AgriEngineering

MDPI

Article

Ghost-Attention-YOLOv8: Enhancing Rice Leaf Disease Detection with Lightweight Feature Extraction and Advanced Attention Mechanisms

Thanh Dang Bui * and Tra My Do Le

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Dai Co Viet Street, Hanoi 100000, Vietnam

* Correspondence: thanh.buidang@hust.edu.vn; Tel.: +84-9158-97699

Abstract: In agricultural research, effective and efficient disease detection in crops is crucial for enhancing yield and sustainability. This study presents a novel approach to improving YOLOv8, a state-of-the-art object detection model, by integrating the Ghost model with three advanced attention mechanisms: Convolutional Block Attention Module (CBAM), Triplet Attention, and Efficiency Multi-Scale Attention (EMA). The Ghost model optimizes feature extraction by reducing computational complexity, while the attention modules enable the model to focus on relevant regions, improving detection performance. The resulting Ghost-Attention-YOLOv8 model was evaluated on the Rice Leaf Disease dataset to assess its efficacy in identifying and classifying various diseases. The experimental results demonstrate significant improvements in accuracy, precision, and recall compared to the baseline YOLOv8 model. The proposed Ghost YOLOv8s with Efficiency Multi-Scale Attention model achieves a parameter count of 5.5 M, a reduction of 4.3 million compared to the original YOLOv8s model, while the accuracy is improved: the mAP@50 metric reaches 95.4%, a 2.3% increase; and mAP@50–95 improves to 62.4%, an increase of 3.7% over the original YOLOv8s. This research offers a practical solution to the challenges of computational efficiency and accuracy in agricultural monitoring, contributing to the development of robust AI tools for disease detection in crops.

Keywords: YOLOv8; Ghost network; CBAM; Triplet Attention; EMA; rice leaf diseases

check for updates

Academic Editors: Ray E. Sheriff and Chiew Foong Kwong

Received: 18 February 2025
Revised: 14 March 2025
Accepted: 17 March 2025
Published: 25 March 2025

Citation: Bui, T.D.; Do Le, T.M. Ghost-Attention-YOLOv8: Enhancing Rice Leaf Disease Detection with Lightweight Feature Extraction and Advanced Attention Mechanisms. *AgriEngineering* **2025**, *7*, 93. https://doi.org/10.3390/agriengineering7040093

Copyright: © 2025 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

Rice is the cornerstone of food security for over half of the world's population, with over 700 million tons produced annually, 90% of which is cultivated and consumed in Asia, including Vietnam [1]. As a critical agricultural commodity, rice plays a pivotal role in the economies of rice producing countries. For instance, Vietnam's rice exports in 2023 amounted to approximately USD 4.4 billion, supplying over 150 countries and meeting increasing demands for higher-quality produce. However, rice production faces significant challenges, particularly from leaf diseases such as rice folder, brown spot, and blast disease, which cause considerable yield losses [2–4]. Blast disease, in particular, poses a severe threat due to its widespread occurrence and potential for catastrophic damage under conducive conditions. Historical data highlight this impact, with documented [5] yield losses of 5–10% in India (1960–1961), 8% in Korea (mid-1970s), and 14% in China (1980–1981). In the Philippines, losses have ranged from 50 to 85%. More recently [6], Vietnam reported an average damage rate of 7% for bacterial stripe disease in 2023, with localized damage reaching 2.5–8.5%. These figures underscore the necessity of effective detection and diagnosis of rice

diseases to enable timely interventions and prevent large-scale crop losses. Despite their critical importance, current disease detection methods primarily rely on manual inspection by agricultural experts, a process that is time consuming, labor intensive, and prone to human error or inconsistency. Additionally, some infected areas on rice leaves are very small, making them hard to detect and significantly reducing the accuracy of identification by humans. Therefore, this often leads to a higher likelihood of misdiagnosis.

Such limitations underscore the growing need for technological innovations to transform disease detection processes in agriculture. Artificial intelligence, particularly deep learning, offers a powerful alternative to traditional methods by enabling precise, automated, and scalable analysis of plant health [7–9]. For example, Chen et al. [10] developed a deep CNN to identify tea plant disease types called LeafNet. Jiang, Feng, et al. [11] used CNNs plus support vector machine (SVM) to classify and predict four rice leaf diseases. A. P. Shaji and H. S [12] proposed using the Resnet-32 model for early detection of rice plant diseases, enhanced by data augmentation to improve classification accuracy. The model achieved 99.55% accuracy on a dataset of 12,684 augmented rice leaf images. Sudhesh, K. M., et al. [13] proposed a dynamic mode decomposition based on attention-driven pre-processing for rice leaf disease identification.

Although the mentioned approaches have shown strong performance in managing crop diseases, they primarily concentrate on either identifying or classifying disease images in a cleaned background dataset. In recent years, deep learning-based computer vision methods, particularly object detection models like YOLO (You Only Look Once), have emerged as promising tools for automating plant disease recognition. The YOLO family of models is widely known for its high speed and accuracy, making it an ideal candidate for real-time agricultural applications. Kiratiratanapruk et al. [14] utilized YOLOv3 to identify and classify six types of rice leaf disease. The model was trained and evaluated on a dataset of 6330 self-collected images, achieving an accuracy of 79.19%. Zhuo, Wang, et al. [15] introduced YOLOv4-CA, a real-time apple detection model. The design utilized MobileNet v3 as the backbone for feature extraction and incorporated depthwise separable convolution in the feature fusion network, achieving an average precision of 92.23%. Haque, Md Ershadul, et al. [16] used a rice leaf disease classification and detection method using YOLOv5, achieving a recognition precision of 90%, recall of 67%, mAP of 76%, and F1-score of 81% on 1500 annotated datasets. An AI-based solution using YOLOv7 to detect tea leaf diseases with 97.3% accuracy was presented by Soeb, Md Janibul Alam, et al. [17], leveraging a dataset of 4000 annotated images.

Among the YOLO series, YOLOv8 is has improved speed, accuracy, and versatility, making it highly effective for detection, segmentation, and keypoint tracking in real-time applications. For example, Qadri, Syed Asif Ahmad, et al. [18] presented an end-to-end deep learning solution for detecting and segmenting plant leaf diseases using the YOLOv8 model, trained on the PlantVillage and PlantDoc datasets. The model achieves high accuracy, with precision and recall reaching 99.8% and 99.3%, respectively. In addition, attention modules play a crucial role in enhancing the accuracy of deep learning models [19–21]. Recently, attention modules have often been integrated into YOLO model to enhance feature learning by emphasizing important parts of an image while suppressing irrelevant or less informative regions. Ma, Baoling, et al. [22] proposed the YOLOv8n-ShuffleNetv2-Ghost-SE model that optimizes apple fruit monitoring in smart orchards by replacing the Conv modules with Ghost modules and the C2f modules with C2fGhost modules in the neck, reducing network computation costs. Additionally, the Squeeze-and-Excitation (SE) modules are embedded in the backbone and neck to enhance feature map representation. The model achieved 94.1% precision and 82.6% recall. Chen, Zuxing, et al. (2024) [23] proposed an improved YOLOv8 method for real-time detection of grape leaf diseases,

incorporating several advancements: the AKConv module, the Coordinate Attention (CA) mechanism, and the CARAFE module. The model achieved an F1-score of 92.4%, mAP50 of 92.8%, and mAP50–95 of 73.8%, with reduced parameters and computational costs. Zamri, Fatin Najihah Muhamad, et al. [24] optimized the YOLOv8n model by integrating the ResCBAM attention module and adding a high-resolution detection head to improve tiny-object detection. The enhanced model, P2-YOLOv8n-ResCBAM, increases mAP from 90.3% to 92.6%. It effectively distinguishes drones from birds and enhances aerial surveillance capabilities. Given these advancements, integrating attention modules into the YOLOv8 model for rice disease detection is also a promising approach to further enhance accuracy.

However, challenges remain in detecting subtle and complex patterns in plant diseases [25], especially in rice leaf disease detection. Common rice leaf diseases such as leaf folders, leaf blast, and brown spot often have characteristics like small spots, small lesions, color changes, and shape variations. Furthermore, other research often just trains the model on simple background datasets [11–13], while the real-life data captured on the field has a more complex background. To enhance the detection accuracy of small objects in real-life images, we propose a lightweight rice leaf disease detection method based on an improved YOLOv8, known as Ghost-Attention-YOLOv8n (GA-YOLOv8n). Firstly, GA-YOLOv8n leverages some advanced attention mechanisms, which allow the model to focus on the most relevant intricate features of the image, thus improving the overall efficiency. We investigated the impact of three distinct attention modules integrated into the Ghost-YOLOv8 framework: Convolutional Block Attention Module (CBAM), Triplet Attention Module, and Efficient Multi-Scale Attention Module (EMA). The selection of these three attention modules is based on their unique characteristics that align with the challenges of rice leaf disease detection. CBAM [26,27] applies both channel-wise and spatial attention to recalibrate the feature maps, allowing the model to focus on the most informative features while suppressing irrelevant ones. This is particularly useful for rice leaf disease detection, where subtle texture differences are key indicators. Triplet Attention [28] captures relationships between features in a triplet-wise manner (height channel, width channel, and spatial dimension), enhancing the model's sensitivity to subtle disease patterns. And EMA [29] captures multi-scale spatial features that help in handling variations in disease severity and scale, a critical aspect for agricultural datasets. In addition, while YOLOv8 is a powerful object detection model, its standard convolutional layers have high computational costs, challenging for real-time application. To optimize the model's size, we utilize the Ghost module [30,31], a lightweight module that reduces computational complexity while preserving detection accuracy. By using fewer parameters while maintaining critical feature information, Ghost-YOLOv8 ensures faster inference times, which is crucial for deployment on edge devices like drones or mobile phones in the field. We trained and tested our proposed model on the Rice Leaf Disease dataset to demonstrate its effectiveness. The Rice Leaf Disease dataset includes 3563 images for 03 common diseases in rice plant, which were collected in the paddy fields of Vietnam National University of Agriculture. By incorporating these attention mechanisms, we hypothesize that Ghost-Attention-YOLOv8 will not only improve detection accuracy but also enhance the model's robustness when detecting diseases in rice leaf images.

The main contributions of this research are as follows:

- By utilizing the lightweight Ghost module, the model reduces computational complexity and parameters while maintaining high accuracy, enabling real-time deployment on edge devices.
- Exploring the integration of three advanced attention modules, CBAM, Triplet Attention, and Efficiency Multi-Scale Attention, each chosen for its potential to enhance feature selection and discrimination in complex natural rice leaf datasets, therefore

providing a detailed comparative analysis of how these attention mechanisms impact the performance of the model in the specific context of rice leaf disease detection.
- The model is evaluated on the Rice Leaf Disease dataset, featuring real-world images with natural backgrounds, ensuring robustness and practical applicability.

The rest of the paper is structured as follows: Section 2 outlines the methodology, including an explanation of the original YOLOv8 and our enhancements to YOLOv8 with Ghost modules and attention mechanisms. Section 3 describes the dataset, the experimental procedures used to test the model, and the results, highlighting the gains in accuracy and computational efficiency. Section 4 discusses the results, compares them with previous studies, and presents aims for future work. Lastly, Section 5 concludes with a summary of the findings and suggestions for future work.

2. Materials and Methods

2.1. Overview of YOLOv8

YOLOv8 [32] introduces a streamlined and efficient architecture for object detection, incorporating key innovations to improve accuracy and reduce computational demands. The backbone of YOLOv8 is based on an enhanced version of CSPDarknet53, using a modular design that performs multiple down-sampling operations. These down-samplings produce feature maps at varying scales, enabling the extraction of rich hierarchical features from the input image.

The backbone utilizes the C2f module, a lightweight and efficient replacement for traditional modules. The C2f module incorporates gradient-parallel connections, which enhance the flow of feature information across layers while maintaining a compact design. It also includes bottlenecks that improve feature extraction without significantly increasing computational complexity. To process input data, the architecture uses a ConvModule, which combines convolutional operations, batch normalization, and the SiLU activation function. This module ensures robust feature extraction and non-linear transformations. Furthermore, the backbone integrates an improved down-sampling module to pool feature maps into adaptive-sized outputs, preserving critical spatial information for downstream tasks.

The neck of YOLOv8 uses a PAN-FPN (Path Aggregation Network and Feature Pyramid Network) structure to integrate features from different levels of the backbone. This design enables the fusion of semantic information from deep layers with fine-grained details from shallow layers, providing a comprehensive representation for object detection.

Finally, YOLOv8 employs a decoupled head structure with two separate branches, one for object classification and another for bounding box regression, to predict the target's location and classification by utilizing the output features from the neck. The general architecture of YOLOv8 is shown in Figure 1.

2.2. Ghost Module

To comprehensively understand input, traditional convolutional layers pass the input tensor through several filters to create new feature maps. This leads to rich and even redundant feature maps, more computation and memory usage, but does not add much value. Therefore, we incorporate "ghost feature maps" into the YOLOv8 architecture, which is faster than standard convolution due to fewer calculations, along with a smaller model size, while still maintaining performance [30,31].

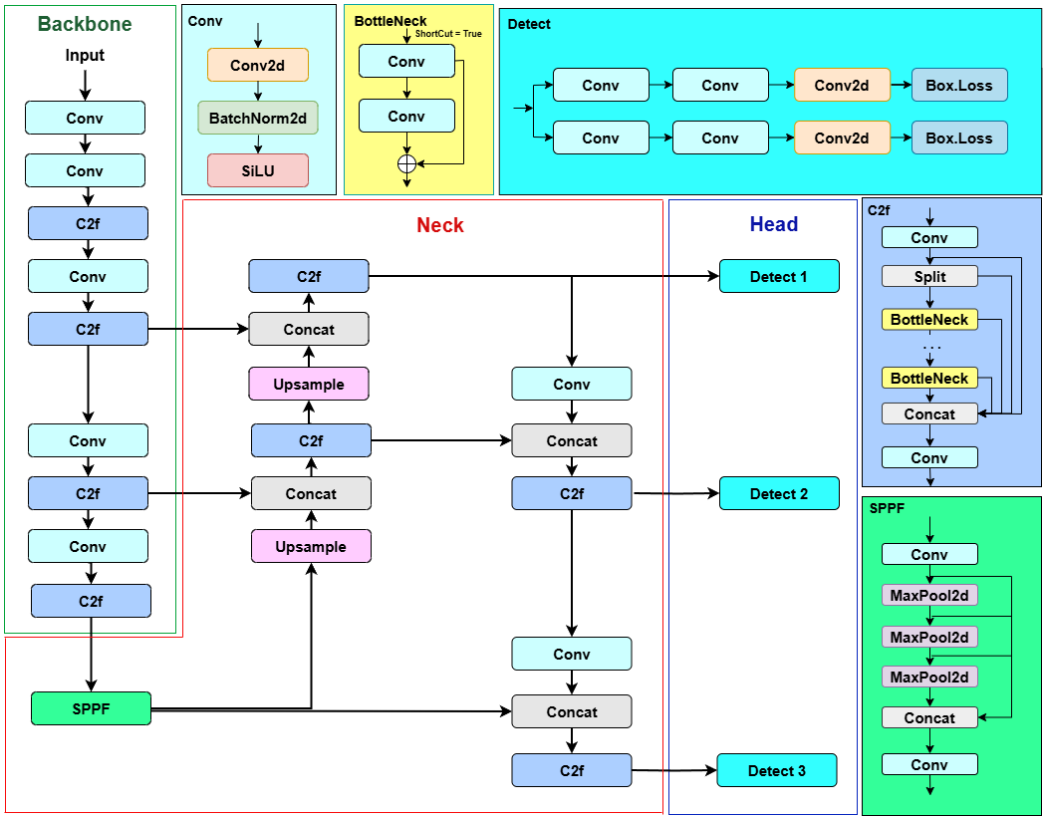


Figure 1. Architecture of original YOLOv8.

To gain these benefits, the Ghost module divides the generated feature maps into two sets. The first set is obtained by generating a smaller number of intrinsic feature maps from standard convolutional files, then from these maps new similar feature maps (the second set) are created by using cheaper linear operations. Finally, two sets of feature maps are stitched, as in the output shown in Figure 2. This makes Ghost convolution more parameter-efficient, suitable for mobile and embedded systems where computation and memory resources are limited.

We represent the input feature map by $X \in R^{C \times H \times W}$; where C, H, and W, respectively, denote the number of channels, and the height and width of the data. The arbitrary conventional convolution layer can be described by the following formula:

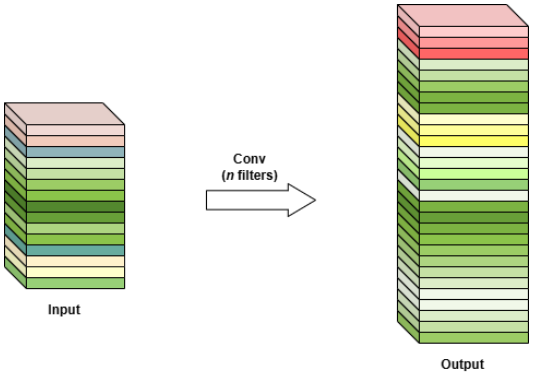
$$Y = X * f + b \tag{1}$$

where $*$ is the convolution operation, b is the bias term, and $f \in R^{C \times k \times k \times n}$ is the convolution filter with kernel size $k \times k$ to generate n feature maps. $Y \in R^{n \times H' \times W'}$ is the output feature map with a number of channels C, data height H', and width W'. Therefore, the number of FLOPs, which represents the computational complexity of traditional convolution, is approximately equal to $C \times H' \times W' \times k \times k \times n$ (simplifying the bias term b). On the other hand, the Ghost module can be described as follows:

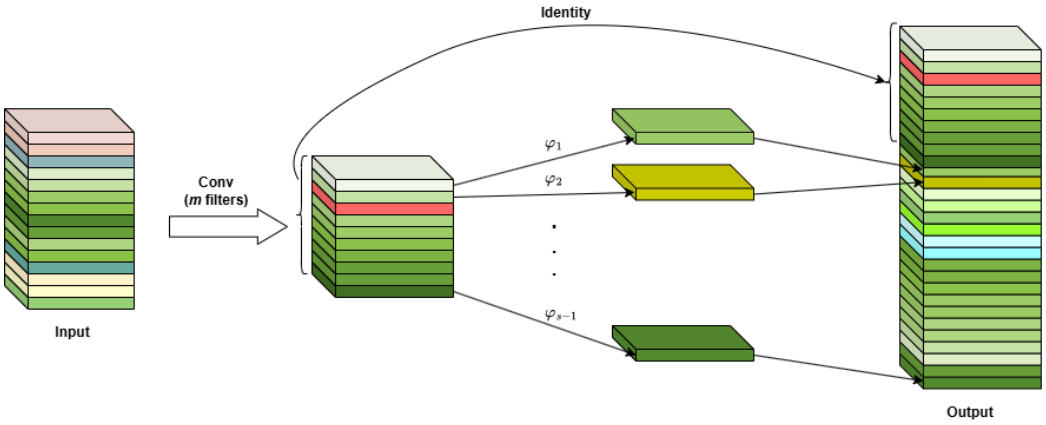
$$Y' = X * f' + b \tag{2}$$

where $f' \in R^{C \times k \times k \times m}$ is the convolution filter, and other parameters such as filter size, stride, and padding remain the same as in conventional convolution. With $m < n$, the output map $Y' \in R^{m \times H' \times W'}$ is the first set mentioned above and has a smaller number of layers than output feature maps Y in traditional convolution. Then, each feature map within Y' undergoes the linear operation φ_i to create several Ghost layers, assuming the

amount of transformation is $s - 1$. There, the final channel map number is $n = m \cdot s$, and the number of FLOPs in the Ghost convolution approximately decreases s times if the average kernel size per linear operation is close to $k \times k$.



(**a**)



(**b**)

Figure 2. Comparison between standard convolutional module and GhostConv module. (**a**) Standard convolutional module; (**b**) GhostConv module.

2.3. Attention Module

2.3.1. CBAM

The Convolutional Block Attention Module (CBAM) [26] is a lightweight network module designed to enhance the performance of networks by leveraging attention mechanisms. Unlike traditional approaches that focus on the depth, width, or cardinality of networks, CBAM improves the representation power by dynamically focusing on the most informative features. It achieves this by sequentially applying two attention modules: one for channel attention, which highlights important features across different channels; and one for spatial attention, which emphasizes relevant spatial locations. This dual-attention mechanism helps the network focus on both important features (by channel) and their locations (spatially), improving information flow and making the model easier to understand. The structure of CBAM is shown in Figure 3.

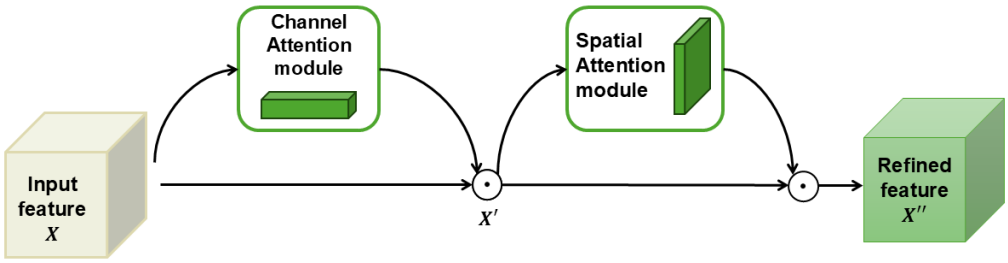


Figure 3. Overview of CBAM (⊙ is the element-wise multiplication matrix).

Given an input feature map $X \in R^{C\times H\times W}$, and final refined output X'', the overall CBAM can be represented as

$$X' = M_c(X) \odot XX'' = M_s(X') \odot X' \tag{3}$$

Channel Attention Module

A channel attention map (Figure 4) is generated by leveraging the relationships between different channels of the feature map. Since each channel acts as a feature detector, channel attention helps identify what is most meaningful in an input image. To compute this attention effectively, it reduces the spatial dimensions of the input tensor using both average pooling and max pooling. Average pooling is applied to assess the target object while max-pooling is to capture additional distinctive features for finer channel-wise attention. After using pooling techniques, two different spatial context descriptors, X^c_{avg} and X^c_{max}, are generated. The next step is passing these two descriptors through a shared multi-layer perceptron (MLP) to compute the final channel attention map $M_c \in R^{c\times 1\times 1}$. The MLP reduces the dimensionality using a hidden layer of size $R^{c/r\times 1\times 1}$, with r being the dimensionality reduction factor. Subsequently, the outputs are aggregated and go through the sigmoid activate function. The calculation process in the channel attention module can be represented by the following equation:

$$M_c(X) = \sigma(MLP(X^c_{avg}) + MLP(X^c_{max})) \tag{4}$$

Spatial Attention Module

After the channel attention step, the spatial attention module takes X' as the input. CBAM utilizes SA to emphasize the location of important information by leveraging the relationships between features in different spatial locations. The structure of the spatial attention module is shown in Figure 5. For this, first average pooling and max-pooling are used along the channel axis, generating two maps X^s_{avg} and X^s_{max}. Following that, the pooled feature maps are concatenated before forwarding them to a convolutional layer and sigmoid function. The process of calculation in the spatial channel module can be presented as

$$M_s(X) = \sigma(Conv[X^s_{avg}; X^s_{max}]) \tag{5}$$

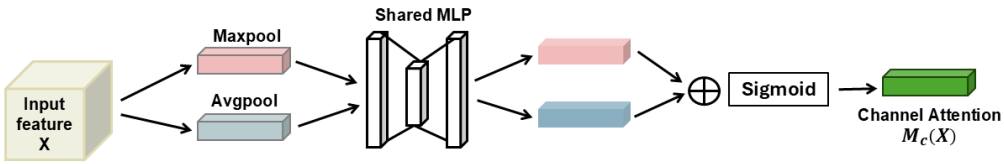


Figure 4. Diagram of channel attention module.

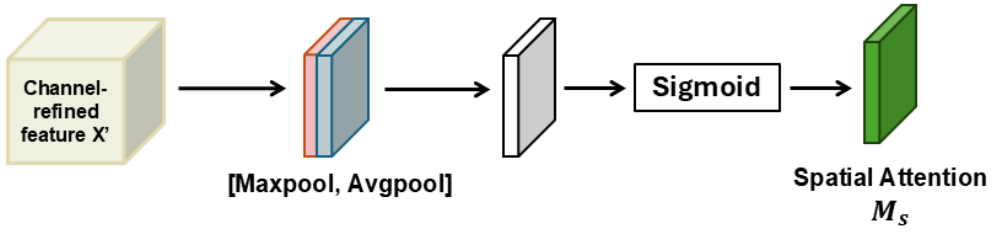


Figure 5. Diagram of spatial attention module.

2.3.2. Triplet Attention

As shown in Figure 6, Triplet Attention [28] is a three-branch module that process an input feature map $X \in R^{C \times H \times W}$ to produce a refined tensor of the same shape. Each branch handles interactions in different dimensions.

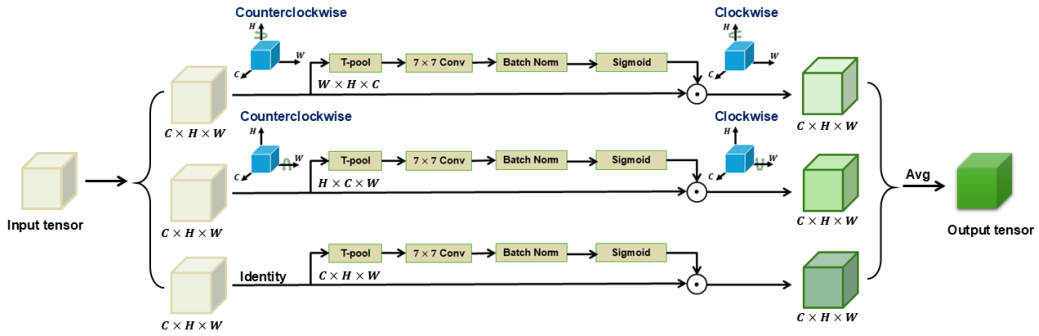


Figure 6. Structure of Triplet Attention module.

The first branch of the Triplet Attention module is utilized to build the relation between the spatial height (H) dimension and the channel (C) dimension. This branch starts by rotating the input X 90 degrees counterclockwise along the H-axis to achieve the permuted tensor $X_H \in R^{C \times H \times W}$. Next, X_H goes through T-pooling, squeezing to $X_H^{'}$, of shape $(2, H, C)$, followed by a convolutional layer of kernel size k×k and batch normalization, yielding $(1, H, C)$. Subsequently, attention weights are generated by using sigmoid activation and applied back to the permuted tensor X_H. At the end of the first branch, X_H is rotated clockwise along the H-axis to regain its original shape. The process of this branch can be represented by the following equation:

$$Y_H = R^{H+}(X_H \cdot \sigma(ConvBN(X_H^{'}))) \quad (6)$$

where R^{H+} represents the clockwise rotation along the H-axis; σ represents the sigmoid activation function; and $ConvBN$ represents the convolution and batch normalization operations.

Similarly, in the second branch, the input feature map X is rotated 90 degrees counterclockwise along the W-axis to obtain the permuted tensor $X_W \in R^{H \times C \times W}$. Next, X_W goes through T-pooling to become $X_W^{'}$, of shape $(2, C, W)$. $X_W^{'}$ then passes through standard convolution with batch normalization, producing a tensor of shape $(1, C, W)$. Attention weights are then computed via sigmoid activation and reapplied to the permuted tensor X_W, which is then rotated back to its original shape, the same as the input X. The process of this branch can be represented by the following equation:

$$Y_W = R^{H+}(X_W \cdot \sigma(ConvBN(X_W^{'}))) \quad (7)$$

In the last branch, the input tensor X is passed directly to T-pooling without permutation, resulting in a tensor X' of shape $(2, H, W)$. This tensor undergoes convolution and batch normalization, then generates attention weights of shape $(1, H, W)$ through sigmoid activation before being applied to the input tensor X. The process of the branch can be represented by the following equation:

$$Y_C = X \cdot \sigma(ConvBN(X'))) \tag{8}$$

Subsequently, the Triplet Attention module uses simple averaging of the three above equations to achieve the final output result. The last equation is as follows:

$$Y = \frac{1}{3}(Y_H + Y_W + Y_C) \tag{9}$$

2.3.3. Efficient Multi-Scale Attention

The Efficient Multi-Scale Attention (EMA) module [29] is designed to enhance feature representation in deep learning, particularly for image classification and object detection. Unlike traditional attention models that reduce channel dimensionality, EMA maintains full dimensionality to improve feature extraction while minimizing computational costs. EMA employs a unique combination of cross-channel and spatial attention mechanisms, working through multi-scale parallel subnetworks that use different convolutional kernels (1×1 and 3×3) to capture a range of spatial features. It operates by dividing channels into groups and processing these subfeatures separately, allowing the model to establish both short- and long-range dependencies. This approach enables EMA to capture both local and global spatial information effectively, making it suitable for high-resolution tasks. In benchmark evaluations on datasets like CIFAR-100, ImageNet-1k, and COCO, EMA demonstrated improvements in accuracy and efficiency over competing modules, such as CBAM and Coordinate Attention. Its design offers practical applications in various computer vision tasks, as it effectively enhances CNN architectures without significantly increasing computational overhead. The general structure of the Efficiency Multi-Scale Attention module is shown in Figure 7.

Feature Grouping and Parallel Subnetworks

Given an input feature map $X_W \in R^{C \times H \times W}$, where C, H, and W are the number of channels, height, and width, it is divided into G subfeatures, with the shape redefined as $C//G \times H \times W$. As large local receptive fields of neurons can collect multi-scale spatial information, the feature maps' attention weight descriptors are gathered by three separate parallel subnetworks, two 1×1 convolutional branches and one 3×3 convolutional branch, which efficiently capture multi-scale feature representation. There exist two 1D global average pooling operations applied to two routes in the 1×1 branch and only one 3×3 kernel is used in the 3×3 branch. Moreover, the operations meant to encode global information along the spatial horizontal and vertical directions are represented, respectively, as the equations below:

$$Z_c^H(H) = \frac{1}{W} \sum_{0 \le j \le W} x_c(H, i) \tag{10}$$

$$Z_c^W(W) = \frac{1}{H} \sum_{0 \le j \le H} x_c(j, W) \tag{11}$$

where x_c indicates the input features at the cth channel. With such processes, EMA captures long-range information. and embeds the positional information into channel-wise interactions spatially. Then, these encoded features are concatenated across a 1×1 convolutional

layer without dimensionality reduction, resulting in the formation of two vectors which are employed by non-linear sigmoid functions to fit the 2D binormal distribution upon linear convolutions. Two attention maps are finally aggregated through multiplication to achieve different cross-channel interactive features in the 1×1 branch. At the same time, the 3×3 kernel in the 3×3 branch captures the local cross-channel interaction to enlarge the feature space.

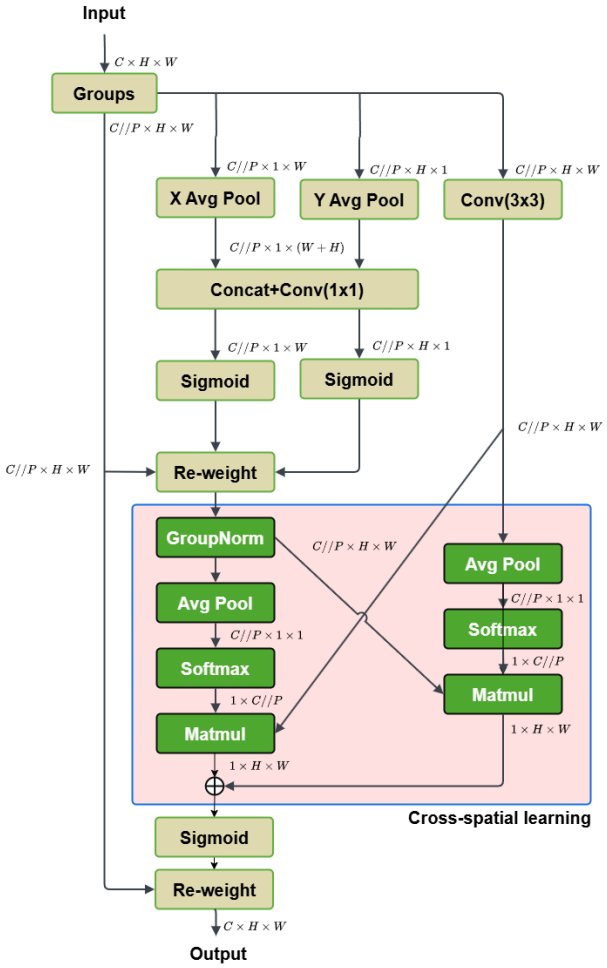


Figure 7. Structure of Efficient Multi-Scale Attention.

Cross-Spatial Learning

For richer feature aggregation, EMA combines outputs from the two branches using cross-spatial learning to establish a final attention map that accounts for pixel-level relationships. In detail, two tensors from different branches are encoded by the 2D global average pooling, formulated as

$$z_c = \frac{1}{H \times W} \sum_{j}^{W} \sum_{i}^{W} x_c(i, j) \tag{12}$$

This is utilized to encode global spatial information and model long-range dependencies. To fit the linear transformations and compute efficiently, the work continues with the application of the natural non-linear function Softmax to both branches. After being directly transformed into the corresponding dimensional shape for the joint activation mechanism of channel features, the outputs are ultimately multiplied with matrix dot-product operations, giving spatial attention maps. After that, the fused feature map is calculated by an aggregation of two spatial attention weight values and passes through a sigmoid activation

to output the final reweighted feature map. It is the same size as X, giving it the ability to stack into modern architectures.

2.4. Improvement of YOLOv8 Network Architecture Design

To improve both the lightweight nature and the accuracy of the YOLOv8 model, this paper proposes the Ghost-Attention-YOLOv8 (GA-YOLOv8) algorithm. Specifically, we replace the standard convolutional layers and C2f modules in the head and neck components of the original YOLOv8 with the GhostConv and C3Ghost modules. These replacements optimize convolutional operations, contributing to a more lightweight model. Furthermore, GA-YOLOv8 integrates modules of the attention mechanism to enhance the detection of small targets. These attention modules are strategically placed before each level's detection head, allowing the model to focus more effectively on fine-grained features. The network architecture of the proposed network structure is described in Figure 8:

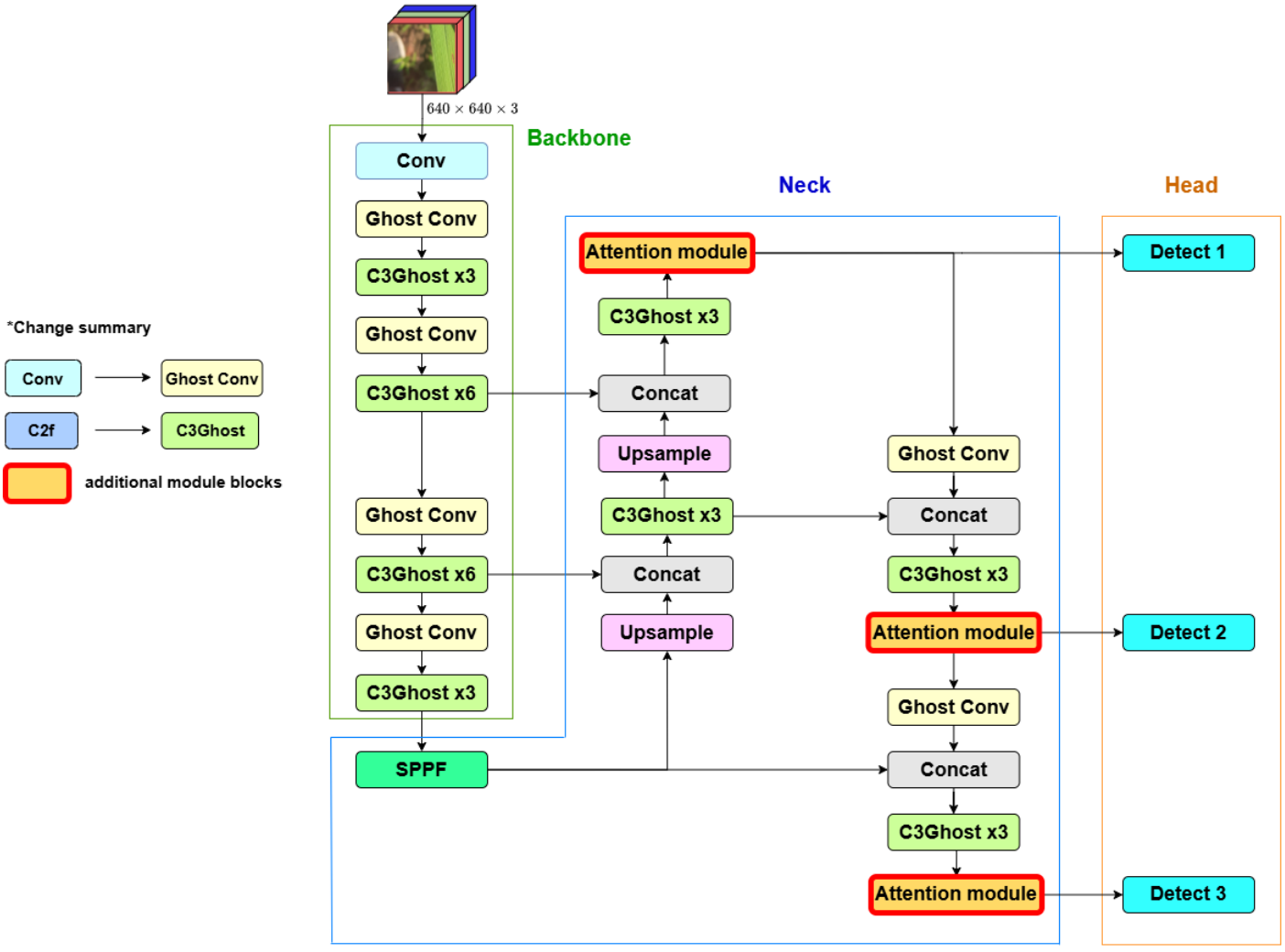


Figure 8. Ghost-Attention-YOLOv8 model architecture. (* part presents the changes).

The primary advantage of GhostConv lies in its ability to reduce computational overhead by generating ghost features that retain the essential information with significantly fewer operations than traditional convolutions, thereby maintaining model performance while improving efficiency. The GhostConv module starts with a standard convolution to produce a set of identity features maps, then divides the output into 2 parts. One part goes through a linear operation layer to generate ghost features. Subsequently, the identity

feature part and ghost feature part are concatenated to form a final feature map. The GhostConv module flowchart is shown in Figure 9.

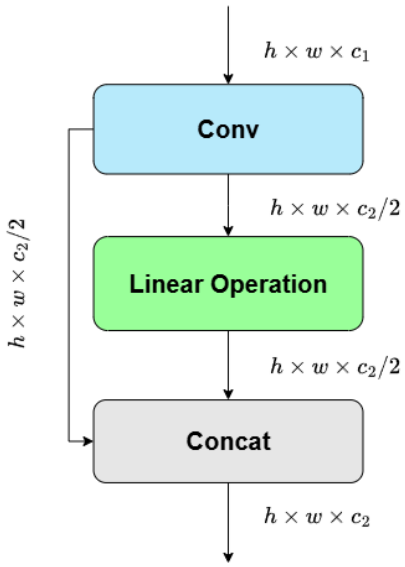


Figure 9. Structure of GhostConv.

The C3Ghost module maintains a good balance between speed and detail, making it suitable for tasks like object detection, where both small details and big picture features matter, especially in devices with limited power. The flowchart of the C3Ghost module is shown in Figure 10. The input is first processed by a standard convolution to extract the identity feature map, then the output is split into two parts. One part uses a Ghost bottleneck structure, which applies GhostConv to generate lightweight ghost features, while the other part uses a traditional convolutional layer to capture more complex features. The outputs from both parts are then concatenated and moved to a convolutional layer to produce the final output.

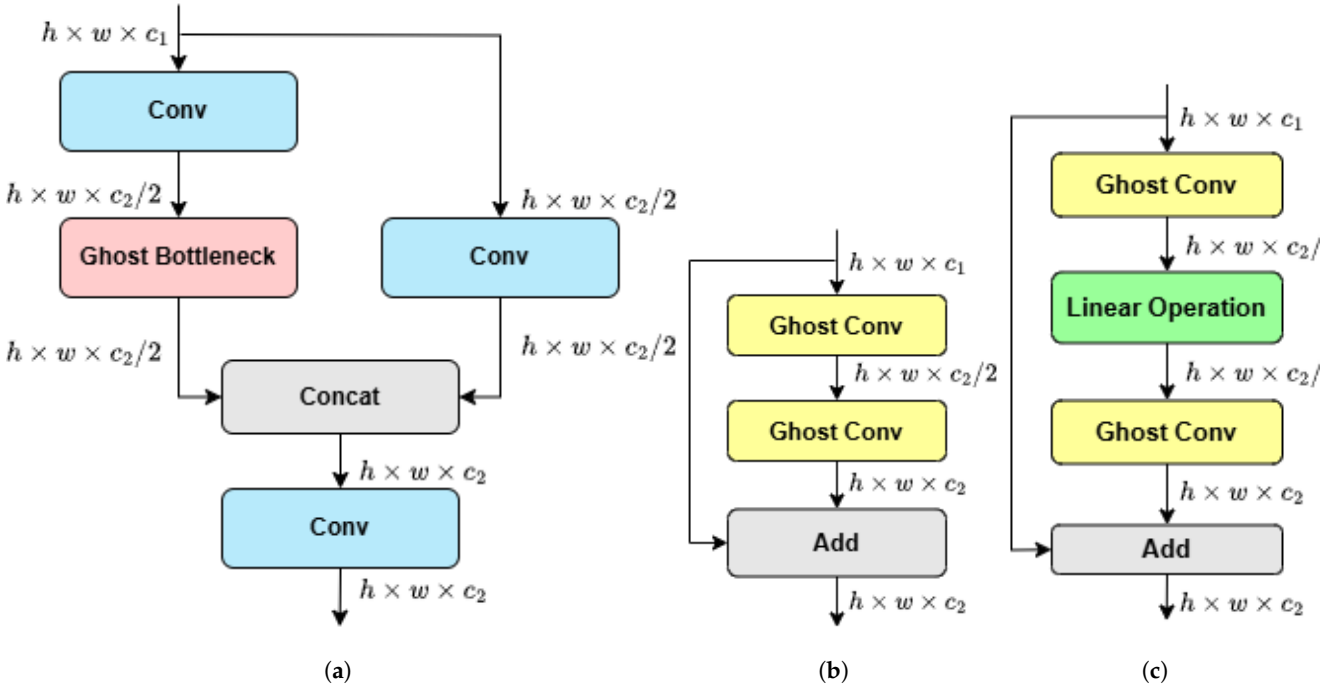


Figure 10. Structure of C3Ghost. (**a**) C3Ghost module. (**b**) Ghost bottleneck (stride = 1). (**c**) Ghost bottleneck (stride = 2).

3. Results and Discussion

3.1. Rice Leaf Disease Dataset Collection and Processing

The dataset used in this study consists of 3563 high-resolution images covering three common rice diseases: leaf folder, rice blast, and brown spot. These images were collected directly from rice fields at the Vietnam National University of Agriculture (VNUA) under real-world agricultural conditions. To ensure a diverse and representative dataset, images were captured at different times of the day and under varying environmental conditions, including changes in lighting and humidity, which can affect the appearance of disease symptoms.

For annotation, we used MakeSense.ai, an open-source labeling tool, to manually annotate disease-affected regions. Each image was labeled with a single bounding box per affected area, following a structured protocol. The annotation process was supervised by agricultural experts from VNUA, who provided guidance on disease classification and symptom differentiation to ensure high accuracy. To minimize bias and ensure dataset balance, the number of labeled samples per class was kept relatively equal: 1225 images for leaf folder, 1188 for rice blast, and 1150 for brown spot. We show some images from our dataset in Figure 11.

To further improve dataset quality and reliability, several quality control measures were implemented. (1) Expert validation: Annotations were reviewed by specialists from VNUA to verify disease classification and bounding box accuracy. (2) Manual inspection: Overexposed or misclassified images were removed to maintain dataset integrity. (3) Balanced dataset split: The dataset was divided into training (80%), validation (10%), and test (10%) subsets, ensuring each class was equally represented in each phase. There are 357 images in the test set, including 121 leaf folder images, 120 leaf blast images, and 116 images of brown spot disease. This dataset provides a realistic and high-quality benchmark for evaluating the Ghost-Attention-YOLOv8 model, ensuring its effectiveness in detecting rice leaf diseases under field conditions.



Figure 11. Images of three diseases: leaf folder, leaf blast, and brown spot.

In YOLOv8, different augmentation techniques are used during training, and their settings can be adjusted in the default.yaml file. In this study, we take advantage of these built-in methods, including horizontal and vertical flipping, translation, and mosaic augmentation, etc. All augmentation methods are listed in Table 1 and visualized in Figure 12. Instead of generating new images, these transformations are applied dynamically at each epoch, which helps increase the variety of training examples without expanding the dataset. This approach allows the model to learn from a wider range of variations, ultimately improving its performance and adaptability to different conditions.

Table 1. Augmentation hyperparameter table.

No.	Name	Fraction	Explanation
1	Hsv_ h	0.015	(float) image HSV—hue augmentation (fraction)
2	Hsv_ s	0.7	(float) image HSV—saturation augmentation (fraction)
3	Hsv_ v	0.4	(float) image HSV—value augmentation (fraction)
4	translate	0.1	(float) image translation (+/− fraction)
5	scale	0.5	(float) image scale (+/− gain)
6	flipud	0.5	(float) image flip up–down (probability)
7	fliplr	0.5	(float) image flip left–right (probability)
8	mosaic	1.0	(float) image mosaic (probability)

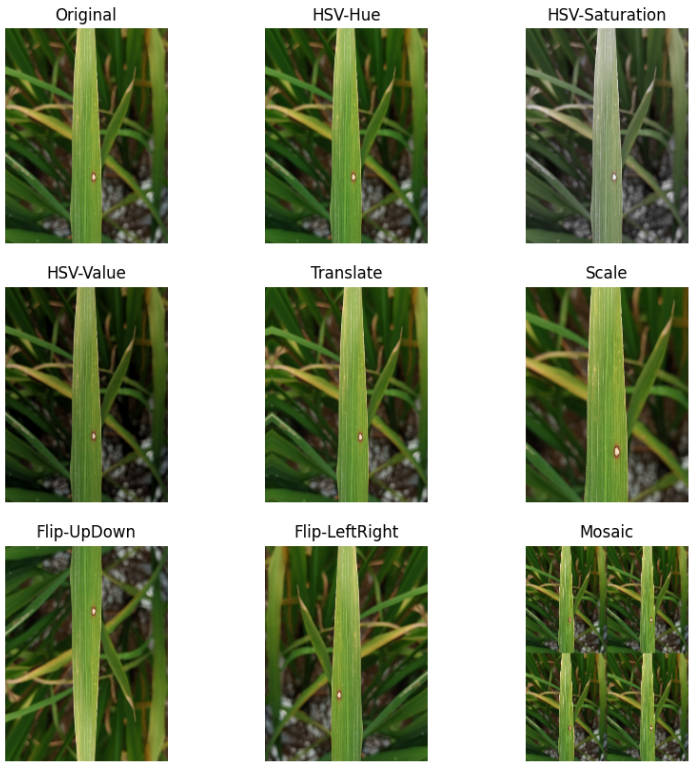


Figure 12. Visualization of augmentation methods used.

3.2. Test Platform and Parameter Settings

Our models were trained, validated, and tested using a computer with an AMD Ryzen 5 2600 processor (6C/12T, 3.4 GHz–3.9 GHz, 16 MB), graphics card, NVIDIA RTX 2070 GPU (8 GB GDDR6), 16 GB RAM. The image input size was automatically resized to 640 × 640 pixels. For the optimizer, Adam was used with a momentum value of 0.937. The initial and final learning rates were set to their default values of 0.01 and 0.0001, respectively. The dataset was trained for 100 epochs with a batch size of 16. Details of these parameters are provided in Table 2.

Table 2. Hyperparameter settings.

No.	Hyperparameter	Value	Explanation
1	Batch_size	16	Number of samples that will be propagated
2	Optimizer	Adam	Optimizer to use
3	Epochs	100	Number of epochs to train for
4	Warmup_epochs	3.0	Warm-up epochs (fraction ok)
5	Lr0	0.01	Initial learning rate
6	Lrf	0.0001	Final learning rate
7	Momentum	0.937	SGDMomentum/Adam beta1
8	box	7.5	Box loss gain
9	Cls	0.5	Class loss gain
10	Dfl	1.5	Distribution loss gain

3.3. Evaluation Metrics

In this study, the performance of the object detection models is assessed using key evaluation metrics, including precision (P), recall (R), F1-score, mean average precision at IoU threshold of 0.5 (mAP@0.5), and mean average precision across multiple IoU thresholds (mAP@50–95). Additionally, network parameters, model size, and inference speed are analyzed to evaluate the model's computational efficiency and real-time applicability.

$$Precision(P) = \frac{TP}{TP + FP} \tag{13}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{14}$$

$$F1 - score = \frac{2 \times P \times R}{P + R} \tag{15}$$

Precision indicates how many of the detected objects are actually correct, helping determine how many of the detected diseased regions are actually correct, reducing the risk of false alarms and unnecessary treatments. Recall measures how well the model detects all relevant objects, ensuring that no infected leaves are overlooked, which is critical for preventing disease spread.

The metric mAP@0.5 represents the mean average precision when the intersection over union (IoU) threshold is set at 0.5, ensuring that predicted bounding boxes significantly overlap with the ground truth. Meanwhile, mAP@50–95 provides a more comprehensive evaluation by averaging precision across multiple IoU thresholds, ranging from 0.5 to 0.95 in increments of 0.05. This metric ensures that the model can precisely differentiate healthy and infected areas under various conditions. In Equations (13)–(15), true positives (TPs) are correctly identified diseased regions, while false positives (FPs) occur when healthy areas are mistakenly classified as diseased. False negatives (FNs) represent actual diseased regions that the model fails to detect, potentially allowing disease spread.

3.4. Evaluation of GA-YOLOv8 with Different Attention Modules

3.4.1. Test Results on Rice Leaf Disease Dataset

The experiment is designed to evaluate the impact of each optimization module on performance. First, we assess the baseline performance of the original YOLOv8s on the Rice Leaf Disease dataset. Next, we enhance the YOLOv8 model by incorporating the Ghost module and evaluate its performance through training and testing. Finally, we further improve the Ghost-YOLOv8 model by embedding the CBAM, Triplet Attention, and EMA attention modules, one at a time, to assess their individual contributions to the model's effectiveness. We evaluate their effectiveness based on key metrics such as

detection accuracy, model efficiency, and computational resource usage, with a particular focus on their ability to distinguish between the three rice diseases in the dataset. The overall performance of each model is evaluated by the metrics precision, recall, F1-score, mAP@0.5, and mAP@50–95, as presented in Table 3.

Table 3. Comparison of different model performance metrics (%).

Model	Precision	Recall	F1-Score	mAP@50	mAP@50–95
YOLOv8s	88.3	90.0	89.1	93.1	58.7
YOLOv8s + Ghost	89.2	91.9	90.5	93.2	59.0
YOLOv8s + CBAM	89.4	91.0	90.2	94.0	58.9
YOLOv8s + CBAM + Ghost	91.9	91.5	91.7	93.8	60.7
YOLOv8s + Triplet Attention	90.0	91.2	90.6	94.9	61.2
YOLOv8s + Triplet Attention + Ghost	93.8	91.4	92.6	95.0	61.6
YOLOv8s + EMA	91.8	91.9	91.8	95.5	62.0
YOLOv8s + EMA + Ghost	91.3	93.4	92.3	95.4	62.4

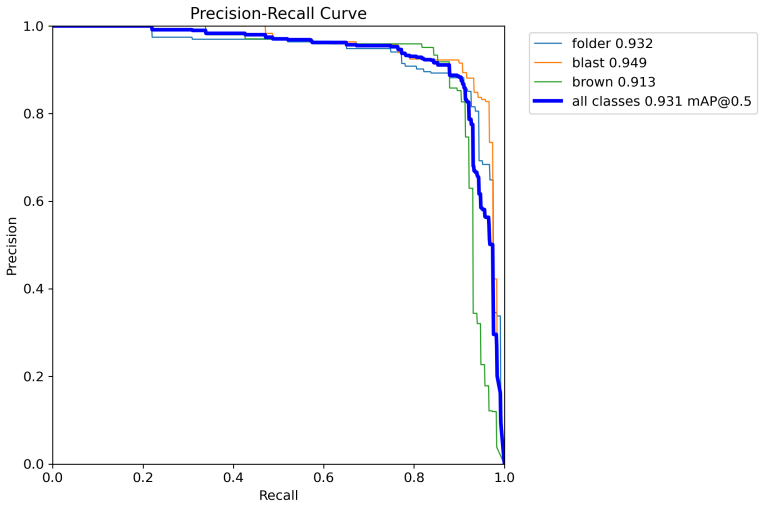
Introducing the Ghost module alongside these attention-based enhancements yielded additional improvements without degrading accuracy. As shown in Table 3, the YOLOv8s + Ghost model recorded an mAP@50 of 93.2% and mAP@50–95 of 59.0%, showing comparable accuracy to the baseline while offering computational efficiency. The CBAM-enhanced model achieved an mAP@50 of 94.0% and mAP@50–95 of 58.9%, while combining CBAM with Ghost modules further increased the mAP@50 to 93.8% and mAP@50–95 to 60.7%, demonstrating that Ghost modules work effectively with attention mechanisms.

The performance improvements of our method over the baseline YOLOv8s model are evident across multiple evaluation metrics. As shown in Table 3, when integrated with Ghost modules, the CBAM, Triplet Attention, and EMA mechanisms contributed 0.7%, 1.9%, and 2.3% increases in mAP@50, respectively. For the mAP@50–95 metric, these enhancements provided respective improvements of 1.0%, 2.9%, and 3.7%, highlighting the benefits of feature refinement and stability mechanisms.

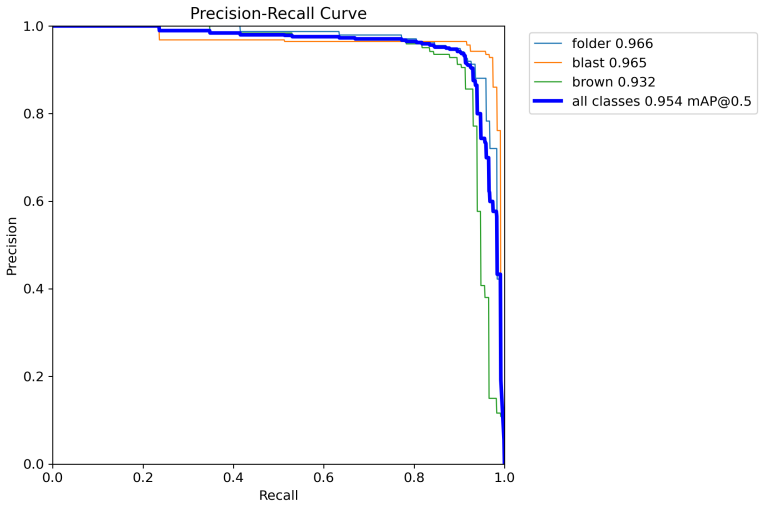
Further improvements were observed with Triplet Attention, which enhanced feature extraction and spatial relationships. YOLOv8s + Triplet Attention achieved an mAP@50 of 94.9% and mAP@50–95 of 61.2%, with its Ghost-enhanced variant reaching an mAP@50–95 of 61.6%, confirming that Ghost modules complement attention-based refinements. The most significant gains were achieved with Efficiency Multi-scale Attention (EMA), which stabilized training and improved generalization. YOLOv8s + EMA recorded an mAP@50 of 95.5% and mAP@50–95 of 62.0%, surpassing all prior configurations. Notably, integrating EMA with Ghost modules resulted in the highest performing model. YOLOv8s + EMA + Ghost achieved the best mAP@50 score of 95.4%, outperforming the original YOLOv8s (93.1%) by 2.3%. Similarly, for mAP@50–95, our proposed model reached 62.4%, marking a 3.7% increase over YOLOv8s (58.7%). The specific metrics for each class of Ghost YOLOv8s with EMA attention module are presented in Table 4. The PR curves of original YOLOv8 and GA-YOLOv8s with the EMA attention module are given in Figure 13.

Table 4. Summarized GA-YOLOv8s model with EMA on Rice Leaf Disease dataset.

Disease	Precision	Recall	mAP@50	mAP@50–95
Leaf Folder	91.9	92.2	96.6	70.0
Leaf Blast	92.2	97.5	96.5	70.4
Brown Spot	95.0	83.1	93.2	46.6
All	93.0	90.9	95.4	62.4



(**a**)



(**b**)

Figure 13. Precision–recall curve comparison: (**a**) YOLOv8s and (**b**) GA-YOLOv8s with EMA.

3.4.2. Comparison of Model Size and Computational Cost Benefits

The integration of Ghost modules and attention mechanisms significantly impacts the model's computational efficiency while maintaining high detection accuracy. As shown in Table 5, the baseline YOLOv8s model has the highest parameter count (9.8 M), memory size (19.9 MB), and FLOPs (23.3 G), indicating higher computational demand.

By incorporating the Ghost module, the parameter count is reduced to 4.6 M, with a memory size of 9.6 MB, and FLOPs decreasing by more than 50% (from 23.3 G to 11.0 G), demonstrating the efficiency of the Ghost-based lightweight design. Additional modifications using attention mechanisms lead to slight variations in complexity. The YOLOv8s + Ghost + CBAM model slightly increases the parameter count (4.9 M) and FLOPs (11.4 G) compared to YOLOv8s + Ghost, while Triplet Attention introduces a more significant increase (5.5 M parameters and 20.6 G FLOPs), reflecting its higher computational cost.

Among all variants, the YOLOv8s + Ghost + EMA model achieves an optimal balance between efficiency and accuracy. Despite a slight increase in parameters (5.5 M) compared to the base Ghost model (4.6 M), it maintains a relatively low memory size (9.4 MB) and achieves the lowest FLOPs (11.3 G) among all attention-enhanced models. These experimental results indicate that EMA effectively enhances feature extraction with

minimal computational overhead, making it a highly efficient solution for real-time rice disease detection.

Table 5. Comparative results on network parameters, memory size, and computation requirements (GFLOPs) for different models.

Model	Params (M)	Size (MB)	FLOPs (G)
YOLOv8s	9.8	19.9	23.3
YOLOv8s + Ghost	4.6	9.6	11.0
YOLOv8s + Ghost + CBAM	4.9	10.1	11.4
YOLOv8s + Ghost + Triplet Attention	5.5	11.4	20.6
YOLOv8s + Ghost + EMA	5.5	9.4	11.3

These results highlight that Ghost modules significantly reduce model complexity, while carefully selected attention mechanisms further improve detection performance with manageable computational costs. Figure 14 depicts the detection results of our proposed Ghost-Attention-YOLOv8 (GA-YOLOv8) model.



Figure 14. Detection results of YOLOv8s + Ghost module + EMA model. The red, green, and blue bounding boxes represent the leaf folder, leaf blast, and brown spot diseases, respectively.

4. Discussion

The integration of Ghost modules has been widely explored in deep learning for balancing computational efficiency and detection accuracy. Han et al. (2020) [30] intro-

duced GhostNet, which demonstrated that redundant feature maps in CNNs could be efficiently generated through lightweight transformations instead of conventional convolutions. This approach significantly reduced computational costs while maintaining strong recognition performance. However, GhostNet was primarily designed for image classification tasks and did not incorporate object detection-specific optimizations or attention mechanisms, limiting its effectiveness in complex real-world applications like agricultural disease detection.

In the field of object detection, Hussein et al. (2023) [33] introduced Ghost-YOLOv8 for traffic sign detection, showing that integrating Ghost modules into YOLO architectures reduces FLOPs and parameters while maintaining high accuracy. However, their study focused on structured urban environments with well-defined objects, which differ significantly from the unstructured, highly variable nature of rice disease symptoms. Similarly, in agricultural applications, Sangaiah et al. (2024) [34] developed the UAV T-YOLO-Rice model, an enhanced Tiny YOLO v4-based detector for rice disease detection. Their model incorporated Ghost modules alongside CBAM, the Sand Clock Feature Extraction Module (SCFEM), and Spatial Pyramid Pooling (SPP) to improve accuracy. While they achieved an mAP of 86% on a custom rice disease dataset, the use of Tiny YOLO v4 presents limitations in feature extraction, multi-scale detection, and overall detection robustness compared to more advanced YOLO architectures.

Our work differs from previous research in that no prior study has explored and investigated the combination of Ghost modules and attention mechanisms (CBAM, Triplet Attention, EMA) in YOLOv8 for specific rice disease detection. Unlike prior works that either focus on reducing computational costs (GhostNet, Ghost-YOLOv8) or enhancing detection accuracy with older YOLO versions (UAV T-YOLO-Rice), our approach achieves a balance between efficiency and feature enhancement. The Ghost module reduces unnecessary computations, while attention mechanisms refine feature extraction, allowing better localization of disease regions in complex rice field environments. Moreover, by leveraging YOLOv8, our model provides state-of-the-art detection performance while remaining lightweight enough for deployment on UAVs and edge devices, making it well suited for real-time rice disease monitoring in Vietnam.

In addition to its lightweight design, the integration of attention mechanisms allows the model to handle variations in disease severity, scale, and appearance more effectively, ensuring reliable performance across diverse agricultural conditions. Our study demonstrates significant improvements in disease detection performance compared to both earlier YOLOv5-based approaches and recent YOLOv8 modifications. First, when comparing to YOLOv5-based models for rice leaf disease detection, our baseline YOLOv8s already provides a substantial enhancement. The YOLOv5 model used in [16] achieved an mAP@50 of 76%, which is significantly lower than our baseline YOLOv8s with mAP@50 of 93.1% and mAP@50–95 of 58.7%. Furthermore, our proposed YOLOv8s + Ghost + EMA model outperforms several enhanced YOLOv8 models applied to the same task. For instance, the Pyramid-YOLOv8 model presented in [32] attained 84.3% mAP@50 for rice leaf blast detection, whereas our model achieves 95.4% mAP@50, an 11.1% improvement. Additionally, our model remains lightweight, with a size of 9.4 MB, whereas their Pyramid-YOLOv8 model has a size of 75.9 MB due to the use of the YOLOv8x backbone, demonstrating that our approach achieves higher accuracy while maintaining efficiency. Compared to [35], the SGDR-enhanced YOLOv8 model achieved 88.6% mAP@50, while our method attains 95.4%, showcasing a 6.8% increase. Similarly, in our previous study [36], the YOLOv8n + Alpha-EIoU model obtained an accuracy of 89.9%, which is 5.5% lower than our current approach. While [36] primarily focused on optimizing loss functions, our modifications incorporate Ghost modules and EMA, enhancing feature extraction while reducing com-

putational complexity. The results highlight the model's potential to assist farmers and agricultural managers in early disease detection and intervention, contributing to improved crop yields and reduced economic losses.

Future work will focus on expanding the scope of the model by incorporating a broader range of crop diseases and larger datasets to enhance generalizability. Additionally, further optimization of the architecture for edge computing devices will be explored to ensure seamless deployment in diverse environments [37,38]. The insights gained from this research underline the transformative role of AI in agriculture, offering a scalable and efficient solution for sustainable farming practices and advancing the field of precision agriculture. By using fewer parameters while maintaining critical feature information, GA-YOLOv8s ensures faster inference times, which is crucial for deployment on edge devices like drones or mobile phones in the field.

5. Conclusions

This study presents the Ghost-Attention-YOLOv8 model, a novel enhancement of the YOLOv8 architecture tailored for the critical task of rice leaf disease detection. By integrating the Ghost module and advanced attention mechanisms—CBAM (Convolutional Block Attention Module), Triplet Attention, and Efficiency Multi-Scale Attention (EMA)—the proposed model addresses key challenges such as computational complexity and the detection of subtle disease patterns. The Ghost module optimizes feature extraction by reducing redundant computations, while the attention mechanisms enable the model to focus on critical features, thereby improving both detection accuracy and robustness in diverse scenarios.

Experimental evaluations on the Rice Leaf Disease dataset demonstrate the effectiveness of the proposed approach. The model achieves significant improvements in detection performance, with an mAP@50 metric of 95.4%, representing a 2.3% increase over the baseline YOLOv8, and an mAP@50–95 of 62.4%, marking a 3.7% improvement. Moreover, the model achieves these results while reducing the parameter count to 5.5 million, a 43% decrease compared to the original YOLOv8s, making it highly suitable for real-time applications. These advancements enable the deployment of the model on resource-constrained edge devices such as drones and mobile phones, supporting practical, field-based agricultural monitoring systems.

Author Contributions: Conceptualization, T.D.B. and T.M.D.L.; methodology, T.D.B. and T.M.D.L.; validation, T.D.B.; formal analysis, T.M.D.L.; investigation, T.D.B.; resources, T.M.D.L.; writing—original draft preparation, T.M.D.L.; writing—review and editing, T.D.B. and T.M.D.L.; visualization, T.M.D.L.; supervision, T.D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors sincerely appreciate the support from the II & IoT Laboratory at the School of Electrical and Electronic Engineering, HUST, as well as the valuable assistance from experts at the Vietnam National University of Agriculture, Vietnam.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
CNN	Convolutional neural network
CBAM	Convolutional Block Attention Module
EMA	Efficiency Multi-Scale Attention
GA-YOLOV8	Ghost-Attention-YOLOv8

References

1. Papademetriou, M.K. Rice production in the Asia-Pacific region: Issues and perspectives. *Bridg. Rice Yield Gap Asia-Pac. Reg.* **2000**, *220*, 4–25.
2. Sebesvari, Z.; Le, T.T.H.; Renaud, F.G. Climate change adaptation and agrichemicals in the Mekong Delta, Vietnam. In *Environmental Change and Agricultural Sustainability in the Mekong Delta*; Springer: Dordrecht, The Netherlands, 2011; pp. 219–239.
3. Nguyen, T.H.; Cassou, E. *An Overview of Agricultural Pollution in Vietnam*; World Bank: Washington, DC, USA, 2017.
4. Phi, A.H. Current Status of Rice Blast in Vietnam and Future Perspectives. *Open Access Libr. J.* **2023**, *10*, 1–17. [CrossRef]
5. Disease and Pest Resistance Rice, International Rice Research Institue. Available online: https://www.irri.org/disease-and-pest-resistant-rice (accessed on 30 January 2019).
6. Hoai Thuong, Actively Prevent Rice Pests, Lai Chau Newspaper. Available online: https://www.vietnam.vn/en/chu-dong-ngan-chan-sau-benh-hai-lua/ (accessed on 13 August 2023).
7. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef]
8. Lee, S.H.; Goëau, H.; Bonnet, P.; Joly, A. New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* **2020**, *170*, 105220. [CrossRef]
9. Ahmad, A.; Saraswat, D.; El Gamal, A. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agric. Technol.* **2023**, *3*, 100083. [CrossRef]
10. Chen, J.; Liu, Q.; Gao, L. Visual tea leaf disease recognition using a convolutional neural net-work model. *Symmetry* **2019**, *11*, 343. [CrossRef]
11. Jiang, F.; Lu, Y.; Chen, Y.; Cai, D.; Li, G. Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Comput. Electron. Agric.* **2020**, *179*, 105824. [CrossRef]
12. Shaji, A.P.; Hemalatha, S. Data augmentation for improving rice leaf disease classification on resid-ual network architecture. In Proceedings of the 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 28–29 January 2022; pp. 1–7.
13. Sudhesh, K.M.; Sowmya, V.; Sainamole Kurian, P.; Sikha, O.K. AI based rice leaf disease identification enhanced by Dynamic Mode Decomposition. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105836.
14. Kiratiratanapruk, K.; Temniranrat, P.; Kitvimonrat, A.; Sinthupinyo, W.; Patarapuwadol, S. Using deep learning techniques to detect rice diseases from images of rice fields. In Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Kitakyushu, Japan, 22–25 September 2020; Springer International Publishing: Cham, Switzerland, 2020; pp. 225–237.
15. Zhuo, W.; Jian, W.; Wang, X.; Jia, S.; Bai, X.; Zhao, Y. Lightweight detection method of apple in natural environment based on improved YOLO v4. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 294–302.
16. Haque, M.E.; Rahman, A.; Junaeid, I.; Hoque, S.U.; Paul, M. Rice leaf disease classification and detection using yolov5. *arXiv* **2022**, arXiv:2209.01579.
17. Soeb, M.J.; Jubayer, M.F.; Tarin, T.A.; Al Mamun, M.R.; Ruhad, F.M.; Parven, A.; Mubarak, N.M.; Karri, S.L.; Meftaul, I.M. Tea leaf disease detection and identification based on YOLOv7 (YO-LO-T). *Sci. Rep.* **2023**, *13*, 6078. [CrossRef] [PubMed]
18. Qadri, S.A.; Huang, N.F.; Wani, T.M.; Bhat, S.A. Plant Disease Detection and Segmentation using End-to-End YOLOv8: A Comprehensive Approach. In Proceedings of the 2023 IEEE 13th International Conference on Control System, Computing and Engi-neering (ICCSCE), Penang, Malaysia, 25–26 August 2023; pp. 155–160.
19. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
20. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
21. Alirezazadeh, P.; Schirrmann, M.; Stolzenburg, F. Improving deep learn-ing-based plant disease classification with attention mechanism. *Gesunde Pflanz.* **2023**, *75*, 49–59. [CrossRef]

22. Ma, B.; Hua, Z.; Wen, Y.; Deng, H.; Zhao, Y.; Pu, L.; Song, H. Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments. *Artif. Intell. Agric.* **2024**, *11*, 70–82. [CrossRef]
23. Chen, Z.; Feng, J.; Yang, Z.; Wang, Y.; Ren, M. YOLOv8-ACCW: Lightweight grape leaf disease detection method based on improved YOLOv8. *IEEE Access* **2024**, *12*, 123595–123608. [CrossRef]
24. Zamri, F.N.; Gunawan, T.S.; Yusoff, S.H.; Alzahrani, A.A.; Bramantoro, A.; Kartiwi, M. Enhanced small drone detection using optimized YOLOv8 with atten-tion mechanisms. *IEEE Access* **2024**, *12*, 90629–90643. [CrossRef]
25. Asif, M.; Kamran, A. Crop stress and its management: Perspectives and strategies. *Crop Sci.* **2012**, *52*, 1968. [CrossRef]
26. 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; pp. 3–19.
27. Chougui, A.; Moussaoui, A.; Moussaoui, A. Plant-leaf diseases classification using cnn, cbam and vision transformer. In Proceedings of the 2022 5th International Symposium on Informatics and Its Applications (ISIA), M'sila, Algeria, 29–30 November 2022; pp. 1–6.
28. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to attend: Convolutional triplet attention module. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 3139–3148.
29. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
30. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1580–1589.
31. Huangfu, Z.; Li, S.; Yan, L. Ghost-YOLO v8: An Attention-Guided Enhanced Small Target Detection Algorithm for Floating Litter on Water Surfaces. *Comput. Mater. Contin.* **2024**, *80*, 3713. [CrossRef]
32. Cao, Q.; Zhao, D.; Li, J.; Li, J.; Li, G.; Feng, S.; Xu, T. Pyramid-YOLOv8: A detection algorithm for precise detection of rice leaf blast. *Plant Methods* **2024**, *20*, 149. [CrossRef]
33. Hussein, M.; Zhu, W.X. A real-time ghost machine learning model built on YOLOv8 for traffic road signs detection and classification in Germany. *Multimed. Syst.* **2024**, *30*, 344. [CrossRef]
34. 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**, *11*, 5201–5216. [CrossRef]
35. Thanh, B.D.; Anh, M.T.; Khanh, G.D.; Dong, T.C.; Huong, N.T. SGDR-YOLOv8: Training Method for Rice Diseases Detection Using YOLOv8. In Proceedings of the International Conference on Advances in Computing and Data Sciences, Velizy, France, 9–10 May 2024; Springer Nature: Cham, Switzerland, 2024; pp. 170–180.
36. Trinh, D.C.; Mac, A.T.; Dang, K.G.; Nguyen, H.T.; Nguyen, H.T.; Bui, T.D. Alpha-EIOU-YOLOv8: An improved algorithm for rice leaf disease detection. *AgriEngineering* **2024**, *6*, 302–317. [CrossRef]
37. Zhang, X.; Cao, Z.; Dong, W. Overview of edge computing in the agricultural internet of things: Key technologies, applications, challenges. *IEEE Access* **2020**, *8*, 141748–141761. [CrossRef]
38. Quach, L.D.; Quoc, K.N.; Quynh, A.N.; Ngoc, H.T.; Thai-Nghe, N. Tomato health monitoring system: Tomato classification, detection, and counting system based on YOLOv8 model with explainable MobileNet models using Grad-CAM++. *IEEE Access* **2024**, *12*, 9719–9737. [CrossRef]

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.