

Article

Not peer-reviewed version

DM-YOLOv8: Improved Cucumber Disease and Insect Detection Model Based on YOLOV8

Ji-Yuan Ding , [Wang-su Jeon](#) , [Sang-Yong Rhee](#) ^{*} , Chang-Man Zou

Posted Date: 20 March 2024

doi: 10.20944/preprints202403.1227.v1

Keywords: component; yolov8s; MultiCat; C2fe; ADC2f; Pest detection



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

DM-YOLOv8: Improved Cucumber Disease and Insect Detection Model Based on YOLOV8

Ji-Yuan Ding ¹, Wang-Su Jeon ² and Sang-Yong Rhee ^{2,*}, Chang-Man Zou ²

¹ Department of IT Convergence Engineering, University of Kyungnam, Changwon 51767, Republic of Korea; jyding100@gmail.com; zouchangman@beihua.edu.cn

² Department of Computer Engineering, University of Kyungnam, Changwon 51767, Republic of Korea; jws2218@naver.com

* Correspondence: syrhee@kyungnam.ac.kr

Abstract: In light of the prevalent pest and disease issues faced by greenhouse cucumbers, a staple vegetable during winter, this study introduces a detection method based on the enhanced YOLOv8s model. This method aims to provide technical support for detecting and classifying pests and diseases in cucumber agricultural production. The model integrates the 'MultiCat' module for multiscale feature fusion and employs the 'C2fe' and 'ADC2f' modules to strengthen spatial and channel attention. The 'Block2d' function also facilitates the choice between average pooling and attention-based spatial pooling. Channel fusion is achieved through additive and multiplicative operations, allowing the model to delve deeper into feature learning. Experimental results confirm that our approach outperforms the original YOLOv8s model in pest detection, particularly excelling in the identification of small-scale and overlapping afflictions.

Keywords: component; yolov8s; MultiCat; C2fe; ADC2f; Pest detection

1. Introduction

In today's globalized world, ensuring food safety and high yields has become a central issue in agricultural research. Cucumber is a member of the Cucurbitaceae family, which has 90 genera and 750 species. It is one of the oldest cultivated vegetable crops grown in almost all temperate countries. It is a warm-loving, frost-tolerant plant that grows best at temperatures above 20°C [1]. Cucumber pests and diseases are one of the main reasons for the decrease in cucumber yield [2]. Outbreaks of agricultural pests not only affect crop production but also the use of pesticides, affecting crop production and increasing both ecological destruction and food safety risks [3]. Therefore, it is particularly important to develop a method that can detect cucumber pests and diseases in a timely and accurate manner.

Traditional methods of detecting pests and diseases on cucumber leaves mainly rely on human visual recognition, that is, directly observing the morphology, texture, and color of leaves with the naked eye [4]. Although this method is simple to operate, its accuracy is affected by the differences in the observer's experience and knowledge, and it is highly subjective, often leading to misdiagnosis and causing irreversible losses to farmers. Therefore, in addition to being time-consuming and costly, it is difficult to achieve the precise detection requirements for cucumber leaf pests. This is not easy to implement in large-scale agricultural production.

With the rapid development of technology, especially in the field of computer vision, deep learning provides a new direction for the identification and detection of pests and diseases in the agricultural field [5]. Against this background, object detection has become a core topic in computer vision research, aiming to accurately identify and locate specific objects in images. Scholars have developed many innovative strategies to achieve this goal. Among them, the one-stage (One-stage) and two-stage (Two-stage) methods stand out and have become the two mainstream technologies in this field.

Representatives of one-stage object detection algorithms include SSD (Liu et al., 2016) [6], Retina Net (Lin et al., 2017) [7], YOLOv4 (Bochkovskiy et al., 2020) [8], YOLOv5 (Jocher et al., 2021) [9], DETR (Carion et al., 2020) [10], FCOS (Tian et al., 2019) [11], and YOLOX (Ge et al., 2021) [12]. In contrast, two-stage object detection algorithms such as R-CNN (Girshick et al., 2014) [13], Fast R-CNN (Girshick, 2015) [14], Faster-RCNN (Ren et al., 2016) [15], Mask R-CNN (He et al., 2017) [16], and Cascade R-CNN (Cai and Vasconcelos, 2018) [17] have a longer computational process.

Two-stage models first generate a series of candidate regions and then use a classifier to refine the classification of these regions. Although they are usually superior in accuracy, the two-step process makes them relatively slow [18,19]. On the other hand, one-stage models directly predict bounding boxes and categories from feature maps without generating candidate areas, providing the advantage of real-time detection. However, this speed sometimes comes at the expense of accuracy. Considering the need to make the model more practical and suitable for deployment on mobile devices, we chose the YOLO model as our object detection algorithm.

To solve the problem of low accuracy in small target detection of one-stage models, this paper proposes a new YOLOv8s model for cucumber pest identification. This study tests the new model on a specially constructed dataset. This study plans to adopt and improve the latest YOLOv8s model for the detection of cucumber pests and diseases. To compensate for the accuracy loss of lightweight models, we also adopted an attention mechanism to assign different weights to each part of the input feature layer, thereby more effectively extracting key features and improving classification performance.

The main contributions of this paper are summarized fellow:

- We propose a lightweight one-stage YOLOv8 model, referred to as the Detail and Multi-scale YOLO Network (DM-YOLO), built upon YOLOv8 for real-time cucumber pest and disease identification. Utilizing the MultiCat module by merging features of different scales, the model's detection capability for pests and diseases of varying sizes on cucumbers is enhanced. We introduced the C2fe module, a modification based on C2f, as a new feature fusion method aiming to more effectively combine multi-scale features. With an attention mechanism based on adaptive average pooling, we constructed a new module named AD-C2f, which intensifies the model's focus on crucial features, thereby increasing detection accuracy and overall model performance.
- We extracted a portion of data from the ai-hub's "Integrated Plant Disease Induction Data" public dataset to construct a new cucumber pest and disease dataset. To ensure data quality, we manually re-annotated the leaves in each image and carefully filtered the original images. After eliminating some unorganized data, we concentrated on two primary cucumber afflictions: downy mildew and powdery mildew, culminating in this optimized new dataset.

2. Related Works

2.1. Traditional Machine Learning Methods

Traditional disease identification methods in machine learning mainly rely on extracting manually designed features from images, such as color, texture, and shape. They then utilize machine learning classifiers, such as Support Vector Machines (SVM), Decision Trees, or K-Nearest Neighbors (KNN), to differentiate between healthy and damaged plants.

Ebrahimi et al. (2017) proposed an insect detection system based on SVM, which was successfully applied to the crop canopy images in strawberry greenhouses, achieving an error rate of less than 2.25% [20].

Mondal et al. (2017) developed a disease recognition system that combined image processing and soft computing techniques. By selecting specific morphological features, they achieved a high-accuracy classification of diseases in okra and bitter melon leaves [21].

Xu et al. (2020) employed BP neural networks and random forest models to detect damages caused by forest caterpillars. The random forest model performed better, emphasizing the importance of balanced sample data [22].

Amirruddin et al. (2020) evaluated the chlorophyll sufficiency levels of mature oil palms using hyperspectral remote sensing technology and classification methods. They achieved high-accuracy chlorophyll classification through a random forest classifier, especially on younger leaves [23].

Traditional machine learning methods might fail to capture complex and advanced patterns in high-dimensional and large-scale data, leading to subpar model performance and accuracy. Moreover, these methods often require extensive preprocessing and feature engineering, adding to the development time and cost. Compared to the adaptive learning capability of deep learning, traditional methods might perform poorly when facing data changes and uncertainties. In summary, although traditional machine learning methods may work well in certain situations, they may not be suitable for tasks requiring capturing complex patterns and relationships in high-dimensional and large-scale data.

2.1. Deep Learning Methods

Deep learning has shown significant advantages in the identification and classification of crop pests and diseases. Unlike traditional machine learning methods, deep learning can automatically learn and extract features from data, alleviating the burden of manual feature design and handling more complex and high-dimensional data. Some advanced deep learning models, like variants of Convolutional Neural Networks and pre-trained models, have been successfully applied to various crop pest and disease identification tasks.

Sethy et al. (2020) successfully identified four rice leaf diseases by combining deep convolutional neural networks and SVM, demonstrating that integrating deep features with SVM can achieve excellent classification, achieving an F1 score of 0.9838 [24].

Yin et al. (2022) successfully developed a grape leaf disease identification method using an improved MobileNetV3 model and deep transfer learning. This method achieved recognition accuracy of up to 99.84% with limited computational resources and dataset size, and the model size was only 30 MB [25].

Sankarshwaran et al. (2023) proposed a new method named Cross Enhanced Artificial Hummingbird Algorithm based on AX-RetinaNet (CAHA-AXRNet) for optimizing rice plant disease detection. This method proved more effective than other existing rice plant disease detection methods, achieving an accuracy of 98.1% [26].

Liu Shiyi et al. (2023) introduced a deep learning algorithm called DCNSE-YOLOv7 based on an improved YOLOv7 algorithm. This significantly enhanced the accuracy of the detection of cucumber leaf pests and diseases, especially in detecting minute features on early-stage diseased leaves. This algorithm showed significant improvement over several mainstream object detection models, providing effective technical support for the precise detection of cucumber leaf pests and diseases [27].

Yang et al. (2023) introduced a tomato automatic detection method based on an improved YOLOv8s model. This method achieved an mAP of 93.4%, meeting the requirements of real-time detection and providing technical support for tomato-picking robots to ensure fast and accurate operations [28].

These studies indicate that through continuous optimization and improvement, deep learning models can achieve remarkable results in crop pest and disease detection and identification.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, and the experimental conclusions that can be drawn.

3. Disease detection overview

3.1. datasets

The data required for this study was sourced from ai-hub's "Integrated Data on Plant Disease Induction" public dataset (Figure 1.). To construct a dataset specifically targeting cucumber pests and diseases, we extracted a subset of data from this collection and deeply integrated it. In order to ensure the model's broad adaptability and robustness, we manually annotated each leaf in the images

and meticulously filtered the original photos. After removing some irregular data, our primary focus was on the two prevalent ailments of cucumbers: downy mildew and powdery mildew. We selected 2,000 images of cucumber pests and diseases to form a new dataset.



Figure 1. Integrated data on plant disease induction.

The entire dataset was split in an 8:2 ratio for training and validation sets, respectively. The specific data distribution is illustrated in Figure 2.

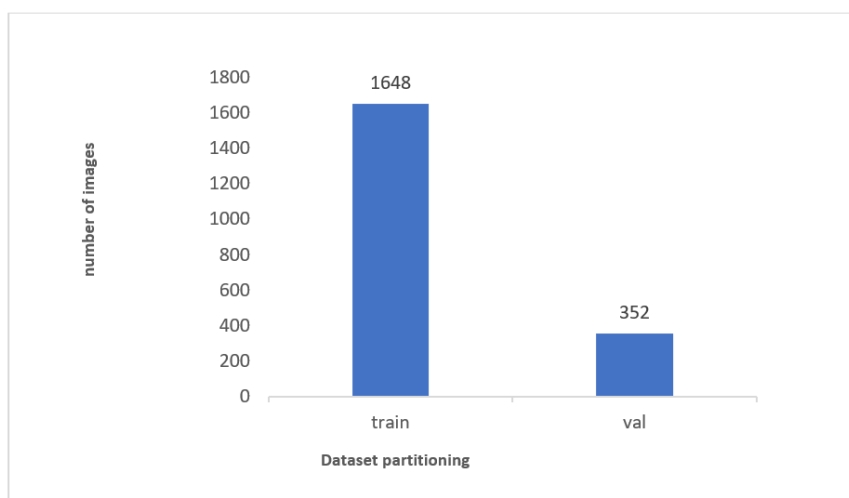


Figure 2. Dataset distribution results.

The dataset is mainly divided into train and val, as shown on the horizontal axis. The vertical axis mainly represents the number of images in each dataset.

3.2. Standard YOLOv8

The YOLOv8 model, a continuation of the YOLO series, sets a new benchmark in state-of-the-art object detection technologies. It integrates numerous advancements and optimizations, boosting its performance and flexibility. Key to its design is the incorporation of varied resolution and scale detection networks, tailored to meet the requirements of diverse use cases. In terms of structure, YOLOv8 has been meticulously refined from its backbone to the head, featuring a gradient flow-enhanced C2f structure and a novel decoupled head design. This results in more autonomous and

effective classification and detection processes. A significant shift in the model's architecture is its move from an Anchor-Based to an Anchor-Free framework, reflecting recent trends in the field. The loss function in YOLOv8 is a combination of the Task-Aligned Assigner and the Distribution Focal Loss, innovative approaches that enable finer detection accuracy across a range of targets. Additionally, YOLOv8 employs strategic data augmentation techniques, notably the omission of Mosaic augmentation in the latter stages of training. The architecture of YOLOv8 is depicted in Figure 3, showcasing its sophisticated design.

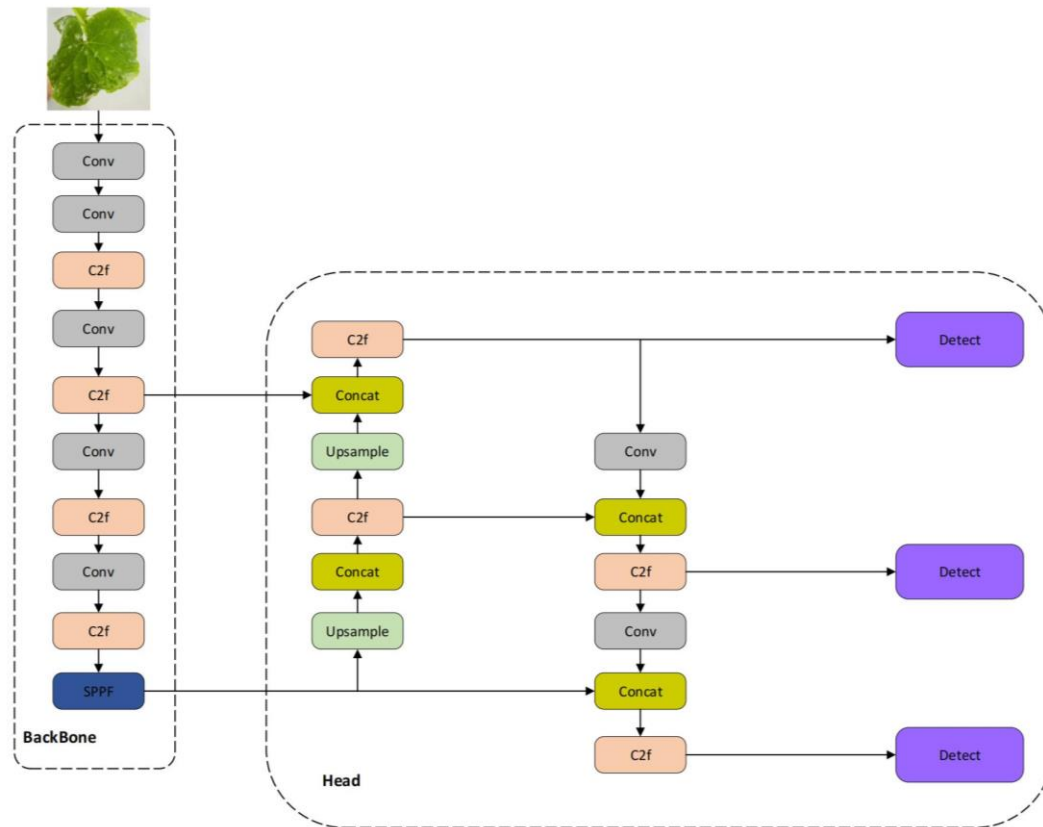


Figure 3. Yolov8 model.

3.3. Proposed YOLOv8

To accommodate the detection of small targets, we proposed and refined the YOLOv8 model as depicted in Figure 4, where one of the c2f modules was altered to an AD-C2f module. A multi-scale feature fusion module, termed MultiCat module, was constructed. By extracting and merging information from three distinct levels of the backbone, the model is capable of capturing and utilizing features from multiple scales and levels. Further optimization and extension were conducted on the c2f basis, forming a C2fe module. Besides conducting feature extraction in the channel dimension, this module also performs in-depth feature processing spatially. This dual operation mode on two dimensions ensures a comprehensive and profound feature extraction by the model.

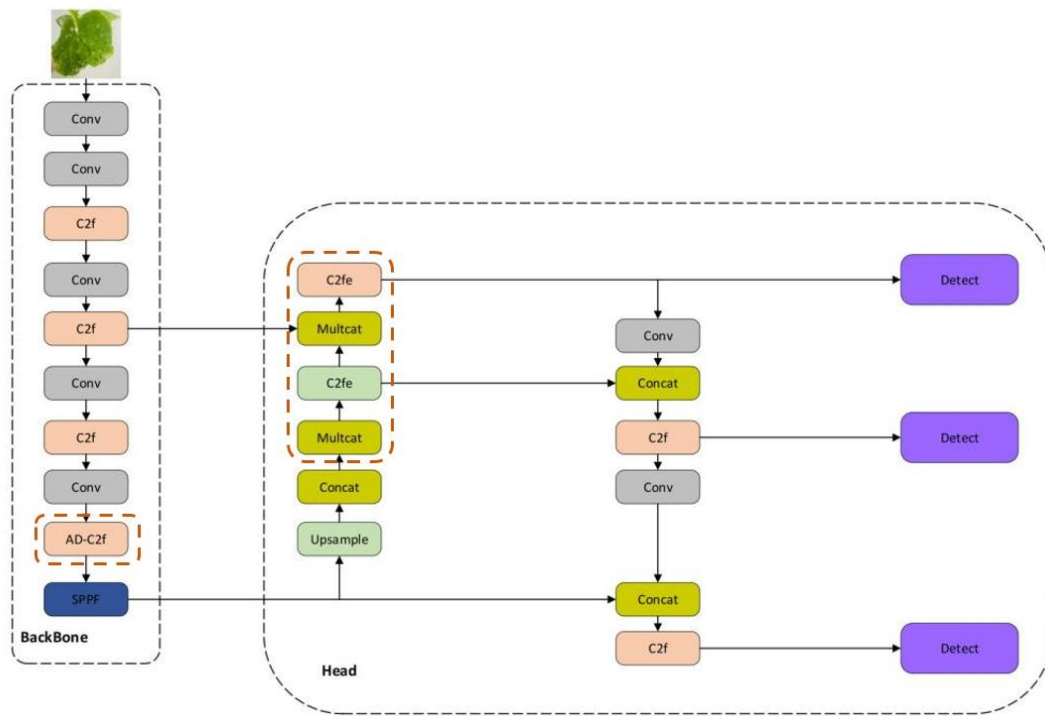


Figure 4. Improved YOLOv8 model.

3.3.1. MultiCat Module

In practical agricultural monitoring, the detection of cucumber diseases and pests faces numerous challenges such as varying lighting conditions, object occlusions, and interference with other diseases or pests. Noise present in images (for instance, background clutter or other irrelevant elements) may undermine the accuracy of disease detection. To address these challenges, our improved model, targeting the cucumber diseases and pests detection problem, specially designed a module named MultiCat (detailed in Figure 5). This module, through meticulous multi-scale feature processing and fusion strategy, aims to favor the model in detecting smaller target diseases and pests on cucumber leaves and is able to integrate rich feature information from the backbone.

Within the MultiCat module, the input feature maps are initially decomposed into three different scaled feature maps, namely L (large scale), M (medium scale), and S (small scale). This decomposition strategy assists the model in capturing disease features on cucumber leaves from various angles and levels. Following this, we apply adaptive max pooling and adaptive average pooling to the large-scale feature map L, to explore disease features both globally and regionally. Simultaneously, we employ nearest neighbor interpolation to upscale the small-scale feature map S, retaining the original data values while altering the feature map dimensions, ensuring all feature maps have the same spatial dimensions. This step aims to preserve the rich local disease feature details. Finally, through channel concatenation operation, we fuse the three scaled feature maps into a unified feature map and return it as the output.

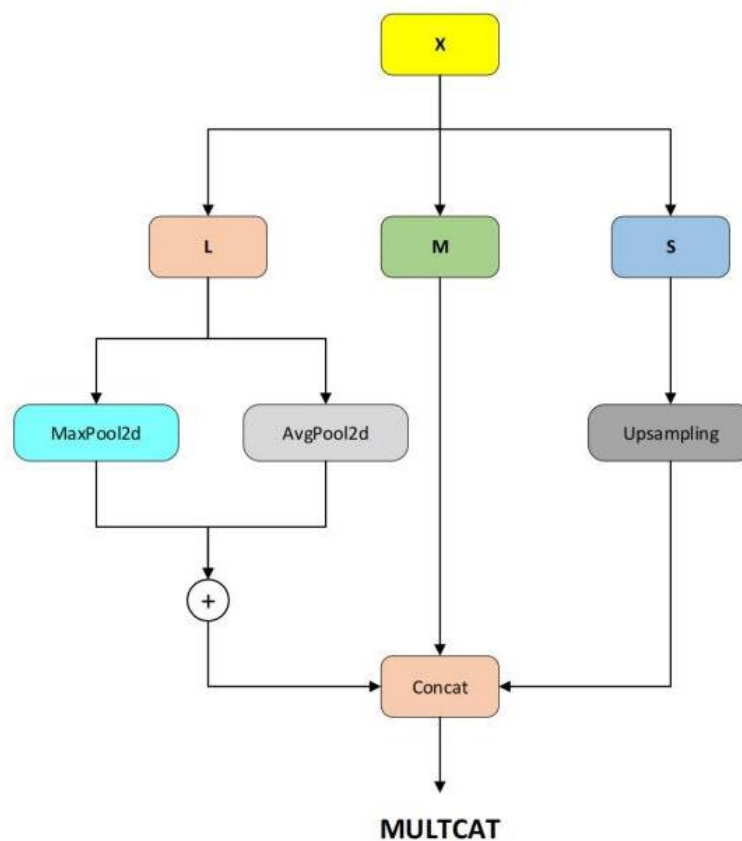


Figure 5. MultiCat Module.

3.3.2. AD-C2f Module

During the cultivation of cucumbers, downy mildew and powdery mildew are two common diseases that significantly impact the growth and yield of the crops. Downy mildew typically forms a white or greyish mold layer on the leaves, while powdery mildew forms a white powdery coating on them. Both diseases exhibit distinct visual characteristics, yet their detection can be challenging due to factors like lighting and occlusions. To address this, our model integrates attention mechanisms and depthwise separable convolution, designing the ADC2F module as illustrated in Figure 6(a), providing robust technical support for accurate identification and localization of cucumber diseases and pests.

The attention mechanism is a technique enabling the model to focus on crucial parts while processing data. It allows the model to assign different "attention" weights to different areas or elements when handling images, texts, or other types of data, selectively focusing on information crucial for the task at hand. For instance, when processing images of cucumber downy mildew and powdery mildew, the attention mechanism helps the model focus on infected areas while ignoring irrelevant background noise. This mechanism has achieved remarkable results in many deep learning applications, including image and video processing, natural language processing, and recommendation systems.

As shown in Figure 6, the structure of the ADc2f module initially decomposes the input feature map into two parts with the same channels ('a' and 'b'). Then, both parts are separately passed through an adaptive average pooling layer and two convolution layers, using a sigmoid activation function to obtain weights in the 'a' 'b' channel directions. The aim is to adaptively filter out more useful features while suppressing useless or interfering features.

Its formula is as follows:

$$\text{attention}_i = \sigma(\text{fc}(\text{pool}(x_i))) \quad (1)$$

In this context, σ represents the Sigmoid function, fc denotes the convolution layer, pool stands for the adaptive average pooling layer, and x_i is a part of the input feature map ('a' or 'b').

By applying attention weights, the module can adjust each part of the input feature map, enabling the model to better focus on the important areas relevant to the task at hand. This process can be represented by the following equation:

$$\text{adjusted}_{i,j} = x_{i,j} * \text{attention}_{i,j} \quad (2)$$

In this manner, the 'AD-C2f' module can leverage the attention mechanism to filter out crucial feature information, thereby enhancing the model's accuracy in identifying and locating cucumber downy mildew and powdery mildew.

As shown in Figure 6, the structure of the ADc2f module, the ConvBlock structure depicted in Figure 6(b) employs depthwise separable convolution to achieve efficient feature extraction. Depthwise separable convolution is a convolutional structure that decomposes traditional convolution operations into two smaller operations (depthwise convolution and pointwise convolution). This decomposition significantly reduces computational load and model parameters while maintaining or even improving the model's performance.

In depthwise separable convolution:

$$\text{out}_{i,j,k} = \text{in}_{i,j,k} * \text{kernel}_{i,j,k} \quad (3)$$

(where (i,j) and k represent the indices for channels, height, and width respectively, and * denotes the convolution operation)

Pointwise Convolution, on the other hand, is essentially a standard convolution operation, with a kernel size of 1×1. It serves to fuse together the feature maps obtained from the depthwise convolution across various channels, generating a new feature map.

$$\text{out}_{i,j} = \text{in}_{i,j,k} * \text{kernel}_{i,j,k} \quad (4)$$

In this context, i, j and k also represent the indices for channels, height, and width respectively, while * denotes the convolution operation.

Through this approach, the ConvBlock module can maintain good feature extraction capability while reducing computation and parameters, providing robust support for the model to capture complex image patterns associated with pest infestation and diseases. This design allows the ADC2f module to accurately identify and locate cucumber pests and diseases with a lower computational burden.

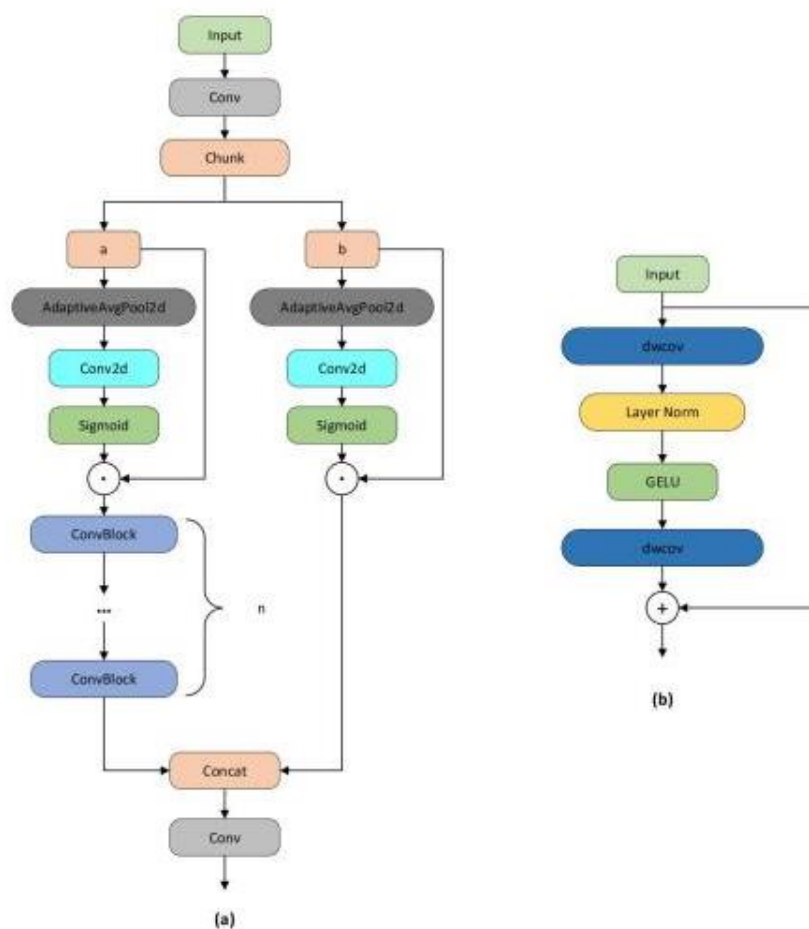


Figure 6. AD-C2f Module.

3.3.2. C2fe (Enhanced) Module

The common diseases during cucumber cultivation, downy mildew and powdery mildew, have a severe impact on the yield. The identification and localization of these diseases are challenging tasks due to their distinct visual characteristics being largely affected by environmental factors. To address this issue, we propose the optimized C2fe module as shown in Figure 7(a). The C2fe module is specifically designed for precise feature extraction, aiming to further enhance the network's feature extraction capability, ensuring improved performance in cucumber disease and pest identification tasks.

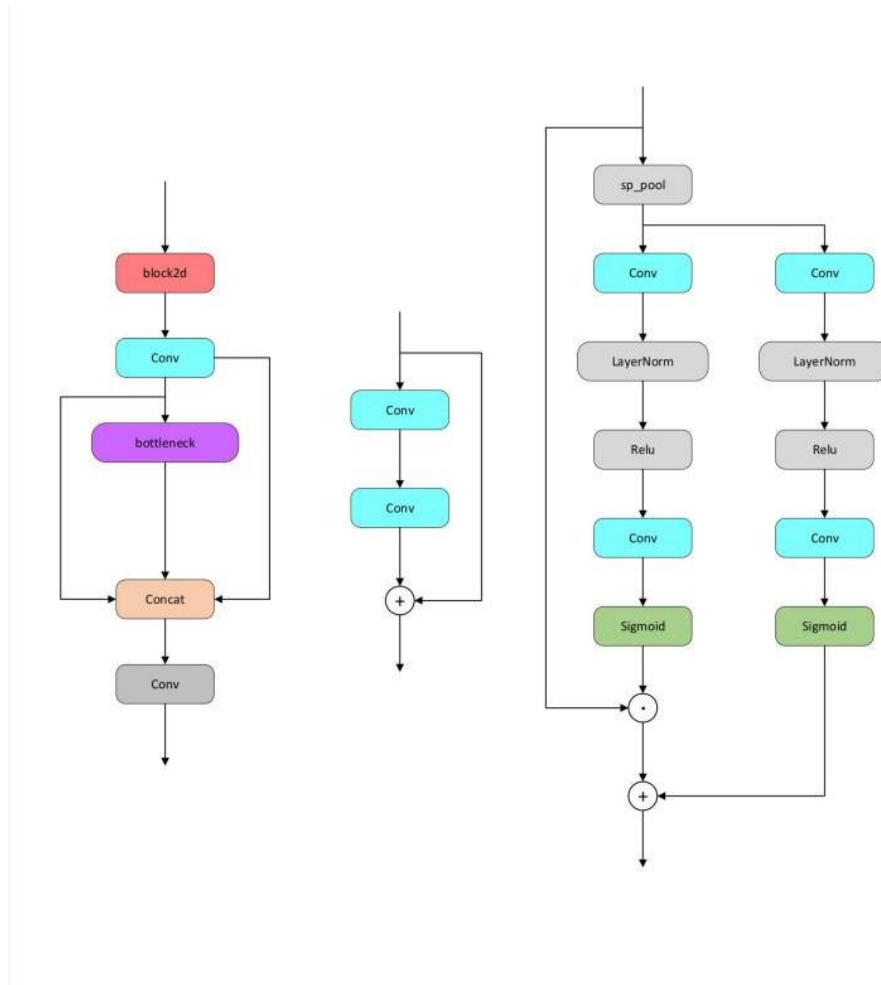


Figure 7. C2fe (Enhanced) Module.

The essence of the C2fe module lies in the integration of the Block2d component (Figure 7(c)), which can accurately pinpoint the core areas of the input feature map through spatial attention mechanisms. This allows the network to concentrate more on the parts with high information density, significantly enhancing the expressiveness of the features.

Moreover, the C2fe module demonstrates a micro-adjustment capability on features, mainly attributed to the introduction of channel_add and channel_mul strategies. These strategies not only make the model more adept at capturing the inter-channel correlations but also deepen the model's representation of features. Through this channel-level fusion, the C2fe module is better at capturing and expressing the correlations among different channels, which is crucial for improving the model's identification and localization accuracy.

In terms of implementation, the C2fe module internally integrates a series of convolution layers and bottleneck layers (Figure 7(b)), working in harmony to further extract and fuse features. The design of these layers aims to capture multi-level and multi-scale information in the input images, enabling the network to obtain a richer and more accurate feature representation.

By introducing the C2fe module, our network architecture becomes more robust and multifunctional, exhibiting outstanding performance in both downy mildew identification and powdery mildew localization tasks.

4. Experiments

4.1. Equipment and Parameter Settings

The experimental operating system used in this study is Windows 10, with PyTorch serving as the framework for developing the deep learning model. The Table 1 is provides specific details regarding the experimental environment.

Table 1. Experimental environment configuration.

Parameters	Configuration
CPU	AMD Ryzen 5 3600 6-Core Processor 3.60 GHz
GPU	NVIDIA GeForce RTX 3060
GPU memory size	32G
Operating systems	Win 10
Deep learning architecture	Pytorch1.9.2 + Cuda11.4 + cudnn 8.2.0

Training Parameter Settings: The image input size is 640×640, batch size is 20, multi-threading is set to 4, the initial learning rate is 0.01, with a total of 120 training iterations (Epochs). The specific parameter settings are shown in Table 2.

Table 2. Training Parameter Configurations.

Parameters	Value
Task	Detect
Epochs	120
Batch size	20
Image Size	640
Optimizer	SGD

4.2. Evaluation Metrics

In cucumber pest and disease detection, to measure the ability and performance evaluation of the YoloV8 model in correctly detecting pests and diseases, several key metrics were primarily selected: accuracy, recall, precision, and mAP.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{mAP} = \frac{\sum_{i=1}^C AP_i}{C} \quad (7)$$

Among them, TP, FP, FN, and TN respectively represent True Positive, False Positive, False Negative, and True Negative. C represents the total number of classes, and AP_i represents the AP value for class i.

4.3. Experimental Results

We compared the improved model with the original YOLOv8 model to evaluate whether our improvements could enhance the performance of the model. To showcase the detection results of the algorithm proposed in this study, we randomly selected images from the test subset for comparison. The specific comparison results are shown in Table 3, and the visual outcomes of the selected images are illustrated in Figure 8.

Table 3. Performance Comparison Between YOLOv8 and Improved YOLOv8 (DM-YOLOv8) Across Different Classes and Metrics (Annotations: A3 as Downy Mildew, A4 as Powdery Mildew).

Algorithm	CLASS	Precision/%	Recall/%	mAP50
Yolov8	All	83.4%	79.4%	87.4%
	A3	87.5%	77.8%	90.3%
	A4	79.3%	81.0%	84.5%
DM-Yolov8	All	84.2%	81.5%	88.25%
	A3	85.1%	83.0%	91.0%
	A4	83.4%	80.0%	85.5%

The DM-YoloV8 model demonstrates superior performance over the original YoloV8 model in key metrics crucial for cucumber pest and disease detection. With a higher recall of 0.81 versus 0.76, DM-YoloV8 effectively reduces the likelihood of missing actual pest and disease instances. Its MAP50 score of 0.90 in the "A3" category, compared to YoloV8's 0.89, reflects better accuracy and precision. Despite a slightly lower frame rate, the significant improvements in recall and MAP50, along with consistent performance across various categories, underscore DM-YoloV8's enhanced suitability and reliability for cucumber pest and disease detection tasks, making it a preferable choice over the original YoloV8 model.

During the training process, we are concerned not only with the model's final performance but also with the training and validation process to ensure the model is progressing in the right direction. For this purpose, we decided to plot some critical metrics during the training and validation process so we could have a visual understanding of the model's learning situation, as shown in Figure 8.

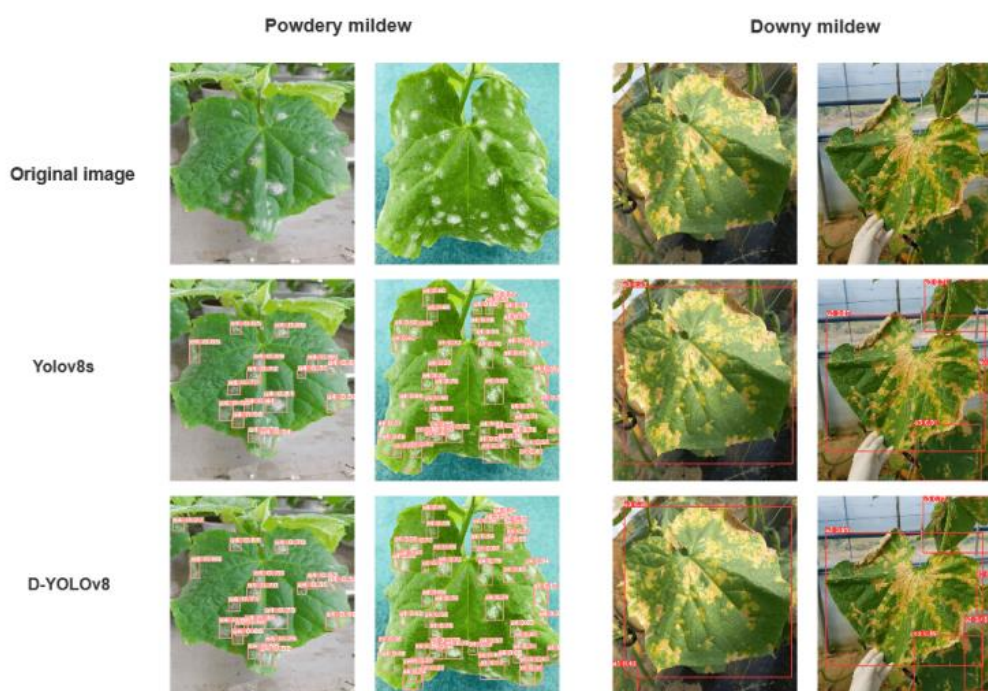


Figure 8. Cucumber Disease Detection and Algorithm Comparison.

In Figure 8, our investigation revealed that the enhanced DM-YOLOv8 model exhibits notable performance enhancements in the detection of foliar diseases, specifically Powdery Mildew and Downy Mildew. When benchmarked against the baseline YOLOv8s model, the DM-YOLOv8 variant demonstrated superior accuracy in bounding box delineation and augmented confidence scores, signaling a refined capability for precise pathogen feature recognition.

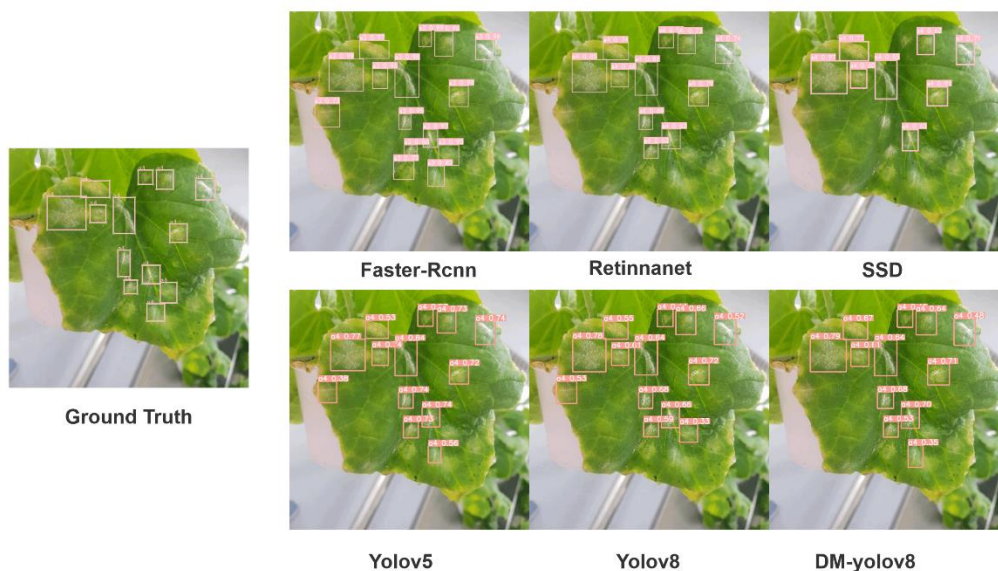
In particular, the DM-YOLOv8 model consistently yielded elevated confidence scores across a multitude of test instances, denoting a heightened proficiency in differentiating healthy leaf tissue from that afflicted by disease. Despite these advancements, the model occasionally generated detection boxes in healthy tissue zones, indicative of potential false positives. Furthermore, there were instances where prominent disease manifestations were not encapsulated within detection boxes, pointing to possible false negatives.

The performance of the DM-YOLOv8 model also varied when processing images characterized by intricate backgrounds and overlapping leaf structures. This variability suggests that the model's robustness in complex visual environments may require additional refinement. Notably, the model exhibited uncertainty in regions of leaf margin and vein convergence, likely attributed to the feature representation similarities between these areas and diseased segments.

Summarily, while the DM-YOLOv8 model demonstrates a distinct advantage in the domain of leaf disease detection, there is an evident imperative for enhancement in minimizing false positives and fortifying detection consistency in multifaceted scenarios. Consequently, this necessitates the development of further optimization strategies to align the model's capabilities with the pragmatic demands of accurate disease detection.

4.4. Comparative Experiments

In our research, we meticulously designed a series of comparative experiments to assess the performance of multiple advanced object detection models in the context of agricultural pest detection tasks. Our study utilized traditional models including Faster R-CNN, Retina Net, SSD, as well as the newer iterations YOLOv5, YOLOv8, and our DM-YOLOv8, testing their proficiency on a dataset of various leaf images with powdery mildew, which served as a realistic simulation of agricultural conditions. The experimental outcomes, illustrated alongside the Ground Truth, not only highlighted the differences in detection accuracy and recall rates among the models but also afforded us insightful revelations into their processing capabilities.



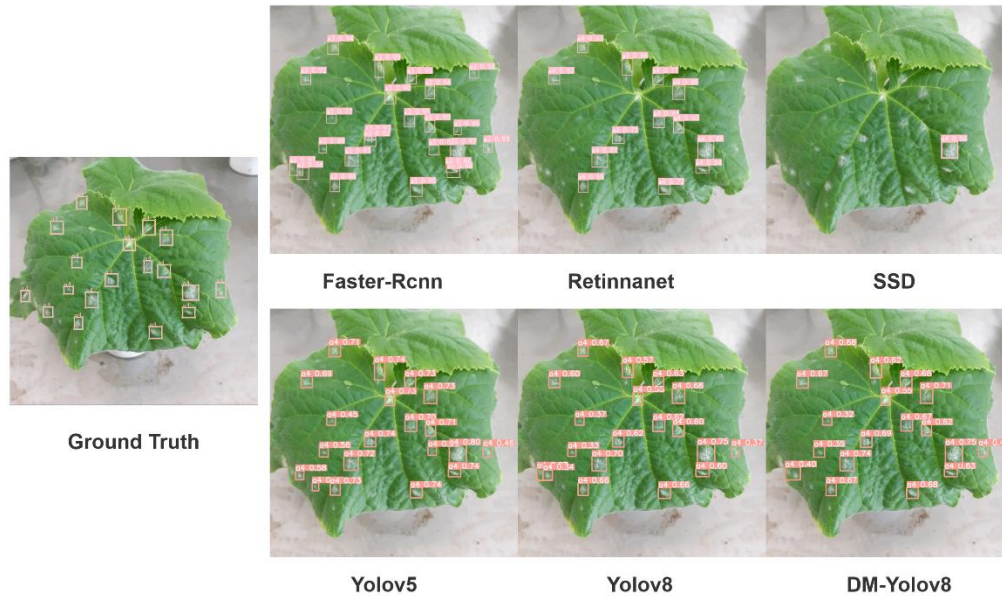


Figure 9. Comparative Analysis of Object Detection Models on Leaf Disease Identification.

We observed that Faster R-CNN excelled in detecting a significant number of lesions, indicating its potential for high-confidence detection. However, this model also exhibited a relatively high false positive rate, notably underperforming in the detection of small-scale lesions. This specific limitation could lead to significant errors in sensitive pest detection tasks where precision is of the utmost importance. In contrast, RetinaNet produced fewer detection boxes than Faster R-CNN. However, it was significantly more focused and precise in accurately delineating the actual lesions, albeit with a problem of missed detections. The SSD model, while demonstrating rapid detection capabilities in specific contexts, showed a marked decrease in performance when tasked with detecting more diminutive and more nuanced features. This decrease in performance could become a hindrance in agricultural applications where the accurate identification of subtle lesions is crucial.

The latest iterations of the YOLO series, namely YOLOv5 and YOLOv8, showcased a relatively high number of detection boxes and improved precision. YOLOv8, while reducing false positives, maintained a high number of detection boxes, further enhancing the accuracy of the model. As an improved version, DM-YOLOv8's results closely matched the Ground Truth, showing significant improvements in precision and recall rates for the identification of small-scale targets.

DM-YOLOv8 stood out in the precise localization of multiple targets, offering a viable solution for agricultural applications requiring high-precision real-time detection with limited computational resources. Nevertheless, the experimental results underscore the necessity for further research to optimize the models' overall performance and validate their robustness under a broader range of real-world conditions.

When conducting a comprehensive comparative analysis of these models, data from Table 4 further corroborated the significant advantage of DM-YOLOv8 in real-time processing capabilities. With a mean Average Precision (mAP) of 88.1%, it surpassed YOLOv5's 87.6%, and its frame rate (FPS) increased significantly to 178.57, much higher than YOLOv5's 153.84, which is particularly crucial for time-sensitive real-time pest monitoring applications. DM-YOLOv8 not only reached new heights in accuracy and speed but also showed significant advantages in model size. Its smaller size, especially compared to the larger model size of Faster R-CNN, makes it an ideal choice for environments with limited computational resources.

Table 4. Comparative Experiment.

Network	Precision/%	Recall/%	mAP50	Pram (MB)	FPS
Yolov5	84.2%	81.0%	87.6%	26.76MB	153.84
Yolov8	83.4%	79.4%	87.4%	11.47MB	181.8
Retinanet	90.63%	60.40%	84.17%	144.84 MB	25.93
SSD	85.16%	30.77%	56.45%	100.27MB	65.8
Faster -Rcnn	43.1%	86.58%	86.58%	108MB	11.94
DM-yolov8	84.2%	80.8%	88.2%	12.2MB	178.57

Through these comparative experiments, the DM-YOLOv8 model demonstrated tremendous potential for real-time agricultural pest detection applications due to its advantages in accuracy, speed, and robust performance in tasks involving small-scale and overlapping targets. While these preliminary results are encouraging, we recognize the need for further validation on a larger scale dataset and under more complex environmental conditions to ensure the efficacy and reliability of the DM-YOLOv8 model in practical agricultural applications.

4.5. Ablation Experiment and result

To systematically discern the contribution of each component in our enhanced YOLOv8 model, we conducted an ablation study. This study aimed to isolate and assess the individual and combined impacts of the C2fe, AD-C2f, and MULT modules on the model's performance. We evaluated the variants of the model on the same dataset to ensure consistency in comparisons. This was achieved by employing uniform training and validation splits, maintaining consistency across all hyperparameters and training conditions. Such methodological rigor ensured that the observed performance variations were exclusively attributable to architectural changes.

This approach allowed for a nuanced understanding of how each module influences the model's effectiveness, particularly in terms of detection accuracy, processing speed, and the ability to handle diverse image conditions. The systematic analysis provided insights into the synergistic effects of the modules when used in combination, offering valuable information on optimizing the model for specific use cases or computational constraints.

Table 5. Ablation experiment results.

Network	Precision/%	Recall/%	mAP50	mAP50-95	FPS
YOLOV8s	83.4%	79.4%	87.5%	52.0%	181.8
DM-YOLOV8	84.2%	80.8%	88.2%	53.0%	178.5
C2fe	84.2%	78.7%	88.1%	53.0%	156.3
ADC2F	82.3%	79.8%	87.8%	51.8%	192.3
MULTCAT	83.7%	79.4%	88.1%	52.9%	172.4
C2fe+ADC2F	83.2%	78.2%	87.4%	52.6%	178.6
C2fe+MULTCAT	83.8%	80.3%	88.0%	53.1%	172.4
ADC2+MULTCAT	83.7%	79.0%	87.5%	52.4%	178.6

As a baseline model, YOLOv8s provided a solid foundation with a precision (P) of 83.4% and a recall (R) of 79.4%. With an IoU threshold of 0.5, its mean Average Precision (mAP50) was 87.5%, while mAP50-95 stood at 52%. The model processed images at a speed of 181.8 FPS, establishing a high-performance benchmark for real-time detection.

Upon the integration of our enhanced modules (DM-YOLOv8), enhancements in both precision and recall were observed. Notably, mAP50-95 increased to 53%, indicating more consistent performance across diverse IoU thresholds. This improvement likely resulted from the addition of the MultiCat module, which enhances the model's ability to detect features at varying resolutions through adaptive pooling and interpolation operations on different-scale feature maps. However,

this gain in precision was achieved at the expense of reduced processing speed, with FPS decreasing to 178.5.

The AD-C2f module, utilizing channel-wise separation, adaptive pooling, and a Sigmoid activation function, established an attention mechanism. This method of processing channels helps emphasize relevant features while suppressing less important ones, which may explain the increase in recall noted during ablation studies. This module's efficiency in computation is also mirrored in its high FPS of 192.3, consistent with the model's performance when the AD-C2f module is employed.

The Block2d module provides spatial attention and channel feature enhancement. Its operation involving channel multiplication and addition after spatial pooling may assist in improving the model's precision. However, this increased complexity might also be the reason for the decrease in FPS when implemented in configurations.

The C2fe module merges Block2d's spatial attention with a series of bottleneck layers for feature processing. It seems to be a more intricate and robust feature extractor, leading to higher scores in mAP50-95 for the enhanced model. However, this added complexity results in a decrease in FPS.

When these modules are combined (e.g., C2fe+AD-C2f, C2fe+MULT, AD-C2f+MULT), the interaction between attention mechanisms and feature processing strategies may either complement or counteract each other, leading to observed shifts in precision, recall, mAP50, mAP50-95, and FPS. Careful balancing of these module combinations is crucial to optimize the model's overall performance.

Our ablation study reveals that the MultiCat and AD-C2f modules are crucial for improving the YOLOv8 model's detection capabilities while maintaining high computational efficiency. Conversely, the Block2d and C2fe modules contribute to increased precision and mAP50-95 but at the expense of higher computational costs. The real-world effectiveness of these modules will hinge on their impact on the precision-speed trade-off, a critical aspect in real-time application deployments.

The Figure 10. presents a visual analysis using Grad-CAM heatmaps to identify two prevalent plant diseases: Powdery Mildew (in the upper images) and Downy Mildew (in the lower images). The first column showcases original leaf images with distinct disease symptoms. Subsequent columns display Grad-CAM heatmaps, where we have specifically selected the C2fe and C2f layers of the network for comparative observation to highlight the network's focal points on features.

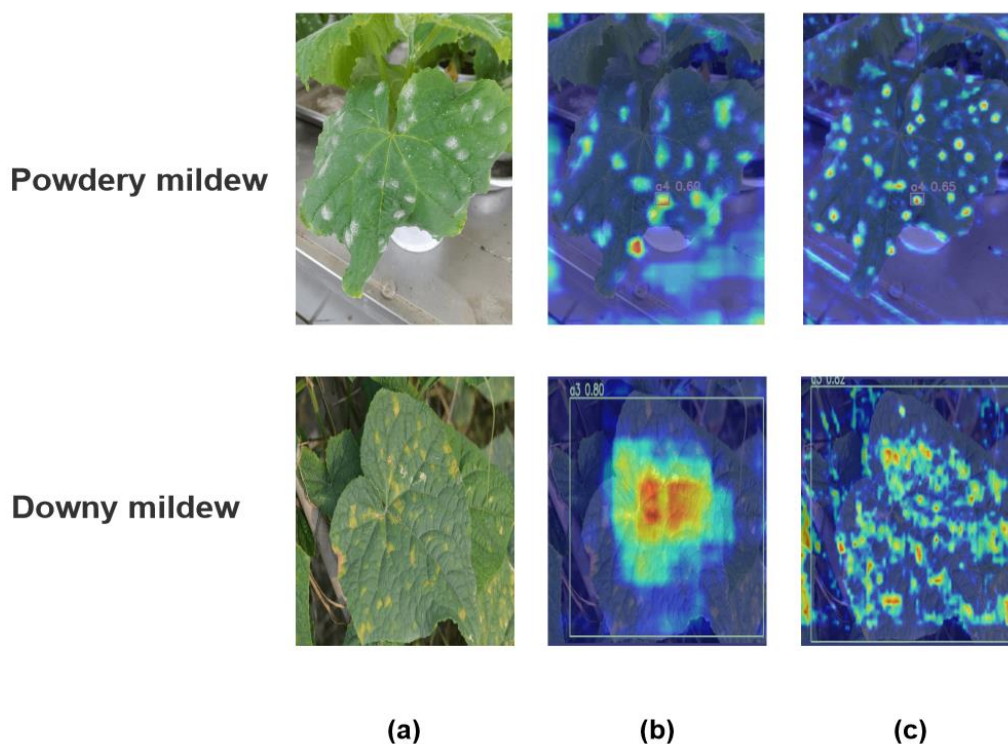


Figure 10. Visual comparison of Grad-CAM outputs. (a) Original images of leaves with Powdery mildew and Downy mildew. (b) C2F heatmaps indicating model-identified areas. (c) C2FE heatmaps show expanded affected regions.

In Figure 11. Grad-CAM images, the integration of the C2fe layer exhibits a more concentrated activation pattern compared to the C2f layer, particularly around disease-specific features. This indicates that the C2FE module contributes to refining the convolutional neural network's focus on key attributes associated with each disease, potentially enhancing diagnostic accuracy. The heatmaps provide visual evidence of the network's ability to localize the most informative features for disease detection, demonstrating the effectiveness of the C2fe layer in honing the model's discriminative power for disease-specific characteristics.

5. Conclusions

This study delved into the detection of diseases and pests in cucumbers within greenhouse environments, successfully proposing an optimized YOLOv8 algorithm tailored for this purpose. Despite the reduced detection accuracy due to interference from complex backgrounds, the introduction of specific modules significantly enhanced the algorithm's feature extraction and representation capabilities. The integration of the C2fe and AD-C2f modules, in particular, collectively strengthened the network's feature extraction prowess, markedly boosting the model's detection capabilities. Additionally, in cucumber disease and pest detection tasks, this model demonstrated a higher recall rate and mean Average Precision (mAP) with an extremely high processing speed while maintaining a relatively small model size compared to other algorithms. These attributes make it an ideal choice for fast and accurate real-time object detection tasks.

In future research, we plan to enhance feature extraction accuracy and detection robustness by introducing more advanced network structures, increasing inter-layer connections, and utilizing deeper networks. We aim to expand the training dataset and adjust and test the algorithm to cater to different types of crops. Integrating the algorithm with hardware platforms such as drones and automated mobile robots, we intend to develop an automated and intelligent disease monitoring system for on-site testing and to optimize the model's real-time application performance. Through these research and development plans, we anticipate not only scientific progress but also significant technological transformation and industrial advancement in practical applications.

Funding: This results was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-003)

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Acknowledgments: This results was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-003)

References

1. Staub, J.E.; Robbins, M.D.; Wehner, T.C. Cucumber. In *Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae*; Prohens, J., Nuez, F., Eds.; Springer: New York, NY, USA, 2008; pp. 241–282.
2. Rur, M.; Rämert, B.; Hökeberg, M.; Vetukuri, RR.; Grenville-Briggs, L. Screening of alternative products for integrated pest management of cucurbit powdery mildew in Sweden. *Eur J Plant Pathol*, vol.150, pp.127–138, 2018. <https://doi.org/10.1007/s10658-017-1258-x>
3. Wan, B.; Zhou, E.; Xiao, P.; Sun, X.; Yang, J. Pesticide overuse in vegetable production: A case study of urban agriculture in city x, China. *Environ. Res. Commun.* vol.5, no. 085012, 2023.
4. Li, Y. Scientific Identification and Control Measures of Several Diseases of Greenhouse Cucumbers. *Grassroots Agricultural Technology Extension*, vol.9, pp. 40-42, 2021.
5. Türkoğlu, M.; Hanbay, D. Plant disease and pest detection using deep learning-based features. *Turkish J. Electr. Eng. Comput. Sci*, vol.27, pp.1636-1651, 2019.

6. 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 European Conference on Computer Vision (ECCV), 2016.
7. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Focal Loss for Dense Object Detection. In Proceedings of the ICCV, 2017.
8. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934, 2020.
9. Jocher, G.; Stoken, A.; Borovec, J.; Fomin, F.F.; Hoang, L.Q.V.; Eskenazi, M.; Anselmo, J.; Zhang, Z.; Borkowski, A.A. Borkowski. YOLOv5. GitHub repository, 2021.
10. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the ECCV, pp.213-219, 2020.
11. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the ICCV, pp.9627-9636, 2019.
12. Ge, X.; Li, Y.; Wang, Z.; Wang, H.; Tang, R.; Luo, P. YOLOX: Exceeding YOLO Series in 2021. arXiv preprint arXiv:2107.08430, 2021.
13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the NeurIPS, 2016.
14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the CVPR, pp.580-587, 2014.
15. Girshick, R. Fast R-CNN. In Proceedings of the ICCV, 2015.
16. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the ICCV, 2017.
17. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into High-Quality Object Detection. In Proceedings of the CVPR, 2018.
18. Sirisha, U.; Praveen, S.P.; Srinivasu, P.N.; Barsocchi, P.; Bhoi, A.K. Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. *Int. J. Comput. Intell. Syst.*, vol.16, no.1, p. 126, 2023. <https://doi.org/10.1007/s44196-023-00302-w>
19. Bacea, D.S.; Oniga, F. "Single stage architecture for improved accuracy real-time object detection on mobile devices," *Image Vision Comput.*, vol. 130, 2023, Art. no. 104613, ISSN 0262-8856, <https://doi.org/10.1016/j.imavis.2022.104613>.
20. Ebrahimi, M.A.; Khoshtaghaza, M.H.; Minaei, S.; Jamshidi, B. "Vision-based pest detection based on SVM classification method," *Comput. Electron. Agric.*, vol. 137, pp. 52-58, 2017, <https://doi.org/10.1016/j.compag.2017.03.016>.
21. Mondal, D.; Kole, D. K.; Roy, K. Gradation of yellow mosaic virus disease of okra and bitter melon based on entropy-based binning and Naive Bayes classifier after identification of leaves. *Computers and Electronics in Agriculture*, vol.142(Part B), pp.485-493, 2017. <https://doi.org/10.1016/j.compag.2017.11.024>
22. Xu, Z.; Huang, X.; Lin, L.; Wang, Q.; Liu, J.; Yu, K.; Chen, C. BP neural networks and random forest models to detect damage by *Dendrolimus punctatus* Walker. *J. For. Res.* Vol.31, pp.107-121, 2020. <https://doi.org/10.1007/s11676-018-0832-1>
23. Amirruddin, A.D.; Muharam, F.M.; Ismail, M.H.; Ismail, M.F.; Tan, N.P.; Karam, D.S. Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (*Elaeis guineensis*) based on frond numbers: Analysis of decision tree and random forest. *Computers and Electronics in Agriculture*, vol.169, pp.105221, 2020. <https://doi.org/10.1016/j.compag.2020.105221>
24. Sathy, 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.*, vol. 173, Art. no. 105527, 2020, ISSN 0168-1699, <https://doi.org/10.1016/j.compag.2020.105527>.
25. Yin, X.; Li, W.; Li, Z.; Yi, L. Recognition of Grape Leaf Diseases Using MobileNetV3 and Deep Transfer Learning. *Int. J. Agric. & Biol. Eng.* Vol.15, no.3, pp.184-194, 2022. <https://doi.org/10.25165/j.ijabe.20221503.7062>.
26. Sankarshwaran, S.P.; Jayaraman, G.; Muthukumar, P.; Krishnan, A. Optimizing Rice Plant Disease Detection with Crossover Boosted Artificial Hummingbird Algorithm Based AX-RetinaNet. *Environ. Monit. Assess.* Vol.195, pp.1070, 2023. <https://doi.org/10.1007/s10661-023-11612-z>.
27. Liu, S.; Hu, B.; Zhao, C. Detection and Identification of Cucumber Leaf Diseases Based Improved YOLOv7. *Trans. Chin. Soc. Agric. Eng.* vol.39, pp.164-172, 2023. <https://doi.org/10.11975/j.issn.1002-6819.202305042>
28. Yang, G.; Wang, J.; Nie, Z.; Yang, H.; Yu, S. A Lightweight YOLOv8 Tomato Detection Algorithm Combining Feature Enhancement and Attention. *Agronomy*, vol.13, pp.1824, 2023. <https://doi.org/10.3390/agronomy13071824>

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.