is the parameter that measures whether the ratio of the true frame is consistent with the predicted frame. The calculation of both is as follows:

$$v = \frac{4}{\pi^2} \left\{ \arctan \frac{w^c}{h^c} - \arctan \frac{w^b}{h^b} \right\}^2 \quad (A5)$$

$$a = \begin{cases} 0, & \text{if } IoU < 0.5 \\ \frac{v}{(1-IoU)+v}, & \text{if } IoU \geq 0.5 \end{cases} \quad (A6)$$

where  $w^c$  and  $h^c$  represent the width and height of the prediction box, and  $w^b$  and  $h^b$  represent the width and height of the real box.

## References

1. Renton, M.; Chauhan, B.S. Modelling crop-weed competition: Why, what, how and what lies ahead? *Crop Prot.* **2017**, *95*, 101–108. [CrossRef]
2. Zhuang, J.; Li, X.; Bagavathiannan, M.; Jin, X.; Yang, J.; Meng, W.; Li, T.; Li, L.; Wang, Y.; Chen, Y.; et al. Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Manag. Sci.* **2022**, *78*, 521–529. [CrossRef] [PubMed]
3. Kanagasingham, S.; Ekpanyapong, M.; Chaihan, R. Integrating machine vision-based row guidance with GPS and compass-based routing to achieve autonomous navigation for a rice field weeding robot. *Precis. Agric.* **2020**, *21*, 831–855. [CrossRef]
4. 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]
5. Reedha, R.; Dericquebourg, E.; Canals, R.; Hafiane, A. Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images. *Remote Sens.* **2022**, *14*, 592. [CrossRef]
6. Peteinatos, G.G.; Weis, M.; Andújar, D.; Ayala, V.R.; Gerhards, R. Potential use of ground-based sensor technologies for weed detection. *Pest Manag. Sci.* **2014**, *70*, 190–199. [CrossRef]
7. García-Santillán, I.D.; Pajares, G. On-line crop/weed discrimination through the Mahalanobis distance from images in maize fields. *Biosyst. Eng.* **2018**, *166*, 28–43. [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. Pulido, C.; Solaque, L.; Velasco, N. Weed recognition by SVM texture feature classification in outdoor vegetable crop images. *Ing. E Investig.* **2017**, *37*, 68–74. [CrossRef]
10. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
11. Gai, R.; Chen, N.; Yuan, H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Comput. Appl.* **2021**, 1–12. [CrossRef]
12. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Anwar, S. Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. *Precis. Agric.* **2021**, *22*, 1711–1727. [CrossRef]
13. Jin, X.; Sun, Y.; Che, J.; Bagavathiannan, M.; Yu, J.; Chen, Y. A novel deep learning-based method for detection of weeds in vegetables. *Pest Manag. Sci.* **2022**, *78*, 1861–1869. [CrossRef]
14. Ying, B.; Xu, Y.; Zhang, S.; Shi, Y.; Liu, L. Weed detection in images of carrot fields based on improved YOLO v4. *Traitement Du Signal* **2021**, *38*, 341–348. [CrossRef]
15. Li, X.; Pan, J.; Xie, F.; Zeng, J.; Li, Q.; Huang, X.; Liu, D.; Wang, X. Fast and accurate green pepper detection in complex backgrounds via an improved YOLOv4-tiny model. *Comput. Electron. Agric.* **2021**, *191*, 106503. [CrossRef]
16. Li, H.; Li, C.; Li, G.; Chen, L. A real-time table grape detection method based on improved YOLOv4-tiny network in complex background. *Biosyst. Eng.* **2021**, *212*, 347–359. [CrossRef]
17. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
18. Gao, C.; Cai, Q.; Ming, S. YOLOv4 object detection algorithm with efficient channel attention mechanism. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China, 25–27 December 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 1764–1770.
19. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—improving object detection with one line of code. In Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.

20. Neubeck, A.; Van Gool, L. Efficient non-maximum suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; IEEE: Piscataway Township, NJ, USA, 2006; Volume 3, pp. 850–855.
21. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [[CrossRef](#)]
22. Wang, L.; Qin, M.; Lei, J.; Wang, X.; Tan, K. Blueberry maturity recognition method based on improved YOLOv4-Tiny. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 170–178.
23. Xu, J.; Li, Z.; Du, B.; Zhang, M.; Liu, J. Reluplex made more practical: Leaky ReLU. In Proceedings of the 2020 IEEE Symposium on Computers and communications (ISCC), Rennes, France, 7–10 July 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 1–7.
24. Chen, Z.; Tian, S.; Yu, L.; Zhang, L.; Zhang, X. An object detection network based on YOLOv4 and improved spatial attention mechanism. *J. Intell. Fuzzy Syst.* **2022**, *42*, 2359–2368. [[CrossRef](#)]
25. Choi, E.; Bahadori, M.T.; Sun, J.; Kulas, J.; Schuetz, A.; Stewart, W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *arXiv* **2016**, arXiv:1608.05745.
26. Schmidt-Hieber, J. Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Stat.* **2020**, *48*, 1875–1897.
27. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
28. Zhou, T.; Fu, H.; Gong, C.; Shen, J.; Shao, L.; Porikli, F. Multi-mutual consistency induced transfer subspace learning for human motion segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10277–10286.
29. Zhong, S.; Chen, D.; Xu, Q.; Chen, T. Optimizing the Gaussian kernel function with the formulated kernel target alignment criterion for two-class pattern classification. *Pattern Recognit.* **2013**, *46*, 2045–2054. [[CrossRef](#)]
30. Ismikhani, H. Ik-means—+: An iterative clustering algorithm based on an enhanced version of the k-means. *Pattern Recognit.* **2018**, *79*, 402–413. [[CrossRef](#)]
31. Eide, A.; Koparan, C.; Zhang, Y.; Ostlie, M.; Howatt, K.; Sun, X. UAV-Assisted Thermal Infrared and Multispectral Imaging of Weed Canopies for Glyphosate Resistance Detection. *Remote Sens.* **2021**, *13*, 4606. [[CrossRef](#)]
32. De Castro, A.I.; Torres-Sánchez, J.; Peña, J.M.; Jiménez-Brenes, F.M.; Csillik, O.; López-Granados, F. An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sens.* **2018**, *10*, 285. [[CrossRef](#)]
33. Hussain, N.; Farooque, A.A.; Schumann, A.W.; McKenzie-Gopsill, A.; Esau, T.; Abbas, F.; Acharya, B.; Zaman, Q. Design and development of a smart variable rate sprayer using deep learning. *Remote Sens.* **2020**, *12*, 4091. [[CrossRef](#)]
34. Wei, H.; Zhang, Q.; Qian, Y.; Xu, Z.; Han, J. MTSDet: Multi-scale traffic sign detection with attention and path aggregation. *Appl. Intell.* **2022**, 1–13. [[CrossRef](#)]
35. Zhang, M.; Xu, S.; Song, W.; He, Q.; Wei, Q. Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion. *Remote Sens.* **2021**, *13*, 4706. [[CrossRef](#)]
36. Kitaev, N.; Kaiser, L.; Levskaya, A. Reformer: The efficient transformer. *arXiv* **2020**, arXiv:2001.04451.
37. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]