Mathematical Biosciences and Engineering

MBE, 20(11): 19341–19359.
DOI: 10.3934/mbe.2023855
Received: 22 July 2023
Revised: 28 September 2023
Accepted: 12 October 2023
Published: 17 October 2023

http://www.aimspress.com/journal/MBE

Research article

A peanut and weed detection model used in fields based on BEM-YOLOv7-tiny

Yong Hua[1], Hongzhen Xu[1,2], Jiaodi Liu[1,2,*], Longzhe Quan[3], Xiaoman Wu[1] and Qingli Chen[1]

[1] College of Mechanical and Control Engineering, Guilin University of Technology, Guilin 541004, China
[2] Key Laboratory of Advanced Manufacturing and Automation Technology (Guilin University of Technology), Education Department of Guangxi Zhuang Autonomous Region, Guilin 541006, China
[3] College of Engineering, Anhui Agricultural University, Hefei 230036, China

* **Correspondence:** Email: Y_H0522@163.com.

Abstract: Due to the different weed characteristics in peanut fields at different weeding periods, there is an urgent need to study a general model of peanut and weed detection and identification applicable to different weeding periods in order to adapt to the development of mechanical intelligent weeding in fields. To this end, we propose a BEM-YOLOv7-tiny target detection model for peanuts and weeds identification and localization at different weeding periods to achieve mechanical intelligent weeding in peanut fields at different weeding periods. The ECA and MHSA modules were used to enhance the extraction of target features and the focus on predicted targets, respectively, the BiFPN module was used to enhance the feature transfer between network layers, and the SIoU loss function was used to increase the convergence speed and efficiency of model training and to improve the detection performance of the model in the field. The experimental results showed that the precision, recall, mAP and F1 values of the BEM-YOLOv7-tiny model were improved by 1.6%, 4.9%, 4.4% and 3.2% for weed targets and 1.0%, 2.4%, 2.2% and 1.7% for all targets compared with the original YOLOv7-tiny. The experimental results of positioning error show that the peanut positioning offset error detected by BEM-YOLOv7-tiny is less than 16 pixels, and the detection speed is 33.8 f/s, which meets the requirements of real-time seedling grass detection and positioning in the field. It provides preliminary technical support for intelligent mechanical weeding in peanut fields at different stages.

Keywords: peanut and weed detection; different weeding periods; BEM-YOLOv7-tiny; precision weeding; deep learning

1. Introduction

Weed competition is an important factor limiting peanut production, and studies have proven that weeds seriously affect the yield of peanuts, with 5 weeds per square meter in peanut fields can reduce peanut yield by 13.89% and 20 weeds can reduce peanut yield by up to 48.31% [1]. There are about 80 species of weeds in Chinese peanut fields, belonging to about 30 families [2]. Field weed management is an important tool for superior peanut production, and weed control in peanut fields is currently mainly based on herbicide spraying, but it can cause irreversible farmland pollution. In compliance with the requirements of precision agriculture and green agricultural production advocated by the World Food and Agriculture Organization, providing an efficient and nondestructive peanut and weed identification and positioning method is a prerequisite and key to achieving precision mechanical weed control in order to achieve effective weed control in the field and ensure green peanut production.

Real-time accurate identification and localization of seedling and weed plants is a prerequisite for automatic weeding in the field. In recent years, machine vision technology has been widely used in the field for plant and weed identification research due to its advantages of low cost and convenience. Traditional machine vision techniques mainly rely on manual extraction of important features such as color, texture and shape to achieve plant and weed recognition, such as Shen [3], Wang [4] and Li [5] for field plant and weed classification and recognition based on single features of images, and Deng et al. [6] for seedling and weed differentiation by multiple feature fusion. However, traditional machine vision techniques take a lot of time to extract important features of the target, and the selection of important features is also affected by human factors, which does not cope with the actual variable field environment.

Deep learning automatically extracts multi-scale and multi-dimensional feature information of seedling and weed targets by virtue of convolutional neural networks, which solves the deficiencies in feature selection of traditional methods. As two commonly used deep learning techniques, image classification and target detection distinguish targets by autonomously learning important features, as Dyrmann et al. [7] and Tao et al. [8] used image classification techniques to classify different plants in the field, and Zong et al. [9] and Zhang et al. [10] based on second-stage target detection techniques Faster R-CNN model and Mask R-CNN models achieved a precision of 91.49% and 94% for field maize crop identification, respectively. Compared with image classification, target detection techniques can provide target location information by outputting detection frame coordinates, and are widely used in field seedling and weed identification and localization research. Xu et al. [11] and Jiang et al. [12] achieved the detection of seedlings and weeds in cotton and corn fields with a precision of more than 91% based on the second-stage target detection model, and the detection time of a single image was 0.26 and 0.98 s, respectively. Li et al. [13] achieved the detection of green pepper in a field with a precision of 96.9% based on the single-stage YOLO target detection technique, and the detection time of a single image was 6.2 ms. The single-stage YOLO model outperformed the second-stage target detection model in real-time detection and was used by Gao [14], Quan [15,16], Gao [17], Ahmad [18], etc. in the field seedling and weed targets detection study. Additionally, the YOLO network is compatible with other networks, such as Zhang et al. [19] and Wang et al. [20] to enhance

the detection performance of YOLOv5 model for field plants and weeds by incorporating different attention mechanisms.

At present, the research on the application of machine vision technology in peanut fields mainly focuses on a certain period, such as Zhang et al. [21] realized the identification of small target weeds in peanut fields during the seedling stage based on the YOLOv4-tiny model, and Lin et al. [22] realized the counting of early peanut seedlings in the field based on YOLOv5. Aiming at the problem that there are many types of weeds in the field, and it is difficult to recognize the many targets for single weed detection, as well as the demand for multiple weeding in the field, the study proposes a BEM-YOLOv7-tiny seedling and weed identification model, which enhances the learning ability of the network for extracting target features and transferring ability between different layers by introducing ECA module and BiFPN module, uses MHSA to enhance the focus on predicted targets at the detection end and uses SIoU loss function to improve the convergence speed and efficiency of the model training process to further enhance the network detection model performance. The model is capable of real-time and accurate detection of peanuts and weeds at each mechanical weeding period, providing a general peanuts and weeds detection and identification model for mechanical weeding at each weeding time in peanut fields, and providing a key technology for real-time and accurate mechanical weeding in peanut fields.

2. Materials and methods

2.1. Data material

2.1.1. Data acquisition

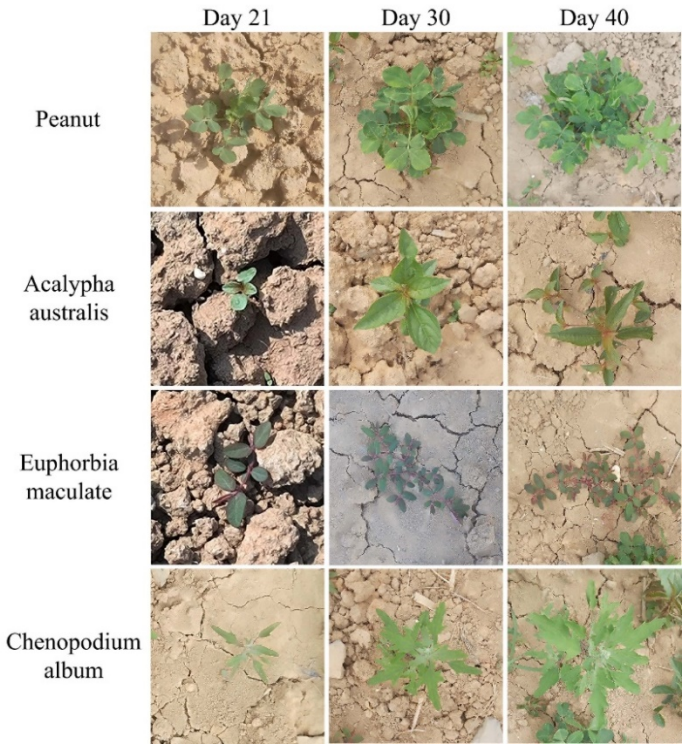


Figure 1. Images of peanut and weeds at different periods.

The two peak weed emergence periods in spring-planted peanut fields are 10–15 days and 35–50 days after sowing, which account for about 50% and 30% of weed occurrence in the whole field [23], respectively. Wann et al. [24] also showed in their study on organic peanut weed control methods that weeds in peanut fields need to be controlled several times to ensure peanut yield. For this reason, in this study, field data were collected for three weeding periods when mechanical weeding could be used in peanut fields to meet the needs of mechanical weeding operations at different weeding stages of peanuts in the field, with the first time image collected on day 21 after sowing (this period is the peanut seedling stage when weeds begin to emerge), the second time image collected on day 30 after sowing (this period is the peanut regrowth stage) and the third time image collected on day 40 days after sowing (at this time, some peanuts began to flower, for the pre-flowering period). The morphological images of peanuts and the three major weeds at different weeding periods are shown in Figure 1. The MV-HS510GC camera was used to capture images vertically downward from the ground height of 40–50 cm according to the requirements of the mechanical weeding device, and the image size was 2056 × 2464 pixels, and later the images were uniformly processed to 640 × 640 pixels.

2.1.2. Data enhancement and labeling

Table 1. Original image composition.

Acquisition period	Training and validation	Testing	Total
Day 21	160	35	195
Day 30	160	35	195
Day 40	280	80	360
Total	600	150	750

Table 2. Dataset after data enhancement.

Dataset	Number
Original images	600
Brightness increase	300
Brightness decrease	300
Mean blur	300
Gaussian blur	300
Total	1800

The composition of the original captured images is shown in Table 1. In order to simulate the real environment in the field as much as possible, the data enhancement method proposed by Liu [25] was used to simulate the effects of different light intensities in the field and vibrations generated by mechanical weeding operations in the field, respectively, and image enhancement was performed by brightness processing and fuzzy processing on the collected original images to ensure the representativeness and diversity of the sample images and to achieve the purpose of enriching the data set and improving the training accuracy of the model. The 600 images were expanded to 1800 images by image enhancement and divided into training and validation sets in the ratio of 7:3, and the remaining 150 images were used as test sets. The brightness adjustment factor in image enhancement is 1.5 and 0.5 times the original image respectively, and the kernel size used for blurring is 5 × 5. The

specific number of image enhancements is shown in Table 2.

The open-source image annotation tool LabelImg was used to manually label the images by drawing tight external rectangular boxes around the plants to produce a dataset in txt data format, and the image annotation schematic is shown in Figure 2. The three annual broadleaf weed targets were uniformly labeled as a single class to reduce the number of weed species detected in the field by referring to the seedling weed detection method proposed by Alessandro et al. [26].

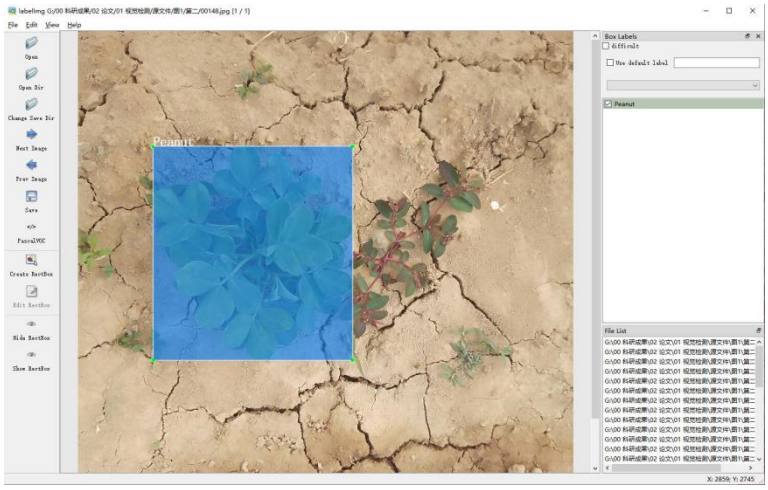


Figure 2. Schematic diagram of target bounding box annotation.

2.2. BEM-YOLOv7-tiny peanut and weed detection algorithm

2.2.1. BEM-YOLOv7-tiny network structure

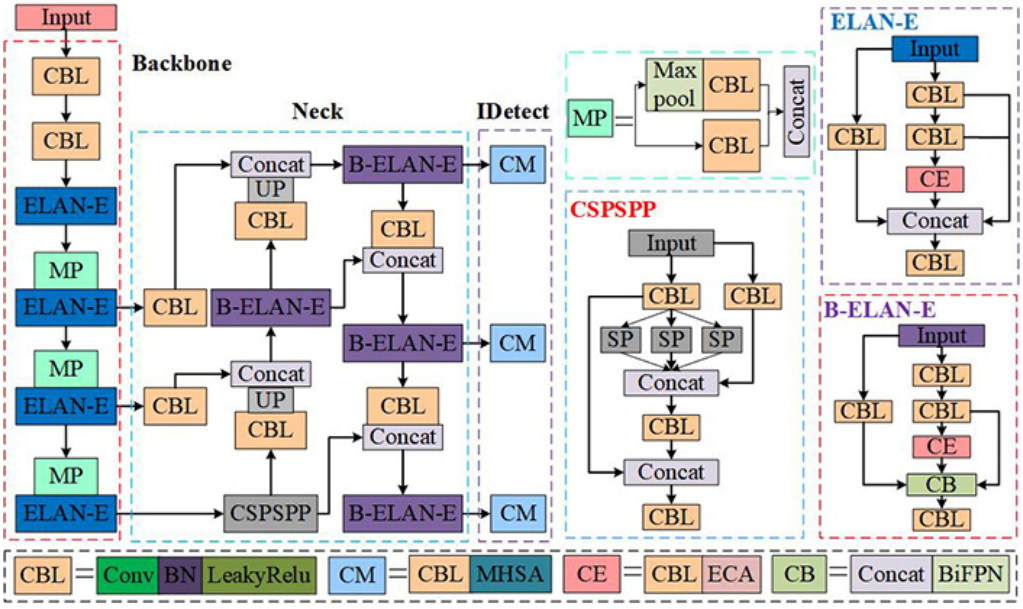


Figure 3. BEM-YOLOv7-tiny model structure.

Based on YOLOv7-tiny, the study enhances the extraction of peanut and weed features in the field at different weeding stages by introducing an ECA attention mechanism in the convolution module and enhances the feature transfer and fusion using BiFPN to improve the extraction and fusion of target features by the network model. Then MHSA is used to enhance the attention to the predicted targets at the detection end to improve the detection ability at the detection end. Finally, the regression accuracy is improved by improving the loss function to accelerate the model convergence. The final network generates three detection layers of different sizes to detect peanut and weed targets, and the structure of the BEM-YOLOv7-tiny algorithm model is shown in Figure 3.

2.2.2. ECA model

ECA [27] is an extremely lightweight channel attention mechanism that adds little complexity to the model but enables the model to focus on more important information. ECA utilizes a local cross-channel interaction strategy without dimensionality reduction to effectively avoid the effect of dimensionality reduction on the channel attention learning effect; the network performance is maintained through proper cross-channel interaction, and significant effect gain can be obtained by only a few parameter adjustments, whose network structure is shown in Figure 4.

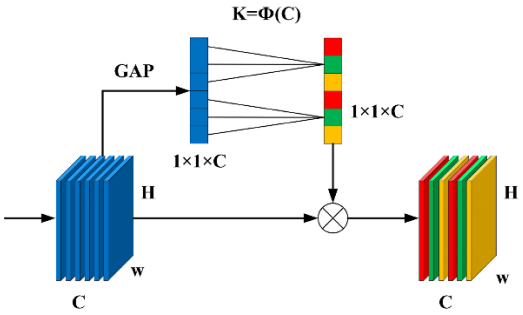


Figure 4. ECA model structure schematic.

In the ECA module, first, the input feature map is globally averaged and pooled to transform the feature map from a matrix of [h,w,c] into a vector of [1,1,c], and the features of each channel are represented by a single value. Second, the adaptive 1D convolution kernel k is calculated based on the number of channels of the feature map, and the 1D convolution of size k is executed to obtain the weights of each channel of the feature map to achieve inter-channel information exchange. Finally, the normalized weights are fused with the original input features to generate a weighted feature map with channel attention.

The ECA is fused into the ordinary convolutional module so that the ordinary convolutional module combined with ECA enhances the focus on important features of peanut and weed targets at different stages. By replacing the original ordinary convolution module with the convolution inserted into the very light ECA module, the network model can avoid the learning of non-target redundant features in images without decreasing the channel dimension and significantly increasing the memory and network depth, while effectively taking into account the important features of peanut and weed targets to improve the detection efficiency of the network model.

2.2.3. The BiFPN model

The original YOLOv7-tiny feature fusion network uses a path aggregation feature pyramid (PAFPN) network, which stitches together the information passed from the top of the feature pyramid network (FPN) with the strong localization information tensor passed from the bottom to the top of the path aggregation network (PANet) to fully fuse different levels of feature information to achieve multi-scale learning. However, tensor splicing is not comprehensive enough for the fusion of feature information in adjacent layers, while nearest neighbor interpolation upsampling does not take into account the two main goals of speed and accuracy in detection tasks, and the fusion network is prone to feature information loss. Unlike PAFPN bi-directional feature fusion, BiFPN has a complex bi-directional fusion structure and adopts cross-scale connectivity to make the network feature fusion information richer. the structure of PANet and BiFPN network is shown in Figure 5.

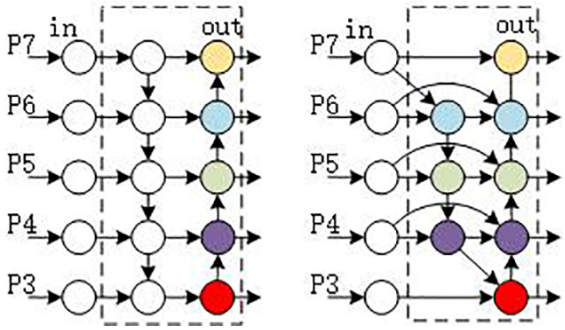


Figure 5. PANet and BiFPN structures.

As can be seen from the figure, based on the PANet structure, the BiFPN structure eliminates nodes with only a single input and adds another edge between the input and output, which not only simplifies the model but also gets more features fused. BiFPN [28] uses the unique bi-directional cross-scale connection and weighted feature fusion to achieve faster and more convenient multi-scale feature fusion by adding an extra for each feature layer weight value to each feature layer, allowing the network to learn the key features with different weight values and enabling the network to fuse features better. In this study, the weighted bidirectional feature pyramid is incorporated into the Neck layer feature information fusion tensor stitching operation to enhance the fusion of features of different sizes between Neck layers.

2.2.4. MHSA model

To improve the performance of the detection head at the output side, focusing on the pre-detection target, the multi-head self-attentive mechanism (MHSA) in the BoT [29] module of the bottleneck converter is used at the output side. MHSA achieves linear dimensionality enhancement by means of full connectivity and implements self-attentive for four heads in parallel, each self-attentive learns multiple weights individually and performs matrix multiplication with the feature weights separately to obtain the corresponding feature output information. Finally, the result information of each head is stitched together to achieve multi-angle learning feature information fusion, the structure of MHSA module is shown in Figure 6.

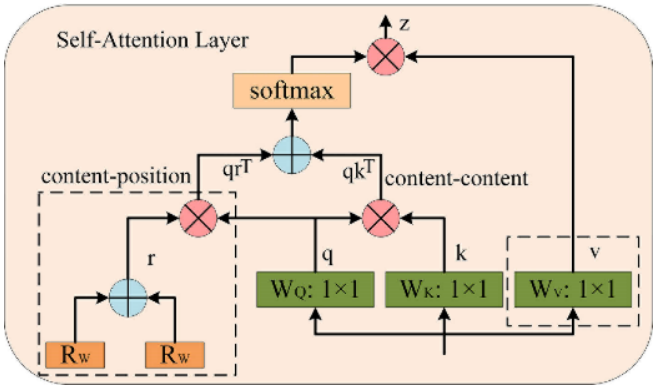


Figure 6. Multi head self-attention mechanism.

MHSA has the property of achieving global attention on the two-dimensional feature map, organically correlates the features in the two stages of network extraction and fusion, and integrates the information under global attention to achieve better detection results. By adding the MHSA module to the three-layer convolution at the detection side, the ordinary convolution is made to fuse the complex feature information after network extraction and fusion by virtue of the global self-attention property of MHSA to improve the performance at the detection side.

2.2.5. SIoU loss function

The original YOLOv7-tiny model uses CIoU [30] as the network localization loss function, defined as shown in Eq (1), with parameters α and v shown in Eqs (2) and (3). Although CIoU considers the overlap area, centroid distance and aspect ratio of the bounding box regression, it does not take into account the mismatch between the actual target and the prediction box, and reflects the relative ratio of image width and height, rather It does not reflect the real difference between image width and height respectively and their confidence levels, which leads to slow and unstable convergence.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha \mathrm{v} \tag{1}$$

$$\alpha = \frac{v}{(1-IoU)+v} \tag{2}$$

$$v = \frac{4}{\pi^2}\left(arc\tan\frac{w^{gt}}{h^{gt}} - arc\tan\frac{w}{h}\right)^2 \tag{3}$$

where, $\rho^2(b,b^{gt})$ is the penalty term, b and b^{gt} represent the centroids of B and B^{gt}, ρ^2 is the Euclidean distance, c is the minimum external rectangular diagonal distance of B and B^{gt}, α is the positive equilibrium parameter, v denotes the consistency of gt with the aspect ratio of the prediction frame, w and w^{gt} represent the width of B and G, h and h^{gt} represent the height of B and G.

Using SIoU [31] as the localization loss function of the network, the influence of distance, angle and shape cost on the model boundary regression is fully considered to make the prediction frame and the actual frame approach the parallel state faster, thus controlling the direction of convergence of the loss function.

The angular cost of SIoU is defined as shown in Eq (4):

$$\Lambda = 1 - 2 \times sin^2\left(arc\sin\left(\frac{z_h}{\sigma}\right) - \frac{\pi}{4}\right) \tag{4}$$

where, z_h is the height difference between the center point of the real frame and the predicted frame, σ is the distance between the center point of the real frame and the predicted frame.

The distance cost of SIoU is defined as shown in Eq (5):

$$\Delta = \sum_{t=x,y}(1 - e^{-\gamma\rho t}) \tag{5}$$

where, $\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2$, $\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2$, $\gamma = 2 - \Lambda$, $b_{c_x}^{gt}$ and $b_{c_x}^{gt}$ are the center coordinates of the real frame, b_{c_x} and b_{c_y} are the center coordinates of the predicted frame, c_w and c_h are the width and height of the smallest outer rectangle of the real frame and the predicted frame.

The shape cost of SIoU is defined as shown in Eq (6):

$$\Omega = \sum_{t=w,h}(1 - e^{-\omega t})^{\theta} \tag{6}$$

where, $\omega_w = \frac{|w - w^{gt}|}{max(w, w^{gt})^2}$, $\omega_h = \frac{|h - h^{gt}|}{max(h, h^{gt})^2}$, w and h are the width and height of the predicted frame, w^{gt} and h^{gt} are the width and height of the real frame, θ is a constant to control the degree of attention to shape loss.

The SIoU, which takes into account the angle, distance and shape costs, is defined as shown in Eq (7). SIoU reduces the probability that the penalty term is zero to occur and makes the function convergence smooth. In this study, SIoU is used as a network loss function to speed up the model convergence, thus improving the regression accuracy and network robustness.

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{7}$$

2.3. Evaluation indicators

To evaluate the performance of the model, the precision rate (P) and recall rate (R) are used as evaluation metrics. They are defined as shown in Eqs (8) and (9).

$$P = \frac{TP}{TP + FP} \times 100\% \tag{8}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{9}$$

where, TP represents the number of correctly detected targets, FP represents the number of incorrectly detected targets, FN represents the number of undetected targets.

The F1 value, as a weighted summed average of precision and recall, can be used to weigh the model's missed and false detection rate for peanut seedling grass targets, as defined in Eq (10).

$$F1 = \frac{2PR}{P + R} \tag{10}$$

The mean average precision (mAP) measures the overall performance of the model at different confidence thresholds. In this paper, we evaluate the model performance using the mean average precision at a threshold of 0.5, defined as shown in Eq (11). The number of parameters (Params), computation volume (FLOPs) and model volume (Volume) are also used to evaluate the complexity of the network model, and FPS (frames per second) is used to evaluate the model detection speed.

$$mAP = \frac{\sum_{N=1}^{N} PR}{N} \tag{11}$$

where, N represents the number of target species in the sample.

2.4. Model training

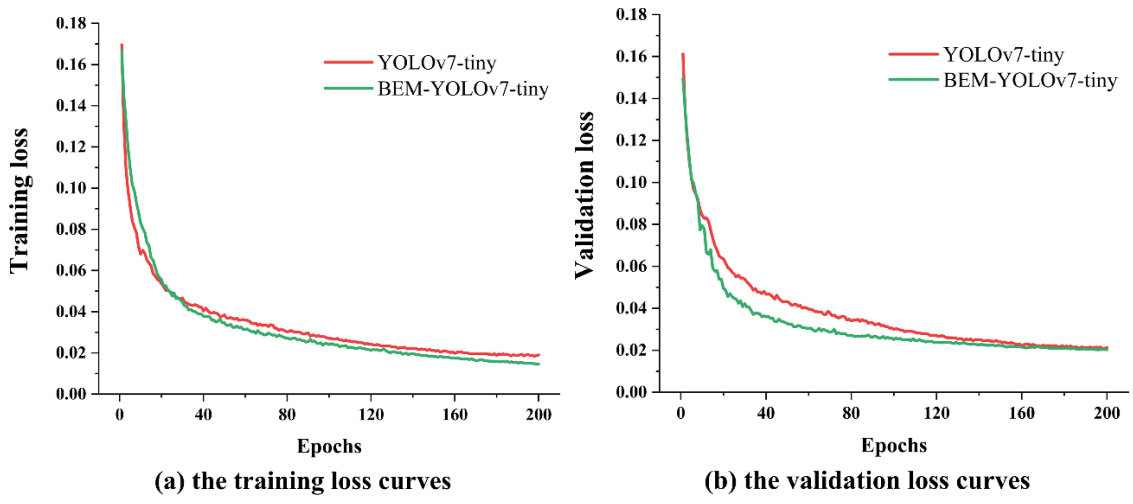
The hardware and software environments for model training and testing are shown in Table 3. The model network image input size is 640 × 640, the initial learning rate of the training process is set to 0.01, the weight decay factor is 0.0005 and the training iteration is 200 times with a batch size of 16.

Table 3. Training and test environment configuration table.

Configuration	Parameter
Operating system	Windows 10
CPU	Inter(R) Core(TM) i7-10700@2.90GHz
GPU	NVIDIA GeForce RTX 3060 12G
Accelerate environment	CUDA
Python	3.8.5
Pytorch	1.7.1

3. Results

3.1. BEM-YOLOv7-tiny model performance analysis



(a) the training loss curves **(b) the validation loss curves**

Figure 7. Loss variation curve of YOLOv7-tiny and BEM-YOLOv7-tiny.

To evaluate the performance of the improved model, the training process before and after the model improvement is compared and analyzed in terms of training loss convergence, model complexity and model performance.

The convergence curves of training loss and validation loss functions during model training are shown in Figure 7(a),(b), respectively. BEM-YOLOv7-tiny has faster loss convergence than YOLOv7-tiny model, and BEM-YOLOv7-tiny has lower values of training and validation loss, which proves that BEM-YOLOv7-tiny has better learning ability and better model performance.

The specific performance of the model for peanut and weed detection is shown in Table 4, the performance of the BEM-YOLOv7-tiny model for weed target detection was significantly improved, and the precision, recall, mAP and F1 values of the model for weed targets were 88.2%, 88.5%, 92.4% and 88.3%, respectively, which were 1.6%, 4.9%, 4.4% and 3.2% better than the original YOLOv7-tiny, respectively. The BEM-YOLOv7-tiny network model outperforms the original YOLIv7-tiny model in detecting multiple weeds as a single target. For all targets, the precision, recall, mAP and F1 values of the BEM-YOLOv7-tiny model were 93.5%, 93.9%, 96.0% and 93.7%, respectively, which were 1.0%, 2.4%, 2.2%, and 1.7% better than the original YOLOv7-tiny, respectively. Although the number of parameters of the BEM-YOLOv7-tiny model increased by 0.9M and the volume increased by 1.8 MB, the computational effort of the network model was reduced.

Table 4. Model detection performance comparison results.

Method	Targets	P/%	R/%	mAP/%	F1/%	Params/M	FLOPs/G	Volume/MB
YOLOv7-tiny	ALL	92.5	91.5	93.8	92.0	6.0	13.2	12.3
	Peanuts	98.4	99.3	99.6	98.8			
	Weeds	86.6	83.6	88.0	85.1			
BEM-YOLOv7-tiny	ALL	93.5	93.9	96.0	93.7	6.9	12.9	14.1
	Peanuts	98.9	99.2	99.6	99.0			
	Weeds	88.2	88.5	92.4	88.3			

In summary, the BEM-YOLOv7-tiny network model had better performance for peanut and weed detection in peanut fields. Compared with the YOLOv7-tiny network model, although the number of parameters and model volume of the network increased, BEM-YOLOv7-tiny had higher mAP and F1 values, indicating that the network model had a lower possibility of missing and false detection of targets, and better detection of peanut and weed targets.

3.2. Performance comparison of improved methods

In order to verify the effectiveness of the improved method on weed detection performance, the effects of different improved modules on the performance of YOLOv7-tiny network were compared and tested, and the evaluation indexes of the weed targets were compared and analyzed in order to better show the magnitude of the improvement in network performance, and the test results are shown in Table 5.

After enhancing the feature extraction ability of the network using the ECA attention mechanism, the mAP and F1 values of the model for the weed target improved by 0.4% and 0.7%, respectively, compared with the original network, indicating the better extraction ability of the ECA attention mechanism for the important features of the weed target. The mAP was improved by 0.6% when using

BiFPN to enhance feature transfer fusion in different layers of the Neck part of the network, indicating the effectiveness of BiFPN in enhancing the feature transfer fusion ability. When the detection side of the network was enhanced using the MHSA module to focus on the predicted targets, although the network detection precision was reduced, the recall and mAP were substantially improved, reducing the network's missed detection of weed targets and demonstrating the enhanced capability of the MHSA module for the network. The overall performance of the network model was further improved by using the SIoU loss function instead of the original network model loss function. Overall, the BEM-YOLOv7-tiny network model co-stacked by the improved method works best, and it has been improved in all indexes relative to the original algorithm, and the network model stability performance is better.

Table 5. Influence of different improved modules on YOLOv7-tiny network.

Method	P/%	R/%	mAP/%	F1/%
YOLOv7-tiny	86.6	83.6	88.0	85.1
YOLOv7-tiny+ECA	87.6	84.1	88.4	85.8
YOLOv7-tiny+ECA+BiFPN	88.3	84.8	90.0	86.5
YOLOv7-tiny+ECA+BiFPN+MHSA	88.0	85.9	91.0	86.9
YOLOv7-tiny+ECA+BiFPN+MHSA+SIoU (BEM-YOLOv7-tiny)	88.2	88.5	92.4	88.3

3.3. Performance comparison of different attention mechanisms

To further verify the advantages of introducing the ECA attention mechanism in the network model, the ECA attention mechanism in the BEM-YOLOv7-tiny model network was replaced with the CBAM attention and SE attention mechanisms respectively for comparison experiments, and the experimental results are shown in Table 6.

Table 6. Performance comparison of different attention models.

Model	P/%	R/%	mAP/%	F1/%	FLOPs/G	Volume/MB
Base-CBAM	92.3	92.5	95.5	92.4	13.0	14.2
Base-SE	93.7	90.1	93.9	91.9	13.0	14.2
BEM-YOLOv7-tiny	93.5	92.5	95.5	93.7	12.9	14.1

The CBAM network uses global maximum pooling to introduce location information in the channel information to enhance the attention of the network to two different channels, and although the mAP value is the same as joining the ECA network, it focuses more on the local information of the target, the F1 value is lower than the ECA model, and the network has a high possibility of missed detection and false detection. the SE network uses the fully connected approach to achieve the information exchange between channels, and the dimensionality reduction adopted leads to the target features are lost, and although the recognition precision is higher than that of ECA network, its recall rate is low, the possibility of network miss detection is high and the mAP is lower than that of ECA

network. The computational volume and model volume of CBAM and SE network models are higher than those of BEM-YOLOv7-tiny. So ECA network is better than other attention modules in this model and more suitable for the peanut and weed detection model.

3.4. Performance comparison of different deep learning models

To verify the superiority of the proposed algorithm for peanut and weed detection performance, it is compared with Faster R-CNN, YOLOv4-tiny and YOLOv5s classical target detection models, where Faster R-CNN uses ResNet50 network. The experimental results are shown in Figure 8.

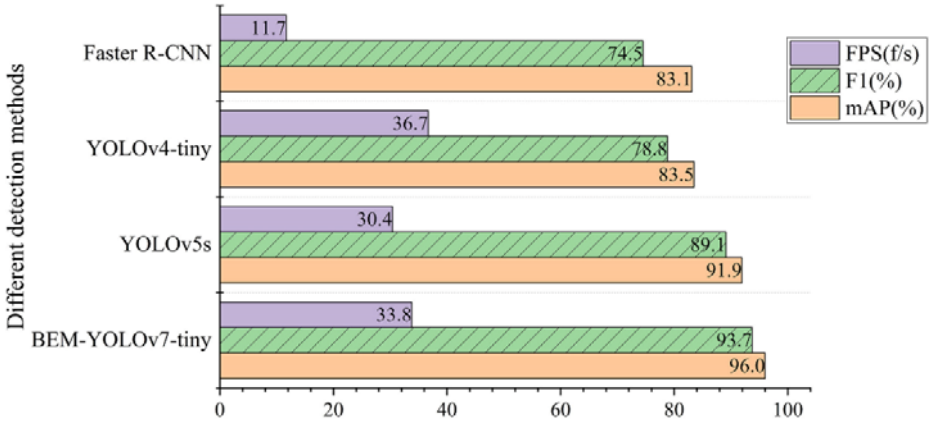


Figure 8. Performance comparison results of different target detection networks.

From the comparison results, it is clear that the BEM-YOLOv7-tiny network model has the best evaluation indexes among all models. Because the dataset classified the three weed targets into a single class, the Faster R-CNN did not build the image feature fusion pyramid, which was not sensitive to the detection of weed targets, and the detection effect was low, with mAP of 83.1% and F1 of only 74.5%, the model was easy to miss and misdetect, and the overall performance was lower than the above YOLO series network models. the BEM-YOLOv7-tiny model was slightly lower than the YOLOv4-tiny network in terms of Although the detection speed is slightly lower than that of the YOLOv4-tiny network, its comprehensive performance is the best and mAP of the model is 12.5% and 4.1% higher than that of the YOLOv4-tiny and YOLOv5s models, respectively.

As shown in Figure 9, a comparison of the test results of three high-performance network models, YOLOv5s, YOLOv7-tiny and BEM-YOLOv7-tiny, is shown, where the blue boxes are the missed detection targets. In the early image detection in the field with small weed targets, the YOLOv5s network is not sensitive to small target weeds in the image, resulting in a large number of small target weed misses, YOLOv7-tiny misses three targets and BEM-YOLOv7-tiny also misses one target. For the obvious case of weed targets, all three networks had good recognition of the targets, but the YOLOv5s and YOLOv7-tiny networks were insensitive to weeds with partial features at the edges. It indicates that BEM-YOLOv7-tiny has better recognition of weeds in peanut fields.

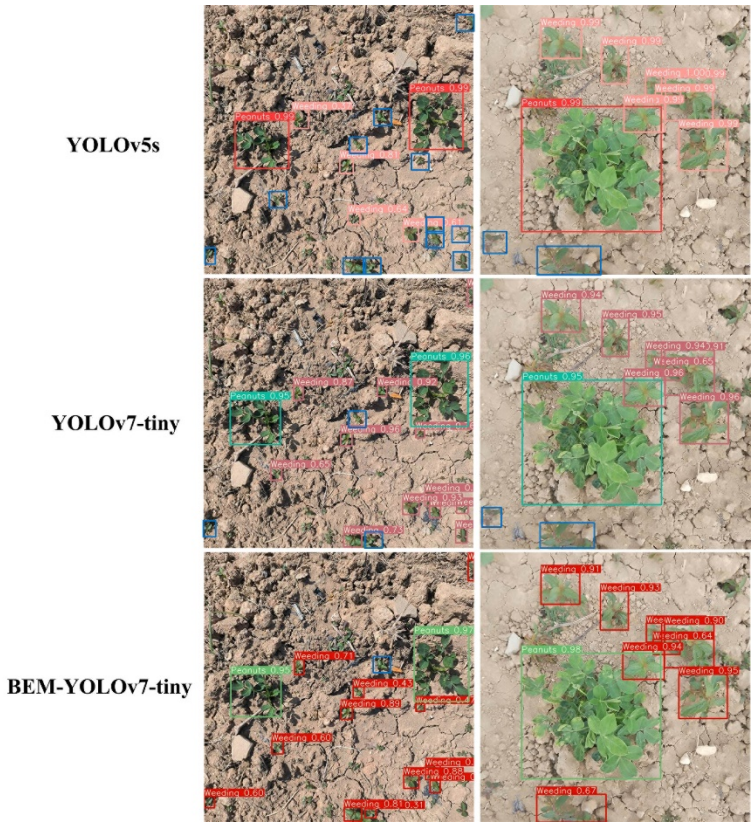


Figure 9. Effect of recognition under different models.

3.5. Analysis of positioning results

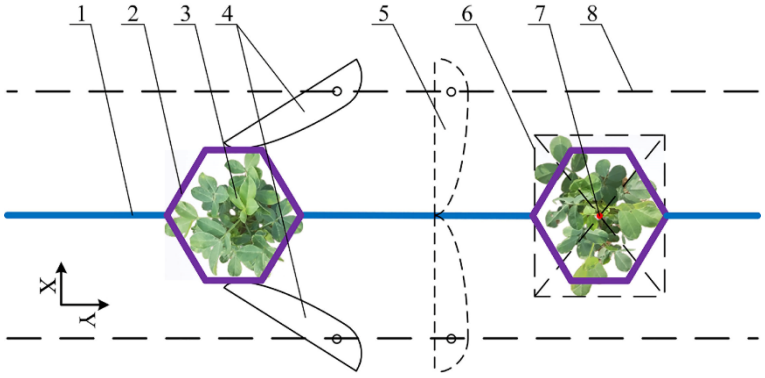


Figure 10. Inter-plant weeding diagram, where: 1. weeding knife tip inter-plant weeding movement track, 2. weeding knife tip avoiding seedling weeding movement track, 3. peanut plant, 4. weeding knife (avoiding seedling weeding state), 5. weeding knife (inter-plant weeding state), 6. detection frame, 7. detection frame positioning center, 8. weeding knife rotation center.

In this study, borrowing from the method proposed by Pérez-Ruíz et al. [32] combining center positioning and hoe control for inter-plant weed removal in field crops, which requires precise extraction of crop center location information, peanut inter-plant weeding showed the working

principle as shown in Figure 10. Unlike lettuce [19], cotton [33] and other crops with prominent targets in the canopy as plant location points, the main branches in the canopy images of peanuts late in growth are difficult to detect due to occlusion, so the method of detecting frame center positioning is used to determine the coordinates of peanut plant location. The results of peanut detection frame center positioning in the test set of 150 images were compared with the actual root positioning coordinates to evaluate the accuracy of the model by detecting frame center positioning.

As shown in Figure 10, Y is the forward direction of the weeder operation, and the movement of the weeding knife in X and Y directions is controlled by rotation to achieve peanut inter-plant weeding and seedling avoidance weeding. Thus, the positioning error is decomposed by X and Y directions, and the offset error d and offset error rate E are proposed to describe the detection frame center positioning effect. d indicates the pixel error between the detection frame center positioning coordinates and the actual root system coordinates of the peanut in X and Y directions, and E indicates the ratio of offset error to the detection frame size in the same direction, which is calculated as in Eqs (12) and (13).

$$d_x = |x_1 - x_2| \quad \text{or} \quad d_y = |y_1 - y_2| \tag{12}$$

$$E_x = \frac{d_x}{\Delta X} \times 100\% \quad \text{or} \quad E_y = \frac{d_y}{\Delta Y} \times 100\% \tag{13}$$

where, (x_1, y_1) represents the actual positioning coordinates of the peanut, (x_2, y_2) represents the positioning coordinates of the inspection center, ΔX and ΔY represent the dimensions of the inspection frame in the X and Y directions.

The offset error and offset error rate results of peanut detection frame center positioning at different stages in the test set are shown in Table 7. The results show that the offset error of detecting frame center positioning in different weeding stages of peanut is less than 16 pixels, and the average offset error rate is less than 7%, i.e., the offset error of peanut detecting frame center positioning is less than 7% of peanut detecting frame size, which meets the requirement of accurate positioning based on peanut canopy detecting frame center to achieve weeding and peanut avoidance.

Table 7. Error test results.

Weeding stage	Offset error/pixel	Offset error rate/%
Day 21	9.61	5.44
Day 30	15.96	6.96
Day 40	10.05	3.82

4. Discussion

In this study, a BEM-YOLOv7-tiny network monitoring model was proposed for peanut and weed identification at different weed control periods in a field environment. The training precision, recall, mAP and F1 values of this model were 93.5%, 93.9%, 96.0% and 93.7%, respectively, which were 1.0%, 2.4%, 2.2% and 1.7% better than the original YOLOv7-tiny network. The precision, recall, mAP and F1 values of the improved model for weeds were 88.2%, 88.5%, 92.4% and 88.3%, respectively, which were improved by 1.6%, 4.9%, 4.4% and 3.2%, respectively. The improved BEM-YOLOv7-tiny network had better performance in peanut and weed detection in fields,

especially for multiple weeds as single target detection. Compared with the Faster R-CNN, YOLOv4-tiny and YOLOv5s deep learning models, the mAP values of BME-YOLOv7-tiny were improved by 12.9%, 12.5% and 4.1%, respectively, and showed good performance in peanut field targets recognition.

Although the BEM-YOLOv7-tiny model has a better recognition effect on peanut and weed, the detection performance is improved and the model computation is reduced a little, but the model size is increased by 1.8 MB to the original YOLOv7-tiny. Facing the problem of small memory space for smart equipment, future considerations to reduce the complexity of the model under the premise of ensuring the model detection effect, such as Liu et al. [34] by introducing ShuffleNet v1 network to lighten the backbone of the model and introduce a lightweight GSConV convolution module, and Li et al. [35] used Ghost Bottleneck structure to achieve a lighter network model and used DW convolution to replace part of the standard convolution to reduce the number of operations, both of which improved the model detection effect while reducing the number of parameters and computation. Arunabha et al. [36] reduced the computational complexity while efficiently detecting multi-scale objects by adding additional feature fusion layers and Swin-Transformer predictor heads to the YOLOv5 network. Future research can further consider model lightweighting while ensuring the model detection effect to reduce the requirement of intelligent embedding equipment.

The BEM-YOLOv7-tiny model detected field weeds with different characteristics of three weeding stages in peanut fields as a single target with good results, indicating the feasibility of the multi-category target combination training detection method, and the detection model can be applied to different weeding stages of peanut. However, there are many kinds of weeds in peanut field, BEM-YOLOv7-tiny only trained on three annual broadleaf weeds and more kinds of weeds are considered to be trained in the future. Additionally, BEM-YOLOv7-tiny was trained only on a single variety of peanuts, data collection should be performed on different varieties of peanuts to enhance the adaptability of the vision system in the peanut field. Promote the application of intelligent equipment in peanut fields.

5. Conclusions

In this paper, a BEM-YOLOv7-tiny model is proposed for the detection of peanut and weed at different weeding stages. ECA is proved to be the best attention mechanism among SE, CBAM and ECA to pay more attention to the target. BiFPN module enhances the feature transfer between different network layers, and MHSA module enhances the attention to the predicted target at the detection end, which improves the performance of the network for the detection of peanut and weed plants in the field. The SIoU loss function is used to improve the model training convergence speed and efficiency.

The precision, recall, mAP and F1 values of the BEM-YOLOv7-tiny model were 93.5%, 93.9%, 96.0% and 93.7%, respectively, which were 1.0%, 2.4%, 2.2% and 1.7% higher than those of the original YOLOv7-tiny model, and the precision, recall, mAp and F1 values for weed detection were 1.6%, 4.9%, 4.4% and 3.2% higher than those of the original YOLOv7-tiny model. Compared to other mainstream detection models, the mAP of the BEM-YOLOv7-tiny model was 12.9%, 12.5% and 4.1% higher than that of the Faster R-CNN, YOLOv4-tiny and YOLOv5, respectively, the network model performed better.

The experimental results of the positioning error show that BEM-YOLOv7-tiny has less than 16 pixels positioning offset error for the center of the peanut detection frame in different periods, and the

detection speed is 33.8 f/s, which meets the demand of real-time detection and positioning of seedlings and grasses. Moreover, the computational volume of the model is 12.9 G, and the volume is only 14.1 MB, which is relatively small in terms of computation and memory requirements for hardware devices, and can be realized for the deployment and application of intelligent hardware devices.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by National Natural Science Foundation of China, grant number 52075092.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. W. Yang, C. H. Cui, Safe weed control technology for peanut fields, *China Sci. Technol. Exp*., **32** (2010), 27.
2. Z. H. Tian, Y. L. Jiang, Y. H. Yin, J. Liang, L. Li, D. W. Liu, Current situation and prospect of weed control in peanut field of southern China, *Shandong Agric. Sci.*, **52** (2020), 162–167. https://doi.10.14083/j.issn.1001-4942.2020.01.031
3. B. G. Shen, S. R. Chen, J. J. Yin, H. P. Mao, Image recognition of green weeds in cotton fields based on color feature, *Trans. Chin. Soc. Agric. Eng.*, **25** (2009), 163–167. https://doi.10.3969/j.issn.1002-6819.2009.06.031
4. H. Y. Wang, J. X. Lü, Identifying corn weed based on texture features and optimized SVM, *Hubei Agric. Sci.*, **53** (2014), 3163–3166+3169. https://doi.10.14088/j.cnki.issn0439-8114.2014.13.110
5. X. F. Li, W. X. Zhu, B. Ji, B. Liu, C. H. Ma, Shape feature selection and weed recognition based on image processing and ant colony optimization, *Trans. Chin. Soc. Agric. Eng.*, **26** (2010), 178–182. https://doi.10.3969/j.issn.1002-6819.2010.10.030
6. X. W. Deng, L. Qi, X. Ma, Y. Jiang, X. S. Chen, H. Y. Liu, et al., Recognition of weeds at seedling stage in paddy fields using multi-feature fusion and deep belief networks, *Trans. Chin. Soc. Agric. Eng.*, **34** (2018), 165–172. https://doi.10.11975/j.issn.1002-6819.2018.14.021
7. M. Dyrmann, H. Karstoft, H. S. Midtiby, Plant species classification using deep convolutional neural network, *Biosyst. Eng.*, **151** (2016), 72–80. https://doi.10.1016/j.biosystemseng.2016.08.024
8. T. Tao, X. Wei, A hybrid CNN-SVM classifier for weed recognition in winter rape field, *Plant Methods*, **18** (2022), 1–12. https://doi.10.1186/s13007-022-00869-z
9. Z. Zong, S. Zhao, G. Liu, Coronal identification and centroid location of maize seedling stage, *Trans Chin. Soc. Agric. Mach.*, **50** (2019), 27–33.
10. W. R. Zhang, H. J. Wen, C. F. Qiao, G. Y. Wang, Mask R-CNN-based method for detecting corn seedlings and corms, *Xinjiang Agric. Sci.*, **58** (2021), 1918–1928. https://doi.10.6048/j.issn.1001-4330.2021.10.020

11. Y. Xu, D. S. Wen, J. P. Zhou, X. P. Pan, Y. Liu, Identification method of cotton seedlings and weeds in Xinjiang based on faster R-CNN, *J. Drain. Irrig. Mach. Eng.*, **39** (2021), 602–607. https://doi.10.3969/j.issn.1674-8530.19.0244
12. H. H. Jiang, C. Y. Zhang, Z. Zhang, W. H. Mao, D. Wang, D. W. Wang, Detection method of corn weed based on mask R-CNN, *Trans. Chin. Soc. Agric. Mach.*, **51** (2020), 220–228+247.
13. X. Li, J. D. Pan, F. P. Xie, J. P. Zeng, Q. Li, X. J. Huang, et al., Fast and accurate green pepper detection in complex backgrounds via an improved YOLOv4-tiny model, *Comput. Electron. Agric.*, **191** (2021), 106503. https://doi.10.1016/j.compag.2021.106503
14. J. F. Gao, A. P. French, M. P. Pound, Y. He, J. G. Pieters, Deep convolutional neural networks for image-based Convolvulus sepium detection in sugar beet fields, *Plant Methods*, **16** (2020), 1–12. https://doi.10.1186/s13007-020-00570-z
15. L. Z. Quan, H. D. Li, H. L. Li, W. Jiang, Z. X. Lou, L. Q. Chen, Two-stream dense feature fusion network based on RGB-D Data for the real-time prediction of weed aboveground fresh weight in a field environment, *Remote Sens.*, **13** (2021), 2288. https://doi.10.3390/rs13122288
16. L. Z. Quan, F. L. Xie, W. Jiang, H. L. Li, H. D. Li, Z. X. Lou, et al., Research on recognition of maize seedlings and weeds in maize mield based on YOLOv4 convolutional neural network, *J. Northeast Agric. Univ.*, **52** (2021), 89–98. https://doi.10.19720/j.cnki.issn.1005-9369.2021.07.011
17. J. X. Gao, F. Tan, J. P. Cui, B. Ma, A method for obtaining the number of maize seedlings based on the improved YOLOv4 lightweight neural network, *Agriculture*, **12** (2020), 2077–0472. https://doi.org/10.3390/agriculture12101679
18. A. Ahmad, Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems, *Comput. Electron. Agric.*, **184** (2021), 106081. https://doi.10.1016/j.compag.2021.106081
19. J. L. Zhang, W. H. Su, H. Y. Zhang, Y. K. Peng, SE-YOLOv5x: An optimized model based on transfer learning and visual attention mechanism for identifying and localizing and visual attention mechanism for identifying and localizing weeds and vegetables, *Agronomy*, **12** (2022), 2061. https://doi.10.3390/agronomy12092061
20. Q. F. Wang, M. Cheng, S. Huang, Z. J. Gai, J. L. Zhang, H. B. Yuan, A deep learning approach incorporating YOLOv5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostratum Dunal seedlings, *Comput. Electron. Agric.*, **199** (2022), 107194. https://doi.10.1016/j.compag.2022.107194
21. H. Zhang, Z. Wang, Y. F. Guo, Y. Ma, W. K. Gao, D. X. Chen, Weed detection in peanut fields based on machine vision, *Agriculture*, **12** (2022), 1541. https://doi.org/10.3390/agriculture12101541
22. Y. D. Lin, T. T. Chen, S. Y. Liu, Quick and accurate monitoring peanut seedlings emergence rate through UAV video and deep learning, *Comput. Electron. Agric.*, **197** (2022), 106938. https://doi.10.1016/j.compag.2022.106938
23. Z. L. Guan, W. Y. Liu, Characteristics of weed occurrence and control in peanut fields, *Modern Agric.*, **10** (2021), 61–62.
24. D. Q. Wann, R. S. Tubbs, Interactive effects of hand weeding, tine and sweep cultivation for weed control in organic peanut production, *Peanut Sci.*, **41** (2014), 124–130. https://doi.10.3146/PS13-15.1
25. H. J. Liu, H. Sun, M. Z. Li, M. Iida, Application of color featuring and deep learning in maize plant detection, *Remote Sens.*, **12** (2020), 2229. https://doi.10.3390/rs12142229

26. D. S. F. Alessandro, D. M. Freitas, G. D. S. Gercina, H. Pistoei, M. T. Folhes, Weed detection in soybean crops using ConvNets, *Comput. Electron. Agric.*, **143** (2017), 314–324. https://doi.10.1016/j.compag.2017.10.027
27. Q. L. Wang, B. G. Wu, P. F. Zhu, P. H. Li, W. M. Zuo, Q. H. Hu, ECA-Net: Efficient channel attention for deep convolutional neural net-works, in *2020 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2020), 11531–11539. https://doi.10.1109/CVPR42600.2020.01155
28. M. X. Tan, R. M. Pang, Q. V. Le, EfficientDet: Scalable and efficient object detection, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 10781–10790. https://doi.10.1109/CVPR42600.2020.01079
29. A. Srinivas, T. Y. Lin, N. Parmar, J. Shlens, P. Abbeel, A. Vaswani, Bottleneck transformers for visual recognition, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 16519–16529. https://doi.10.1109/CVPR46437.2021.01625
30. Z. H. Zheng, P. Wang, W. Liu, J. Z. Li, R. G. Ye, D. W. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in *Proceedings of the AAAI conference on Artificial Intelligence*, **34** (2020), 12993–13000. https://doi.10.1609/aaai.v34i07.6999
31. Z. Gevorgyan, SIoU loss: More powerful learning for bounding box regression, preprint, arXiv:2205.12740.
32. M. Perez-Ruiz, D. C. Slaughter, F. A. Fathallah, C. J. Gliever, B. J. Miller, Co-robotic intra-row weed control system, *Biosyst. Eng.*, **126** (2014), 45–55. https://doi.016/j.biosystemseng.2014.07.009
33. C. Wang, S. He, H. Wu, G. Teng, C. Zhao, Identification of growing points of cotton main stem based on convolutional neural network, *IEEE Access*, **8** (2020), 208407–208417. https://doi.10.1109/ACCESS.2020.3038396
34. H. H. Liu, Y. M. Fan, H. Q. He, K. H. Hui, Improved YOLOv7-tiny's object detection lightweight model, *Comput. Eng. Appl.*, **59** (2023), 1–11. https://doi.10.3778/j.issn.1002-8331.2302-0115
35. X. Li, C. Wang, B. Li, Z. P. Guo, Q. L. Li, Z. Y. Li, Steel surface defect detection algorithm based on improved YOLOv5, *J. Air Force Eng. Univ.*, **23** (2022), 26–33. https://doi.10.3969/j.issn.1009-3516.2022.02.005
36. A. M. Roy, J. Bhaduri, DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism, *Adv. Eng. Inf.*, **56** (2023), 102007. https://doi.org/10.1016/j.aei.2023.102007

AIMS Press