

Article

# A Lightweight Algorithm for Recognizing Pear Leaf Diseases in Natural Scenes Based on an Improved YOLOv5 Deep Learning Model

Jianian Li <sup>1,2,3</sup> and Zhengquan Liu <sup>1</sup> and Dejin Wang <sup>1,2,3,\*</sup>

<sup>1</sup> Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China; jianianli@kust.edu.cn (J.L.); zq\_liu2023@163.com (Z.L.)

<sup>2</sup> Yunnan Provincial Field Scientific Observation and Research Station on Water-Soil-Crop System in Seasonal Arid Region, Kunming University of Science and Technology, Kunming 650500, China

<sup>3</sup> Yunnan Provincial Key Laboratory of High-Efficiency Water Use and Green Production of Characteristic Crops in Universities, Kunming University of Science and Technology, Kunming 650500, China

\* Correspondence: wangdj@kust.edu.cn; Tel.: +86-187-9360-3688

**Abstract:** The precise detection of diseases is crucial for the effective treatment of pear trees and to improve their fruit yield and quality. Currently, recognizing plant diseases in complex backgrounds remains a significant challenge. Therefore, a lightweight CCG-YOLOv5n model was designed to efficiently recognize pear leaf diseases in complex backgrounds. The CCG-YOLOv5n model integrates a CA attention mechanism, CARAFE up-sampling operator, and GSConv into YOLOv5n. It was trained and validated using a self-constructed dataset of pear leaf diseases. The model size and FLOPs are only 3.49 M and 3.8 G, respectively. The mAP@0.5 is 92.4%, and the FPS is up to 129. Compared to other lightweight models, the experimental results demonstrate that the CCG-YOLOv5n achieves higher average detection accuracy and faster detection speed with a smaller computation and model size. In addition, the robustness comparison test indicates that the CCG-YOLOv5n model has strong robustness under various lighting and weather conditions, including frontlight, backlight, sidelight, tree shade, and rain. This study proposed a CCG-YOLOv5n model for accurately detecting pear leaf diseases in complex backgrounds. The model is suitable for use on mobile terminals or devices.

**Keywords:** leaf disease; coordinate attention; CARAFE; GSConv; YOLOv5; lightweight

**Citation:** Li, J.; Liu, Z.; Wang, D. A Lightweight Algorithm for Recognizing Pear Leaf Diseases in Natural Scenes Based on an Improved YOLOv5 Deep Learning Model. *Agriculture* **2024**, *14*, 273. <https://doi.org/10.3390/agriculture14020273>

Academic Editors: Zhen Li, Tao Wen and Shilei Lyu

Received: 24 December 2023

Revised: 3 February 2024

Accepted: 5 February 2024

Published: 7 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Pear leaf diseases significantly reduce fruit quality and yield [1]. Accurate detection is crucial for effective treatment of leaf diseases. However, detecting leaf diseases can be time-consuming, labor-intensive, and prone to inaccuracy when using the naked eye [2]. Fortunately, with the development of computer technology, image recognition technology has shown potential for the efficient detection of plant diseases [3]. High flexibility and real-time automatic identification are required to accurately detect leaf diseases in complex and variable growth environments [4]. However, the diagnostic process is often interfered with complex background information, resulting in poor model performance [5]. Therefore, automatic detection of leaf diseases remains a challenge in complex natural scenes.

In recent decades, researchers have tried to improve the image recognition capacity for plant disease detection by combining image processing technology with machine learning algorithms [6–10]. For example, Zhang et al. [11] extracted apple disease features using a genetic algorithm and correlation-based feature selection after basic image processing with the HIS, YUV, and grayscale models, and then identified apple diseases using

an SVM classifier with an accuracy rate of more than 90%. Almadhor et al. [7] extracted diverse and informative feature vectors with the color (RGB, HSV) histogram and texture (LBP) feature, and then identified four guava diseases by using the advanced machine. However, this technology has been limited by artificially designed features, such as unstable extraction and susceptibility to complex natural backgrounds [10,12,13]. Recently, deep learning algorithms, such as the two-stage detection model and one-stage detection model, have been widely used in the field of plant disease diagnosis, providing a faster and more accurate detection algorithm. The two-stage detection model extracts features from generated candidate boxes, and then classifies and regresses the object, while the one-stage detection model directly classifies and regresses the object. For example, Bari et al. [14] improved a Faster R-CNN model (a two-stage detection model) for detecting three rice leaf diseases, with an average detection accuracy of 98.7% under a single background. Xue et al. [15] proposed an improved GC-Cascade R-CNN model (a two-stage detection model) to effectively detect four types of pear diseases with an accuracy rate of 89.4% and an FPS of 5 under a single background. Roy et al. [16] improved a YOLOv4 model (a one-stage detection model) to detect four types of tomato diseases, with an accuracy rate of 89.4% and an FPS of 70.19 under complex backgrounds. Li et al. [17] constructed an MTC-YOLOv5n model (a one-stage detection model) to detect three types of pumpkin diseases under complex backgrounds, with an average detection accuracy of 84.9% and an FPS of 143. Many experiments have shown that the one-stage detection model has slightly lower detection accuracy, but its detection speed is faster than that of the two-stage detection model [18–23]. Therefore, the one-stage detection model has more advantages for plant disease detection applied in actual agricultural production and mobile terminals.

Attention mechanisms imitate the finding of salient areas in complex scenes by human visual behavior [24]. This can strengthen the ability of an object detection algorithm by focusing on a specific area in the image, thus improving object location and identification in complex environments [25]. For example, Qi et al. [26] embedded SE attention mechanisms into YOLOv5 to detect tomato disease in the natural background. Zhang et al. [27] introduced ECA attention and hard-wish activation functions in YOLOX to detect five cotton diseases in natural backgrounds. Song et al. [28] embedded a CBAM attention mechanism into a YOLOv3 network to detect maize leaf blight infestation in a field scene. De Moraes et al. [29] integrated a CBAM attention mechanism into a YOLOv7 network to detect nine papaya fruit diseases, with an accuracy rate of 86.2% under the complex backgrounds. However, although the above models can effectively alleviate the interference of the natural background by providing channel or spatial information, they neglect model size and cannot acquire long-range dependency information. Coordinate attention mechanisms have made a breakthrough in classification performance by improving the extraction of global information [30]. Therefore, these attention mechanisms can provide a new perspective on extracting pear leaf diseases from complex backgrounds.

This study aimed to build a lightweight model of disease detection for detecting pear leaf disease lesions in complex natural backgrounds. The CCG-YOLOv5n integrates a CA attention mechanism, CARAFE up-sampling operator, and GSConv convolution module into YOLOv5n, realizing rapid detection of pear leaf diseases. This model can provide a technological basis for diagnosing pear leaf diseases applied to mobile terminals and subsequent disease control.

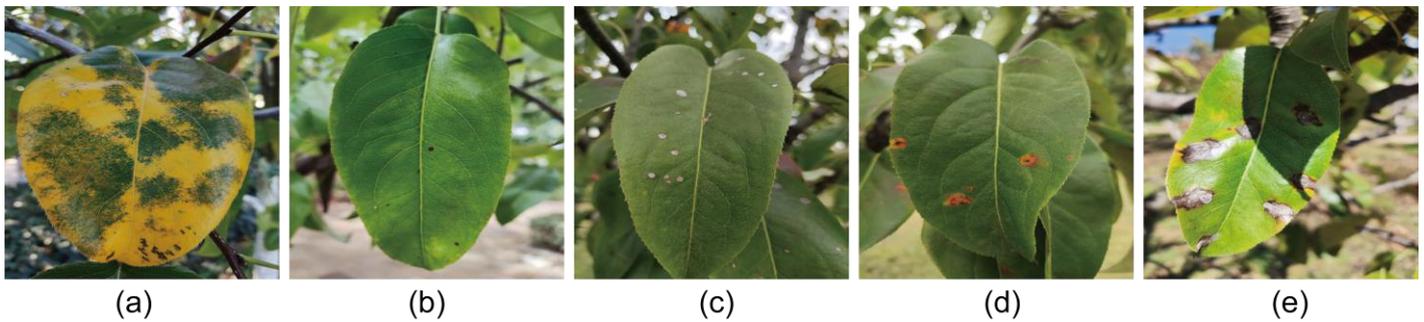
## 2. Materials and Methods

### 2.1. Dataset Construction

#### 2.1.1. Data Collection

The image dataset contains 3408 images, including mosaic, black rot, leaf spot, rust, and anthrax (Figure 1). All images were taken with a native smartphone camera from June to September in a pear orchard in Chenggong District, Kunming City. The images were taken from downlight, backlight, and sidelight angles under cloudy, sunny, and rainy

conditions. All images were clipped to a uniform size (640 × 640 pixels) and were categorized and labeled according to the leaf disease type (Table 1).



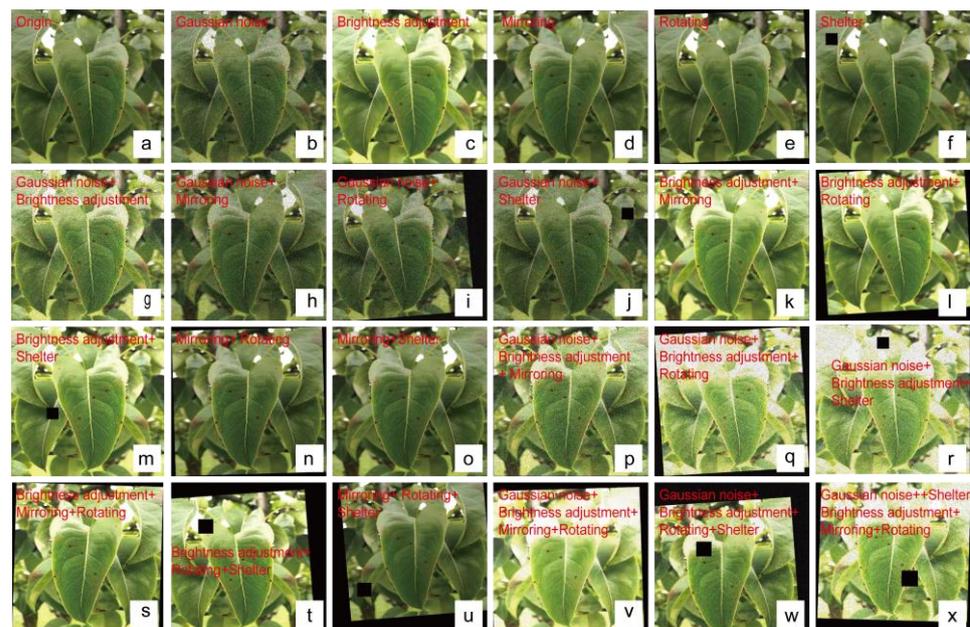
**Figure 1.** Five types of common pear leaf diseases: (a) mosaic; (b) black rot; (c) leaf spot; (d) rust; (e) anthrax.

**Table 1.** Dataset of five pear leaf diseases.

Leaf Disease Type	Original Dataset (Training Dataset/Test Dataset)	Enhanced Dataset (Training Dataset/Test Dataset)
Rust	634 (511/123)	2167 (2044/123)
Anthrax	679 (546/133)	2317 (2184/133)
Black rot	543 (439/104)	1860 (1756/104)
Leaf spot	675 (543/132)	2604 (2172/132)
Mosaic	877 (704/173)	2989 (2816/173)
Total	3408 (2743/665)	11,637 (10,972/665)

### 2.1.2. Image Enhancement

To enrich dataset diversity and avoid overfitting, image enhancement adopts one or more of the following methods: Gaussian noise, brightness adjustment, mirroring, rotating, and shelter (Figure 2). After image enhancement, the total number of images in the original training dataset increased from 2743 to 10,972, while the test dataset remained invariant (Table 1).



**Figure 2.** Partial example of image data enhancement of pear leaf disease: (a) original image, (b–f) once enhanced image, (g–o) twice enhanced image, (p–u) three times enhanced image, (v,w) four times enhanced image, (x) five times enhanced image. The black square indicates the shelter area.

## 2.2. Detection Algorithm of Pear Leaf Disease

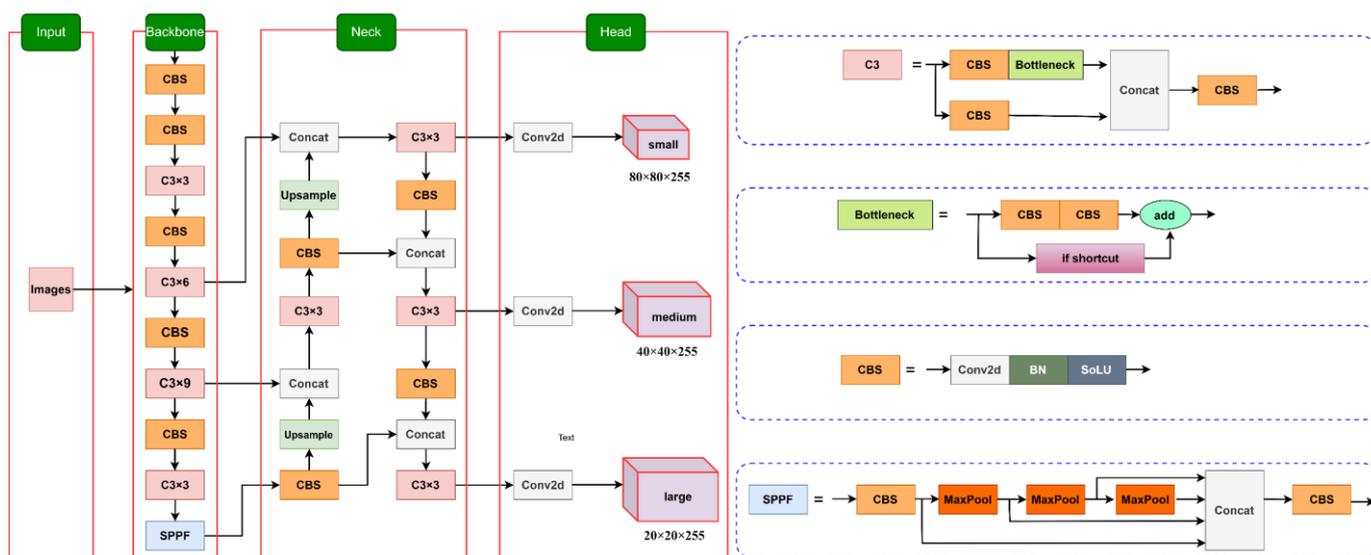
### 2.2.1. Baseline Model Selection

In actual plant disease detection, lightweight models work well in mobile terminals or devices. For this reason, we selected the YOLOv5n, YOLOv6n, YOLOv7-ting, and YOLOv8n from the YOLOv5, YOLOv6, YOLOv7, and YOLOv8 series, respectively. After training on the pear leaf disease dataset, the four selected models were evaluated using the matching test dataset. Table 2 displays the four models with similar accuracy (mAP@0.5: 87.6–88.6%) but with obviously different FLOPs and model sizes. Among them, the YOLOv5n has the smallest model size (3.74 M) and FLOPs (4.1 G). Consequently, YOLOv5n can serve as a baseline model for pear leaf disease detection.

**Table 2.** Test results for four lightweight models on the pear leaf disease dataset.

Model	Model Size (MB)	FLOPs (G)	mAP@0.5 (%)
YOLOv6n	8.27	11.8	87.6
YOLOv7-ting	11.70	13.2	87.7
YOLOv8n	5.94	8.1	88.1
YOLOv5n	3.74	4.1	88.6

The YOLOv5n network architecture comprises four main modules: input, backbone network, neck network, and head (Figure 3). The input module performs preprocessing tasks, such as mosaic data augmentation, adaptive anchor box calculation, adaptive image scaling, etc. The backbone network module extracts object features using CBS (Conv+BachNorm+SiLU), C3, and SPPF. The neck network module enhances these features using a path aggregation network (PANet). The head module decodes the feature maps to output the classification and location of the detected objects.



**Figure 3.** The network architecture of the YOLOv5n algorithm.

### 2.2.2. Improvement of the YOLOv5n Model

To improve the detection of pear leaf diseases, an improved YOLOv5n model (CCG-YOLOv5n model) is proposed by integrating the YOLOv5n algorithm with CA (Figure 4). The specific procedures are as follows:

- (1) Creating the C3CA module. At the 4th and 6th layers of the backbone network, the C3CA module is created by integrating the CA into the BottleNeck of the C3 module (Figure 5). This module can enhance the valuable feature information within the network and improve the feature extraction of pear leaf features to reduce interference with background information.

- (2) Adding the up-sampling operator. The CARAFE up-sampling operator is integrated into the neck layer. This operation can expand the receptive field to better capture target information and improve target accuracy.
- (3) Replacing Conv with GSConv. The Conv is replaced by GSConv in the neck network layer. This process can strengthen feature fusion, improve image representation, and reduce the parameters and computational cost.

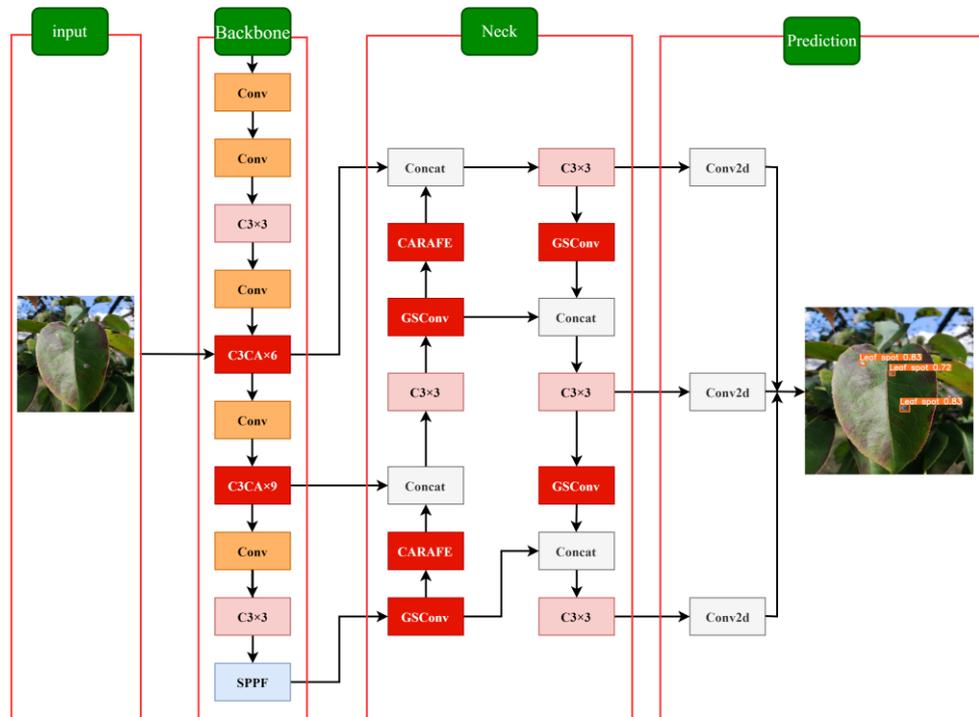


Figure 4. Network architecture of the CCG-YOLOv5n model.

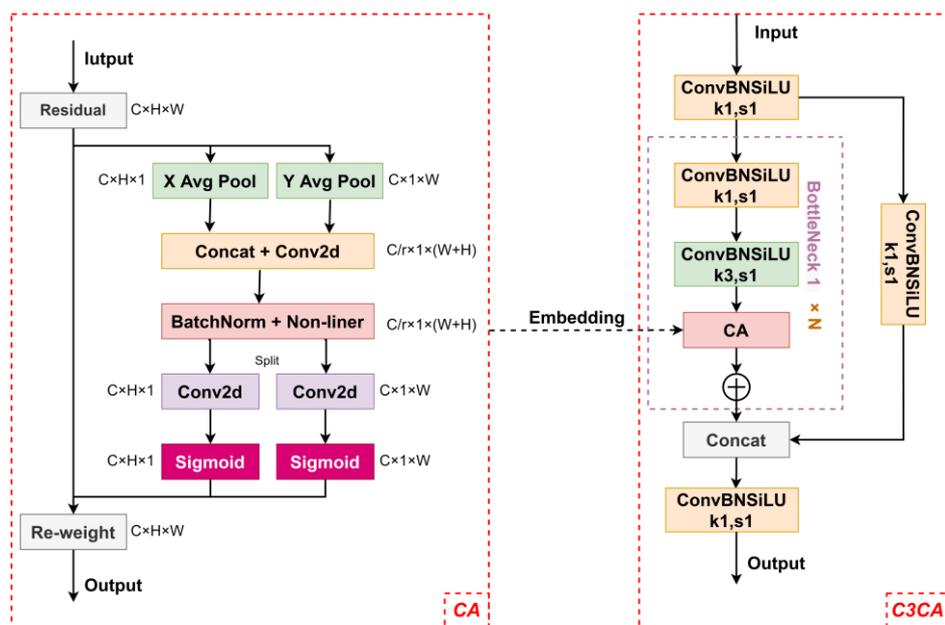


Figure 5. Architecture of C3CA module formation.

1. Coordinate attention (CA) mechanism

To improve the object recognition accuracy of pear leaf disease in natural environments, CA is integrated with the YOLOv5n network. CA is a lightweight attention

mechanism that can strengthen object features and weaken the interference of background information [31]. A long-range dependency or coded channel can be used with the location information. The integration operation involves two steps: global information embedding and coordinate attention generation (Figure 5).

### (1) Embedding global information

To capture inter-channel relations and location information using CA, the original global pooling operation is decomposed and transformed into two one-dimensional features using the encoding operations described in Equation (1).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

where  $z_c$  is the global pooling operation related to the  $c$ th channel and  $x_c(i, j)$  is a component of the input  $X$ .

The input feature maps, which have a shape of  $C \times H \times W$ , are pooled channel by channel using pooling kernels of dimensions  $(H, 1)$  and  $(1, W)$  in the  $X$  and  $Y$  directions, respectively. The pooled input feature maps produce a feature map with  $C \times H \times 1$  and  $C \times 1 \times W$  shapes. Therefore, the outputs  $z_c^h(h)$  and  $z_c^w(w)$  of the  $c$ th channel can be expressed as follows:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq w} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w) \quad (3)$$

The above two transformations aggregate the features in two spatial directions. This output enables the attention module to accurately capture and store location information from one spatial direction to another, thereby improving the network's ability to precisely locate the target object.

### (2) Coordinate Attention Generation

Attention is generated by using the globally represented features in Equation (1). The corresponding results, spliced by Equations (2) and (3), were transformed by a convolutional algorithm to generate the feature map.

$$f = \delta(F_1[z^h, z^w]) \quad (4)$$

where  $F$  is the tensor that is divided into two separate tensors ( $f^h$  and  $f^w$ ) along the spatial dimension. After upgrading the dimension using a  $1 \times 1$  convolution, the separate tensors combine with the sigmoid activation function to obtain the final attention vectors  $g^h$  and  $g^w$ :

$$g^h = \sigma(F_h(f^h)) \quad (5)$$

$$g^w = \sigma(F_w(f^w)) \quad (6)$$

Finally, the attention mechanism module outputs the attention weights  $w_c(i, j)$  by expanding the  $g^h$  and  $g^w$ , as shown in Equation (7).

$$w_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

## 2. CARAFE up-sampling operator

The YOLOv5n model uses the nearest-neighbor interpolation operator to interpolate the original pixels. This operator has a simple algorithm and a small calculation cost. However, it only determines the up-sampling kernel using the spatial positions of pixels, without utilizing the semantic information of the feature map. As a result, the perceptual range is limited, and semantic information is inadequately captured. In addition, noise can weaken the representation ability of objects during interpolation. To improve the

YOLOv5n up-sampling algorithm, we propose a lightweight CARAFE up-sampling module.

The CARAFE up-sampling technique can dynamically generate adaptive kernels using only a small number of parameters. It can expand the receptive field, remain lightweight, and optimize the utilization of surrounding information [32]. The CARAFE up-sampling module comprises a prediction module and a content-aware module (Figure 6). The specific computation processes are as follows:

- (1) Channel compression. To reduce the parameter number and computing cost for subsequent steps, the feature map is compressed from  $H \times W \times C$  to  $H \times W \times C_m$ , where  $C_m$  represents the number of compression channels and is set to 64.
- (2) Content encoding and up-sampling kernel prediction. The up-sampling kernel (size:  $\sigma H \times \sigma W \times k_{up}^2$ ) is obtained and predicted by using a  $K_{encoder} \times K_{encoder}$  convolutional layer. Here,  $k_{up}$  and  $K_{encoder}$  are set to 5 and 3, respectively.
- (3) Up-sampling normalization. The above predicted up-sampling kernel is normalized by the Softmax function.
- (4) Content-aware feature reorganization. The convolution operation is performed by combining the predicted up-sampling kernel mentioned above with the input features.

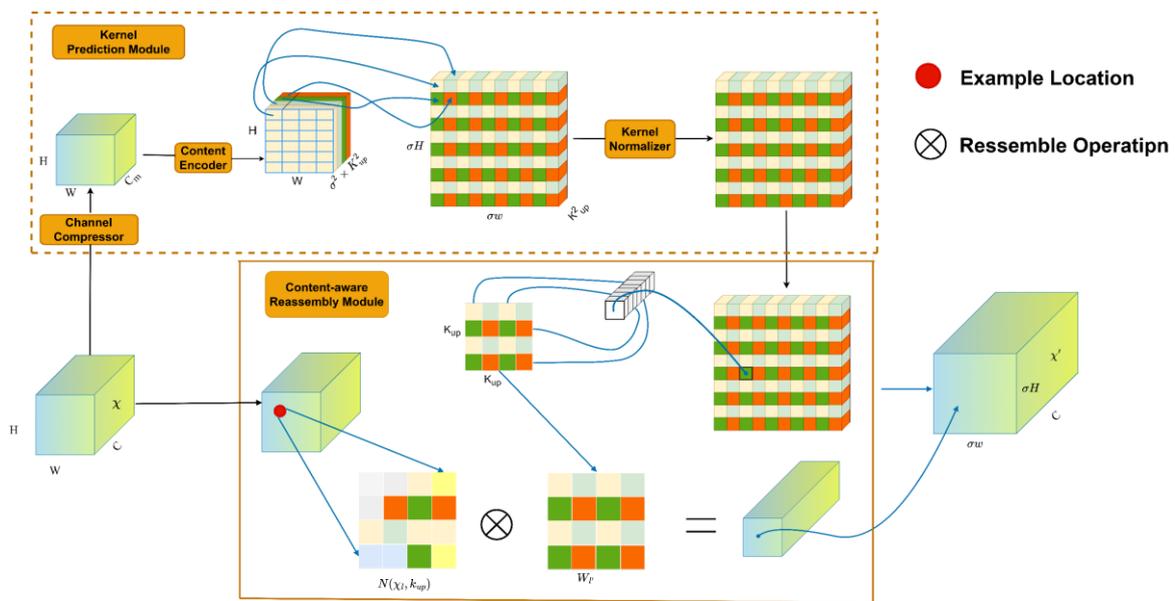


Figure 6. CARAFE calculation flowchart.

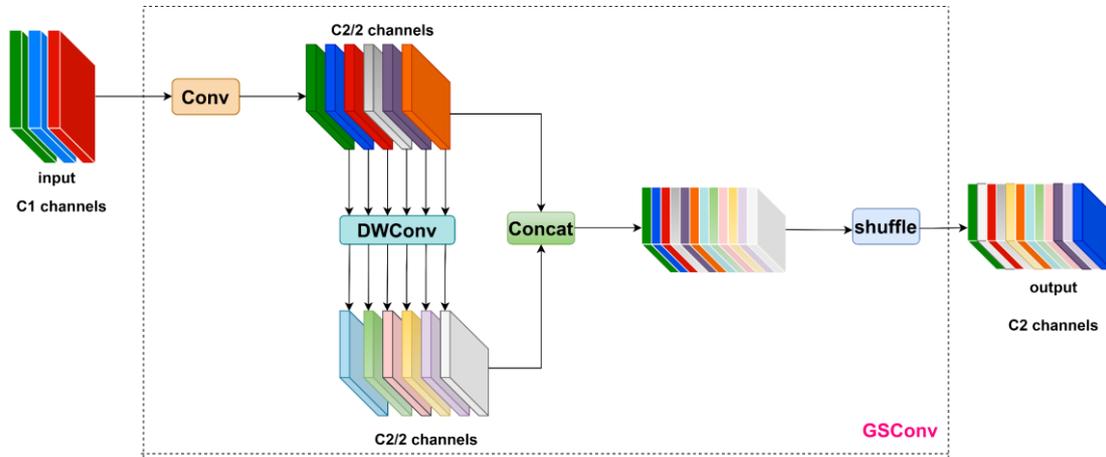
### 3. GSConv module

Disease detection requires a smaller model size and higher algorithm processing speed using edge devices. Deep convolution can reduce the number of parameters and the computational complexity of the model, but it is only a single-channel convolution. For this reason, deep convolution cannot change the number of channels during operation and lacks feature fusion. To solve this problem of lack of feature fusion, the GSConv module was proposed by Li et al. (2022) [33]. The GSConv module mainly includes a Conv module, a DWConv module, a Concat module, and a Shuffle module (Figure 7). The construction steps of the GSConv module are as follows:

- (1) The input feature map with a channel number of  $C_1$  has been processed by standard convolution and depth-separable convolution (DSC) to produce two types of feature maps with a channel number of  $C_2/2$ .
- (2) These two feature maps are concatenated to obtain and output an object feature map with a channel number of  $C_2$ .

- (3) The channel with a number of  $C_2$  is uniformly shuffled to strengthen the feature fusion and improve the representability of the image feature.

In this work, the GSConv module is integrated into the neck module of YOLOv5 to minimize semantic information loss caused by spatial compression.



**Figure 7.** GSConv architecture.

### 2.3. Equipment Environment

The model was trained and tested using the PyTorch 1.13.0 deep learning framework on a Windows 11 system. The hardware devices included a 12th generation Intel(R) Core(TM) i7-12700k@3.6GHz processor, 64 GB of memory, from Intel Corporation, Santa Clara, California, USA, and an NVIDIA GeForce RTX3090 graphics card with 24 GB of video memory from NVIDIA, Santa Clara, California, USA. The software included Cuda 11.6, cudnn 8.6.0, and python 3.9.13.

The training parameters of the model are shown in Table 3. In the process of training, the initial learning rate was set to 0.01, and the learning rate was decreased by the cosine annealing strategy. Additionally, the neural network parameters were optimized using the stochastic gradient descent (SGD) method. Here, the momentum value and weight decay index score were set to 0.937 and 0.0005, respectively. The image batch size was 32, the training epoch was 250, and the input image resolution was  $640 \times 640$  pixels.

**Table 3.** Parameter setting for training procedures.

Parameter	Value	Parameter	Value
Batch size	32	Initial learning rate	0.01
Epochs	250	Momentum	0.937
Input size	$640 \times 640$	Weight decay	0.0005
Optimizer	SGD		

### 2.4. Model Evaluation

Considering the requirements of pear disease detection in natural environments, models are evaluated by the model size, average precision (AP), mean average precision (mAP), floating point operations (FLOPs), and frames per second (FPS).

Model size is the required space for model storage, depending on the parameter number. Smaller model sizes are more convenient to embed in a mobile terminal.

AP is defined as the area surrounded by the precision–recall ( $P$ - $R$ ) curve, with recall as the  $x$ -axis and accuracy as the  $y$ -axis, expressed by Equation (8).

$$AP = \int_0^1 (P \cdot R) dR \quad (8)$$

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

where precision ( $P$ ) and recall ( $R$ ) are defined by Equations (9) and (10), respectively.  $TP$ ,  $FP$ , and  $FN$  are the numbers that represent the target being detected correctly, incorrectly, and missed, respectively.

The mAP is the value when the IoU is set to 0.5. The calculation process is shown in Equation (11).

$$\text{mAP} = \frac{\sum_{i=1}^C AP_i}{C} \quad (11)$$

FLOPs represent the number of floating-point multiplication and addition operations in the model. The lower the FLOPs, the less computation and execution time the model requires.

Given the real-time detection speed in the application scenario, FPS represents the number of pictures processed per second.

### 3. Results and Discussion

#### 3.1. Performance Comparison of the Attention Mechanisms

To effectively evaluate four self-constructed attention mechanism modules (C3CBAM, C3ECA, C3SE, and C3CA), performance comparison experiments were conducted based on a dataset of pear leaf disease. In the baseline model, the model size, FLOPs, and mAP@0.5 were 3.74 MB, 4.1 G, and 88.6%, respectively (Table 4). Compared with the baseline model, the values of the other four models dropped by 2.94–3.74% in model size, fell by 7.32% in FLOPs, and increased by 0.5–1.2% in mAP@0.5, respectively (Table 4). The findings indicate that the four self-constructed attention mechanism modules can effectively reduce the model size and calculation cost, as well as enhance the average accuracy of the model, and the C3CA has the highest average detection accuracy.

The SE and ECA modules only focus on the channel of the feature maps but ignore the spatial features. The CBAM module improves the SE module and can simultaneously obtain channel information and spatial features. However, the CA attention module comprehensively considers the spatial information, the channel features, and the long-term dependence. This module reduces natural interference, which can help to collect more accurate location information. Therefore, the CA attention mechanism is more appropriate for this study.

**Table 4.** Performance comparison of four attention mechanism modules.

Attention Mechanism	Model Size (MB)	FLOPs (G)	mAP@0.5 (%)
Base (YOLOv5n)	3.74	4.1	88.6
C3CBAM	3.63	3.8	89.1
C3ECA	3.61	3.8	89.1
C3SE	3.60	3.8	89.3
C3CA	3.60	3.8	89.8

#### 3.2. Ablation Experiments

Ablation experiments were conducted to assess the impact of the improved models, which take YOLOv5n as the baseline model and gradually add the CA attention mechanism, CARAFE up-sampling operator, and GSConv convolution module. The baseline model was improved by adding the C3CA, CAREFE, or GSConv modules. After improvement, the base model (YOLOv5n) included configurations with a single module (YOLOv5n\_1, YOLOv5n\_2, and YOLOv5n\_3), two modules (YOLOv5n\_4 and

YOLOv5n\_5), and a three-module model (CCG-YOLOv5n) (Table 5). Compared to YOLOv5n, mAP@0.5 in YOLOv5n\_1 and YOLOv5n\_3 increased by 1.2% and 0.7%, respectively. Their model size fell by 2.9% and 5.6%, while FLOPS decreased by 7.3% and 2.5%, respectively. This indicates that adding the C3CA module or GSConv module can improve model recognition accuracy while reducing the model size and number of parameters. However, mAP@0.5 in YOLOv5n\_2 increased by 1.1%, and its model size and FLOPS increased by 2.9% and 2.4%, respectively. The mAP@0.5 in YOLOv5n\_4, YOLOv5n\_5, and YOLOv5n\_6 increased by 2.9%, 1.4%, and 1.6%, respectively. In addition, mAP@0.5 in the CCG-YOLOv5n model increased by 3.8% and was up to 92.4%; while the model size and FLOPs decreased by 6.7% to 3.49M and 7.3% to 3.8G, respectively. The results illustrate that adding modules can improve detection accuracy by algorithm superposition. CCG-YOLOv5n exhibits the best comprehensive detective performance in detective accuracy, model size, and FLOPs. Therefore, CCG-YOLOv5n can be an optimal model for detecting pear leaf diseases.

**Table 5.** Results of the ablation experiments.

Model	C3CA	CAREFE	GSConv	Model Size (MB)	FLOPs (G)	mAP@0.5 (%)
Base (YOLOv5n)				3.74	4.1	88.6
YOLOv5n_1	√			3.63	3.8	89.8
YOLOv5n_2		√		3.85	4.2	89.7
YOLOv5n_3			√	3.53	4.0	89.3
YOLOv5n_4	√	√		3.69	3.9	91.5
YOLOv5n_5	√		√	3.43	3.7	90.0
YOLOv5n_6		√	√	3.59	4.1	90.2
CCG-YOLOv5n	√	√	√	3.49	3.8	92.4

Notes: √ indicates the use of the module.

### 3.3. Performance Comparison of Different Mainstream Algorithms

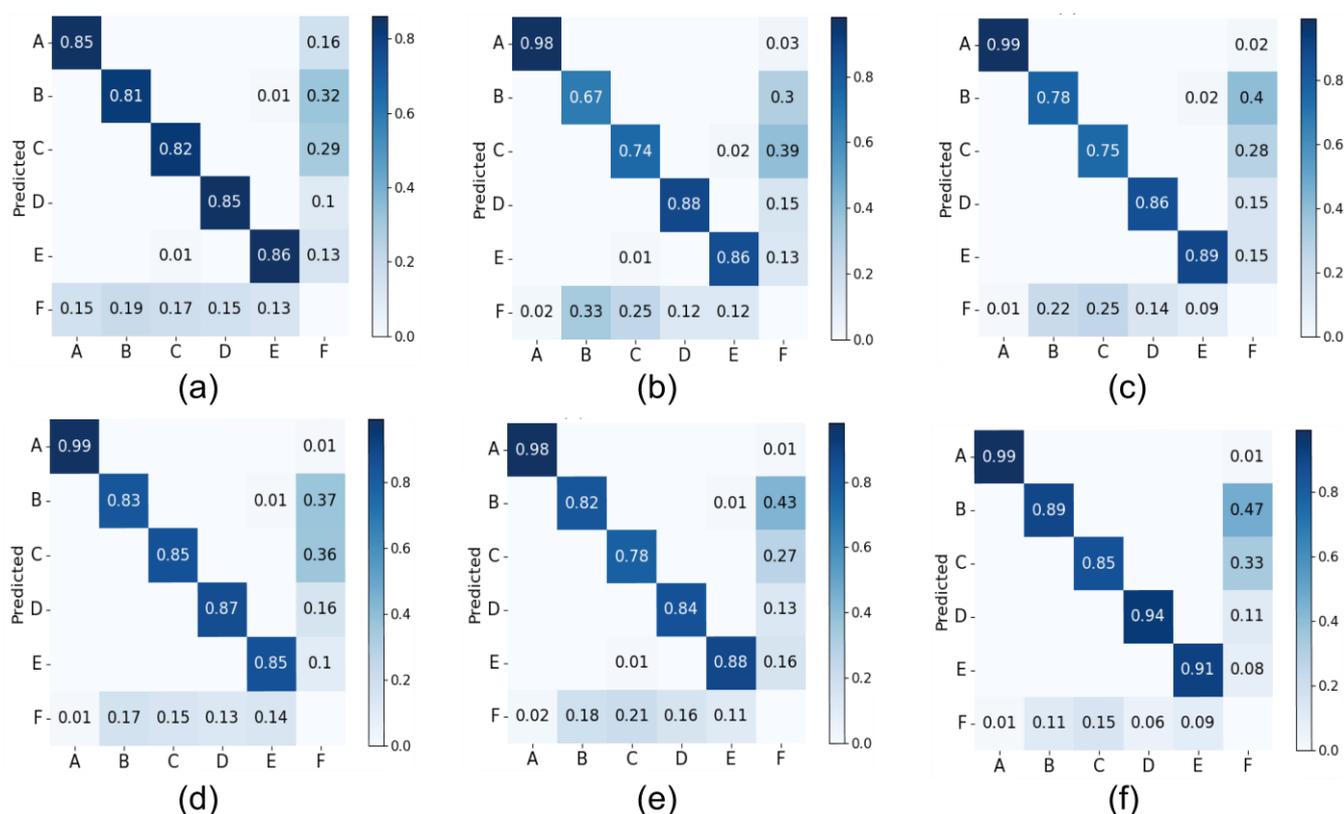
To validate the superiority of the CCG-YOLOv5n model, a series of models were tested on our self-constructed dataset for pear leaf disease shown in Table 6. Throughout the training process, we ensured consistency in the model parameters. Subsequently, these models were evaluated using an independent test dataset. The CCG-YOLOv5n model exhibited the highest mAP@0.5 (92.4%) and FPS (129), and the lowest model size (3.49 M) and FLOPs (3.8 G). Compared with the other five models, the CCG-YOLOv5n model increased by 3.4% to 9.7% in mAP@0.5 and reduced by 1.21 MB to 85.34 MB in model size. This result indicates that the CCG-YOLOv5n model has a higher average accuracy and is better suited for pear leaf disease detection on mobile terminals or mobile devices.

**Table 6.** Performance comparison of different detection algorithms.

Methods	Model Size (MB)	FLOPs (G)	mAP@0.5 (%)	FPS
YOLOv3-ting	33.16	13.0	82.7	122
YOLOv4-ting	22.40	6.8	83.3	125
MTC-YOLOv5n [17]	4.70	6.1	87.1	124
YOLOv5s	13.70	15.8	89.0	120
GC-Cascade R-CNN [15]	88.83	312.4	88.5	7
CCG-YOLOv5n	3.49	3.8	92.4	129

Figure 8 reveals the distinction between the confusion matrix of CCG-YOLOv5n and the other five mainstream single-stage detection models. Except for the CCG-YOLOv5n model, the other models exhibit slight inter-class misclassification. For example, Figure 8a–f show a little anthracnose black spot that is misrecognized as a black spot. Figure 8a,b,e show some rust and brown spots that are misidentified as anthracnose. Due to the

fact that leaf lesions (especially at the edge) are similar to their environmental background, the leaf disease is easily misrecognized. The main reasons for misrecognition are the similar color attributes of leaf diseases, and incomplete feature extraction lacking long-range information dependency. The last row of the matrix represents undetected diseases. Compared to the other five detection models, the CCG-YOLOv5n model has the lowest misrecognition ratio, the lightest color, and the highest recognition rate, which is 0.99, 0.89, 0.85, 0.94, and 0.91 for mosaic, black spot, leaf spot, rust, and anthrax, respectively. It is concluded that the CCG-YOLOv5n model can accurately identify pear leaf diseases despite the interference of natural backgrounds.



**Figure 8.** Confusion matrix of different mainstream algorithms. (a) YOLOv3-tiny; (b) YOLOv4-tiny; (c) MTC-YOLOv5n; (d) YOLOv5s; (e) GC-Cascade R-CNN; (f) CCG-YOLOv5n. The inner A–F of the confusion matrix denote mosaic, black rot, leaf spot, rust, anthrax, and background, respectively.

Figure 9 shows the test results of leaf disease detection for the CCG-YOLOv5n model and the other five mainstream one-step detection models. Compared with the other five detection models, the CCG-YOLOv5n model improves the detection accuracy for more effective detection of leaf disease (especially at the leaf edge and image edge) in complex environments. The CCG-YOLOv5n model has a higher detection accuracy for mosaic detection. It can reduce the false and missed detection of anthracnose, as well as the missed detection of spot rot and rust. In addition, the CCG-YOLOv5n model can identify the black spots and reduce the missed detection of early small black diseases. This result validates the earlier confusion matrix test.



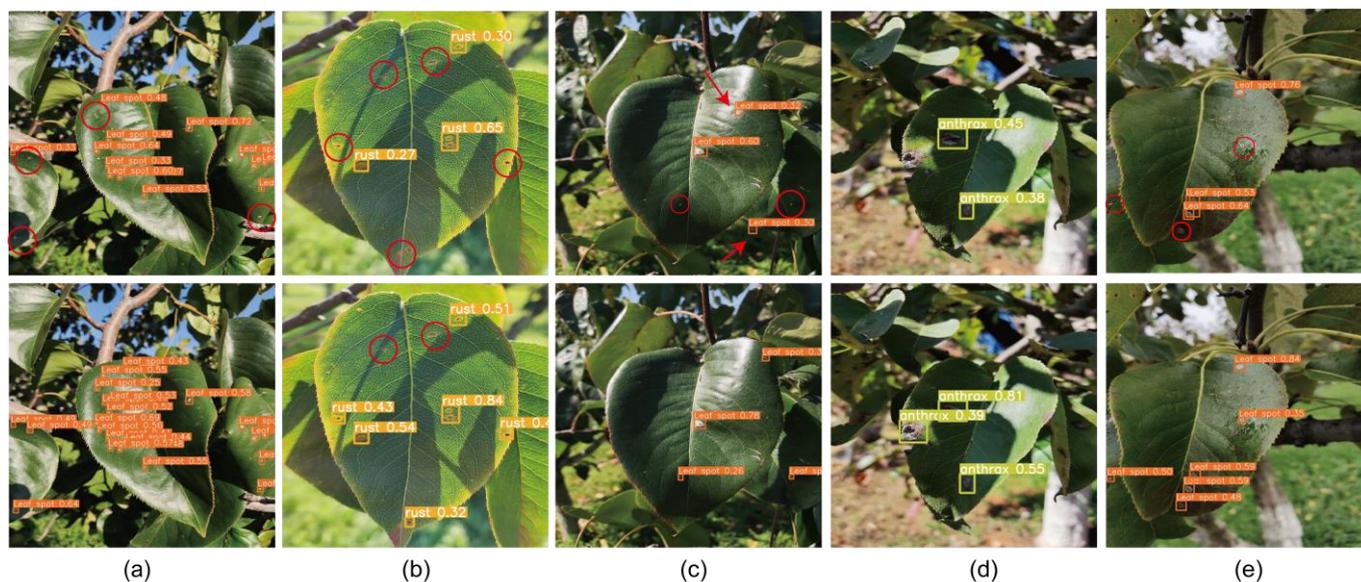
**Figure 9.** Detection results of different models on the test images. The first to sixth rows show the detection results of YOLOv3-tiny, YOLOv4-tiny, MTC-YOLOv5n, YOLOv5s, GC-Cascade R-CNN, and CCG-YOLOv5n, respectively. The first to fifth columns are pear leaf diseased leaves of mosaic,

anthrax, rust, leaf spot, and black rot, respectively. The red circle and red arrow represent the missing and false detections, respectively.

### 3.4. Robustness Comparison

The detection of pear leaf diseases is affected by the various noises in the interference environment, such as shooting angles (i.e., frontlight, backlight, and sidelight), tree shade, rainfall, etc. This required that the algorithm have strong robustness to ensure detection accuracy with the inference of natural external environments. To compare the detection performance between YOLOv5n and CCG-YOLOv5n under the five interference environments, we analyzed the random test results of the five leaf diseases shown in Figure 10. Compared with the baseline model YOLOv5n, the CCG-YOLOv5n model has a higher detection accuracy and a lower false and missing detection ratio under frontlight, backlight, sidelight, tree shade, and rainy conditions. The YOLOv5n cannot detect the small scab lesions due to front light interference, as the scab lesions are confused by intense light. The YOLOv5n is affected by the side light and misidentifies small light spots as brown lesion spots. The YOLOv5n model misses a lesion location in the leaf margin that is interfered with by shadow due to insufficient feature information and ambient noise.

In brief, the YOLOv5n model exists for false and missing detection of pear leaf disease as a result of insufficient feature information extraction of the convolutional network interfering with a complex background. To solve this issue, the CCG-YOLOv5n model integrates a CA module into the backbone, and GARAFE and GSConv modules into the neck layer, respectively. The integration of the CA module attenuates the background noise while focusing on important plaque features by reusing the feature information of the network. Meanwhile, the integration of the CARAFE module expands the receptive field and improves the ability to detect objects to better capture object information. In addition, the replacement of GSConv improves the extraction capacity and reduces the model size and computational cost. Therefore, the improved CCG-YOLOv5n strengthens the robustness of the algorithm and is more suitable for pear leaf disease detection in natural environments.



**Figure 10.** Comparison of algorithm robustness: (a) downlight; (b) backlight; (c) sidelight; (d) tree shade; (e) rain. The first row and the second row represent robustness test images for the baseline model YOLOv5n and the improved model CCG-YOLOv5n, respectively. The red circle and red arrow represent the missing and false detection, respectively.

#### 4. Conclusions

This study proposes a lightweight model (CCG-YOLOv5n) for recognizing pear leaf diseases in complex backgrounds. The CCG-YOLOv5n model exhibits a greater advantage in model size (3.49 M), FLOPS (3.8 G), mAP@0.5 (92.4%), and FPS (129). This model obviously improves the detection of pear leaf diseases under complex conditions. It shows strong robustness under the conditions of frontlight, backlight, sidelight, tree shade, and rain. This proposed model possesses lightweight features with a low hardware equipment requirement, faster detection speed, and higher accuracy. It is suited for pear leaf disease detection under cloudy, sunny, and rainy conditions. This model lays the foundation for the detection of pear tree leaf diseases in actual natural scenes. However, it is limited in detecting leaf diseases across different weather conditions and growth cycles due to the dataset constraints. Further research should consider expanding the model to include more plant diseases and different weather conditions, while also enhancing the sensitivity and automation of multi-target disease recognition.

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

**Funding:** This research was funded by the Yunnan Revitalization Talent Support Program (no. KKRD202223052).

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

**Data Availability Statement:** The dataset and code 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 organization.

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

#### References

- Liu, L.; Tao, H.; Fang, J.; Zheng, W.; Wang, L.; Jin, X. Identifying Anthracnose and Black Spot of Pear Leaves on Near-infrared Hyperspectroscopy. *Transactions of the Chinese Society for Agricultural Machinery* **2022**, *53*, 221–230.
- Xue, W.; Yi, W.; Kang, Y.; Xu, Y.; Dong, C. Recognition of pear leaf small anthracnose spot based on multi-resolution and multi-class feature fusion. *Journal of Nanjing Agricultural University* **2021**, *44*, 982–992.
- Su, P.; Li, H.; Wang, X.; Wang, Q.; Hao, B.; Feng, M.; Sun, X.; Yang, Z.; Jing, B.; Wang, C.; et al. Improvement of the YOLOv5 Model in the Optimization of the Brown Spot Disease Recognition Algorithm of Kidney Bean. *Plants-Basel* **2023**, *12*, doi:10.3390/plants12213765.
- Yang, F.; Li, F.; Zhang, K.; Zhang, W.; Li, S. Influencing factors analysis in pear disease recognition using deep learning. *Peer-to-Peer Networking and Applications* **2021**, *14*, 1816–1828, doi:10.1007/s12083-020-01041-x.
- Li, S.-f.; Li, K.-y.; Qiao, Y.; Zhang, L.-x. Cucumber Disease Detection Method Based on Visible Light Spectrum and Improved YOLOv5 in Natural Scenes. *Spectroscopy and Spectral Analysis* **2023**, *43*, 2596–2600, doi:10.3964/j.issn.1000-0593(2023)08-2596-05.
- Ahmad, N.; Asif, H.M.S.; Saleem, G.; Younus, M.U.; Anwar, S.; Anjum, M.R. Leaf Image-Based Plant Disease Identification Using Color and Texture Features. *Wireless Personal Communications* **2021**, *121*, 1139–1168, doi:10.1007/s11277-021-09054-2.
- Almadhor, A.; Rauf, H.T.; Lali, M.I.U.; Damasevicius, R.; Alouffi, B.; Alharbi, A. AI-Driven Framework for Recognition of Guava Plant Diseases through Machine Learning from DSLR Camera Sensor Based High Resolution Imagery. *Sensors* **2021**, *21*, doi:10.3390/s21113830.
- Kundu, N.; Rani, G.; Dhaka, V.S.; Gupta, K.; Nayak, S.C.; Verma, S.; Ijaz, M.F.; Wozniak, M. IoT and Interpretable Machine Learning Based Framework for Disease Prediction in Pearl Millet. *Sensors* **2021**, *21*, doi:10.3390/s21165386.
- Liu, Y.; Wang, Z.; Wang, R.; Chen, J.; Gao, H. Flooding-based MobileNet to identify cucumber diseases from leaf images in natural scenes. *Computers and Electronics in Agriculture* **2023**, *213*, doi:10.1016/j.compag.2023.108166.
- Guo, Y.; Zhang, J.; Yin, C.; Hu, X.; Zou, Y.; Xue, Z.; Wang, W. Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming. *Discrete Dynamics in Nature and Society* **2020**, *2020*, doi:10.1155/2020/2479172.
- Zhang, C.; Zhang, S.; Yang, J.; Shi, Y.; Chen, J. Apple leaf disease identification using genetic algorithm and correlation based feature selection method. *International Journal of Agricultural and Biological Engineering* **2017**, *10*, 74–83, doi:10.3965/j.ijabe.20171002.2166.
- Yin, C.; Su, Y.; Pan, M.; Duan, J. Detection of the quality of famous green tea based on improved YOLOv5s. *Transactions of the Chinese Society of Agricultural Engineering* **2023**, *39*, 179–187.

13. Deng, J.; Yang, C.; Huang, K.; Lei, L.; Ye, J.; Zeng, W.; Zhang, J.; Lan, Y.; Zhang, Y. Deep-Learning-Based Rice Disease and Insect Pest Detection on a Mobile Phone. *Agronomy-Basel* **2023**, *13*, doi:10.3390/agronomy13082139.
14. Bari, B.S.; Islam, M.N.; Rashid, M.; Hasan, M.J.; Razman, M.A.M.; Musa, R.M.; Ab Nasir, A.F.; Majeed, A.P.P.A. A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework. *Peerj Computer Science* **2021**, doi:10.7717/peerj-cs.432.
15. Xue, W.; Cheng, R.; Kang, Y.; Huang, X.; Xu, Y.; Dong, C. Pear Leaf Disease Spot Counting Method Based on GC-Cascade R-CNN. *Transactions of the Chinese Society for Agricultural Machinery* **2022**, *53*, 237-245.
16. Roy, A.M.; Bose, R.; Bhaduri, J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. *Neural Computing & Applications* **2022**, *34*, 3895-3921, doi:10.1007/s00521-021-06651-x.
17. Li, S.; Li, K.; Qiao, Y.; Zhang, L. A multi-scale cucumber disease detection method in natural scenes based on YOLOv5. *Computers and Electronics in Agriculture* **2022**, *202*, doi:10.1016/j.compag.2022.107363.
18. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object Detection in 20 Years: A Survey. *Proceedings of the Ieee* **2023**, *111*, 257-276, doi:10.1109/jproc.2023.3238524.
19. Zhou, T.; Yu, Z.; Cao, Y.; Bai, H.; Su, Y. Study on an infrared multi-target detection method based on the pseudo-two-stage model. *Infrared Physics & Technology* **2021**, *118*, doi:10.1016/j.infrared.2021.103883.
20. Guirguis, K.; Abdelsamad, M.; Eskandar, G.; Hendawy, A.; Kayser, M.; Yang, B.; Beyerer, J.; Ieee. Towards Discriminative and Transferable One-Stage Few-Shot Object Detectors. In Proceedings of the 23rd IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, 2023 Jan 03-07, 2023; pp. 3749-3758.
21. Ale, L.; Zhang, N.; Li, L. Road Damage Detection Using RetinaNet. In Proceedings of the IEEE International Conference on Big Data (Big Data), Seattle, WA, 2018 Dec 10-13, 2018; pp. 5197-5200.
22. Hamza, R.; Chtourou, M. Comparative Study on Deep Learning Methods for Apple Ripeness Estimation on Tree. In Proceedings of the 21st International Conference on Intelligent Systems Design and Applications (ISDA), World Wide Web, ELECTRONETWORK, 2022 Dec 13-15, 2021; pp. 1325-1340.
23. Li, Z.; Tian, X.; Liu, X.; Liu, Y.; Shi, X. A Two-Stage Industrial Defect Detection Framework Based on Improved-YOLOv5 and Optimized-Inception-ResnetV2 Models. *Applied Sciences-Basel* **2022**, *12*, doi:10.3390/app12020834.
24. Brauwiers, G.; Frasincar, F. A General Survey on Attention Mechanisms in Deep Learning. *Ieee Transactions on Knowledge and Data Engineering* **2023**, *35*, 3279-3298, doi:10.1109/tkde.2021.3126456.
25. Yao, M.; Min, Z. Summary of Fine-Grained Image Recognition Based on Attention Mechanism. In Proceedings of the 13th International Conference on Graphics and Image Processing (ICGIP), Yunnan Univ, Kunming, PEOPLES R CHINA, 2022 Aug 18-20, 2021.
26. Qi, J.; Liu, X.; Liu, K.; Xu, F.; Guo, H.; Tian, X.; Li, M.; Bao, Z.; Li, Y. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Computers and Electronics in Agriculture* **2022**, *194*, doi:10.1016/j.compag.2022.106780.
27. Zhang, Y.; Ma, B.; Hu, Y.; Li, C.; Li, Y. Accurate cotton diseases and pests detection in complex background based on an improved YOLOX model. *Computers and Electronics in Agriculture* **2022**, *203*, doi:10.1016/j.compag.2022.107484.
28. Song, B.; Lee, J.; Ieee. Detection of Northern Corn Leaf Blight Disease in Real Environment Using Optimized YOLOv3. In Proceedings of the IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Electr Network, 2022 Jan 26-29, 2022; pp. 475-480.
29. de Moraes, J.L.; Neto, J.d.O.; Badue, C.; Oliveira-Santos, T.; de Souza, A.F. Yolo-Papaya: A Papaya Fruit Disease Detector and Classifier Using CNNs and Convolutional Block Attention Modules. *Electronics* **2023**, *12*, doi:10.3390/electronics12102202.
30. Yang, L.; Zhang, F.; Wang, P.S.-P.; Li, X.; Meng, Z. Multi-scale spatial-spectral fusion based on multi-input fusion calculation and coordinate attention for hyperspectral image classification. *Pattern Recognition* **2022**, *122*, doi:10.1016/j.patcog.2021.108348.
31. Hou, Q.; Zhou, D.; Feng, J.; Ieee Comp, S.O.C. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Electr Network, 2021 Jun 19-25, 2021; pp. 13708-13717.
32. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D.; Ieee. CARAFE: Content-Aware ReAssembly of FEatures. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, SOUTH KOREA, 2019 Oct 27-Nov 02, 2019; pp. 3007-3016.
33. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. **2022**, arXiv:2206.02424, doi:10.48550/arXiv.2206.02424.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.