Article

Rice Diseases Identification Method Based on Improved YOLOv7-Tiny

Duoguan Cheng, Zhenqing Zhao and Jiang Feng

Article

Rice Diseases Identification Method Based on Improved YOLOv7-Tiny

Duoguan Cheng, Zhenqing Zhao and Jiang Feng *

College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China;
18703860154@163.com (D.C.); zzq725@126.com (Z.Z.)
* Correspondence: fj345@163.com

Abstract: The accurate and rapid identification of rice diseases is crucial for enhancing rice yields. However, this task encounters several challenges: (1) Complex background problem: The rice background in a natural environment is complex, which interferes with rice disease recognition; (2) Disease region irregularity problem: Some rice diseases exhibit irregular shapes, and their target regions are small, making them difficult to detect; (3) Classification and localization problem: Rice disease recognition employs identical features for both classification and localization tasks, thereby affecting the training effect. To address the aforementioned problems, an enhanced rice disease recognition model leveraging the improved YOLOv7-Tiny is proposed. Specifically, in order to reduce the interference of complex background, the YOLOv7-Tiny model's backbone network has been enhanced by incorporating the Convolutional Block Attention Module (CBAM); subsequently, to address the irregularity issue in the disease region, the RepGhost bottleneck module, which is based on structural reparameterization techniques, has been introduced; Finally, to resolve the classification and localization issue, a lightweight YOLOX decoupled head has been proposed. The experimental results have demonstrated that: (1) The enhanced YOLOv7-Tiny model demonstrated elevated F1 scores and mAP@.5, achieving 0.894 and 0.922, respectively, on the rice pest and disease dataset. These scores exceeded the original YOLOv7-Tiny model's performance by margins of 3.1 and 2.2 percentage points, respectively. (2) In comparison to the YOLOv3-Tiny, YOLOv4-Tiny, YOLOv5-S, YOLOX-S, and YOLOv7-Tiny models, the enhanced YOLOv7-Tiny model achieved higher F1 scores and mAP@.5. The improved YOLOv7-Tiny model boasts a single image inference time of 26.4 ms, satisfying the requirement for real-time identification of rice diseases and facilitating deployment in embedded devices.

Keywords: rice diseases; image identification; YOLOv7-Tiny; object detection

1. Introduction

Rice is one of the most widely grown food crops in the world. However, the frequent occurrence of rice diseases has had a serious impact on the yield and quality of rice, posing a great threat to food security [1]. Despite the continuous development of agricultural science and technology, the diagnosis of crop diseases in most areas still relies on the traditional method of manual identification, which is not only inefficient but also fails to meet the needs of modernized agricultural production [2]. Therefore, it is particularly important to study efficient and accurate artificial intelligence rice disease recognition algorithms. Currently, crop disease detection primarily relies on two types of data: RGB images and hyperspectral images. Hyperspectral imaging-based techniques for crop disease severity detection capture detailed spectral data by recording the spectral reflectance of crops at various wavelengths. The quality of this data is then enhanced through preprocessing steps such as denoising, correction, and normalization. Subsequently, key features are extracted from the optimized data, and machine learning models are developed and trained to analyze these features, ultimately facilitating the detection of crop diseases. For instance,

Zhang [3] employed hyperspectral imaging and SVM to classify the severity of rice leaf blasts at different growth stages, achieving a classification accuracy of 94.75%. Similarly, Cao [4] and colleagues utilized hyperspectral imaging and SDC-3DCNN to detect bacterial blight in rice, reaching an accuracy of 95.4427%. Although hyperspectral imaging offers high precision in crop disease recognition, it typically focuses on identifying a single disease and the high costs of equipment and maintenance limit its widespread adoption, particularly for resource-constrained small-scale farms and research institutions. In contrast, RGB image-based techniques for crop disease severity detection are relatively cost-effective and do not involve complex data processing, making them a suitable choice for detecting rice diseases in this study.

In recent years, target detection algorithms have been extensively applied in agricultural production. These algorithms are typically classified into two groups: the two-stage method and the single-stage method for detecting targets [5]. The two-stage target detection algorithm initially generates a series of anchor boxes by utilizing common region selection methods, including Selective Search and the Region Proposal Network. Subsequently, it inputs the generated anchor boxes into a convolutional neural network for feature extraction and classification regression. Typical detection algorithms comprise R-CNN [6]. Fast R-CNN [7], Faster R-CNN [8], and Mask R-CNN [9]. Yuanqin Zhang et al. [10] developed an improved Faster R-CNN rice ear detection network aimed at addressing the problem of identifying small target rice ears, achieving an average accuracy of 80.3%. Zhenguo Zhang et al. [11] implemented safflower filament detection utilizing Faster R-CNN combined with an attention mechanism, resulting in an average recognition accuracy of 91.49%. Gaoliang Zhang et al. [12] proposed a rice stalk cross-section parameter detection network based on Mask R-CNN, yielding an average accuracy of 94.37%. Although the aforementioned two-stage target detection algorithm, which is based on generating candidate frames, exhibits high detection accuracy, its deployment on mobile devices is challenging due to the large number of network model parameters and its slow detection speed. The single-stage algorithm for object detection bypasses the creation of candidate frames, directly computing both the object's class probability and its positional coordinates. Notable algorithms encompass SSD (Single Shot MultiBox Detector) [13] and YOLO (You Only Look Once) [14]. Regarding the SSD algorithm, Lin et al. [15] developed a rice planthopper recognition algorithm employing SSD and dictionary learning, achieving a recognition accuracy of 89.3%. With respect to the YOLO algorithm, Xiong et al. [16] designed a multi-scale convolutional neural network named Des-YOLOv3, tailored for identifying citrus at night, attaining an average accuracy rate of 90.75%. Wang et al. [17] developed a rice disease recognition algorithm utilizing YOLOv4-Tiny, achieving an average recognition accuracy of 81.79%. Sun et al. [18] incorporated phantom convolution and attention modules into YOLOv5s for the recognition of apple fruit diseases. Aziz et al. [19] proposed an improved YOLO to classify diseased rice leaves with 94% accuracy. Sangaiah et al. [20] proposed a T-yolo-Rice rice disease detection model, and the mAP reached 86%. In contrast to the two-stage algorithm for object detection, the single-stage variant, which depends on regression, not only diminishes the parameter count of the model but also improves its performance in real-time scenarios. However, these algorithms struggle with detecting small targets and are susceptible to the issue of missing detection.

Currently, rice disease detection has yielded good results but continues to face the following challenges: (1) Complex background problem: rice grows in a natural environment, and the complex background can interfere with rice disease identification; (2) Disease region irregularity problem: In rice disease detection, diseases like bacterial blight, rice blast, and brown spot are characterized by irregular shapes and varying target region sizes, complicating their detection process; (3) Classification and localization problem: rice disease recognition utilizes the same features for classification and localization tasks; however, the features for classification and localization are spatially misaligned, which affects the training results. To address these challenges, the study focused on five diseases as research subjects and enhanced the YOLOv7-Tiny model. For the complex background issue, the

Convolutional Block Attention Module (CBAM) has been integrated into the YOLOv7-Tiny model, enhancing focus on the disease regions and minimizing the impact of complex backgrounds on rice disease recognition. To tackle the issue of irregular disease regions, the RepGhost bottleneck module (RG-bneck) has been introduced to the YOLOv7-Tiny model, thereby enhancing its capability to extract features from irregular disease regions through structural reparameterization technology. Subsequently, a lightweight YOLOX decoupled head has been proposed to enhance the model's classification and localization accuracy. Finally, to accelerate convergence, the model has adopted a transfer learning approach.

In the remainder of this paper, the methods for rice disease image acquisition and identification are detailed in Section 2, the experimental results are presented in Section 3, the discussion is provided in Section 4, and the conclusions are outlined in Section 5.

## 2. Materials and Methods

2.1. Dataset

Utilizing the self-built dataset and two public datasets, this article has constructed a dataset containing 5 types of diseases. The self-built dataset was sourced from the Shuangmajiatun paddy field, Acheng District, Harbin City, Heilongjiang Province. Data collection was conducted using a OnePlus smartphone (The phone was made by Shenzhen OnePlus Science and Technology Co., Ltd, and purchased in Harbin, China.), spanning from 1 August to 10 September. The dataset included a total of 887 images, including bacterial blight, rice blast, brown spot, rice tungro, and rice false smut. In order to enrich the dataset, 613 rice disease images from the public rice disease dataset [21] were also used in this study. The images obtained through the above methods underwent further screening to remove duplicates and misclassified images from the original public dataset, resulting in the construction of a rice pest and disease dataset containing 1500 images. Figure 1 showcases a range of rice disease images.

| (a) | (b) | (c) | (d) | (e) |
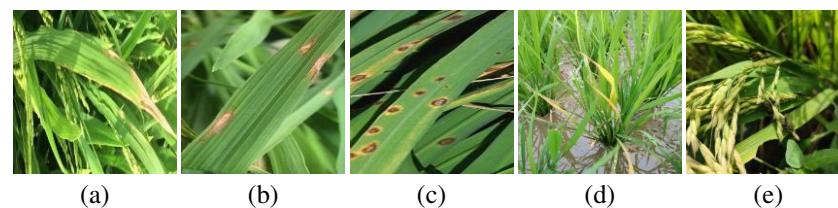
Figure 1. Image samples of rice diseases dataset. (**a**) Bacterial blight; (**b**) Rice blast; (**c**) Brown spot; (**d**) Rice tungro; (**e**) Rice false smut.

2.2. Data Preprocessing

In this research, prior to training the model, images were annotated using LabelImg v1.8.6 [22] according to the Pascal VOC dataset's annotation format. In order to enhance the generalization ability of the model and avoid overfitting, the data are augmented by offline and online data enhancement [18]. To improve the model's generalization ability and ensure sample space consistency, offline data augmentation operations—including noise addition, panning, cropping, flipping, and random luminosity adjustment—were conducted on the images, increasing the image count to 10,500. The offline data enhancement diagram is shown in Figure 2. The data were partitioned into training, validation, and test sets in a 6:2:2 ratio. To enrich the image background and enhance model training efficiency, an online data augmentation strategy was employed to conduct Mixup and Mosaic enhancement operations on the input image data during the training process. This online data augmentation strategy eliminates the need for additional storage space for the enhanced image data, thus conserving storage resources and offering high flexibility. Table 1 enumerates the number and labeling specifics of the images related to rice pests and diseases subsequent to their preprocessing.

| (a) | (b) | (c) | (d) | (e) | (f) |
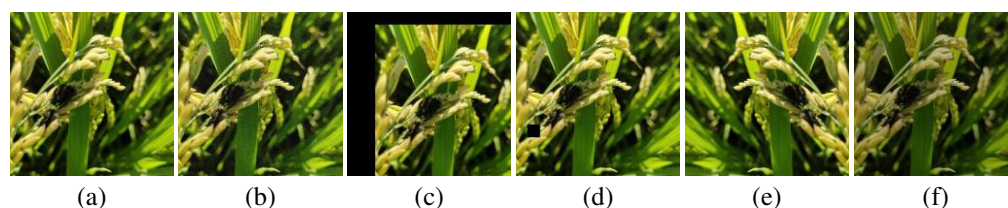
Figure 2. Data enhancement diagram. (**a**) Original image; (**b**) noise addition; (**c**) panning; (**d**) cropping; (**e**) flipping; (**f**) random luminosity adjustment.

Table 1. Overview of rice pests and diseases dataset.

Category	Number of Original Samples	Number of Enhancement Samples	Label
Bacterial blight	198	1386	0
Rice blast	290	2030	1
Brown spot	469	3283	2
Rice tungro	293	2051	3
Rice false smut	250	1750	4
Total	1500	10,500	

2.3. YOLOv7 Network Architecture

In this study, a recognition model for rice diseases was developed, utilizing the YOLOv7-Tiny [23] framework as its basis. YOLOv7-Tiny is characterized by a streamlined model structure and fast inference and is suitable for scenarios with limited resources. However, the recognition accuracy of the model for rice diseases in complex backgrounds is still to be improved and needs to be further optimized and improved.

Figure 3 illustrates the architecture of the YOLOv7-Tiny model, primarily composed of Input, Backbone, Neck, and Prediction layers. The input side initially preprocesses the image, which primarily entails data augmentation and adaptive anchor frame calculation to ensure uniform scaling of the RGB image in order to fulfill the input size requirement of the backbone network. The backbone network is primarily composed of three modules: CBL, T-ELAN, and MP. The CBL module comprises convolution, batch normalization, and LeakyReLU activation functions. The E-ELAN (Extended efficient layer aggregation networks) module is an extension proposed on the basis of the ELAN [24] module. The E-ELAN maintains the network's initial gradient trajectory while improving its capacity to learn. This enhancement is achieved by integrating various computational blocks designed for distinct feature groups, thereby enabling the network to acquire a wider array of features. The T-ELAN module is a streamlined version of the ELAN module, with two fewer convolution operations relative to the ELAN module. MP denotes the Maximum Pooling Layer, primarily employed to reduce the image's dimensions by half in both length and width and to extract the maximal value information from the local region. The neck network of YOLOv7-Tiny employs Feature Pyramid Networks and Path Aggregation Network architectures. It comprises the CBL module, T-ELAN module, MP module, and SPPCSP module. The prediction layer possesses three detection branches responsible for detecting targets of different sizes and generating the predicted class probabilities and location information.
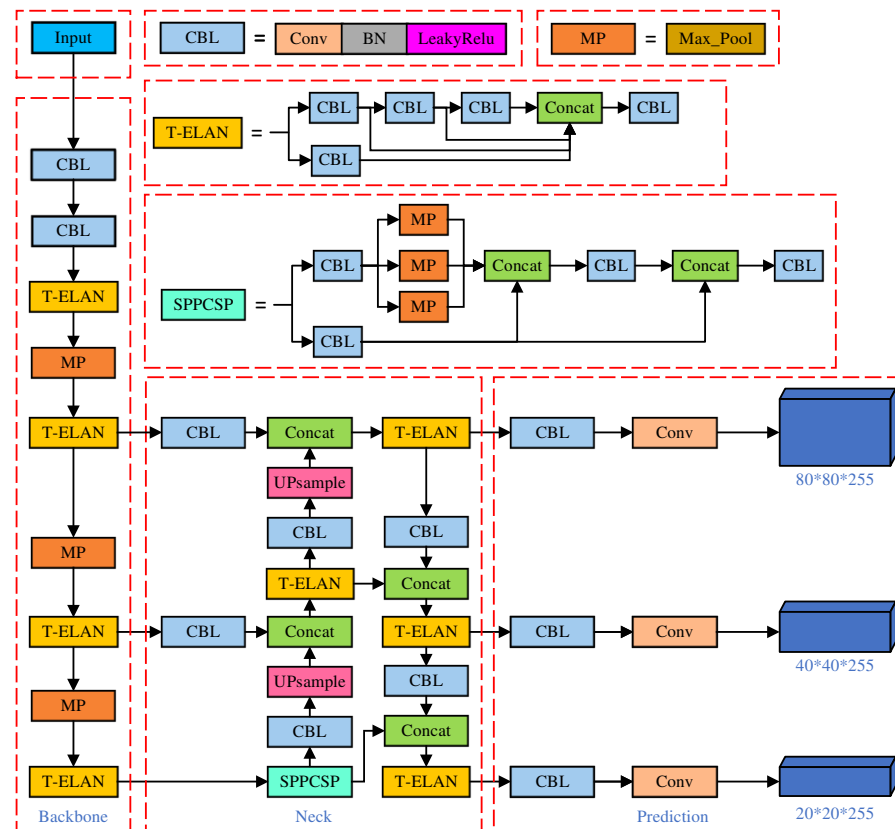
Figure 3. YOLOv7-Tiny Structure of the target detection network.

2.4. Convolutional Block Attention Module

To address the problem that the rice disease dataset constructed in this study has a complex background that interferes with rice disease identification, the Convolutional Block Attention Module (CBAM) [25] and the T-ELAN module in the YOLOv7-Tiny model backbone network were combined to form the improved C-T-ELAN module, the structure of which is shown in Figure 4. By assigning weights to the spatial dimension and channel dimension of input features, CBAM improves the network's attention to rice disease feature information, thereby reducing the influence of complex background on rice disease recognition to a certain extent.
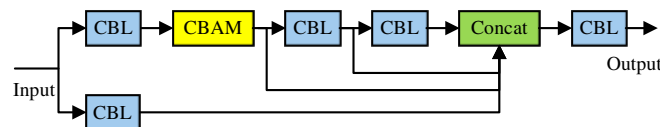

Figure 4. C-T-ELAN Model structure.

CBAM, a proficient and compact attention module, comprises both a channel attention module and a spatial attention module. Its architecture is depicted in Figure 5. The channel attention module receives the input feature map $F \in R^{C \times H \times W}$, performs global maximum pooling and global average pooling to generate two one-dimensional feature vectors, which are fed into the multilayer perceptron, and then performs element-by-element summing and sigmoid function activation of the feature vectors output by the perceptron to obtain the normalized channel attention weights matrix $M_c \in R^{C \times 1 \times 1}$, and finally multiplies the channel attention weights matrix with the input feature map $F \in R^{C \times H \times W}$ to obtain the adjusted feature map $F_a \in R^{C \times H \times W}$. The spatial attention module takes the feature map $F_a \in R^{C \times H \times W}$ as input and performs global maximum pooling and global average pooling to obtain two feature vectors, performs convolution operation on the spliced features of these two vectors and activates them by a sigmoid function to obtain the normalized

spatial attention weight matrix $M_s \in R^{1 \times H \times W}$, and finally multiplies the spatial attention weight matrix with the input feature map $F_a \in R^{C \times H \times W}$ to obtain the adjusted feature map $F_b \in R^{C \times H \times W}$. The CBAM is calculated as shown in Equations (1) and (2). $\otimes$ denotes dot product.

$$F_a = M_c(F) \otimes F \tag{1}$$
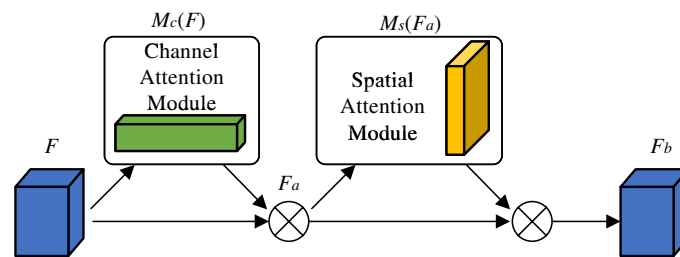
$$F_b = M_s(F_a) \otimes F_a \tag{2}$$

Figure 5. Convolutional Block Attention Module structure.

2.5. RepGhost Bottleneck Module

Aiming at the problem of irregular rice disease regions, this paper introduces the RepGhost bottleneck (RG-bneck) module [26] so that it replaces one of the convolution operations in the C-T-ELAN module in the model backbone network and thus constructs the improved R-C-T-ELAN module, which in turn improves the model feature extraction capability. Its structure is shown in Figure 6.
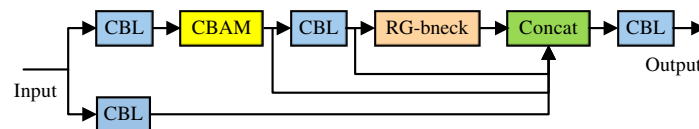

Figure 6. R-C-T-ELAN Model structure.

The RepGhost bottleneck Module represents a hardware-efficient architecture developed through advanced structure-heavy parameterization techniques aimed at enhancing model training accuracy while preserving fast inference speed. The structure of the RepGhost bottleneck Module is depicted in Figure 7. dconv is the Depthwise convolutional layer, SBlock is the Shortcut block, DS is the Downsample layer, and SE is the Squeeze-and-Excitation block [27]. During the training phase, the input feature map undergoes processing via two branches. One branch comprises the jump-connection layer, SBlock; meanwhile, in the other branch, the input feature map initially traverses the RepGhost module, which includes convolution, depth-separable convolution, a Batch Normalization layer, and a ReLU activation function. Subsequently, it moves through an intermediate layer, follows through the RepGhost module without the ReLU activation function, and ultimately, this branch combines with the jump connection layer. During the inference phase, the module converts the depth-separable convolution and the batch normalization layer into equivalent depth-separable convolutions via parameter fusion. The RepGhost bottleneck Module utilizes structural reparametrization to synthesize and amalgamate different feature maps, thereby reducing the loss of feature information and consequently enhancing the model's accuracy during the training phase. Furthermore, the module adopts the add operation over the inefficient ConCat operation to enhance the model's inference speed.

(a) RepGhost bottleneck Module training   (b) RepGhost bottleneck Module inference
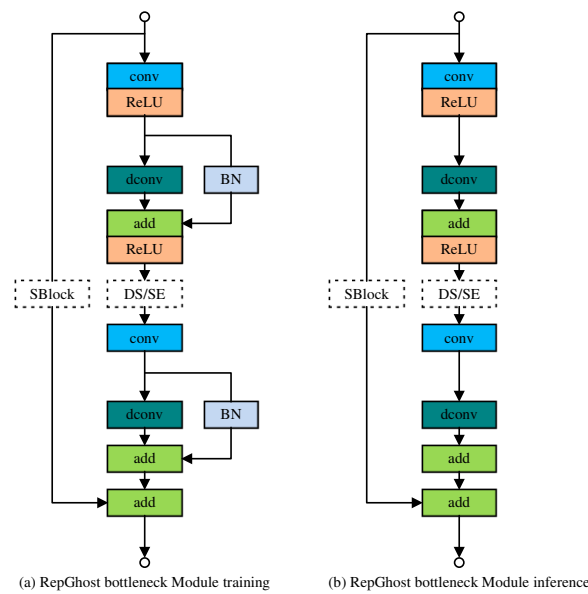
Figure 7. RepGhost bottleneck Module structure.

2.6. Improved Decoupled Head

The detection head of Yolov7-Tiny employs identical features for both classification and localization tasks. However, the features for classification and localization exhibit spatial misalignment, which may influence the training results [28]. To enhance the accuracy of classification and localization further, the detection head of Yolov7-Tiny has been substituted with the Decoupled Head of the YOLOX [29] model.

The layout of the YOLOX decoupled detection head is showcased in Figure 8. Here, 'anchor' indicates the number of anchor frames, and initially, the input feature map passes through a 1×1 convolutional layer, resulting in the creation of two divergent branches. In the first branch, to execute the classification task, the feature map undergoes two 3×3 convolutions followed by one 1×1 convolution. In the second branch, aimed at localization and confidence tasks, the feature map is subjected to two 3×3 convolutions, subsequently followed by two parallel 1×1 convolutions. Executing the above steps effectively decouples distinct feature channels for the classification, localization, and confidence tasks, thereby diminishing the prediction error stemming from the task differences. The YOLOX decoupled detection head considerably increases the model's parameters while enhancing the accuracy of classification and localization. To achieve equilibrium between precision and rapidity, this paper implements a lightweight Depthwise over-parameterized depthwise convolutional layer (Do-DConv) [30] to substitute for the 3×3 convolution in the YOLOX decoupled detection head, and consequently develops a lightweight decoupled head, DD-Head.
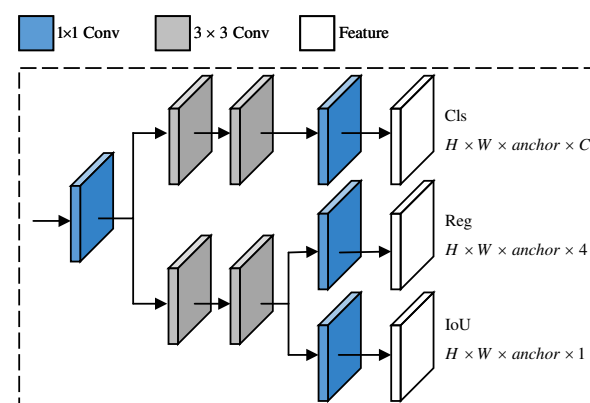
Figure 8. Decoupled head structure.

Do-DConv is composed of a pair of profound convolution operations. Its operational mechanics are depicted in Figure 9. In this context, C_{in} symbolizes the quantity of channels present in the input feature map. At the same time, D_{mul} and D_{mul}^W represent the depth multiplier. Additionally, $M \times N$ signifies the dimensions of the sensory field associated with the deep convolution kernel. The whole computation process is shown in Equations (3) and (4). $D^T \in R^{C_{in} \times (M \times N) \times D_{mul}}$ represents the transpose of the deep convolution kernel D, W^T denotes the transpose of the deep convolution kernel W, $W^* \in R^{C_{in} \times (M \times N) \times D_{mul}^W}$ denotes the deep convolution kernel, $W^{*T} \in R^{D_{mul}^W \times (M \times N) \times C_{in}}$ denotes the transpose of the deep convolution kernel W^*, $P \in R^{C_{in} \times (M \times N)}$ denotes the patch on the inout feature map, $O \in R^{C_{in} \times D_{mul}^W}$ denotes the patch on the output feature map, $\circ$ denotes the deep convolution operator.

$$W^* = D^T \circ W^T \tag{3}$$
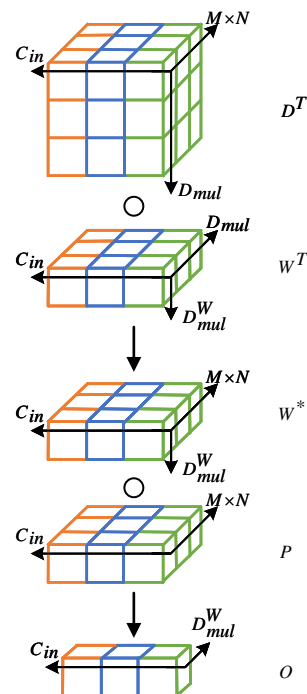
$$O = W^{*T} \circ P \tag{4}$$

Figure 9. Calculation flow chart of Do-DConv.

2.7. Transfer Learning

Transfer learning efficiently alleviates the problems of protracted convergence and excessive fitting that arise in training for rice diseases. It accomplishes this by channeling knowledge from the original domain to the newly established target domain [31]. In this research, a transfer learning strategy was employed, designating the VOC2007 dataset as the source domain while utilizing the rice diseases dataset as the destination domain. A base network was initially pre-trained on the VOC2007 dataset; subsequently, the learned feature parameters were transferred to the target network to train the rice diseases dataset within the target domain. Given the low similarity between the VOC2007 dataset and the rice diseases dataset, the transfer learning method involving retraining all layer parameters in the target domain after loading the source domain weights was selected, and the transfer learning process is depicted in Figure 10.
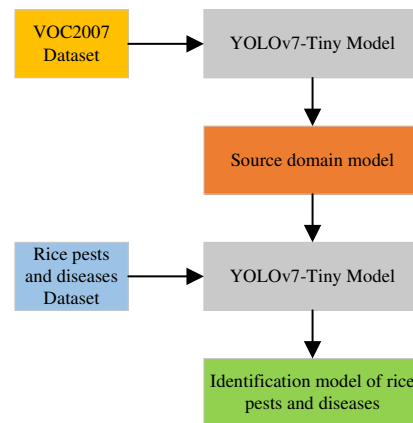
Figure 10. Flow chart of transfer learning.

3. Results

3.1. Experimental Settings

This study utilizes the open-source PyTorch deep learning framework for the development and enhancement of models. Furthermore, the training and evaluation of these models are performed on a system running Windows 11. The hardware configuration is built around an AMD Ryzen 7 5800H processor (The processor's manufacturer is Advanced Micro Devices, Inc., purchased in Harbin, China). It is complemented by an NVIDIA GeForce RTX 3060 laptop (The manufacturer of the GPU is NVIDIA Corporation, purchased in Harbin, China). GPU equipped with 6 GB of video memory. To accelerate network training, a GPU was used for acceleration, utilizing CUDA version 11.3. During training, the input image size was consistently resized to 640×640 pixels, and the batch size was configured at 16. Stochastic Gradient Descent (SGD) was utilized as the optimizer, streamlining the training procedure of the neural network. The SGD momentum parameter was configured at 0.937, and the weight decay parameter was established at 0.0005. The initial learning rate was established at 0.01, and a warm-up strategy was employed: training commenced with a learning rate of 0.0001 for the initial three epochs. Subsequently, the learning rate was reset to its original value of 0.01.

3.2. Test Evaluation Indicators

In the context of recognizing rice disease targets in complex environments, both the precision and real-time performance of the detection network must be taken into consideration. To accurately assess the model's performance in rice disease recognition, this study employs six widely recognized performance evaluation metrics for target detection algorithms: Precision, Recall, F1 Score, Mean Average Precision, Single-Image Inference Time, and Model Parameter Size. Among these metrics, Mean Average Precision correlates with both accuracy and recall, with its specific formula being as follows:

- P stands for Precision, which represents the proportion of accurately identified positive results among all classified as positive.

$$P = \frac{TP}{(TP + FP)} \tag{5}$$

- R denotes Recall, which reflects the proportion of true positive outcomes relative to the overall number of genuine positive instances.

$$R = \frac{TP}{(TP + FN)} \tag{6}$$

- AP indicates average precision.

$$AP = \int_0^1 P(R)dR \tag{7}$$

- mAP denotes the average precision mean.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{8}$$

- F1 denotes the harmonic mean of precision and recall.

$$F1 = \frac{2PR}{P + R} \tag{9}$$

TP represents the number of accurately classified positive samples by the model, while FP indicates the number of instances erroneously tagged as positive, and FN refers to the number of samples incorrectly identified as negative. AP@0.5 signifies the AP calculated at an IoU threshold of 0.5, and mAP@0.5 indicates the mAP computed at an IoU threshold of 0.5. AP@0.5 and mAP@0.5 served as evaluation metrics.

3.3. Ablation Test and Analysis of Results

To validate the effectiveness of various improvements to the YOLOv7-Tiny model in augmenting its overall performance, ablation experiments were carried out. In these experiments, Precision, Recall, F1 Score, Average Precision Mean, Model Parameter Count, and Single Image Inference Time were primarily employed as evaluation metrics.

In the model shown in Table 2, "C" indicates the inclusion of the C-T-ELAN module, "R" indicates the inclusion of the R-C-T-ELAN module, and "D" signifies the introduction of the improved DD-Head. The C-YOLOv7-Tiny model emerges from the integration of CBAM into the T-ELAN module within the backbone network of YOLOv7-Tiny. When compared with the baseline YOLOv7-Tiny model, its number of parameters increases by only 0.1 MB, and its mAP@.5 is enhanced by 1.0 percentage points. To enhance the model's recognition accuracy while maintaining inference speed, the R-C-YOLOv7-Tiny model was developed by optimizing the baseline YOLOv7-Tiny model with CBAM and the RepGhost bottleneck module. Upon comparison with the baseline YOLOv7-Tiny model, its number of parameters decreases by 1.1 MB, and its F1 score and mAP@.5 experience improvements of 1.9 and 1.6 percentage points, respectively. To enhance the accuracy of model classification and localization, the improved D-R-C-YOLOv7-Tiny model is introduced, incorporating the enhanced DD-Head along with the CBAM and RepGhost bottleneck module. Compared to the baseline YOLOv7-Tiny model, mAP@.5 improved by 2.2 percentage points. At the same time as the inference time of the enhanced model sees a marginal increase, it continues to meet the demands of real-time image processing.

Table 2. The results of Ablation experiments.

Model	F1 Score	mAP@.5	Parameters/MB	Inference Time/ms
YOLOv7-Tiny	0.863	0.90	11.7	16.3
C-YOLOv7-Tiny	0.878	0.910	11.8	20.1
R-C-YOLOv7-Tiny	0.882	0.916	10.6	20.0
D-R-C-YOLOv7-Tiny	0.894	0.922	12.2	26.4

To visually demonstrate the C-YOLOv7-Tiny model's effectiveness proposed in this research, a class activation map for rice diseases was produced, with the findings depicted in Figure 11. The red circles indicate the response areas. As illustrated by the figure, for rice blast, brown spot, and rice false smut, the YOLOv7-Tiny model focuses more on the background information, whereas the C-YOLOv7-Tiny model focuses more on

the disease subjects, thus mitigating background interference and enhancing the model's feature extraction capability.


(a)   (b)   (c)
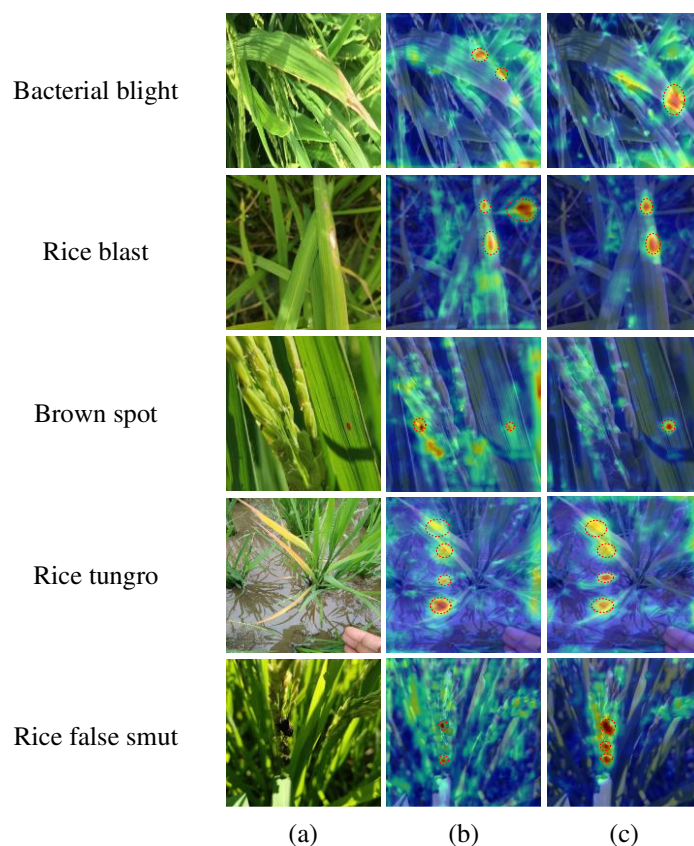
Figure 11. Class activation diagram of the model on rice diseases. (**a**) origin image; (**b**) Class activation map for YOLOv7-Tiny; (**c**) Class activation map for C-YOLOv7-Tiny.

The mAP@.5 evaluation results of training 200 epochs for each model are shown in Figure 12. The figure illustrates that there is no marked steep rise or fall in the mAP@.5 of each model, indicating a consistent performance. Compared to the baseline model YOLOv7-Tiny, the improved model demonstrates enhanced performance; the mAP@.5 increases more rapidly during the first 150 epochs of training and remains close to the optimal value throughout the last 50 epochs. The mAP@.5 of the improved models has significantly improved relative to the baseline model, suggesting that the modifications applied to the YOLOv7-Tiny model in this paper have been effective. In order to further analyze the performance of the models on specific rice disease categories, AP@.5 performance tables of each model under different rice disease categories were made, and the specific results were shown in Table 3.

Table 3. AP@.5 performance of each model under different disease categories.

| Category	YOLOv7-Tiny	C-YOLOv7-Tiny	R-C-YOLOv7-Tiny	D-R-C-YOLOv7-Tiny
Bacterial blight	0.802	0.840	0.880	0.919
Rice blast	0.892	0.897	0.895	0.890
Brown spot	0.877	0.893	0.892	0.861
Rice tungro	0.986	0.990	0.987	0.982
Rice false smut	0.942	0.932	0.926	0.959
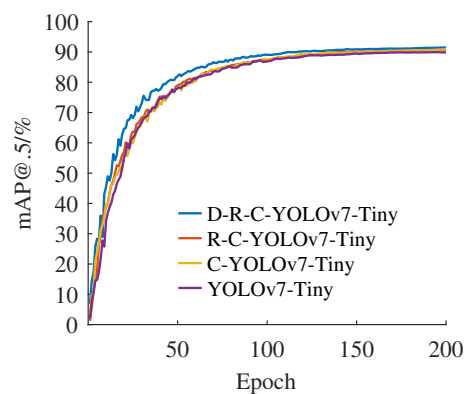
Figure 12. Comparison of model mAP@.5 under different strategies.

From Table 3, it can be seen that the YOLOv7-Tiny model showed high recognition ability on rice tungro and rice false smut, with AP@.5 of 0.986 and 0.942, respectively. However, it had low recognition rates of 0.802, 0.892, and 0.877 for bacterial blight, rice blast, and brown spot, respectively. Compared to the baseline model YOLOv7-Tiny, model C-YOLOv7-Tiny showed a more significant improvement in the recognition of bacterial blight and brown spots. Model R-C-YOLOv7-Tiny showed improved recognition performance for bacterial blight compared to model C-YOLOv7-Tiny. Compared with the baseline model YOLOv7-Tiny, the improved model D-R-C-YOLOv7-Tiny proposed in this paper improves 11.7 and 1.7 percentage points on bacterial blight and rice false smut, respectively, and the accuracy distribution of its model is more balanced, which can more accurately fulfill the tasks of classifying and locating disease targets.

3.4. Evaluation of Various Target Detection Algorithms

To further evaluate the performance of the D-R-C-YOLOv7-Tiny model proposed in this study, precision, recall, mAP@.5, and single-image inference time were used as measures for comparative tests with the YOLOv3-Tiny [32], YOLOv4-Tiny [33], YOLOv5-S [34], YOLOX-S, and YOLOv7-Tiny models. The horizontal and vertical axes of the P-R curve are the recall and precision, respectively, which can reflect the comprehensive performance of the target detection network. Figure 13 illustrates the P-R curves of each comparative model on the rice pest validation and test sets. On both the test and validation sets, the curve of the D-R-C-YOLOv7-Tiny model approaches the coordinate (1,1) at the equilibrium point (where precision rate equals recall rate), demonstrating that its detection accuracy surpasses that of the other five target detection models.
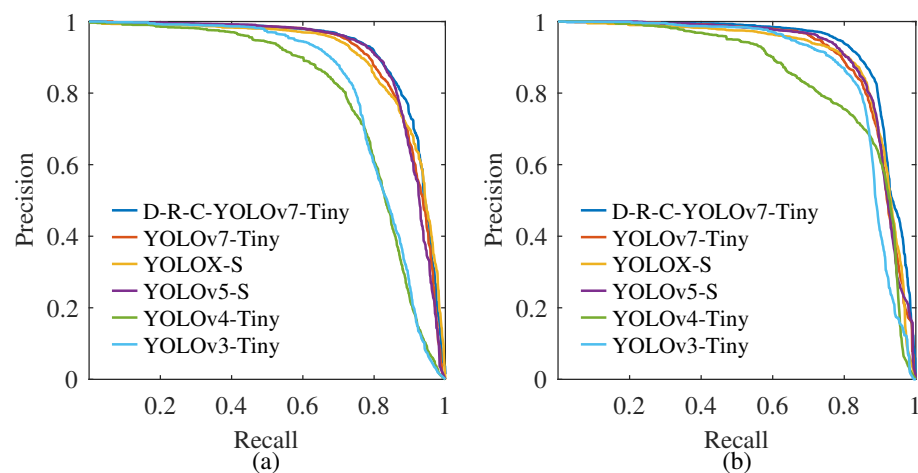
Figure 13. The P-R curves of each comparison model. (**a**) P-R curves for each comparison model on the validation set; (**b**) P-R curves for each comparison model on the test set.

　　Table 4 presents the precision, recall, and single-image inference time metrics for various target detection models. Regarding model detection accuracy, the D-R-C-YOLOv7-Tiny model outlined in this study significantly outperforms the YOLOv7-Tiny model. It achieves mAP scores that are 2.4, 1.8, 7.4, and 5.0 percentage points higher than those of the YOLOX-S, YOLOv5-S, YOLOv4-Tiny, and YOLOv3-Tiny models, respectively. In the context of model detection speed, the D-R-C-YOLOv7-Tiny model exhibits an inference time of 26.4 ms. Although this represents an increase compared to YOLOv7-Tiny, YOLOv5-S, YOLOv4-Tiny, and YOLOv3-Tiny, it nonetheless meets the criteria for rapid identification of pests and diseases. Moreover, the D-R-C-YOLOv7-Tiny model adeptly maintains a balance between detection precision and inference velocity, resulting in optimal overall performance.

Table 4. Comparative analysis of the performance of various target detection models.

Model	Precision	Recall	F1 Score	mAP@.5	Inference Time/ms
YOLOv3-Tiny	0.911	0.810	0.857	0.872	10.1
YOLOv4-Tiny	0.874	0.775	0.821	0.848	11.9
YOLOv5-S	0.929	0.851	0.888	0.904	23.7
YOLOX-S	0.887	0.840	0.862	0.898	27.8
YOLOv7-Tiny	0.925	0.828	0.873	0.90	16.3
D-R-C-YOLOv7-Tiny	0.928	0.862	0.893	0.922	26.4

4. Discussion

4.1. Model Performance

　　In order to reduce the interference of complex background on rice disease recognition, this research integrates the Convolutional Block Attention Module with the T-ELAN module in the YOLOv7-Tiny model's backbone network, resulting in the advanced C-T-ELAN module. The experimental results showed that the C-T-ELAN module enhanced the model's attention to rice disease subjects and improved the model's recognition ability. To enhance the model's capability to recognize irregular diseases, the RG-bneck module was used to replace a convolution operation in the C-T-ELAN module, and an improved R-C-T-ELAN module was proposed. The analysis of test outcomes reveals that the R-C-T-ELAN module is effective in decreasing the model's parameter count while enhancing its detection accuracy. In order to reduce the impact of spatial misalignment of the classification and positioning features of the Yolov7-Tiny detection head on the training results, a lightweight DD-Head is proposed. The test results show that DD-Head can effectively improve the classification and positioning accuracy of network models.

　　The improved D-R-C-YOLOv7-Tiny model was compared with the classical target detection model. The results showed that the scores of mAP@.5 and F1 of the improved D-R-C-YOLOv7-Tiny model were 0.922 and 0.894, respectively, which showed higher detection performance and could identify rice diseases more effectively.

4.2. Future Work

　　This study amassed a moderately sized dataset of rice disease images. Future plans involve expanding both the variety and quantity of rice disease images in the dataset, aiming to encompass a comprehensive range of categories. This is intended to enhance the model's adaptability to complex field environments. Additionally, while the operational speed of the proposed model meets real-time detection standards, there remains potential for further improvement. Going forward, the model will undergo further optimization to develop a lightweight version with enhanced detection accuracy.

5. Conclusions

　　In this study, the D-R-C-YOLOv7-Tiny model is proposed, which is an advanced rice disease recognition model that strikes a balance between accuracy and reasoning speed. The

Convolutional Block Attention Module, RepGhost bottleneck module, and T-ELAN module are integrated into the backbone network of the YOLOv7-Tiny model, which improves the recognition accuracy of the model. In addition, the proposed DD-Head significantly alleviates the spatial misalignment problem in the classification and localization tasks performed by the YOLOv7-Tiny detection head during training. The experimental results show that our improved model is superior to the benchmark model in terms of the validity of the rice disease dataset.

References

1. Reinke, R.; Kim, S.M.; Kim, B.K. Developing japonica rice introgression lines with multiple resistance genes for brown planthopper, bacterial blight, rice blast, and rice stripe virus using molecular breeding. *Mol. Genet. Genom.* **2018**, *293*, 1565–1575. [CrossRef] [PubMed]
2. Kong, S.; Li, J.; Zhai, Y.; Gao, Z.; Zhou, Y.; Xu, Y. Real-Time Detection of Crops with Dense Planting Using Deep Learning at Seedling Stage. *Agronomy* **2023**, *13*, 1503. [CrossRef]
3. Zhang, G.; Xu, T.; Tian, Y. Hyperspectral imaging-based classification of rice leaf blast severity over multiple growth stages. *Plant Methods* **2022**, *18*, 123. [CrossRef] [PubMed]
4. Cao, Y.; Yuan, P.; Xu, H.; Martínez-Ortega, J.F.; Feng, J.; Zhai, Z. Detecting Asymptomatic Infections of Rice Bacterial Leaf Blight Using Hyperspectral Imaging and 3-Dimensional Convolutional Neural Network With Spectral Dilated Convolution. *Front. Plant Sci.* **2022**, *13*, 963170. [CrossRef] [PubMed]
5. Gong, H.; Liu, T.; Luo, T.; Guo, J.; Feng, R.; Li, J.; Ma, X.; Mu, Y.; Hu, T.; Sun, Y.; et al. Based on FCN and DenseNet Framework for the Research of Rice Pest Identification Methods. *Agronomy* **2023**, *13*, 410. [CrossRef]
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
7. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems*; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2015; Volume 28.
9. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
10. Zhang, Y.; Xiao, D.; Che, H.; Liu, Y. Rice Panicle Detection Method Based on Improved Faster R-CNN. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 10. (In Chinese with English Abstract)
11. Zhang, Z.; Shi, R.; Xing, Z.; Guo, Q.; Zeng, C. Improved Faster Region-Based Convolutional Neural Networks (R-CNN) Model Based on Split Attention for the Detection of Safflower Filaments in Natural Environments. *Agronomy* **2023**, *13*, 2596. [CrossRef]
12. Zhang, G.; Liu, Z.; Liu, M.; Fang, P.; Chen, X.; Liang, X. Automatic Detection of Rice Stem Section Parameters Based on Improved Mask R-CNN. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 281–289. (In Chinese with English Abstract)
13. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 21–37.
14. 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
15. Lin, X.; Zhang, J.; Xu, X.; Zhu, S.; Liu, D. Recognition and Classification of Rice Planthopper with Incomplete Image Information Based on Dictionary Learning and SSD. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 165–171. (In Chinese with English Abstract)

16. Xiong, J.; Zheng, Z.; Liang, J.; Zhong, Z.; Liu, B.; Sun, B. Citrus Detection Method in Night Environment Based on Improved YOLO v3 Network. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 8. (In Chinese with English Abstract)
17. Wang, Y.; Lin, J.; Wang, S. Early rice disease recognition method based on YOLOv4-tiny model. *Jiangsu Agric. Sci.* **2023**, *51*, 147–154. (In Chinese with English Abstract)
18. Sun, F.; Wang, Y.; Lan, P.; Zhang, X.; Chen, X.; Wang, Z. Identification of apple fruit diseases using improved YOLOv5s and transfer learning. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 11. (In Chinese with English Abstract)
19. Aziz, F.; Ernawan, F.; Fakhreldin, M.; Adi, P.W. YOLO Network-Based for Detection of Rice Leaf Disease. In Proceedings of the 2023 International Conference on Information Technology Research and Innovation (ICITRI), Jakarta, Indonesia, 16 August 2023; pp. 65–69. [CrossRef]
20. Sangaiah, A.K.; Yu, F.N.; Lin, Y.B.; Shen, W.C.; Sharma, A. UAV T-YOLO-Rice: An Enhanced Tiny Yolo Networks for Rice Leaves Diseases Detection in Paddy Agronomy. *IEEE Trans. Netw. Sci. Eng.* **2024**, 1–16. [CrossRef]
21. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* **2020**, *175*, 105527. [CrossRef]
22. Lin, T. LabelImg. [EB/OL]. Available online: https://github.com/tzutalin/labelImg (accessed on 15 May 2023).
23. 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 (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
24. Wang, C.Y.; Liao, H.Y.M.; Yeh, I.H. Designing Network Design Strategies Through Gradient Path Analysis. *arXiv* **2022**, arXiv:2211.04800.
25. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
26. Chen, C.; Guo, Z.; Zeng, H.; Xiong, P.; Dong, J. RepGhost: A Hardware-Efficient Ghost Module via Re-parameterization. *arXiv* **2022**, arXiv:2211.06088.
27. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
28. Song, G.; Liu, Y.; Wang, X. Revisiting the Sibling Head in Object Detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
29. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
30. Cao, J.; Li, Y.; Sun, M.; Chen, Y.; Lischinski, D.; Cohen-Or, D.; Chen, B.; Tu, C. DO-Conv: Depthwise Over-Parameterized Convolutional Layer. *IEEE Trans. Image Process.* **2022**, *31*, 3726–3736. [CrossRef] [PubMed]
31. Gao, Y.; Mosalam, K.M. Deep transfer learning for image-based structural damage recognition. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 748–768. [CrossRef]
32. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
33. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
34. Glenn, J. YOLOv5 by Ultralytics. [EB/OL]. Available online: https://github.com/ultralytics/yolov5 (accessed on 1 May 2023).