agronomy

MDPI

Article

Cabbage Transplantation State Recognition Model Based on Modified YOLOv5-GFD

Xiang Sun [1,2,3,†], Yisheng Miao [1,2,3,†], Xiaoyan Wu [2,4], Yuansheng Wang [1,2,3], Qingxue Li [1,2,3], Huaji Zhu [1,2,3,*] and Huarui Wu [1,2,3,*]

1 Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; sunx@nercita.org.cn (X.S.); miaoys007@nercita.org.cn (Y.M.); wangys007@nercita.org.cn (Y.W.); liqx007@nercita.org.cn (Q.L.)
2 National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; wuxy007@nercita.org.cn
3 Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing 100125, China
4 School of Computer and Electronic Information, Guangxi University, Nanning 530000, China
* Correspondence: zhuhj007@nercita.org.cn (H.Z.); wuhr007@nercita.org.cn (H.W.)
† These authors contributed equally to this work.

Abstract: To enhance the transplantation effectiveness of vegetables and promptly formulate subsequent work strategies, it is imperative to study the recognition approach for transplanted seedlings. In the natural and complex environment, factors like background and sunlight often hinder accurate target recognition. To overcome these challenges, this study explores a lightweight yet efficient algorithm for recognizing cabbage transplantation states in natural settings. Initially, FasterNet is integrated as the backbone network in the YOLOv5s model, aiming to expedite convergence speed and bolster feature extraction capabilities. Secondly, the introduction of the GAM attention mechanism enhances the algorithm's focus on cabbage seedlings. EIoU loss is incorporated to improve both network convergence speed and localization precision. Lastly, the model incorporates deformable convolution DCNV3, which further optimizes model parameters and attains a superior balance between accuracy and speed. Upon testing the refined YOLOv5s target detection algorithm, improvements were evident. When compared to the original model, the mean average precision (mAP) rose by 3.5 percentage points, recall increased by 1.7 percentage points, and detection speed witnessed an impressive boost of 52 FPS. This enhanced algorithm not only reduces model complexity but also elevates network performance. The method is expected to streamline transplantation quality measurements, minimize time and labor inputs, and elevate field transplantation quality surveys' automation levels.

Keywords: the state of cabbage transplantation; target detection; deep separable convolution; YOLOv5s

Citation: Sun, X.; Miao, Y.; Wu, X.; Wang, Y.; Li, Q.; Zhu, H.; Wu, H. Cabbage Transplantation State Recognition Model Based on Modified YOLOv5-GFD. *Agronomy* **2024**, *14*, 760. https://doi.org/10.3390/agronomy14040760

Academic Editor: Yang Zhu

Received: 24 January 2024
Revised: 28 February 2024
Accepted: 4 March 2024
Published: 8 April 2024

1. Introduction

Vegetables occupy a pivotal position in people's daily dietary intake. Year after year, China has witnessed a steady increase in vegetable sowing areas and production. Specifically, in 2022, China's vegetable sowing area is anticipated to reach approximately 22,375,000 hectares, with a projected production of 787,052,000 tons. Among the various vegetable planting methods, the practice of seedling transplantation stands out as a widely adopted approach. Roughly half of China's vegetable varieties rely on this method, emphasizing the importance of transplantation quality [1]. This is because the quality of transplantation directly impacts the subsequent growth of vegetables and, ultimately, their yield. Consequently, it becomes crucial to identify and assess the state of transplanted seedlings in a timely manner. This not only allows for the timely detection of transplantation issues but also provides valuable data references for subsequent replanting efforts. Currently, however, the majority of transplantation statistics rely on manual methods. This

approach is not only time-consuming and labor-intensive but also lacks real-time accuracy. When dealing with extensive planting fields, collecting such statistics becomes even more challenging. Therefore, there is an urgent need for efficient and accurate methods to assess transplantation quality, ensuring the healthy growth of vegetables and optimizing overall yield.

The inevitable occurrences of blockage, seedling injury, seedling clamping, seedling dropping, etc., may lead to planting quality problems such as seedling exposure, seedling burying, seedling injury, empty holes, and inverted planting in the transplanting process [2]. Traditionally, seedling transplanting status identification revolved around cavitation and collapse [3–5]. As transplanting machinery explores a higher degree of automation, vegetable transplanting has been able to greatly reduce the occurrence of missed planting. At this stage of practice, it is found that different plot information will lead to transplantation due to the mechanical parameters not being suitable for the emergence of exposed seedlings, buried seedlings, etc., and, when not adjusting the machinery in a timely manner to deal with the situation, it is very easy to cause the seedlings to die of exposure to the sun, resulting in additional economic losses. Up to now, there is a lack of statistical research regarding transplanting work and effects, and there is no systematic intelligent solution, with continued reliance on manual identification and help from experienced persons necessary for operation, so the efficiency is very low.

In contrast, intelligent identification and image processing technology to realize the identification and analysis of vegetable state after transplanting will greatly reduce the identification time and improve the identification efficiency. At present, the rapid and simultaneous development of computer technology, image recognition technology, transportation, reconnaissance, security, and other fields has resulted in great overall progress [6–8], and the application of agriculture is also characterized by good performance, but improvements are still necessary regarding the volume and speed of the model [9–11].

Therefore, in order to better automate the statistics of cabbage transplantation and reduce the need for labor and other resources, in this paper, we propose an algorithm based on the YOLOv5s model for recognizing the transplanted seedling status in the natural environment of the field, which provides a referable, lightweight, and efficient method for the research of recognizing the transplanted seedling status of cabbage. The main contributions are as follows:

(1) FasterNet [12] is used as the backbone network of the model to improve the stability of the model in complex environments and the feature extraction ability of small targets, and to reduce the model complexity;
(2) At the same time, the Global Attention Mechanism attention mechanism [13] is introduced to improve the attention of the network to the features of different scales so as to improve the detection efficiency of the model and the recognition accuracy of small targets, to improve the attention of the model to the features of transplanted seedlings, and to reduce the interference of the background and light;
(3) The deformable convolution DCNV3 operator [14] is introduced to reduce the model size and speed up the model inference; the data fitting is accelerated with the help of regression loss function EIoU [15].

2. Materials and Methods

2.1. Image Acquisition

The image data were collected from the National Precision Agriculture Base in Changping District, Beijing, China. The collection device is an Android smartphone, with a unified data collection height of 1 m. All images were taken at a vertical ground angle. We took a large number of images of cabbage in its natural environment in different time periods and light conditions. These images contain the growth state of the cabbage in both high- and low-light conditions, further enriching our dataset. After transplanting, the plant spacing of cabbage was 35 cm and the row spacing was 45 cm. In order to control the data

balance of images of cabbage seedlings in different categories, the images were collected under different lighting conditions at different times.

After screening, 1070 cabbage images under different lighting conditions were ultimately retained. The image format was JPG, and the size was uniformly processed to 651 × 868 pixels. Finally, divide the dataset into training, validation, and testing sets in an 8:2 ratio. Three types, which are buried seedlings, exposed seedlings, and normal seedlings, were marked: the buried seedling is buried by the soil; the substrate of the exposed cabbage seedlings is exposed to the soil; the normal seedling is intact and upright. Using the annotation tool LabelImg, the outer rectangular frame of the cabbage target was drawn in the cabbage image to realize the manual annotation of cabbage. The image was labeled according to the smallest rectangle around the cabbage to ensure that the rectangle contained as little background area as possible. A total of 252 samples of cabbage exposed seedlings, 180 samples of buried seedlings, and 598 samples of normal seedlings were annotated, and label files in YOLO format were created.

2.2. Data Processing

The acquired image data were screened and data augmentation operations performed on the data. The data augmentation methods used in this paper are (1) random image rotation: the flip angle is (90°, −90°); (2) random cuts: cropping down a square area of 100 × 100 pixels at a random position on the image; (3) random brightness: adjusting the brightness of the image randomly in the range of (0.35–1) randomly. The above methods are used to eliminate the differences in scale and position of cabbages in the test set and training set, and to improve the imbalance of data. After the above data augmentation methods to expand the original data, the final number of images obtained is 5350, of which the number of images in the training set is 4070, and the number of images in the test set is 1080. An example of the processed data is shown in Figure 1:

(**a**) (**b**) (**c**)

Figure 1. Examples of processed data. (**a**) Random image rotation; (**b**) random cuts; (**c**) random brightness.

2.3. Algorithmic Training Assessment Indicators

The values of precision (*P*), recall (*R*), and mean accuracy (*mAP*) were chosen for the training experiments as the performance evaluation indexes of the cabbage transplant seedling state model, and a higher value indicates better performance of the model, in which the formulas of precision, recall, and *mAP* are as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$\mathrm{m}AP = \frac{\sum P}{N(Class)} \quad (3)$$

True Positive (*TP*) is the number of samples that are actually positive and predicted to be positive by the model, False Positive (*FP*) is the number of samples that are actually negative but predicted to be positive by the model, False Negative (*FN*) is the number of samples that are actually positive but predicted to be negative by the model, and *P* (precision) is the average of the prediction accuracies for the same sample. False Negative (*FN*) indicates the number of samples that are actually positive but predicted by the model

as negative samples, and precision (P) is the average of the prediction accuracies for the same sample. The number of parameters, the number of floating-point operations (GFLOPs), and the size of bytes occupied by the model file are chosen as the criteria for the lightness of the model, and, the lower the value is, the better the lightness of the model is. The number of parameters is determined by the structure of the network model, and the number of floating-point operations is the number of calculations that the model needs to perform.

2.4. Object Detection Algorithm

Before deep learning was involved in this field, the traditional target detection ideas included region selection, manual feature extraction, and classifier classification. Because the manual feature extraction method often has difficulty meeting the diversified features of the target, the traditional method has not been a good solution to the problem of target detection. After the rise of deep learning, neural networks can automatically learn powerful feature extraction and fitting ability from a large number of data, so many excellent target detection algorithms have emerged.

At present, object detection algorithms based on deep learning are mainly divided into two categories: two-stage object detection and single-stage object detection. Compared with the single-stage target detection algorithm, the two-stage target detection algorithm first extracts the candidate box according to the image, and then makes a secondary correction based on the candidate region to obtain the detection point result, with higher detection accuracy but slower detection speed. The first work of this kind of algorithm is R-CNN, and then Faster R-CNN improves it. Because of its excellent performance, Faster R-CNN is still a very competitive algorithm in the field of target detection. Compared with the two-stage object detection algorithm, the single-stage object detection method directly calculates the image to generate the detection result; the detection speed is low, but the detection accuracy is low. The representative of this kind of algorithm is YOLO, SSD, although the prediction accuracy is not as good as the two-stage target detection algorithm; because of the fast running speed, YOLO has become the mainstream of target detection research. Zhao, Q. et al. proposed an exemplary method for enhancing tire specification character recognition within the YOLOv5 network [16]. YOLO not only demonstrates outstanding performance in various industries but also exhibits exceptional capabilities in the agricultural sector. It plays a significant role in the recognition of apple blossoms [17] and the cultivation of tomatoes [18].

2.5. Target Detection Model Based on YOLOv5s

YOLOv5 is a single-stage target detection algorithm open-sourced by Ultralytics, whose main features are fast speed, high accuracy, and the ability to detect and recognize multiple objects in an image in a relatively short period of time. The network structure of the YOLOv5 model is mainly divided into four parts, namely Input, Backbone, Neck, and Head. Input part carries out preparatory work, such as mosaic data enhancement of the image; Backbone part uses structures such as SPPF, C3, etc., to extract features from the input image; Neck part draws on the structure of PANet and performs shallow feature fusion on feature maps of large, medium, and small scales; Head part is responsible for predicting the processed feature maps. According to the above division of the four modules of the model, the structure diagram of the entire model was drawn. The network structure of YOLOv5 is shown in Figure 2:

There are four versions of the YOLOv5 model series: x, m, l, and s. YOLOv5s is the version with the smallest depth and lowest complexity in the series [19]. Compared with YOLOv5s, the remaining three versions have better detection performance, their model complexity increases sequentially, and their real-time performance is weakened. On the basis of realizing high-precision recognition, YOLOv5s also has the advantage of lightweight real-time, which can be applied to scenarios requiring target detection under resource constraints, greatly facilitating the transplantation and application of the model.

Therefore, this paper will discuss the key techniques and how to build the YOLOv5s target detection model in detail. At the same time, YOLOv5 is an excellent single-stage target detection algorithm. In order to verify the feasibility of identifying the transplanting state of cabbage, the calculation cost and recognition accuracy of the model are assessed regarding popular single-stage detection algorithms YOLOv3-tiny [20], YOLOv4-tiny [21], and YOLOv7-tiny [22] compared with the traditional two-stage model Faster-RCNN [23] framework in the experimental part.

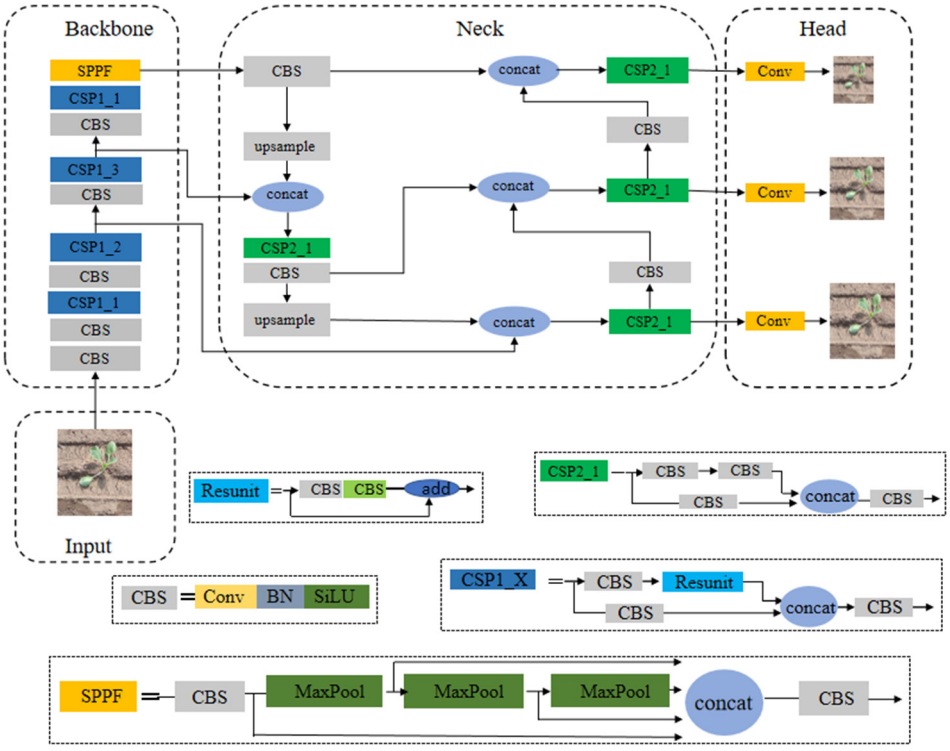

Figure 2. The YOLOv5 network structure.

2.6. Target Detection Model Based on Improved YOLOv5s

2.6.1. FasterNet

Since the transplanted seedlings are in a more complex natural environment, the background may contain interference information such as light, soil impurities, etc., which affects the feature extraction. The backbone network of YOLOv5s, CSPNet, may be prone to lose the target in complex environments and has relatively weak target localization ability and average feature expression ability. In order to improve the feature extraction and expression ability and reduce the influence of complex environment, this paper replaces the backbone network of YOLOv5s with the lighter and more efficient FasterNet network.

The core idea of FasterNet is to improve the feature expression ability and the coverage of the sensory field while maintaining light weight and high speed. The FasterNet network is a four-layered stage. The first stage is preceded by an embedding layer (an embedding layer, a 4 × 4 conv, with a step size of 4), which converts the high-dimensional discrete inputs into a low-dimensional continuous vector representation, greatly reducing the number of parameters in the network. This layer structure converts the high-dimensional discrete inputs into a low-dimensional continuous vector representation, which greatly reduces the number of parameters in the network and improves the feature representation. The remaining stages are preceded by the merging layer (a regular 2 × 2 Conv with a step size of 2), which serves to spatially downsample the feature maps and boost the channel dimensions. Each stage includes several FasterNet blocks, which are the core of the FasterNet concept. FasterNet blocks consist of partial convolution (PConv), conv_1 × 1, BN, and Relu, and PConv can extract spatial features more effectively by reducing redundancy

accounting and memory visiting, and the structure of FasterNet blocks is shown in Figure 3:

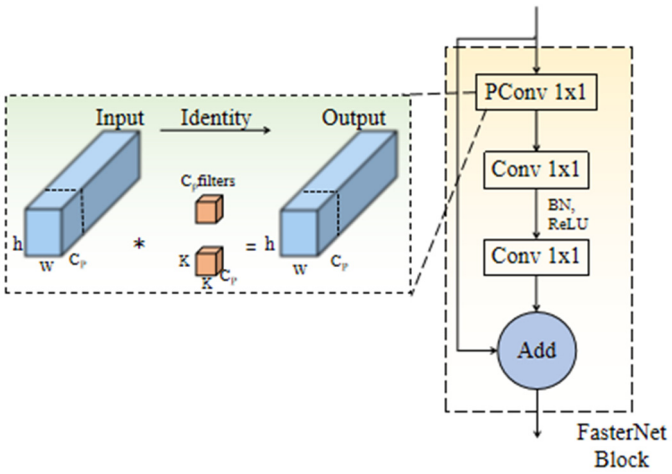

Figure 3. FasterNet blocks structure diagram.

2.6.2. Attention Module

The small size of some transplanted seedlings results in a relatively large proportion of the background and a different transplanted state of the seedlings in images of different sizes; at the same time, when in the exposed state, the substrate of the cabbage seedlings will be semi-naked or completely naked to the surface of the soil, and the substrate is extremely similar to the soil in contact with the substrate, and the accuracy of the identification of the substrate influences the identification of the transplanted seedling's exposed state; overlap may also occur between cabbage plants. In order to reduce the interference of redundant information and improve the model's attention to the transplanted seedlings as a whole and reduce the influence of target overlap on recognition results, the Global Attention Mechanism (GAM) was added to the original model, which reduces the loss of information and enhances the global feature interactions through the channel and spatial dual attention modules. The GAM performs channel attention first, then spatial attention, and can gradually refine the feature representation at different levels, allowing the model to understand the channel and spatial structure of the data more comprehensively, thereby improving the model's performance on complex tasks. This sequential attention model can effectively improve the representation learning ability and generalization ability of the model. The structure of the GAM is shown in Figure 4 below:

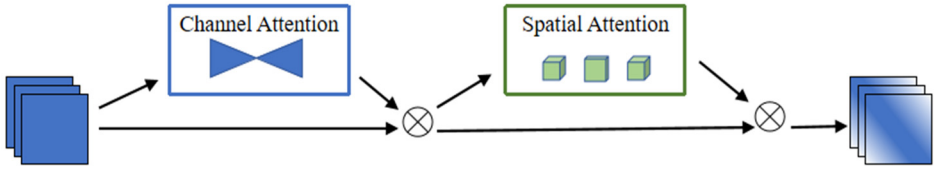

Figure 4. The overview of GAM.

Its channel attention sub-module uses a three-dimensional arrangement to retain information in three dimensions and a two-layer multilayer perceptron (MLP) to amplify cross-dimensional channel-space dependencies; in the spatial attention sub-module, two convolutional layers were used for spatial information fusion in order to focus on the spatial information, and the pooling operation was deleted to further preserve the feature mapping. Eventually, the model is made to pay more attention to the important information in the image to enhance the information extraction ability for cabbage seedlings, which in turn improves the detection performance of the model.

2.6.3. Loss Function

The original IoU of YOLOv5s is CIoU [19].The problem of CIoU loss is that the length and width of the anchor cannot be increased or decreased at the same time, which inhibits the optimization of the model. In order to improve the convergence speed and localization accuracy of the model, the loss function of the model is replaced with EIoU loss. Firstly, the EIoU loss is illustrated, which can be divided into 3 parts: IoU loss L_{IOU} + distance loss + aspect ratio loss. The definition of the formula is as follows:

$$\begin{aligned} L_{EIOU} &= L_{IOU} + L_{\text{dis}} + L_{asp} \\ &= 1 - IOU + \frac{\rho^2(b,b^{gt})}{(w^c)^2+(h^c)^2} + \frac{\rho^2(w,w^{gt})}{(w^c)^2} + \frac{\rho^2(h,h^{gt})}{(h^c)^2} \end{aligned} \tag{4}$$

where w^c and h^c represent the width and length of the minimum enclosing frame, gt represents the ground truth, the true bounding box of the target. In object detection tasks, ground truth refers to the bounding box of the real target annotated in the dataset, which means that the EIoU loss can directly minimize the difference between the width and height of anchor and gt, resulting in faster convergence and better localization. Replacing the CIoU in the Head part of the model with the EIoU loss can make the model locate the target better, improve the localization accuracy of the model, and accelerate the convergence of the model at the same time.

2.6.4. Deformable Convolution v3

In order to further accelerate the model training, reduce the model parameters, make the model have better portability, and at the same time improve the recognition effect on the target, the C3 layer of the model was replaced with the deformable convolution operator (DCNV3). This can to some extent scale down the model complexity added by the C3 layer convolution and improve the model's recognition and convergence effect on small targets. The reason for this is that the shape of the convolution kernel in deformable convolution is not fixed and can adaptively change according to the content of the target in the image. This flexible mapping allows for better coverage of the detected targets so as to capture more useful characterization information, while the operator is also more efficient in terms of computational effort and memory. The following Equation (5) shows the formalization of this convolutional description:

$$y(p_0) = \sum_{g=1}^{G} \sum_{k=1}^{K} w_g m_{gk} x_g (p_0 + p_k + \Delta p_{gk}) \tag{5}$$

where G identifies the number of aggregation groups, for the gth group, $w_g \in R^{C \times C'}$, $C' = C/G$ denotes the location-irrelevant projection weights; K denotes the number of sampling points, denotes the modulation factor of the kth sampling point, and is normalized along the dimension K by softmax.

The feature maps output from the 17th, 20th, and 23rd layers of YOLOv5s are used for small, medium, and large target detection, respectively, but, due to the relatively large difference in the area occupied by different sizes of cotton in the data picture, the small targets will be lost after multiple convolutions, resulting in missed detection. In order to improve the detection accuracy of small-sized cotton bolls and solve the problems of rough target location and even loss, based on the original algorithm, all the 17-layer, 20-layer, and 23-layer C3 layers were replaced with deformable convolution, which improves the model's identification and localization accuracy of the target. The improved model is named YOLOV5s-GFD; the model is divided into four parts: Input, Backbone, Neck, and Head. The Backbone of the model is fasternet, and the Neck part uses the DCNV3_C3 layer as a deformable convolution operator, and its network structure is shown in Figure 5:

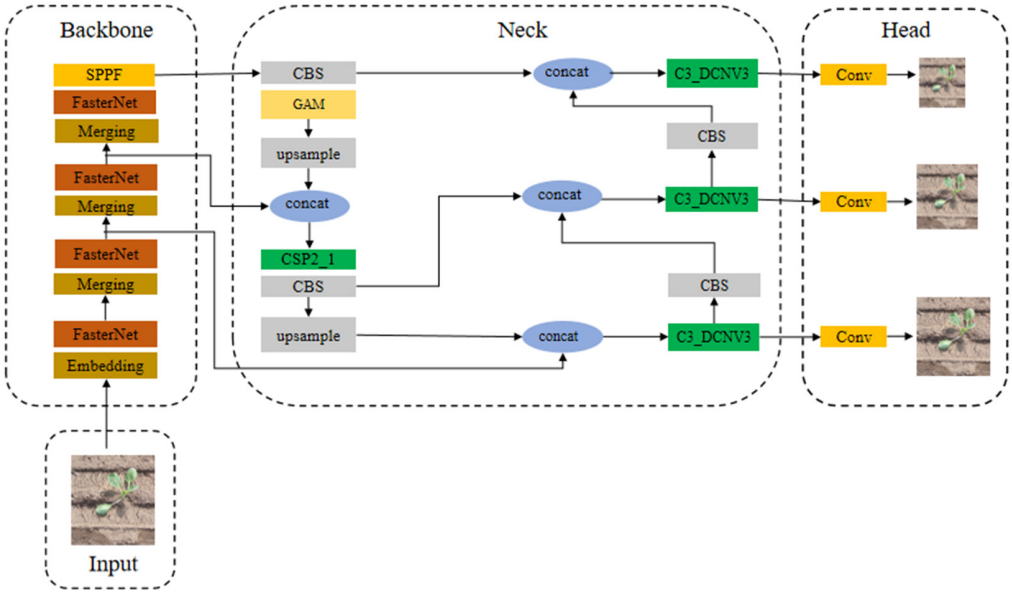

Figure 5. The YOLOv5-GFD network structure.

3. Results

3.1. Training Platform

All the experiments in this paper were conducted in the deep learning framework Pytorch 1.12, Python 3.7, and Cuda 10.2. The processor used for the experiments is Intel(R) Core(TM) i9-9820X CPU @ 3.30 GHz with 20 cores; the graphics card is RTX 2080 Ti with four blocks. The dataset is 5350 images after data enhancement, divided according to 8:2, and the input image size is 640 × 640, and the detection model recognizes the buried, exposed, and qualified status of transplanted cabbage seedlings.

3.2. Model Training

3.2.1. Algorithm Training Parameter

The input image size of this model is 640 × 640, the batch size is 16, and the number of iterations is 300. The learning rate for training is set to 0.01, momentum factor is set to 0.937, while weight decay is set to 0.0005.

3.2.2. Training Results

The dataset images were fed into the improved model for training, a total of 300 epochs were trained, BatchSize was set to 16, and the results of training the network are shown in Figure 6.

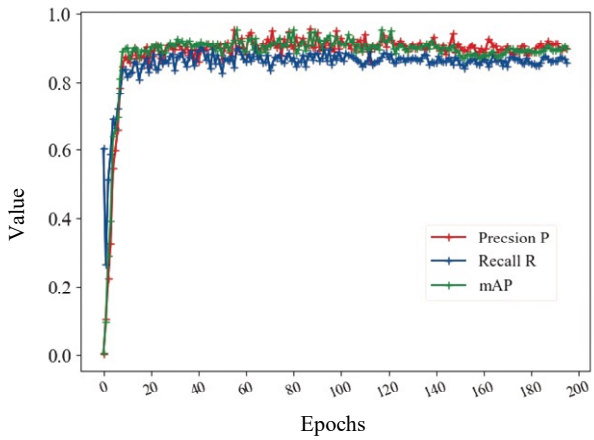

Figure 6. Trends in P, R, and mAP for YOLOv5s-GFD.

As can be seen from the above figure, the learning efficiency of the model at the beginning of the training period performs very well, the values grow relatively fast, and a good precision and recall can be achieved in a few tens of iterations, which gradually smooths out in the later iterations.

3.3. Ablation Test

The improved model incorporates FasterNet, introduces the GAM attention mechanism, and uses deformable convolution to replace ordinary convolution. In order to verify its optimization effect on the model, ablation experiments were carried out on the improved model with the same parameter settings as the training process, and the obtained results of the ablation experimental verification are shown in Table 1.

Table 1. Results of model ablation experiments.

FasterNet	EMA	C3_DCNV3	Focal-Eiou	P/%	R/%	FLOPs/G	mAP/%
×	×	×	×	92.0	87.8	15.8	92.1
√	×	×	×	93.0	88.8	11.2	93.6
×	×	×	√	93.5	89.9	15.8	93.9
√	×	×	√	94.7	89.1	11.2	93.8
√	×	√	√	93.9	89.5	11.2	94.6
√	√	√	√	95.3	88.5	11.5	95.4

As can be seen from Table 1, "×" represents the absence of the corresponding module mentioned above, while "√" represents the inclusion of the corresponding module mentioned above. compared to YOLOv5s, the average precision of the model with the addition of FasterNet increases by 1.5 percentage points, and the recall also increases by 1 percentage point, while FLOPs decrease by 4.6 Gs, and there is a slight decrease in mAP, which suggests that the adoption of FasterNet as the network backbone for YOLOv5s can reduce the model complexity while improving the positioning accuracy and recognition accuracy of the model. Compared to YOLOv5s, the model using EIoU as the loss function has an increase of 1.8 percentage points in mAP and 2.1 percentage points in recall, indicating that the addition of Focal-EIoU can increase the fitting speed and improve the model recognition accuracy. The model also improves the precision rate by 1.9 percentage points and the average precision by 0.6 percentage points compared to YOLOv5s, so replacing the C3 layer of the original model for the DCNV3 operator reduces the model complexity while ensuring the model detection performance. The improved model finally improves the precision by 3.3 percentage points and recall by 0.7 percentage points over the original model, YOLOv5s, and the model size increases by 0.3 G, but the mAP grows by 0.8 percentage points, which is significant growth, and there is still a reduction by 4.9 G in FLOPs over the original model. It can be concluded that the improved model, YOLOv5s-GFD, in this paper reduces the number of model parameters while steadily increasing the recognition accuracy of the model, which is more advantageous in terms of detection performance and magnitude compared to the original model.

3.4. Comparison of Different Network Model Training

In order to verify the reliability of the algorithm proposed in this chapter, the current popular target detection algorithms (YOLOv3-tiny, YOLOv4-tiny, Faster-RCNN, and YOLOv7-tiny) were compared with the algorithms proposed in this paper. The experimental results are shown in Table 2.

The experimental results show that, compared with other popular target detection models and the YOLOv3-tiny, YOLOv4-tiny, YOLOv7-tiny, and Faster-RCNN target detection models, the detection accuracy of YOLOv5s-FDN is improved by 9.1%, 6%, 1.6%, and 0.7%, respectively. The model is reduced by 33.9%, 51.2%, 89.3%, and 6.5%, and the accuracy and volume of the model are more advantageous. The detection speed is 140 FPS, which meets the real-time line requirements of identification tasks. Compared

with the best lightweight models, the accuracy of YOLOv7-tiny and YOLOv5s-GFD is improved by 2.2 percentage points. The detection speed of the improved model is slightly slower than that of YOLOv7-tiny, but the average accuracy is increased by 0.7%, and the improved model is 1% higher than Faster-RCNN, which has the highest recognition accuracy. Therefore, in terms of overall performance, the enhanced YOLOv5s-GFD maintains a superior balance between detection accuracy and speed while maintaining its light weight and efficiency, indicating that the addition of FasterNet and DCNV3 not only makes the model lighter but also helps the model to obtain more feature details. Thus, high-precision identification of cabbage transplanting state was realized.

Table 2. Comparative experimental results of different models.

Model	FLOPs/G	P/%	mAP/%	T1/fps
YOLOv3-tiny	17.4	83.4	86.3	115
YOLOv4-tiny	23.6	92.3	89.4	114
Faster-RCNN	108.0	94.3	93.8	82
YOLOv7-tiny	12.3	93.1	94.7	189
YOLOv5s-GFD	11.5	95.3	95.4	140

3.5. Analysis of Identification Results

In order to visualize the detection effect of the model, the same datasets were used for the improved model and the original model, and the detection results are shown in Figure 7. The red box represents the model's recognition of the exposed seedling, the green box represents the model's recognition of the buried seedling, and the blue box represents the recognition of the normal transplanted seedling.

Judging from the identification results, our model can accurately identify the transplanting states of the three seedlings, while the original model, YOLOv5s, easily misses detection when multiple plants are planted, and the identification accuracy of seedlings with different transplanting states is not high, and it may misidentify dry weeds in the background as cabbage, indicating that it is susceptible to background interference. This is because the Non-Maximum Suppression (NMS) algorithm it uses will only select the bounding box with the highest confidence when dealing with multiple overlapping bounding boxes and suppress the others. This strategy may result in some bounding boxes being missed because they are overwritten by bounding boxes with higher confidence. The improved model with the addition of GAM can learn more critical feature information in the overlapping region so as to pay more attention to such important information when deciding the final detection result and reduce missed detection in the case of overlap. The detection accuracy can be maintained above 0.8, and the recognition accuracy and localization accuracy perform much better. It can be concluded that the improved model, YOLOV5s-GFD, in this study can enhance detection performance in the case of occluded and smaller targets, and its robustness is also higher.

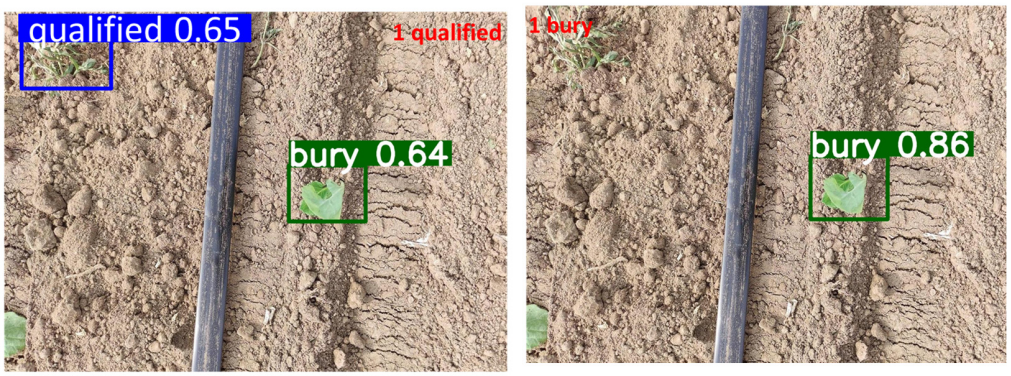

Figure 7. *Cont.*

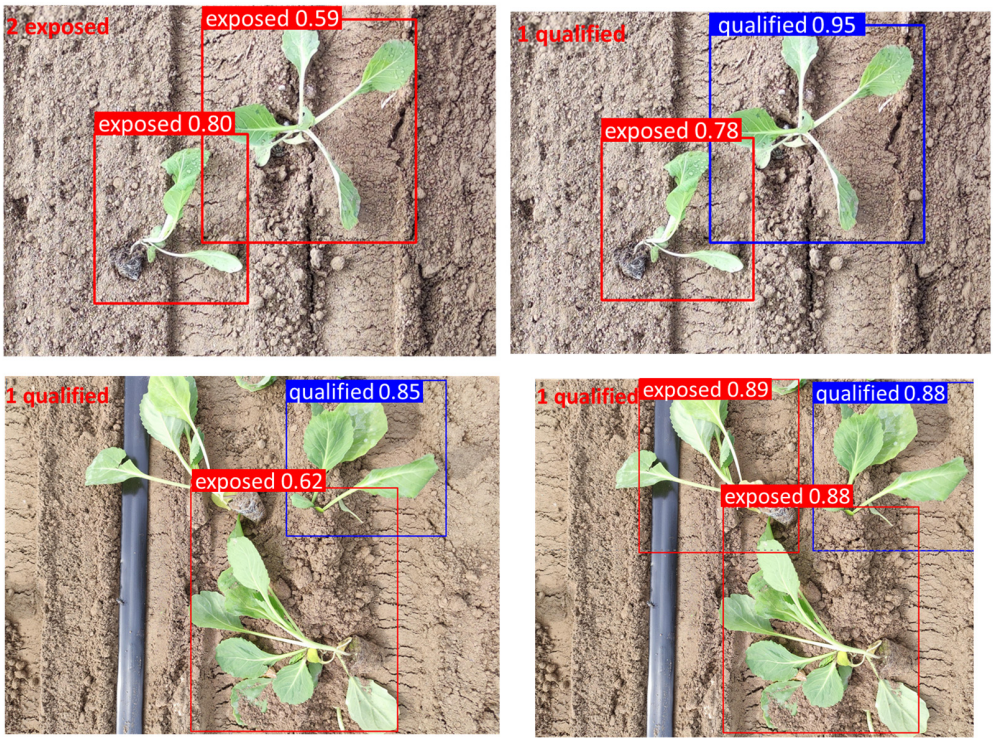

Figure 7. Comparison of selected assays between YOLOV5s and YOLOV5s-GFD.

4. Conclusions

In order to be able to realize the rapid identification of dewy and buried seedlings of cabbages after transplanting, and to assist the decisionmaking of automated planting plan in the field, in this study, a research method is proposed based on the YOLOv5s network model to realize the efficient detection of transplanted cabbage status. Aiming at the problems of small individual cabbages, complex field environment, low recognition accuracy of YOLOv5s, and large model arithmetic, this study proposes an improved model based on YOLOv5s. By introducing the GAM attention mechanism to improve the attention to small targets and reduce the interference of environmental factors, FasterNet is used as the target extraction network, and deformable convolution is added to improve the detection performance of the network and reduce the complexity of the model.

The improved model can achieve 95.3% recognition accuracy, 88.5% callback rate, and 95.6% mAP value for cabbage transplantation. Its mAP value was improved by 3.5 percentage points relative to the original model. The model detection speed is 140 FPS, which can meet the real-time line requirements of recognition tasks. The complexity of the model is also low, and FLOPS is 11.5 G, which can match an environment with low computing resources and facilitate model transplantation from other hardware devices. At the same time, the efficient identification of cabbage transplanting state can provide data guidance and analysis for the transplanting work, help to achieve better transplanting results, and then promote the intelligent and automatic transplanting workflow.

Although the comprehensive performance of the model is good, there are still some problems. There is still room for improvement in the recognition accuracy of the improved model. In the case of background interference and small targets, improving the recognition accuracy of the model is still the direction of future development. At the same time, in order to better evaluate the quality of the transplanting work, it is necessary to continue to expand the types of dataset samples and annotations.

Author Contributions: Conceptualization, X.S. and H.W.; methodology, X.S., Y.W. and H.Z.; software, X.W., Q.L. and H.Z.; validation, Y.M. and H.W.; Data curation, H.Z.; writing—original draft preparation, X.S. and Y.M.; writing—review and editing, X.W., X.S. and H.Z.; project administration, Y.M. and H.Z.; funding acquisition, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Key Research and Development Program of China under Grant 2023YFD2001205, the China Agriculture Research System of MOF and MARA Grant CARS-23-D07. Huaji Zhu and Huarui Wu are the co-corresponding authors of this paper.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy policy of the authors' institution.

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

References

1. Cui, Z.; Guan, C.; Yang, Y.; Gao, Q.; Chen, Y.; Xiao, T. Research Status of Mechanized Transplantation Technology and Equipment for Vegetables. *China J. Agric. Mach. Chem.* **2020**, *41*, 85.
2. Jiang, Z. *Design and Experiment of Online Monitoring System for Planting Quality of Rape Carpet Seedling Transplanter*; Chinese Academy of Agricultural Sciences: Beijing, China, 2021.
3. Zhao, D.; Zhao, H. Research on Image Recognition Technology for Missing and Drift Seedlings Based on CNN Algorithm. *Softw. Guide* **2020**, *19*, 230–233.
4. Wang, C.; Guo, X.; Xiao, B.; Du, J.; Wu, S. An automatic measurement method for the number of missing maize plants in seedling stage based on image stitching. *J. Agric. Eng.* **2014**, *30*, 148–153.
5. Jiang, Z.; Zhang, M.; Wu, J.; Jiang, L.; Li, Q. Real time monitoring method for transplanting and missed planting of rapeseed blanket seedlings—Based on video image stitching. *Agric. Mech. Res.* **2022**, *30*, 189–195.
6. Al-qaness, M.A.A.; Abbasi, A.A.; Fan, H.; Ibrahim, R.A.; Alsamhi, S.H.; Hawbani, A. An improved YOLO-based road traffic monitoring system. *Computing* **2021**, *103*, 211–230. [CrossRef]
7. Degadwala, S.; Vyas, D.; Chakraborty, U.; Dider, A.R.; Biswas, H. Yolo-v4 deep learning model for medical face mask detection. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 209–213.
8. Peng, H.; Zhang, Y.; Yang, S.; Song, B. Battlefield image situational awareness application based on deep learning. *IEEE Intell. Syst.* **2019**, *35*, 36–43. [CrossRef]
9. Hasan, A.S.M.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [CrossRef]
10. Wang, M.; Zhang, Q. Research progress on deep learning based image recognition technology in crop disease and pest identification in China. *Chin. Veg. J.* **2023**, *3*, 22–28.
11. Yang, W.; Liu, T.; Tang, X.; Xu, G.; Ma, Z.; Yang, H.; Wu, W. Progress in Plant Phenomics Research under the Background of Smart Agriculture. *J. Henan Agric. Sci.* **2022**, *51*, 1–12.
12. Chen, J.; Kao, S.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 12021–12031.
13. Liu, Y.; Shao, Z.; Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv* **2021**, arXiv:2112.05561.
14. Wang, W.; Dai, J.; Chen, Z.; Huang, Z.; Li, Z.; Zhu, X.; Hu, X.; Lu, T.; Lu, L.; Li, H.; et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023.
15. Yang, Z.; Wang, X.; Li, J. EIoU: An improved vehicle detection algorithm based on vehiclenet neural network. *J. Phys. Conf. Series* **2021**, *1924*, 012001. [CrossRef]
16. Zhao, Q.; Wei, H.; Zhai, X. Improving Tire Specification Character Recognition in the YOLOv5 Network. *Appl. Sci.* **2023**, *13*, 7310. [CrossRef]
17. Shang, Y.; Zhang, Q.; Song, H. Application of deep learning based on YOLOv5s in natural scene apple flower detection. *J. Agric. Eng.* **2022**, *38*, 222–229.
18. Zhang, J.; Bi, Z.; Yan, Y.; Wang, P.; Hou, C.; Lv, S. Rapid recognition of greenhouse tomatoes based on attention mechanism and improved YOLO. *J. Agric. Mach.* **2023**, *54*, 236–243.
19. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12993–13000. [CrossRef]
20. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

21. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13029–13038.
22. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
23. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.