Article

Improved YOLO v7 for Sustainable Agriculture Significantly Improves Precision Rate for Chinese Cabbage (*Brassica pekinensis* Rupr.) Seedling Belt (CCSB) Detection

Xiaomei Gao [1], Gang Wang [1,*], Jiangtao Qi [1], Qingxia (Jenny) Wang [2,3], Meiqi Xiang [1], Kexin Song [1] and Zihao Zhou [1]

[1] College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China; gaoxm22@mails.jlu.edu.cn (X.G.); qijiangtao@jlu.edu.cn (J.Q.); xiangmq20@mails.jlu.edu.cn (M.X.); songkx23@mails.jlu.edu.cn (K.S.); zhouzh22@mails.jlu.edu.cn (Z.Z.)

[2] School of Business, University of Southern Queensland (UniSQ), Springfield Central, QLD 4300, Australia; jenny.wang2@unisq.edu.au

[3] Centre for Applied Climate Sciences, University of Southern Queensland (UniSQ), Darling Heights, QLD 4350, Australia

* Correspondence: gw611004@jlu.edu.cn

Abstract: Precise navigation in agricultural applications necessitates accurate guidance from the seedling belt, which the Global Positioning System (GPS) alone cannot provide. The overlapping leaves of Chinese cabbage (*Brassica pekinensis* Rupr.) present significant challenges for seedling belt fitting due to difficulties in plant identification. This study aims to address these challenges by improving the You Only Look Once (YOLO) v7 model with a novel approach that decouples its network head deriving from the Faster-Regions with Convolutional Neural Network (Faster R-CNN) architecture. Additionally, this study introduced a BiFormer attention mechanism to accurately identify the centers of overlapping Chinese cabbages. Using these identified centers and pixel distance verification, this study achieved precise fitting of the Chinese cabbage seedling belt (CCSB). Our experimental results demonstrated a significant improvement in performance metrics, with our improved model achieving a 2.5% increase in mean average precision compared to the original YOLO v7. Furthermore, our approach attained a 94.2% accuracy in CCSB fitting and a 91.3% Chinese cabbage identification rate. Compared to traditional methods such as the Hough transform and linear regression, our method showed an 18.6% increase in the CCSB identification rate and a 17.6% improvement in angle accuracy. The novelty of this study lies in the innovative combination of the YOLO v7 model with a decoupled head and the BiFormer attention mechanism, which together advance the identification and fitting of overlapping leafy vegetables. This advancement supports intelligent weeding, reduces the reliance on chemical herbicides, and promotes safer, more sustainable agricultural practices. Our research not only improves the accuracy of overlapping vegetable identification, but also provides a robust framework for enhancing precision agriculture.

Keywords: YOLO v7; decoupled head; BiFormer attention mechanism; Chinese cabbages seedling belt; fitting algorithm

1. Introduction

Using machine vision to locate and identify seedling belts is integral to improved agricultural machinery navigation. Whereas real-time kinematic Global Positioning System (GPS) commonly reports 2 cm accuracy, sub-centimeter accuracy is necessary to control and operate agricultural machinery [1]. The belt-fitting line is an improved method for the automatic navigation of agricultural machinery [2]. The acquisition of belt-fitting lines can be delineated into two categories according to the navigation method: satellite signals and visual identification [3,4]. Satellite methods are common with many agricultural original equipment

manufacturers (OEMs) having an integrated product available. A visual identification system detects the positions of field crops or seedlings, relaying this information to the autonomous navigation system [5]. This facilitates real-time and precise navigation adjustments during agricultural machinery operations, such as spraying or harvesting [6]. V7Seedling belt fitting is a technique that is dependent on the actual conditions in the field rather than relying on positioning from satellites. This technique has shown accuracy can reach the centimeter level or even higher [7]. Due to the small spacing between many horticultural row crops, an identification error of a few centimeters can lead to misjudgment of the resultant agricultural operations [8]. Misidentification of weeds can lead to increased use of herbicides or laser energy, which can increase the risk of environmental pollution and harm the protection of soil, water, and biodiversity, as well as sustainable agricultural development.

Researchers have successfully identified seedling belts for economically important crops like corn and wheat [9–11], such as crop belt detection using a vision system [11]. Other studies have been able to identify corn-stabilizing root features based on vertical projections [12]. Zhai et al. [13] used multi-crop-belt detection algorithms with binocular vision and reported accurate recognition results. However, the overlap between the leaves of mature Chinese cabbage (*Brassica pekinensis* Rupr.) planted at 30 cm spacing makes seeding belt identification difficult. Further, juvenile weeds are small compared to Chinese cabbages and often superimpose on each other [14]. This results in displays of dispersed leaves and clustering during various growth stages. This phenomenon seriously affects the application of seedling belt identification technology in Chinese cabbage fields. As a result, there are relatively few studies of Chinese cabbage seedling belt (CCSB) identification. Nevertheless, Chinese cabbage is consumed globally [15]. Therefore, further study of CCSB identification technology is necessary to improve production efficiency.

Accurately identifying weeds from Chinese cabbages is a challenge in seedling belt identification. Commonly used machine learning identification algorithms include the You Only Look Once (YOLO) series [16–21] and Faster-Regions with Convolutional Neural Network (Faster R-CNN) [22]. The YOLO series employs end-to-end neural networks for real-time prediction using bounding boxes and class probabilities [16–21]. YOLO v7 also employs coupled heads (the feature map output by the convolutional layer is directly fed into several fully connected layers or convolutional layers to generate the output of the target position and class), restricting its versatility and functionality for the CCSB [23]. The coupled head systems require feeding feature maps directly into multiple fully connected or convolutional spatial data layers. These layers are responsible for outputting target position and category information. This process not only demands a significant number of parameters and computational resources, but also is prone to overfitting [24]. Therefore, decoupled heads have been introduced to address these issues. Decoupled heads extract target position and category information separately. Machine learning approaches process through distinct network branches before merging [25]. The advantages include efficiently reducing parameters and computational complexity while improving the model's generalization and robustness [26]. Meanwhile, Faster R-CNN integrates the strengths of the Fast R-CNN and Region Proposal Network (RPN) in a unified network structure, training distinct models for the two tasks (target position and category) [27]. The focus of our research is exploring the use of Faster R-CNN for decoupled design to address the issue of leaf occlusion in Chinese cabbage crops.

In the realm of artificial intelligence and machine learning, attention refers to the ability of a model to selectively focus on certain parts of the input [28]. The attention mechanism poses another challenge in CCSB identification. The transformer is crucial for image identification [29]. It relies on attention, and therefore can significantly affect the model accuracy rate [30]. The YOLO series uses both channel attention and spatial attention [16]. Channel attention computes weights for the entire feature map, but limits long-range dependencies. In the field of computer vision, spatial attention involves the process of selectively attending to specific spatial regions within an image. Spatial attention is computationally complex. To reduce the computational complexity of spatial attention,

researchers developed the BiFormer attention mechanism [31]. BiFormer uses adaptive queries for content-aware sparse patterns, reducing computation and memory complexity [31], unlike vision transformers with the deformable attention mechanism [32] and the deformable patch-based transformer attention mechanism [33]. BiFormer achieved efficient and accurate image identification through overlapping patch embedding, patch merging, and increased channels [34]. Therefore, introducing the BiFormer attention mechanism into our CCSB identification algorithm can optimize the network and improve efficiency when solving the problem of leaf overlap between Chinese cabbage and weeds.

The fitting algorithm is crucial for modeling the spatial distribution of Chinese cabbage paths based on identification results, providing a vital foundation for agricultural machinery navigation route planning. The Hough transform [35], linear regression [36], Blob analysis [37], and stereo vision [38] are primary fitting algorithms. The Hough transform is a technique used in image analysis and computer vision, whose primary use is to detect geometric shapes in images [35]. Linear regression fits data points onto a straight line, minimizing the error between the fitted line and the data point [36]. Blob analysis represents crop distribution by center points of connected regions with identical pixels in binary images [39]. Stereo vision leverages the height difference between crops and weeds above the ground to distinguish and identify crop belts [38]. The growth characteristics of Chinese cabbages, such as their shorter and scattered nature, make it challenging to accurately detect changes using the Hough transform and linear regression. Meanwhile, the overlapping leaves during the rosette stage also lead to errors with Blob analysis. Furthermore, the minimal height difference between Chinese cabbages and weeds affects stereo vision fitting. Therefore, it is essential to design a fitting algorithm tailored to the unique growth characteristics of Chinese cabbages.

This study compared the performance of Faster R-CNN, YOLO v3, and YOLO v7. The features of each method are summarized in Table 1, illustrating the comparative performance and specific improvements introduced in our proposed method. To further enhance YOLO v7, this study made two key modifications: first, this study decoupled the originally coupled heads in YOLO v7; second, this study introduced the BiFormer attention mechanism to optimize the network, enabling content-aware sparse patterns. These enhancements allowed for the precise identification of Chinese cabbage and weeds. Additionally, this study developed a fitting algorithm specifically designed to match the growth characteristics of Chinese cabbage for seedling belt identification. This algorithm not only automatically fits the distribution path of Chinese cabbage based on the recognition algorithm, but also achieves a smaller fitting error than other existing algorithms.

Table 1. Comparison of each model.

Models	Novel Features	Reference
Faster R-CNN	Introduces the Region Proposal Network (RPN) for faster and more accurate two-stage object detection using an anchor mechanism.	Ren, He, Girshick and Sun [22]
YOLO v3	Employs multi-scale predictions and the Darknet-53 feature extraction network for efficient and accurate single-step detection.	Redmon and Farhadi [18]
YOLO v7	Optimizes network architecture and training strategies with a new loss function for faster and more accurate object detection.	Wang, Bochkovskiy and Liao [21]
Improved YOLO v7	Separates classification and regression tasks while using adaptive BiFormer attention to enhance detection accuracy and feature representation.	This paper

2. Material and Methods

2.1. Dataset Preparation

The photographs of Chinese cabbages were captured at the Zhanlin Green Agricultural Picking Garden in Changchun City, Jilin Province, China (125°12′033″ E, 43°59′027″ N) from 5 September to 10 September 2023. The Chinese cabbages were initially sown in seedbeds and later transplanted when they reached the 4–6 leaf stage. During transplantation, Chinese cabbages were spaced 40–45 cm apart from each other and 60–70 cm apart between rows. At the time of capturing the image, Chinese cabbages were 7–10 days post-transplanting and in the seedling stage. The seedling period is defined before 7–8 leaves according to Shanmuganathan and Benjamin [40]. Figure 1 depicts the equipment used to capture images, wherein a CMOS RGB industrial camera (SY011HD-V1, Sichuan Weixin Vision Technology Co., Ltd., Chengdu, China) with a resolution of 1920 × 1080 was vertically affixed to an automatically movable trolley at a height of 65 cm above the ground. The imaging area covered 65 × 110 cm, excluding the tires and body of the trolley. To ensure sample randomness, datasets were collected both in the morning and afternoon on sunny and cloudy days. As the image acquisition equipment automatically moved at a constant speed of 0.4 m/s and took a photo every 2 s, a total of 5466 images were captured. For image labeling, the software Labelme (version 5.2.1, relying on Anaconda3 software for implementation) was utilized, and the acquired images were manually labeled to create jston files identifying Chinese cabbages and weeds. Among the total images collected, 4000 Chinese cabbage images were employed to train models, while 1466 images were used to validate the models.
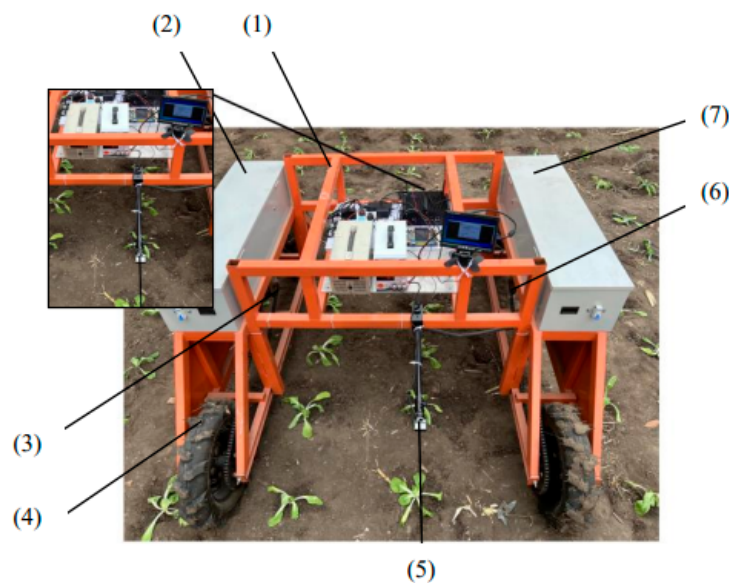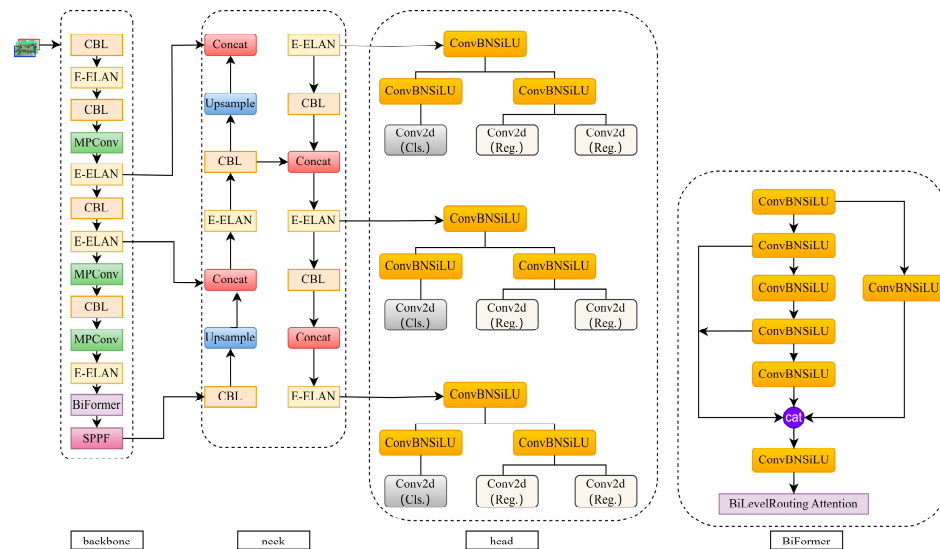

Figure 1. Field image acquisition equipment: (1) vehicle frame, (2) control box (**left**), (3) brushless motor (**left**), (4) wheels, (5) RGB industrial camera, (6) brushless motor (**right**), (7) control box (**right**). The field image acquisition equipment, operating at a constant speed of 0.4 m/s and capturing an image every 2 s, resulted in a total of 5466 images collected, serving as valuable visual data for analyzing and monitoring models in the Chinese cabbage field.

2.2. Integration of Decoupled Head and BiFormer Attention Mechanism to YOLO v7

2.2.1. Overview of Decoupled Head

The decoupling head is a method of the target detection model [41]. It can improve the ability of the model to detect small targets by decoupling different size target detection tasks. The decoupling head focuses on targets of different sizes, especially small ones, to improve detection accuracy [42]. It breaks the object detection task into multiple decoupled heads, each of which is responsible for processing targets of a specific size, thus enabling the model to detect and identify targets of different sizes more accurately [43]. The overlapping weeds and their small sizes make it challenging to distinguish them from normal Chinese

cabbage, and Region-CNN (R-CNN) and Single Shot MultiBox Detector algorithms may have difficulty detecting them accurately. However, the decoupling head can focus on processing small targets [44], so it can improve the detection precision rate and recall rate of Chinese cabbage and weeds. Therefore, decoupling heads are very important for weed detection in Chinese cabbage fields. In this way, the model can better identify and distinguish between normal Chinese cabbage and weeds, thereby improving the planting quality and yield of Chinese cabbage. Figure 2 shows the decoupled head network of the improved YOLO v7 structure diagram and the improved YOLO v7 working mechanism. Specifically, the decoupling head separates the classification and regression branches into two independent parallel subnetworks, where the classification subnetwork predicts the category label of the target (as shown in the black rectangle of Figure 2). The regression subnetwork predicts the boundary box coordinates and the object confidence score of the target (as shown in the red rectangle of Figure 2).

Figure 2. The decoupled head network of improved YOLO v7 structure diagram. The decoupled head network structure separates the original detection head into two distinct branches: one for class label prediction (as shown in the black rectangle) and the other for bounding box coordinate prediction (as shown in the red rectangle), enabling independent predictions and enhancing the model's capability to handle variations in Chinese cabbages and weeds.

2.2.2. Overview of BiFormer Attention Mechanism

Figure 3 depicts the BiFormer attention mechanism network of the improved YOLO v7 structure diagram. BiFormer, a visual transformer for bi-level routing attention, is used to capture remote dependencies in input sequences [31]. Based on a sparse attention mechanism using dynamic query awareness, BiFormer combines the advantages of the transformer and pyramid network (as shown in the right picture of Figure 3). It is widely used in object detection because of the different positions of objects in images and the complex relationship with external factors. In addition, the BiFormer attention mechanism enhances feature extraction by allocating more attention to areas containing small objects [45]. At the same time, it addresses this problem using sparse attention methods to dynamically focus on important features at multiple levels [46].

Figure 3. The BiFormer attention mechanism network of the improved YOLO v7 structure diagram. The BiFormer attention mechanism showcases the integration of bi-level routing attention (as shown in the right picture), enabling YOLO v7 to effectively capture both local and global dependencies in the input feature map.

2.2.3. Integrate Algorithm

The original YOLO v7 integrates the feature extraction and pixel prediction processes in the same network and realizes classification and positioning at the same time through fusion and sharing. However, the focus on classification and positioning is different [47]. Classification pays more attention to the texture content of the target, whereas positioning is more focused on edge information. The decoupled head decouples the two processes and deals with them separately. As shown in Figure 2, this study modified the head network layer of the original YOLO v7 for decoupling, extracted the target location and category information, which were learned separately through different network branches and finally fused. The class label branch predicts the probabilities of different Chinese cabbage and weed classes present in the image, while the bounding box coordinate branch predicts the coordinates (x, y, width, height) of the bounding boxes representing the locations of Chinese cabbages and weeds [41]. The implementation of algorithms to decouple the original YOLO v7 and incorporate the BiFormer attention mechanism involves several steps. First, the image is input into the YOLO network for feature extraction. The decoupling head then processes the features separately for classification and localization: the classification branch predicts the target class probabilities, while the localization branch predicts the bounding box coordinates. The outputs are subsequently fused.

The BiFormer attention mechanism employs coarse-grained region partitioning and fine-grained token-to-token attention to filter out irrelevant pairs and apply detailed attention to the retained regions. This enhances the network's focus on significant features, resulting in more accurate output images. The main BiFormer attention mechanism eliminates the most irrelevant key–value pairs at a coarse region level, resulting in only a small portion of routed regions remaining. Then, fine-grained token-to-token attention is applied in the union of these routed regions [31]. As shown in Figure 4, the key steps mainly include the following contents:

Initially, queries (Q), keys (K), and values (V) are taken as inputs, and an attention function transforms each query into a weighted sum of values, where the weights are computed as normalized dot products between the query and corresponding keys. Q, K, and V are derived as linear projections of the same input.

Next, region partition and input projection are performed. Given a 2D input feature map, it is divided into S × S non-overlapped regions, each containing feature vectors.

Subsequently, region-to-region routing with a directed graph is carried out. The attending relationship, i.e., the regions that should be attended for each given region, is determined by constructing a directed graph.

Token-to-token attention is then applied. Using the region-to-region routing index matrix Ir, fine-grained token-to-token attention can be implemented. For each query token in a region I, it attends to all key–value pairs residing in the union of k-routed regions indexed with Ir (i, 1), Ir (i, 2), . . ., Ir (i, k).

Finally, the output images are generated.

H, W, and C are input images of height, width, and channel; Q, K, and V refer to queries, keys, and values, respectively.
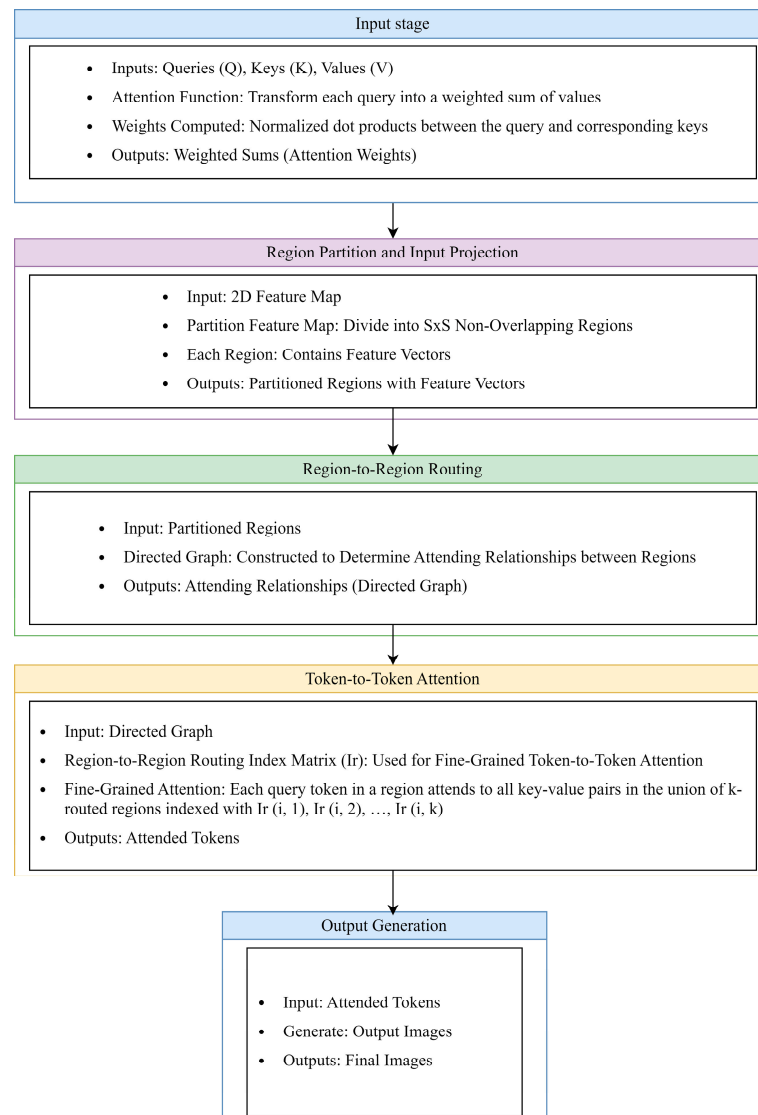
Figure 4. Integration of the BiFormer attention mechanism to YOLO v7. Firstly, Q, K, and V are taken as inputs, and an attention function transforms each query into a weighted sum of values. Next, the input is partitioned into non-overlapping regions, and region-to-region routing with a directed graph is carried out. Token-to-token attention is then applied using the region-to-region routing index matrix. Finally, output images are generated.

2.3. Chinese Cabbages Seedling Belt-Fitting Algorithm (CCSBFA)

Chinese cabbages usually grow shorter. The leaves are numerous, wide, and concentrated near the ground, and the whole cabbage is relatively scattered [48]. Therefore, the Hough transform and linear regression may not accurately detect these changes because

of the intricate spatial arrangement of Chinese cabbage leaves. At the same time, during the rosette stage, overlapping Chinese cabbage leaves lead to a single center point representing multiple Chinese cabbages, resulting in significant errors when using Blob analysis. Additionally, the minimal height difference between Chinese cabbages and weeds results in significant errors when using stereo vision for fitting. Therefore, this study proposed a new improved algorithm. This algorithm aims to overcome the limitations in the fitting of the CCSB and improve the accuracy of the CCSB fitting algorithm (CCSBFA), which can only automatically fit the distribution path of Chinese cabbages based on the recognition algorithm. The framework diagram for this CCSBFA is depicted in Figure 5.
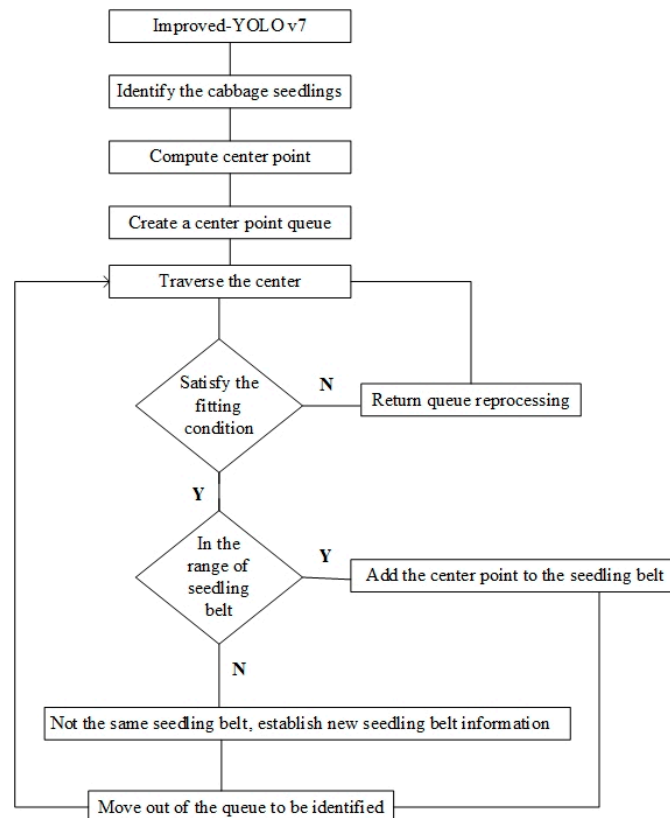
Figure 5. Flow chart of the Chinese cabbage seedling belt-fitting algorithm (CCSBFA). The improved CCSBFA utilizes the recognition algorithm to calculate the number of Chinese cabbages and constructs a center point queue for traversal. The algorithm accurately identifies and groups seedlings into their respective belt-fitting center points of Chinese cabbages.

The process begins by calculating the expected number of Chinese cabbages and the total number of seedling belts in the image using the recognition algorithm. Then, a center point queue is constructed, which traverses through the center points of each Chinese cabbage in the image (Figure 5).

The fitting conditions used to determine the same CCSB were established based on the average Chinese cabbage spacing of 65 cm and half of the row spacing of 20 cm measured in the field. These measurements allowed for a proportional relationship between the camera image and the actual terrain to be established. Specifically, as shown in Figure 6, a row spacing of 65 cm between Chinese cabbages corresponds to 200 pixels on the camera, while a half-row spacing of 20 cm corresponds to 60 pixels on the camera. In Figure 6, the x direction is defined as the forward direction of the trolley, while the y direction is perpendicular to the movable trolley's forward direction. The algorithm utilized this coordinate system to determine if two Chinese cabbages belong to the same seedling belt. In the y direction, the algorithm checks if the distance between the center point of a plant in the queue and the center point of the next Chinese cabbage is less than 60 pixels. If this

condition is satisfied, the algorithm further checks if the distance between the two center points in the x direction is less than 200 pixels. Both of these fitting conditions must be met simultaneously to determine that the Chinese cabbages belong to the same seedling belt. If the conditions are not satisfied, the queue is discarded, and the algorithm starts creating a new CCSB by traversing all the points again.
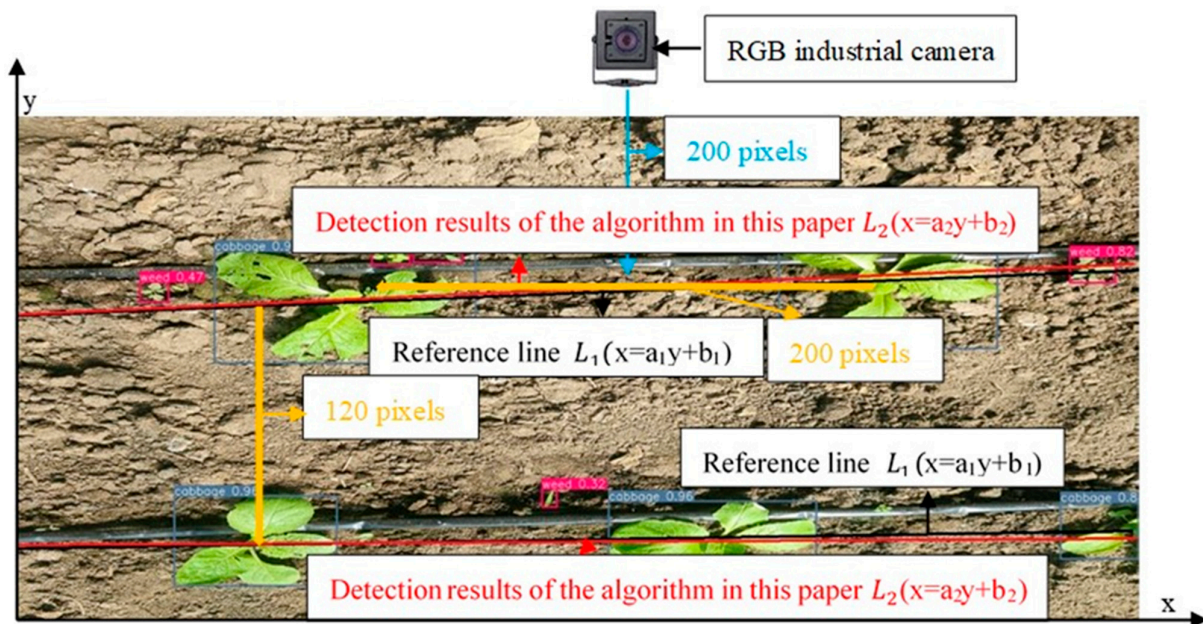
Figure 6. Schematic diagram of angle error of the Chinese cabbage seedling belt (CCSB). The black line L_1 (x = a_1y + b_1) represents the centerline of the artificially marked seedling belt and the red line L_2 (x = a_2y + b_2) corresponds to the centerline of the detection algorithm proposed in this paper.

The implementation of this improved CCSBFA enables the precise fitting of the CCSB. To quantitatively assess the algorithm's efficacy in detecting the CCSB, this study adopted the method proposed by Jiang et al. [49]. As illustrated in Figure 6, the angle between the CCSB detected by the algorithm and the manually marked reference line was employed as the deviation angle accuracy to evaluate the algorithm's accuracy in the CCSB detection. Specifically, the images were labeled manually. The CCSB extracted by the algorithm was then compared using the manually labeled centerline as a baseline. Finally, different fitting algorithms were evaluated using the identification rate and deviation angle as indicators. In this paper, 300 random images were used to evaluate the accuracy of the fitting algorithms.

The angle between the two lines is denoted as θ. A smaller θ indicates a higher fitting accuracy of the system. θ is illustrated in Equation (1):

$$\theta = \arctan\left(\frac{|a_1 - a_2|}{1 + a_1 \times a_2}\right) \tag{1}$$

where a_1 and a_2 represent the slope of the lines.

The average error angle is calculated by averaging all the error angles. The calculation formula is shown in Equation (2):

$$\bar{\theta} = \frac{\sum_{w=1}^{N_t} \theta_w}{N_t} \tag{2}$$

where $\bar{\theta}$ is the average error angle (°), N_t is the total number of image samples, and θ_w is the linear error angle of the wth image sample (°).

The correct fitting ratio (R) is calculated as follows in Equation (3):

$$R = \frac{n}{N_t} \tag{3}$$

where R is the correct proportion of the detection CCSB, and n is the number of images for detecting the CCSB. For example, if the detection is 100 images, which is within 5° of the target angles of 60 images, then R is equal to 60%.

2.4. Model Training Environment and Performance Evaluation

The computing hardware environments were as follows: Intel i7-14700K core processor, 3.20 GHz main frequency, 16 GB RAM, and NVIDIA GeForce RTX 4080 graphics processor with 16 GB graphics memory. The operating system was Windows 10, with Cuda 10.2, torch 1.12.0, torchvision 0.13.0, cuDNN 7.6.5, and Python 3.10.4. The model was trained for 300 epochs, with a starting learning rate of 1×10^{-5} and a learning rate momentum of 0.6. A batch size of 4 was used, and the input image size was set to 640×640, with 50 iterations. The 'Adam' optimizer was utilized to optimize the network, ensuring continuous learning rate adjustments and preventing overfitting during training models.

This paper applied various metrics to evaluate model performance. Precision rate, recall rate, F1score, and mean average precision (mAP) were used to evaluate model performance. Specifically, the F1-Curve was utilized to measure the performance of the binary model by adjusting the threshold to achieve a balance between the precision rate and recall rate for an optimal F1score. The mAP@$_{0.5}$ (at a threshold of 0.5, averaging AP for all categories across all images) and Frames Per Second (the number of frames transmitted per second, FPS) evaluate the performance of the model. Furthermore, this paper used radar maps to comprehensively evaluate various models. Meanwhile, a larger area enclosed by the metrics indicates better overall performance of the models. The formulas for these metrics are as follows:

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

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

$$\text{F1}_{\text{score}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{AP} = \sum \left(\text{Recall}_{(i+1)} - \text{Recall}_{(i)} \right) \times \text{Precision}_{(i+1)} \tag{7}$$

where AP is short for average precision.

$$\text{mAP} = \frac{\sum \text{APc}}{\text{N}} \tag{8}$$

where N is the total number of classes, and APc is the AP of class C.

To compare the accuracy and effectiveness of different target detection models, and find a suitable detection model for the identification of Chinese cabbages and weeds, comparative and ablation experiments were carried out in this paper. Comparative experiments evaluate the performance of different algorithms by comparing them on the same dataset. Specifically, three models were used to conduct the comparative experiments, and the same Chinese cabbages and weeds datasets were trained and verified simultaneously. Precision rate, recall rate, F1score, mAP@$_{0.5}$, FPS, and radar map area were used to evaluate the advantages and disadvantages of different models. Three object detection algorithms, Faster R-CNN, YOLO v3, and YOLO v7, were selected. The advantage of Faster R-CNN is that it integrates feature extraction, proposal, bounding box regression, and classification into one network, which significantly improves the comprehensive performance, especially the precision rate. YOLO v3 and YOLO v7 are both real-time object detection algorithms. YOLO v3 detects targets on multiple feature layers of different scales to improve detection accuracy. YOLO v3 used Darknet-53 to obtain better target classification capabilities and used convolution nuclei of different sizes to extract features of different scales to obtain

better robustness [18]. YOLO v7 introduced multi-scale feature fusion to further improve the accuracy of the model. In addition, YOLO v7 improved the ability to capture target information at different scales and semantics by fusing and weighting features at different levels [21]. At the same time, in the corn and wheat field of weeds identification, YOLO v3, YOLO v7, and Faster R-CNN have achieved important results and have been widely used [50,51]. Therefore, this study has chosen these three models for in-depth comparison to explore their performance, recognition accuracy, and practical application effects in the Chinese cabbage field.

Furthermore, to improve the detection precision rate of YOLO v7, the original detection head was decoupled and the BiFormer attention mechanism was introduced in the backbone. At the same time, to evaluate the effects of the two methods on the improvement of YOLO v7, ablation experiments were conducted with precision rate, recall rate, mAP@$_{0.5}$, FPS, and radar area as performance evaluation indexes. The ablation test consisted of four groups: (1) YOLO V7-a, the original YOLO v7; (2) YOLO V7-b: YOLO v7 with decoupled head; (3) YOLO v7-c: YOLO v7 adds BiFormer attention mechanism on the backbone C3 module; (4) YOLO v7-d: YOLO v7 with decoupled head and BiFormer attention mechanism.

3. Results and Discussion

3.1. Comparison Experiment

In the comparative experiment assessing object detection models, YOLO v7 demonstrated superior overall performance compared to YOLO v3 and Faster R-CNN. As shown in Figure 7, YOLO v7's area was the largest at 14.525×10^{-7}, followed by YOLO v3 at 9.468×10^{-7}, and Faster R-CNN at 8.942×10^{-7}. A larger radar map area indicates better overall performance, affirming YOLO v7's leading position. Despite its smaller radar map area, Faster R-CNN excelled in precision rate, surpassing YOLO v3 by 1.1% and YOLO v7 by 2.11%. This advantage is attributed to the inclusion of the Region Proposal Network (RPN) [52], which generates candidate frames with diverse sizes and aspect ratios, enhancing precision in detecting objects with varying scales and shapes [22]. YOLO v7, while slightly underperforming YOLO v3 in precision rate (1.01% lower) and recall rate (1.46% lower), excelled in other critical metrics. YOLO v7 improved its F1score by 4%, mAP@$_{0.5}$ by 25.82%, and frames per second (FPS) by 48%, due to its use of multi-scale feature fusion and attention mechanisms. These enhancements allow YOLO v7 to effectively capture detailed information and process data more efficiently.

Figure 8 shows partial detection results of Chinese cabbage and weed datasets using Faster R-CNN, YOLO v3, and YOLO v7. Chinese cabbages identified by the model are selected in the blue box. The orange box selects part of the identified weeds. The selected part of the black rectangle box is the result of incorrect identification; that is, the Chinese cabbages are identified as weeds or the weeds are identified as Chinese cabbages. The red rectangular box is selected as an unidentified weed or Chinese cabbage. As shown in Figure 8, Faster R-CNN has the least false recognition and unrecognized cases in the process of object detection among the three models, followed by YOLO v3 and YOLO v7. This is consistent with the results of the above comparative experiment. In addition, unrecognized weeds or Chinese cabbages appear in YOLO v3 and YOLO v7, indicating that there is still room for optimization in target detection. YOLO v7 is an improvement on YOLO v3 and has advantages in the processing speed that the YOLO v3 does not have, while there is a small difference in precision rate between YOLO v7 and YOLO v3. The overall performance of YOLO v7 is better than that of YOLO v3. Therefore, YOLO v7 was selected for improvement in recognizing Chinese cabbages. Meanwhile, compared with Faster R-CNN and YOLO v7, Faster R-CNN has an excellent precision rate (73.61%) since it is a region-based convolutional neural network. Faster R-CNN is mainly composed of the RPN and the full connection layer for object detection. The RPN is used to generate candidate target boxes, and the full connection layer is used to classify and regress these candidate boxes [22]. Therefore, the advantages of Faster R-CNN can be combined into

YOLO v7, which is manifested in the decoupling head design of the original YOLO v7 by using Faster R-CNN.
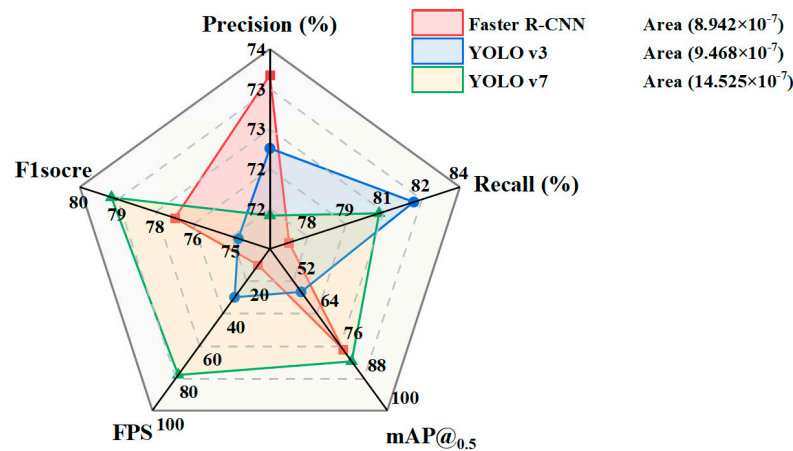
Figure 7. Comparison experiment results based on radar map. Figure 7 presents a comparison of different models' performance in identifying Chinese cabbages and weeds. YOLO v7 achieved the highest overall performance, with values of precision rate, recall rate, F1score, mAP@$_{0.5}$, and FPS, and radar areas are 71.5%, 80.6%, 79, 81.8%, 78%, and 14.525×10^{-7}, respectively.
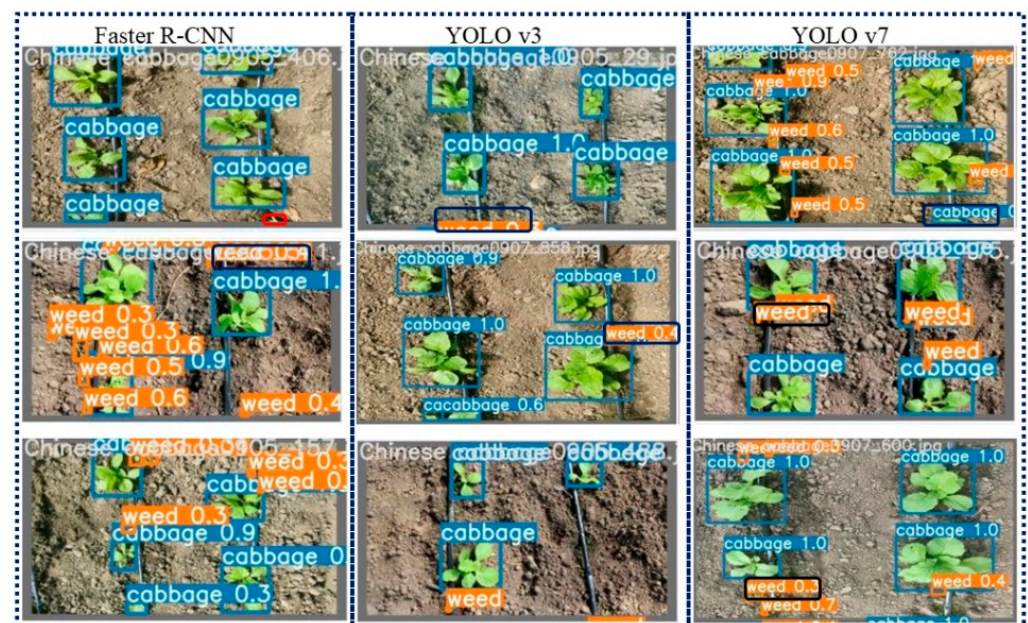

Figure 8. Partial detection results of Chinese cabbage and weed datasets using Faster R-CNN, YOLO v3, and YOLO v7. Faster R-CNN has the least false recognition and unrecognized images in the process of object detection among the three types of models, followed by YOLO v3 and YOLO v7.

3.2. Ablation Experiment

The ablation experiments, as illustrated in Figure 9, demonstrate a marked improvement in the overall performance of YOLO v7-b compared to its predecessor, YOLO v7-a. The key enhancement in YOLO v7-b is the introduction of the decoupled head, which led to significant performance gains. Specifically, YOLO v7-b shows a 15.9% improvement in precision rate over YOLO v7-a [53]. This improvement is largely due to the decoupled head structure, which allows for more flexible processing of different scales and semantic information, thereby enhancing multi-scale feature fusion capability. This flexibility enables more accurate pixel-level predictions, directly contributing to the increased precision rate. In addition to precision rate, other performance metrics also saw notable improvements

with YOLO v7-b. The F1score increased by 1%, the recall rate by 0.7%, and mAP@$_{0.5}$ by 0.5% [54]. The decoupled head integrates feature information from various scales and adds branches at different levels of the backbone network, which enhances the segmentation ability for multi-scale targets and improves the recall rate [55]. This multi-scale target detection and improved segmentation accuracy also contribute to the higher mAP@$_{0.5}$. The F1score of YOLO v7-b is 1% higher than that of YOLO v7-a. The decoupled heads better retain details and edge information, leading to improved segmentation accuracy [56]. Moreover, YOLO v7-b achieves a superior balance between model efficiency and inference accuracy compared to YOLO v7-a, making it more precise in identifying Chinese cabbages and weeds. In summary, the decoupled head significantly enhances YOLO v7's physical-level accuracy and enables parallel processing at the network level. These enhancements make YOLO v7-b more efficient and accurate in identifying Chinese cabbages and weeds, affirming its superiority over YOLO v7-a [42].
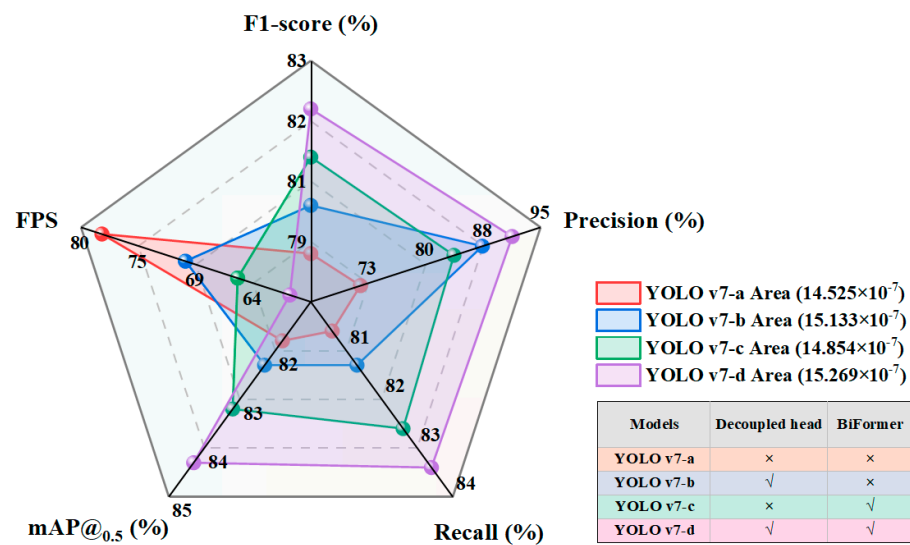
Figure 9. Ablation experiment results based on the radar map. YOLO v7-d, incorporating both the decoupled head and BiFormer attention mechanism, achieves the highest scores across various metrics including precision rate, recall rate, mAP@$_{0.5}$, F1score, FPS, and radar area, with values of 91.3%, 83.4%, 82.0%, 84.3%, 60%, and 15.269×10^{-7}, respectively.

The ablation experiments presented in Figure 9 demonstrate that both YOLO v7-b and YOLO v7-c outperform YOLO v7-a, with significant enhancements in various performance metrics. YOLO v7-b, with its decoupled head, improves the precision rate by 15.9%, recall rate by 0.7%, F1score by 1%, and mAP@$_{0.5}$ by 0.5%, due to its enhanced multi-scale feature fusion and segmentation capabilities [31]. Similarly, YOLO v7-c, which incorporates the BiFormer attention mechanism, achieves improvements of 12.2% in the precision rate, 2% in recall rate, 2% in F1score, and 1.4% in mAP@$_{0.5}$ [31]. The BiFormer mechanism enhances the model's focus on crucial features of small targets like weeds and leverages a pyramid network structure to better capture multi-scale target features [46]. These advancements enable YOLO v7-c to more accurately identify and detect Chinese cabbages and weeds, demonstrating the effectiveness of integrating attention mechanisms and advanced structural designs in improving object detection performance [57].

Figure 9 shows the ablation experiment results based on a radar map. YOLO v7-d, which integrates both the decoupled head and BiFormer attention mechanism, achieves the highest scores across various metrics including precision rate (91.3%), recall rate (83.4%), mAP@$_{0.5}$ (82.0%), F1score (84.3%), FPS (60%), and radar area (15.269×10^{-7}). Figure 10 displays partial detection results of Chinese cabbage and weed datasets using YOLO v7-a, YOLO v7-b, YOLO v7-c, and YOLO v7-d. The color box meanings in Figure 10 are consistent with those in Section 3.1. As shown, YOLO v7-d significantly improved the accuracy of identifying Chinese

cabbages and weeds compared with YOLO v7-a, with no identification errors or unrecognized cases in YOLO v7-d.
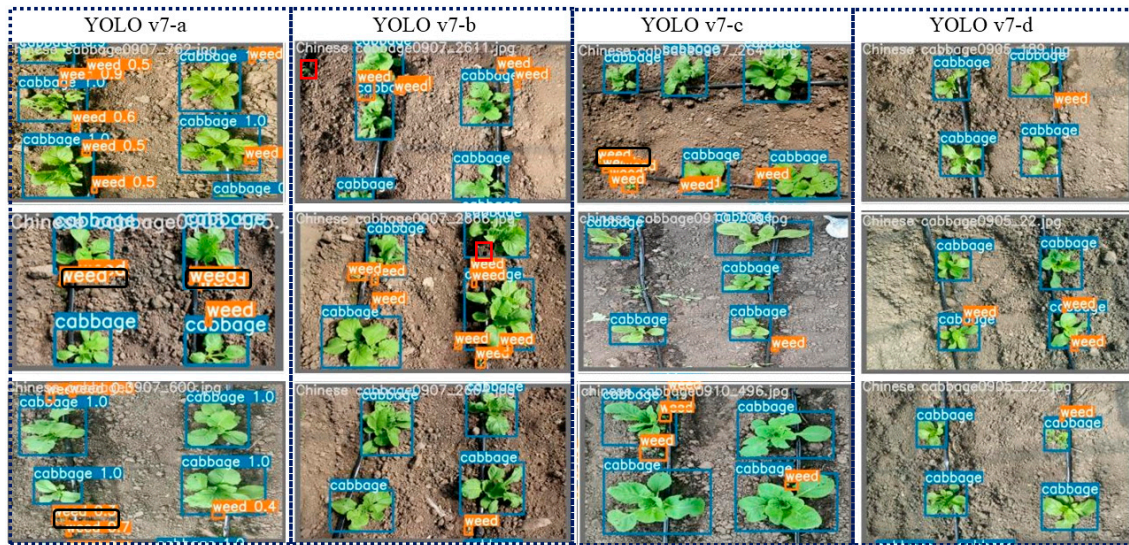
Figure 10. Partial detection results of Chinese cabbage and weed dataset by YOLO v7-a, YOLO v7-b, YOLO v7-c, and YOLO v7-d. Compared with YOLO v7-a, YOLO v7-d achieves the greatest improvement in precision rate, recall rate, F1socre, mAP@$_{0.5}$, and radar area, with improvements of 3%, 19.8%, 2.8%, 2.5%, and 0.744×10^{-7}, respectively.

The ablation experiments, as shown in Figure 9, highlight the advancements made with different modifications to YOLO v7. Modifying the decoupled head alone increased the precision rate by 15.9%, while adding the BiFormer attention mechanism alone increased it by 12.2%. However, neither adjustment alone achieved optimal performance. By integrating both the decoupled head structure and BiFormer attention mechanism, YOLO v7-d achieved substantial improvements: the precision rate increased by 19.8%, F1score by 3%, recall rate by 2.8%, and mAP@$_{0.5}$ by 2.5%. Although YOLO v7-d's FPS is lower compared to other models, it still operates effectively in real-time. The combination of these enhancements ensures high accuracy in detecting Chinese cabbages and weeds, providing a balanced trade-off between accuracy and speed. Consequently, YOLO v7-d is well-suited for real-time applications where precise object detection is critical, offering superior performance across key metrics despite a slight reduction in processing speed [58].

The improvements are attributed to two key aspects: firstly, decoupling the head enhances the detection ability for small targets, which typically have lower resolution and less key feature information, making them challenging for models to detect and identify correctly [59]. YOLO v7-d, with decoupled heads, can better focus on Chinese cabbage and weeds of different sizes, thereby improving detection the precision and recall rate for small targets (YOLO v7-a and YOLO v7-d). Secondly, the BiFormer attention mechanism helps YOLO v7-d better understand global information and context, improving the recognition precision rate for Chinese cabbage and weed detection [21]. In object detection tasks, global information and the context around the target are crucial for accurate location and classification. The BiFormer attention mechanism helps the model capture this information more effectively, allowing for more accurate environmental and contextual understanding and more confident predictions (YOLO v7-c and YOLO v7-d) [60]. As demonstrated in Figures 9 and 10, coupling the decoupled head and the BiFormer attention mechanism allows them to complement each other, further enhancing overall performance. This combination optimizes feature representation and learning ability, making the model more efficient in object detection tasks. This coupling leverages the advantages of both components, improving the model's performance in various complex scenarios. It also simplifies the model structure, reducing complexity and improving training and reasoning

efficiency [21]. Thus, coupling the decoupled head and the BiFormer attention mechanism leads to better overall performance.
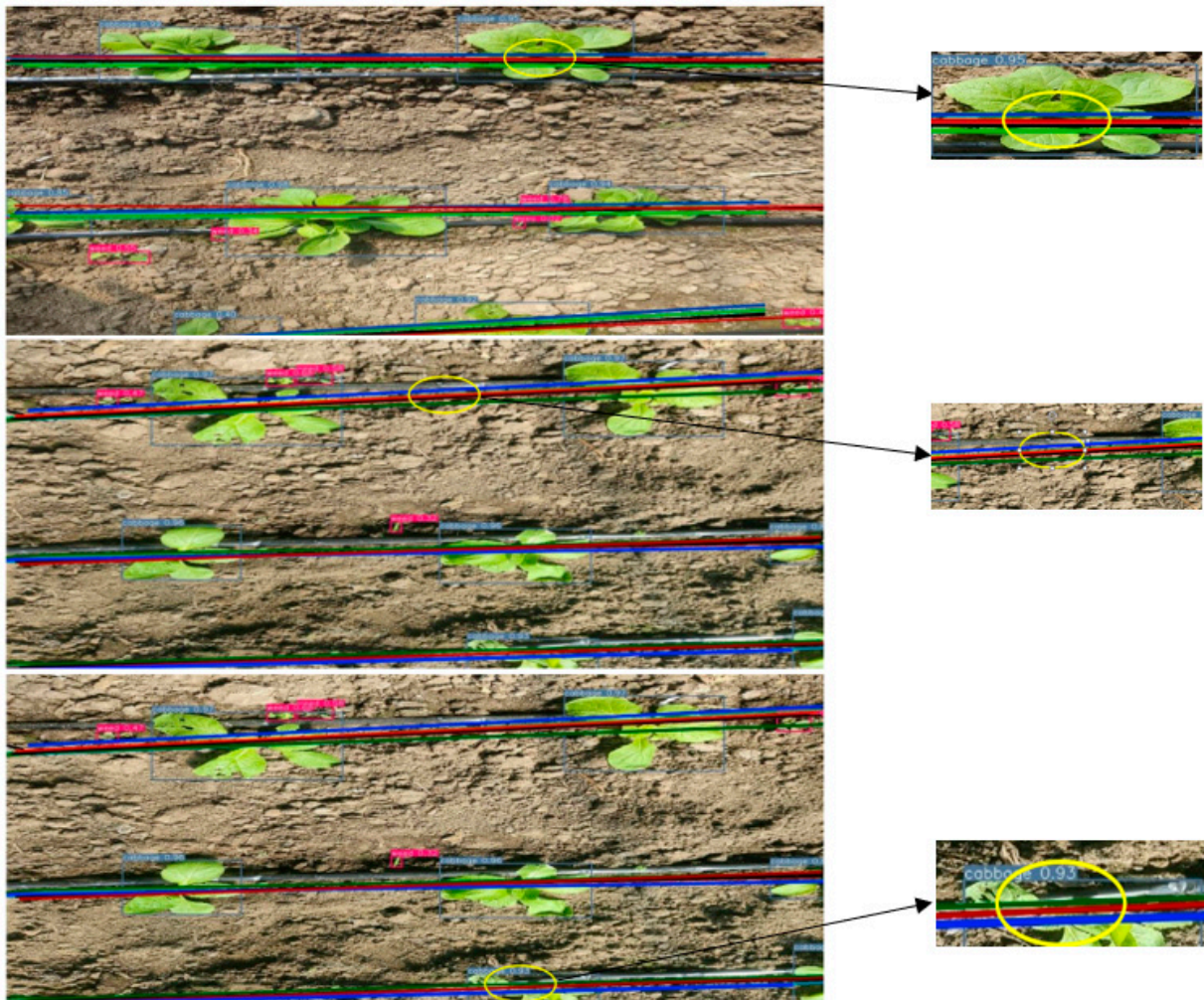
3.3. The Results of the Chinese Cabbage Seedling Belt-Fitting Algorithm (CCSBFA)

Based on YOLO v7-d, the identified Chinese cabbages were fitted to provide the basis for the autonomous navigation of agricultural machines. In this paper, the CCSBFA is compared with the existing fitting algorithm, mainly including the Hough transform and linear regression. The fitting algorithms' accuracy was evaluated using 300 random images, and the results are presented in Figure 11. The identification rate of the CCSB proposed in this paper is 94.2%, with a deviation angle accuracy of 95%. In comparison, the Hough transform and linear regression achieved an accuracy of 78.0% and 73.2% for the CCSB, respectively. At the same time, the deviation angle accuracy was 79.2% and 75.6%, respectively. The Hough transform has a good fitting effect for specific shapes such as lines and circles and is suitable for fitting problems with definite geometric features [61]. The linear regression algorithm is simple, intuitive, fast, and suitable for real-time fitting and large-scale data processing [58]. It also has a good fitting effect for data with obvious linear relationships, but a poor fitting effect for data with non-linear or complex shapes. However, CCSB recognition needs to consider the morphology and characteristics of the Chinese cabbages in the images, which is different from the Hough transform and linear regression. Specifically, the morphology of Chinese cabbage is complex and varied, and the features are irregular [15]. The Hough transform and linear regression make it difficult to accurately locate and identify the location of the CCSB. In addition, the characteristics of Chinese cabbage seedlings include leaf color and texture differences. These features need to be analyzed and identified by more complex image processing and machine learning algorithms. As a result, the recognition accuracy of the proposed algorithm is an average of 18.6% higher than that of the two algorithms, with a deviation angle accuracy that is 17.6% higher, on average. When the deviation angle was small (less than 5°), the fitting effect was good, and it could be applied to agricultural machinery navigation [62]. Consequently, the algorithm exhibited superior performance in identifying the CCSB and demonstrates significant advantages in deviation accuracy over the Hough transform and linear regression.
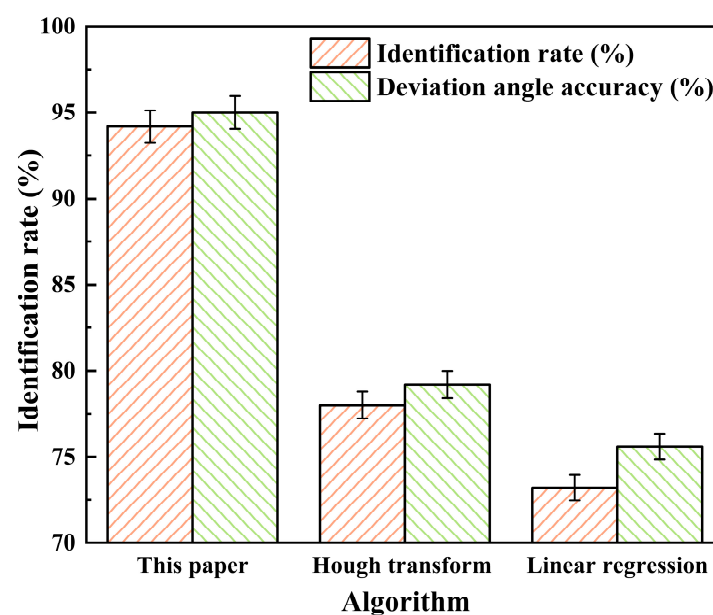
Figure 11. Comparison of algorithms for Chinese cabbage seedling belt-fitting algorithm (CCSBFA) accuracy. Figure 11 presents a comparison of various algorithms' performance in identifying the Chinese cabbage seedling belt (CCSB). The algorithm proposed in this paper achieves the highest identification rate (94.2%) and deviation angle accuracy (95.0%) among the evaluated algorithms, surpassing the Hough transform and linear regression.

Figure 12 shows that after target detection and recognition of YOLO v7-d, three fitting algorithms were used to fit the identified CCSB. Among them, the artificially labeled line, Hough transform fitting line, linear regression fitting line, and CCSB fitting line are represented by black, blue, green, and red colors, respectively. As shown in Figure 12, the red line and the black line almost coincide, while the blue line and the green line have a large deviation from the black line, of which the blue line and the black line have the largest deviation. This shows that the Hough transform and linear regression are significantly affected by the central point of Chinese cabbages. The fitting results are not significant (as shown in the yellow ellipse in Figure 12). At the same time, due to the large distance between Chinese cabbages, the Hough transform and linear regression make mistakes in the fitting process of the next Chinese cabbage, resulting in a large fitting error of the CCSB. However, the CCSBFA accurately describes the spatial distribution of Chinese cabbages, effectively reflecting the CCSB information and its agreement with the actual data. In addition, the high CCSB recognition rate and deviation angle accuracy also indicate that the CCSBFA has a high accuracy for CCSB detection [63]. This has important implications for the full automation of Chinese cabbage cultivation that relies on an accurate CCSBFA.

Figure 12. The results of the Chinese cabbage seedling belt-fitting algorithm (CCSBFA). Three images (92nd, 2582nd, and 4257nd) were randomly selected for mapping in this paper. The red line, the black line, the blue line, and the green line represent the algorithm presented in this paper, the manually labeled Chinese cabbage seedling belt (CCSB), the results of linear regression, and the Hough transform results of the CCSBFA, respectively. The red line and the black line are nearly coincident, while the blue line and the green line deviate significantly from the black line, with the blue line showing the largest deviation from the black line.

4. Conclusions

This study aimed to enhance the precision and efficiency of Chinese cabbage (*Brassica pekinensis* Rupr.) and weed detection in agricultural fields by comparing three advanced object detection algorithms: Faster R-CNN, YOLO v3, and YOLO v7. Faster R-CNN was selected for its decoupled design and integrated processes, contributing to a higher precision rate. YOLO v3 and YOLO v7 were chosen for their real-time capabilities, having demonstrated success in corn and wheat weed identification. This comparative evaluation assessed their performance and practical application in Chinese cabbage fields. YOLO v7, an improvement over YOLO v3, offers better overall performance and processing speed. However, Faster R-CNN excels in precision rate due to its region-based convolutional neural network architecture. The decoupled head design of YOLO v7, combined with the precision advantages of Faster R-CNN, addresses the challenges of leaf occlusion and smaller weeds in Chinese cabbage identification. Furthermore, the BiFormer attention mechanism enhances the network structure to solve the issue of overlapping leaves, enabling effective and precise image recognition. Our research introduced an innovative recognition algorithm that integrates YOLO v7 with a decoupled head and BiFormer attention mechanism to improve the precision rate and mean average precision (mAP) in Chinese cabbage seedling belt (CCSB) recognition. Experimental results demonstrated the superior performance of the improved YOLO v7, achieving a maximum mAP@$_{0.5}$ of 84.3%, surpassing the performance of other models like Faster R-CNN, YOLO v3, and the original YOLO v7. This enhancement has significant implications for automating and improving the efficiency of Chinese cabbage farming operations. Additionally, this study proposed a novel Chinese cabbage seedling belt-fitting algorithm (CCSBFA), which was compared to the Hough transform and linear regression. The CCSBFA algorithm achieved an 18.6% higher recognition accuracy and a 17.6% higher deviation angle accuracy, on average, compared to the other methods. The CCSBFA effectively addresses the limitations of traditional algorithms, enabling precise CCSB fitting.

In summary, the recognition algorithm based on YOLO v7, combined with a decoupled head and BiFormer attention mechanism, offers an accurate and efficient solution for recognizing Chinese cabbages and weeds. This accurate identification and fitting of CCSB supports intelligent weeding, reduces the reliance on chemical herbicides, and improves the quality and safety of agricultural products. This research contributes to computer vision and agricultural technology, providing valuable insights and paving the way for further advancements in crop monitoring and guidance systems.

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

References

1. He, X.; Zhang, D.; Yang, L.; Cui, T.; Zhong, X. Design and experiment of a GPS-based turn compensation system for improving the seeding uniformity of maize planter. *Comput. Electron. Agric.* **2021**, *187*, 106250. [CrossRef]
2. Thomas, M.; Reger, J.; Stumpenhausen, J.; Bernhardt, H. Lidar and radar enable the next generation of dairy cattle feeding. *Appl. Eng. Agric.* **2022**, *38*, 207–217.
3. Gao, X.; Li, Y.; Bao, J. Efficient carrier acquisition and tracking for high dynamic and weak satellite signals. *J. Commun.* **2016**, *11*, 644–652. [CrossRef]
4. Diao, Z.; Yan, J.; He, Z.; Zhao, S.; Guo, P. Corn seedling recognition algorithm based on hyperspectral image and lightweight-3D-CNN. *Comput. Electron. Agric.* **2022**, *201*, 107343. [CrossRef]
5. Cury, A.; Crémona, C. Pattern recognition of structural behaviors based on learning algorithms and symbolic data concepts. *Struct. Control Health Monit.* **2012**, *19*, 161–186. [CrossRef]
6. Xie, B.; Jin, Y.; Faheem, M.; Gao, W.; Liu, J.; Jiang, H.; Cai, L.; Li, Y. Research progress of autonomous navigation technology for multi-agricultural scenes. *Comput. Electron. Agric.* **2023**, *211*, 107963. [CrossRef]
7. Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Zou, X. Recognition and Localization Methods for Vision-Based Fruit Picking Robots: A Review. *Front. Plant Sci.* **2020**, *11*, 510. [CrossRef] [PubMed]
8. Sugahara, K.; Nanseki, T.; Fukatsu, T. Verification of a Prototype System to Recognize Agricultural Operations Automatically based on RFID. In Proceedings of the World Conference on Agricultural Information and IT, IAALD AFITA WCCA 2008, Tokyo, Japan, 24–27 August 2008; Tokyo University of Agriculture: Fuchu, Japan, 2008.
9. Liu, F.; Yang, Y.; Zeng, Y.; Liu, Z. Bending diagnosis of rice seedling lines and guidance line extraction of automatic weeding equipment in paddy field. *Mech. Syst. Signal Process.* **2020**, *142*, 106791. [CrossRef]
10. Montalvo, M.; Pajares, G.; Guerrero, J.M.; Romeo, J.; Guijarro, M.; Ribeiro, A.; Ruz, J.J.; Cruz, J.M. Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Syst. Appl.* **2012**, *39*, 11889–11897. [CrossRef]
11. Zhang, X.; Li, X.; Zhang, B.; Zhou, J.; Tian, G.; Xiong, Y.; Gu, B. Automated robust crop-row detection in maize fields based on position clustering algorithm and shortest path method. *Comput. Electron. Agric.* **2018**, *154*, 165–175. [CrossRef]
12. Song, Y.; Liu, Y.; Liu, L.; Zhu, D.; Chen, L. Extraction Method of Navigation Baseline of Corn Roots Based on Machine Vision. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* **2017**, *48*, 38–44.
13. Zhai, Z.; Zhu, Z.; Du, Y.; Song, Z.; Mao, E. Multi-crop-row detection algorithm based on binocular vision. *Biosyst. Eng.* **2016**, *150*, 89–103. [CrossRef]
14. Deng, W.; Huang, Y.; Zhao, C.; Chen, L.; Wang, X. Bayesian discriminant analysis of plant leaf hyperspectral reflectance for identification of weeds from cabbages. *Afr. J. Agric. Res.* **2016**, *11*, 551–562.
15. Gao, P.P.; Zhang, X.M.; Xue, P.Y.; Dong, J.W.; Dong, Y.; Zhao, Q.L.; Geng, L.P.; Lu, Y.; Zhao, J.J.; Liu, W.J. Mechanism of Pb accumulation in Chinese cabbage leaves: Stomata and trichomes regulate foliar uptake of Pb in atmospheric PM2.5. *Environ. Pollut.* **2022**, *293*, 118585. [CrossRef] [PubMed]
16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
17. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
18. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
19. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
20. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
21. 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.
22. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
23. Wu, W.; Wu, X.; Cai, Y.; Zhou, Q. Deep coupling neural network for robust facial landmark detection. *Comput. Graph.* **2019**, *82*, 286–294. [CrossRef]
24. Piotrowski, A.P.; Napiorkowski, J.J. A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling. *J. Hydrol.* **2013**, *476*, 97–111. [CrossRef]
25. Zhao, W.; Alwidian, S.; Mahmoud, Q.H. Adversarial training methods for deep learning: A systematic review. *Algorithms* **2022**, *15*, 283. [CrossRef]
26. Zhu, Y.; Wu, Y.; Liu, Q.; Guo, T.; Qin, R.; Hui, J. A backward control based on σ -Hopf oscillator with decoupled parameters for smooth locomotion of bio-inspired legged robot. *Robot. Auton. Syst.* **2018**, *106*, 165–178. [CrossRef]
27. Xiao, Y.; Wang, X.; Zhang, P.; Meng, F.; Shao, F. Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information. *Sensors* **2020**, *20*, 5490. [CrossRef] [PubMed]
28. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]

29. Zheng, Z.; Hu, Y.; Guo, T.; Qiao, Y.; He, Y.; Zhang, Y.; Huang, Y. AGHRNet: An attention ghost-HRNet for confirmation of catch-and-shake locations in jujube fruits vibration harvesting. *Comput. Electron. Agric.* **2023**, *210*, 107921. [CrossRef]

30. Chen, D.; Wang, D.; Hu, H. Transformer with Sparse Self-Attention Mechanism for Image Captioning. *Electron. Lett.* **2020**, *56*, 764–766.

31. Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R. BiFormer: Vision Transformer with Bi-Level Routing Attention. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 10323–10333.

32. Xia, Z.; Pan, X.; Song, S.; Li, L.; Huang, G. Vision Transformer with Deformable Attention. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.

33. Liu, Y.; Jiang, P.-T.; Petrosyan, V.; Li, S.-J.; Bian, J.; Zhang, L.; Cheng, M.-M. Del: Deep embedding learning for efficient image segmentation. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; p. 870.

34. Yang, F.; Wang, T.; Wang, X. Student Classroom Behavior Detection based on YOLOv7-BRA and Multi-Model Fusion. *arXiv* **2023**, arXiv:2305.07825.

35. Illingworth, J.; Kittler, J. A survey of the Hough transform. *Comput. Vis. Graph. Image Process.* **1988**, *44*, 87–116. [CrossRef]

36. Guerrero, J.M.; Ruz, J.J.; Pajares, G. Crop rows and weeds detection in maize fields applying a computer vision system based on geometry. *Comput. Electron. Agric.* **2017**, *142*, 461–472. [CrossRef]

37. Billingsley, J.; Schoenfisch, M. The successful development of a vision guidance system for agriculture. *Comput. Electron. Agric.* **1997**, *16*, 147–163. [CrossRef]

38. Lazaros, N.; Sirakoulis, G.C.; Gasteratos, A. Review of stereo vision algorithms: From software to hardware. *Int. J. Optomechatron.* **2008**, *2*, 435–462. [CrossRef]

39. Wang, S.; Yu, S.; Zhang, W.; Wang, X. The identification of straight-curved rice seedling rows for automatic row avoidance and weeding system. *Biosyst. Eng.* **2023**, *233*, 47–62. [CrossRef]

40. Shanmuganathan, V.; Benjamin, L. The influence of sowing depth and seed size on seedling emergence time and relative growth rate in spring cabbage (*Brassica oleracea* var. capitata L.). *Ann. Bot.* **1992**, *69*, 273–276. [CrossRef]

41. Wang, H.; Jin, Y.; Ke, H.; Zhang, X. DDH-YOLOv5: Improved YOLOv5 based on Double IoU-aware Decoupled Head for object detection. *J. Real-Time Image Process.* **2022**, *19*, 1023–1033. [CrossRef]

42. Yuan, G.; Liu, G.; Chen, J. A Decoupled YOLOv5 with Deformable Convolution and Multi-scale Attention. In *Knowledge Science, Engineering and Management*; Memmi, G., Yang, B., Kong, L., Zhang, T., Qiu, M., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 3–14.

43. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.

44. Li, J.; Zhang, Y.; Qian, C. The enhanced resource modeling and real-time transmission technologies for Digital Twin based on QoS considerations. *Robot. Comput.-Integr. Manuf.* **2022**, *75*, 102284. [CrossRef]

45. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef] [PubMed]

46. Yang, Z.; Feng, H.; Ruan, Y.; Weng, X. Tea Tree Pest Detection Algorithm Based on Improved Yolov7-Tiny. *Agriculture* **2023**, *13*, 1031. [CrossRef]

47. Bek, M.K.; Shaheen, E.M.; Elgamel, S.A. Classification and Mathematical Expression of Different Interference Signals on a GPS Receiver. *Navigation* **2015**, *62*, 23–37. [CrossRef]

48. Grunicheva, E.A. *Studies on Chinese Cabbage in Glasshouses and Frames*; CABI Digital Library: Wallingford, UK, 1970.

49. Jiang, G.; Wang, Z.; Liu, H. Automatic detection of crop rows based on multi-ROIs. *Expert Syst. Appl.* **2015**, *42*, 2429–2441. [CrossRef]

50. Liu, S.; Jin, Y.; Ruan, Z.; Ma, Z.; Gao, R.; Su, Z. Real-Time Detection of Seedling Maize Weeds in Sustainable Agriculture. *Sustainability* **2022**, *14*, 15088. [CrossRef]

51. Quan, L.; Feng, H.; Lv, Y.; Wang, Q.; Zhang, C.; Liu, J.; Yuan, Z. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R–CNN. *Biosyst. Eng.* **2019**, *184*, 1–23. [CrossRef]

52. Zheng, Z.; Hu, Y.; Li, X.; Huang, Y. Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks. *Comput. Electron. Agric.* **2023**, *215*, 108469. [CrossRef]

53. Li, Z.; Liu, Y.; Li, B.; Hu, W.; Wu, K.; Wang, P. SDTP: Semantic-aware Decoupled Transformer Pyramid for Dense Image Prediction. *arXiv* **2021**, arXiv:2109.08963. [CrossRef]

54. Pan, M.; Xia, W.; Yu, H.; Hu, X.; Cai, W.; Shi, J. Vehicle Detection in UAV Images via Background Suppression Pyramid Network and Multi-Scale Task Adaptive Decoupled Head. *Remote Sens.* **2023**, *15*, 5698. [CrossRef]

55. Sinha, A.; Dolz, J. Multi-scale self-guided attention for medical image segmentation. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 121–130. [CrossRef] [PubMed]

56. Li, X.; Li, X.; Zhang, L.; Cheng, G.; Shi, J.; Lin, Z.; Tan, S.; Tong, Y. Improving semantic segmentation via decoupled body and edge supervision. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XVII 16. pp. 435–452.

57. Li, Z.; Ouyang, B.; Qiu, S.; Xu, X.; Cui, X.; Hua, X. Change Detection in Remote Sensing Images Using Pyramid Pooling Dynamic Sparse Attention Network with Difference Enhancement. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 7052–7067. [CrossRef]

58. Sklar, M.G.; Armstrong, R.D. A linear programming algorithm for the simple model for discrete chebychev curve fitting. *Comput. Oper. Res.* **1983**, *10*, 237–248. [CrossRef]
59. Fan, Y.; Chen, X.; Xie, J.; Fu, Z. An Algorithm for Detecting the Integrity of Outer Frame Protection Net on Construction Site Based on Improved SSD. *J. Phys. Conf. Ser.* **2021**, *1827*, 012168. [CrossRef]
60. Andreon, S.; Gargiulo, G.; Longo, G.; Tagliaferri, R.; Capuano, N. Wide Field Imaging. I. Applications of Neural Networks to object detection and star/galaxy classification. *Mon. Not. R. Astron. Soc.* **2010**, *319*, 700–716. [CrossRef]
61. Koufogiannis, E.T.; Sgouros, N.P.; Ntasi, M.T.; Sangriotis, M.S. Grid reconstruction and skew angle estimation in Integral Images produced using circular microlenses. In Proceedings of the 2013 18th International Conference on Digital Signal Processing (DSP), Fira, Greece, 1–3 July 2013. [CrossRef]
62. Åstrand, B.; Baerveldt, A.-J. A vision based row-following system for agricultural field machinery. *Mechatronics* **2005**, *15*, 251–269. [CrossRef]
63. Bahrampour, S.; Ray, A.; Sarkar, S.; Damarla, T.; Nasrabadi, N.M. Performance comparison of feature extraction algorithms for target detection and classification. *Pattern Recognit. Lett.* **2013**, *34*, 2126–2134. [CrossRef]