An Improved Pest and Disease Detection Algorithm Based on YOLOv10

Binghui Liu[1*]
Xinjiang Teacher's College
Email: 17796721009@163.com
Funded by the Xinjiang Uygur Autonomous Region Natural Science Foundation (No. 2022D01A252).

***Abstract*—Existing cotton pest and disease detection algorithms face challenges in adapting to variations in target sizes and complex scenarios, with an inherent trade-off between detection accuracy and speed during the algorithm optimization process. To overcome these limitations, we propose MADN-YOLOv10, an enhanced algorithm based on YOLOv10 that introduces three key innovations: the Multi-Scale Dilated Fusion Attention (MDFA) module strengthens the feature representation of key pest and disease characteristics while improving adaptability to dynamic and complex environments, thereby enhancing detection precision; the DualConv module reduces the model's complexity and shortens the detection time, making the model better suited for crop field pest and disease detection scenarios, and even allowing it to perform optimally on embedded systems with limited hardware resources; the MPDIoU function can mitigate bounding box distortion caused by the significant variability of pest and disease samples and improve the model's robustness during detection. When evaluated on the CottonInsect dataset, our MADN model achieved remarkable results, with a mean Average Precision (mAP) of 95.9% (a 2.3% improvement) and an accuracy 2.0% higher than that of mainstream alternative algorithms. Additionally, we compressed the parameters to 4.0MB and reduced the inference time to 0.9 milliseconds, representing a 30% speed increase compared to the original model.**

***Index Terms*—Image Processing, YOLOv10, Cotton Pests and Diseases**

I. Introduction

Cotton is a core crop in China's agricultural economy, but it is frequently invaded by pests and diseases, which constrains its yield and fiber quality, and impacts the economic returns and sustainable development of the entire industry chain [1]. Effective control of cotton pests and diseases is crucial for safeguarding food security and promoting agricultural economic development. There is an urgent need for rapid and accurate identification of pests and diseases species and infection levels to adopt effective control measures. Traditional identification relies on manual labor, which is time-consuming, low in accuracy, and prone to missing the best control timing, while also being costly and inefficient. The diverse species and forms of pests and diseases make traditional methods ineffective in processing complex images.

Therefore, the development of an automated and intelligent cotton pest and disease recognition system based on computer vision technology is of utmost importance for promoting the sustainable development of the cotton industry and enhancing the efficiency and accuracy of pest and disease control.

In response to the aforementioned challenges, we propose a lightweight cotton pest and disease detection network based on YOLOv10, named MADN. This network not only improves the accuracy and robustness of cotton pest and disease detection but also reduces the computational complexity and parameter size of the model. The main contributions of this paper include:

- To address the challenge of identifying small targets in cotton pest and disease-affected areas, we have innovatively proposed the MDFA module. This module not only enhances the feature representation of critical cotton pest and disease characteristics but also improves adaptability to dynamically changing and complex scenarios, while simultaneously boosting the accuracy of pest and disease target detection.
- To better adapt the model to the scenario of crop field pest and disease detection, we have introduced the DualConv module, which reduces the model's complexity and thus significantly shortens the detection time. This enables the model to perform optimally in crop field pest and disease detection scenarios, even on embedded systems with limited hardware resources.
- To address the issue of bounding box distortion caused by significant sample variability and to enhance the model's robustness during detection, we have introduced the MPDIoU function. By enabling the recognition of both macroscopic and microscopic pest and disease characteristics, this enhancement not only bolsters the model's robustness but also streamlines the extraction of unnecessary features from pest and disease targets, further reducing the number of parameters in the network model.
- On the publicly available CottonInsect insect image dataset, our proposed MADN model has exhibited exceptional performance, outperforming all other mainstream algorithms. Specifically, the mean Average Precision (mAP) has been improved by 2.3 percentage points, achieving an impressive 95.9%, and the accuracy has also increased by 2.8%. Furthermore, we have successfully compressed the model parameters to 4.0MB, significantly boosting the inference speed. The processing time has been sharply reduced from its original value to a mere 0.9 milliseconds, representing an approximate 30% speed-up compared to the original model.

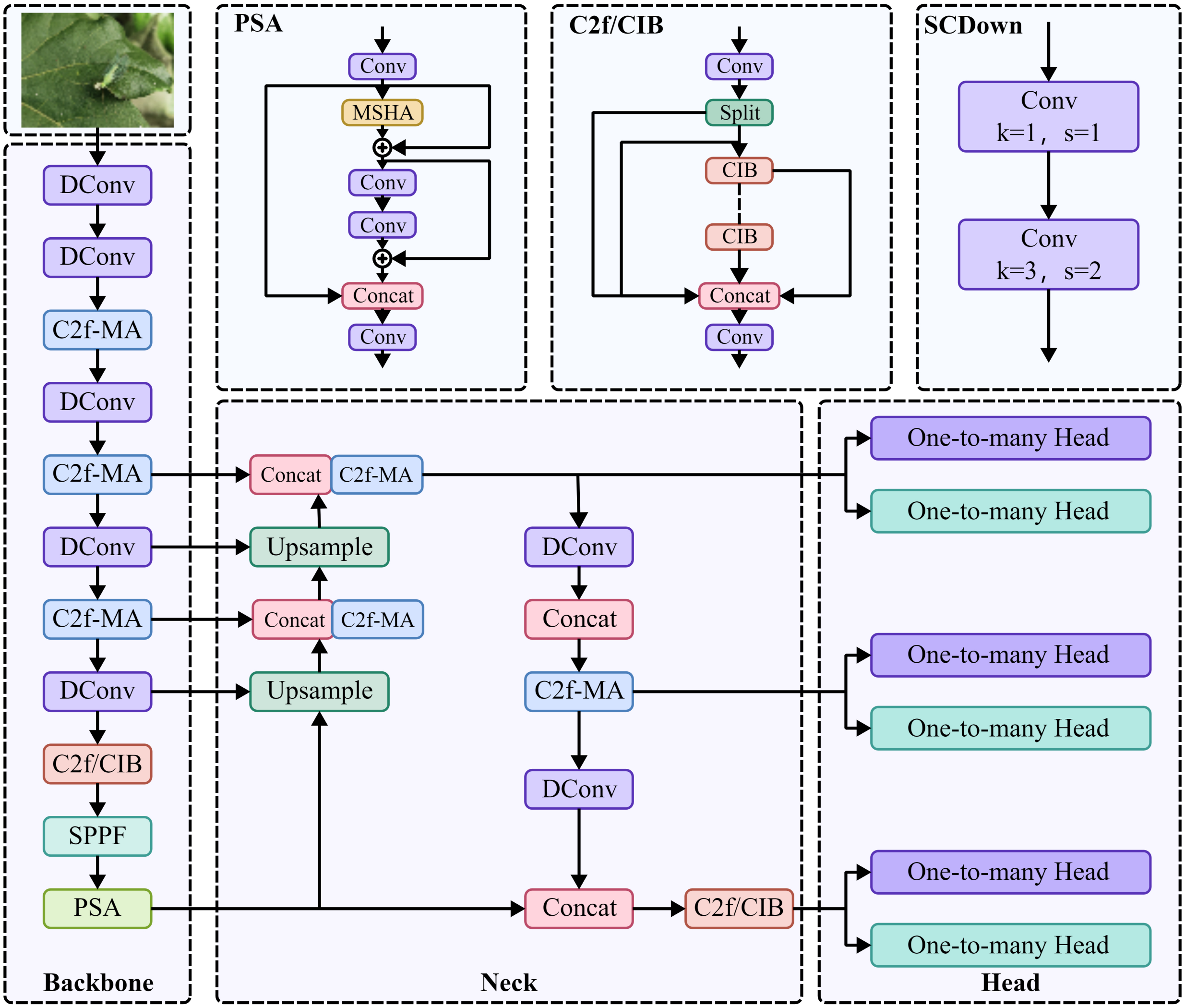

Fig. 1: MADN Overall Architecture

II. Related Work

In recent years, with the rise of spectral information analysis and remote sensing imagery technology, numerous scholars have made significant contributions in this field. Zhao Liang et al. [2] constructed a pest diagnostic model based on sensitive spectral bands, achieving the identification and extraction of cotton mite infestations by integrating spectral information analysis with remote sensing imagery technology. Dilixati Yimamu et al. [3] realized rapid and accurate identification and monitoring of multiple pest species in cotton fields through various algorithms. Meanwhile, the application of deep learning technology in the agricultural sector has gradually deepened, offering novel solutions for cotton pest and disease identification. Wang Zenglong et al. [4] designed a cotton pest identification method based on the Faster-RCNN model, training it with a natural environment image database containing three typical pests: red spiders, cotton bollworms, and aphids. This improved convolutional neural network (CNN) approach laid a methodological foundation for the development of intelligent pest identification systems. In 2023, Zhang Nannan et al. [5] introduced the CBAM attention mechanism between the backbone and head networks of YOLOv7, constructing the CBAM-YOLOv7 model, which significantly enhanced the detection performance of cotton pests and diseases compared to the original YOLOv7. Liu Runfei et al. [6] adopted the latest one-stage object detection algorithm, YOLOv8, and obtained a more accurate cotton pest detection model by incorporating the deformable convolution v3 operator (DCNv3) and adding the efficient channel attention mechanism (ECA).

However, deep learning still faces numerous challenges in the field of cotton pest and disease identification, including difficulties in recognizing key features of cotton pests, high similarity in the appearance of pests and diseases, severe interference from complex backgrounds, and challenges in model training and optimization.

To address the aforementioned challenges in cotton pest and disease identification, this study innovatively proposes a precise cotton pest identification solution based on the YOLOv10 deep learning model. As a cutting-edge version of the YOLO series, YOLOv10 demonstrates significant performance improvements in pest identification tasks with its superior recognition accuracy and unprecedented detection speed. This study conducted in-depth optimization and fine-tuning of the YOLOv10 model, aiming to further enhance the accuracy and processing speed of cotton pest identification.

We anticipate that this innovative approach will provide new strategies and tools for the effective monitoring and control of cotton pests and diseases, thereby promoting the development of intelligent precision agriculture to deeper and broader levels. This is not only an innovative attempt to the existing agricultural management model but also makes important contributions to the construction of sustainable agricultural ecosystems in the future.

III. Materials And Methods

A. MADN Network Structure

The YOLOv10 network architecture is mainly composed of four key components: Input, Backbone, Neck, and Detect Head [7]. The Input component enhances the data through scaling and normalization. The Backbone and Neck utilize the C2f structure and PAN (Path Aggregation Network) structure to accelerate network convergence and perform feature extraction. The Detect Head adopts a dual-head design, namely the One-to-many Head and the One-to-one Head, for object localization and classification. The overall architecture of our proposed MADN is depicted in Figure 1. Building upon YOLOv10, MADN incorporates our proposed C2f-MA and DualConv modules. These modules reduce redundant computations, more effectively capture disease and pest features, and enhance the model's accuracy. Additionally, we introduce the MPDIoU function, which not only improves the model's robustness but also simplifies the extraction of unnecessary features from disease and pest targets, further reducing the number of parameters in the network model.

B. C2f-MA

To address the challenge of identifying minute targets in critical areas infested by cotton pests and diseases, there is an urgent need to optimize and improve the neck network of the model. Due to the influence of weather conditions and the movement of mobile image acquisition devices, the captured target images may suffer from geometric distortions, which severely impact detection performance. By incorporating our innovatively proposed Multi-scale Dilated Fusion Attention Module (MDFA),this module aims to enhance the representation of key features of cotton pests and diseases by leveraging multiple dilation rates and integrating channel and spatial attention mechanisms. This design meets the demand for simultaneously capturing detailed information and extensive contextual information within images, which is crucial for the detection and identification of cotton pests and diseases in complex scenarios.

The network architecture is illustrated in Figure 2.

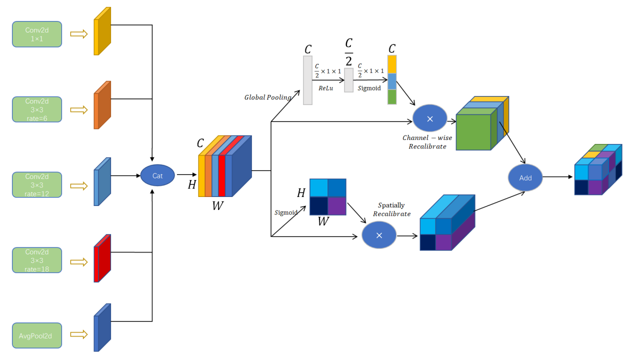

Fig. 2: MDFA Module

Our MDFA architecture achieves feature extraction at different scales through five parallel convolutional branches, each configured with distinct dilation rates to expand the receptive field and capture spatial information across varying ranges. In the second part, we perform channel and spatial feature merging and calibration. A comprehensive feature map synthesized from features of different scales extracted by the five distinct branches in the first part is subjected to parallel calibration by two attention mechanisms: the channel attention mechanism is applied above, while the spatial attention mechanism is applied below.

Finally, the output feature maps from the channel and spatial attention mechanisms are fused through element-wise addition, resulting in an enhanced feature map. At this stage, the feature maps that have undergone channel and spatial calibration respectively are added element-wise to the original merged feature map to integrate and enhance relevant features. The enhanced feature map is then passed through a 1x1 convolutional layer for dimensionality reduction and integration, generating the final output feature map.

By concatenating the output features of multi-scale dilated convolutions along the channel dimension and adjusting weights through a combination of channel and spatial attention mechanisms, our MDFA module is capable of fusing features from multiple perspectives. The integration of different dilated convolutional layers and global features ensures that the network focuses not only on local details but also on global information. This design results in a more comprehensive output from the network, enhancing the model's capability for cotton pest and disease detection in complex scenarios. The final 1x1 convolution for feature dimensionality reduction not only reduces subsequent computational overhead but also integrates information from different branches, generating a more effective feature representation for the ultimate task.

C. DualConv

To better adapt the model to the scenario of pest and disease detection in farmland, lightweight design and high efficiency are indispensable. However, the Conv architecture in YOLOv10 typically has high demands on memory and computation, making it impractical for embedded systems with limited hardware resources. To address this, we introduce DualConv (Dual Convolution) to replace the traditional Conv, thereby constructing a lightweight deep neural network. DualConv [8] combines the advantages of group convolution and heterogeneous convolution. Some convolutional kernels perform both 3×3 and 1×1 convolution operations simultaneously, while others only perform 1×1 convolution. It consists of group convolution and heterogeneous convolution components. Its structure is illustrated in Figure 3.

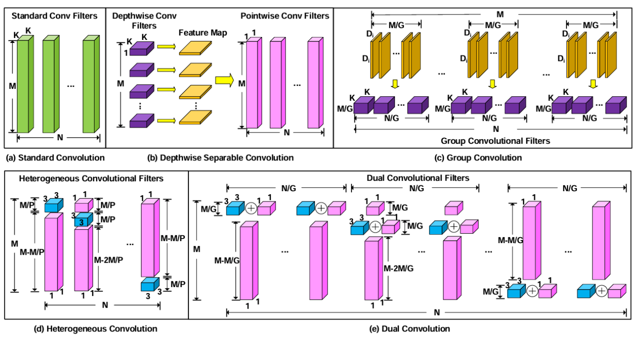

Fig. 3: A Comparison Between DualConv Convolution and Other Convolutional Methods

The concept of GroupConv was first introduced in AlexNet. Due to the limited GPU performance at that time, the model was split across two GPUs for training. GroupConv divides the convolutional filters into G groups, and the input feature map channels are also divided into G groups. Each group of convolutional filters processes the corresponding group of input feature map channels. Since each group of convolutional filters is only applied to the corresponding input channel group, the computational cost of convolution is significantly reduced.

However, channel information is not shared across different groups. That is, the output feature map channels of different groups only receive information from the corresponding input channel groups. This impedes the information flow between channels of different groups and reduces the feature extraction capability of GroupConv. To solve this problem, DualConv incorporates a channel shuffling operation to enhance information exchange between channels of different groups.

HetConv (Heterogeneous Convolution) includes both 3×3 and 1×1 convolutional kernels within a single convolutional filter. The heterogeneous filters are arranged in a shifted manner, with the 3×3 convolutional kernels being discretely arranged. The 3×3 and 1×1 convolutional kernels alternate within the filter. By using heterogeneous convolution, the computational complexity of the original 3×3 standard convolution can be reduced by 3 to 8 times. However, the heterogeneous design inherently disrupts the continuity of cross-channel information integration and negatively impacts the preservation of complete information from the input feature maps. Therefore, this strategy may reduce the network's accuracy. To address this, multiple filters are arranged in a circularly shifted manner to interleave the 3×3 and 1×1 convolutional kernels.

In DualConv, the number of convolutional filter groups G is used to control the proportion of K * K convolutional kernels in a convolutional filter. For a given G, the proportion of combined simultaneous convolutional kernels with size of (K*K +1*1) is 1/G of all channels, while the proportion of the remaining 1*1 convolutional kernels is (1-1/G). Therefore, in a dual convolutional layer composed of G convolutional filter groups, the number of FLOPs for the combined convolutional kernels is:

$$FL_{CC} = (D_2^o \times K^2 \times M \times N + D_2^o \times M \times N)/G \tag{1}$$

and the number of FLOPs for the remaining 1*1 pointwise convolutional kernels is:

$$FL_{PC} = (D_2^o \times M \times N) + (1 - 1/G) \tag{2}$$

The total number of FLOPs is:

$$FL_{DC} = FL_{CC} + FL_{PC} = D_2^o \times K^2 \times M \times N/G + D_2^o \times M \times N \tag{3}$$

Comparing the computational cost (FLOPs) of dual convo lutional layer with that of standard convolutional layer, the computational reduction ratio $R_{DC/SC}$ is:

$$R_{DC/SC} = \frac{FL_{DC}}{FL_{SC}} = \frac{1}{G} + \frac{1}{K^2} \tag{4}$$

As demonstrated in Table 4, by introducing DualConv, the computational cost and number of parameters of deep neural networks have been significantly reduced, making the model more lightweight and efficient, which allows it to better suit the scenario of detecting pests and diseases in farmland.

D. MPDIoU

In the object detection task for cotton pests and diseases, the design of the loss function significantly impacts the accuracy of detection results. Its primary role is to optimize the positional discrepancy between predicted bounding boxes and ground-truth annotations, enabling the model to generate predictions that more closely approximate the true object positions—thereby enhancing detection precision. The YOLOv10s model employs the Complete Intersection over Union (CIoU) loss function, which, due to its relatively ambiguous handling of aspect ratio variations, may lead to suboptimal optimization outcomes. Additionally, in practical detection scenarios, significant inter-sample variability poses challenges for the model's adaptive adjustments to diverse detection targets, thereby slowing convergence.

To address these limitations, this study replaces CIoU with the Minimum Point Distance-based Intersection over Union (MPDIoU) [9] loss function. By redesigning the loss metric to incorporate the minimum Euclidean distance between opposing corner points (i.e., top-left and bottom-right corners) of predicted and ground-truth bounding boxes, MPDIoU effectively mitigates detection distortions caused by high sample variability and improves model robustness. This approach reduces the degrees of freedom in the loss function, streamlining the optimization process and enhancing adaptability to diverse pest/disease morphologies.

The detailed computation of the Minimum Point Distance-based Intersection over Union (MPDIoU) loss is formulated as follows:

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \tag{5}$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \tag{6}$$

$$MPDIoU = \frac{A \cap B}{A \cup B} + \frac{d_1^2}{w^2 + h^2} + \frac{d_2^2}{w^2 + h^2} \tag{7}$$

where A denotes the ground-truth bounding box, and B represents the predicted bounding box. $(x_1^A, y_1^A), (x_2^A, y_2^A)$ are the top-left and bottom-right corners of the bounding box, respectively. $(x_1^B, y_1^B), (x_2^B, y_2^B)$ are the top-left and bottom-right corners of the predicted bounding box, respectively. d_1 is the distance between the top left corners of the two boxes; d_2 is the distance between the lower right corners of the two frames. w represents the width of the image; h represents the height of the image.

IV. Experiment And Analysis

A. Introduction to the Dataset

In 2023, Yang Manxian and colleagues from Xinjiang Agricultural University published CottonInsect: an openly accessible image dataset for cotton field insect identification research [10]. This dataset provides fundamental data for image classification and object detection of cotton field insects, holding significant practical application value for advancing farmland development, research on cotton pests and diseases control, and enhancing cotton production. They constructed an image dataset of insects in Xinjiang cotton fields.

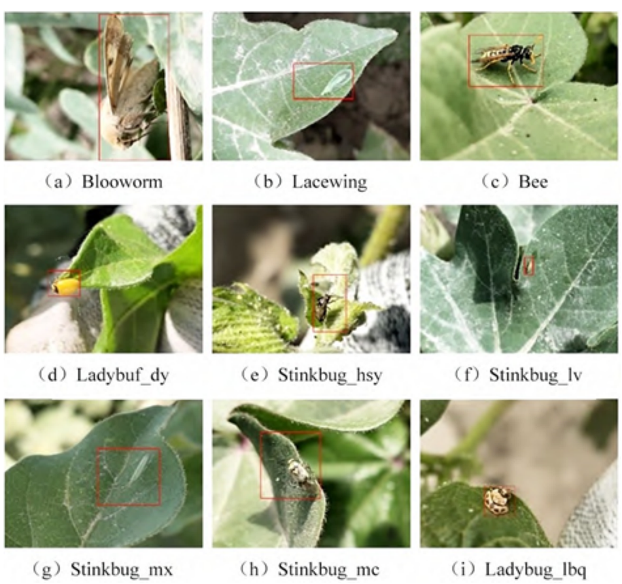

Fig. 4: Schematic diagram of partial dataset

The dataset encompasses 13 common cotton insect species, including the cotton bollworm, ladybugs, and alfalfa weevil, among others, totaling 6,451 images with raw data amounting to 24 GB. Some of the data are presented in Figure 4. The image data were collected at various times, locations, and in different environments, capturing images of insects in various poses and growth stages in real-world settings. Through manual screening and annotation, the data were categorized into two formats: one for image classification and another for object detection. Ultimately, this dataset can provide foundational data for tasks such as image classification and object detection of cotton insects in cotton fields.

In the field of deep learning, the preprocessed dataset is scientifically divided into: training set, validation set, and test set, as detailed in Table 1. The training set bears the responsibility of extracting image features; the validation set plays a crucial role in preventing overfitting; and the test set is used to accurately evaluate the model's accuracy on an independent dataset.

TABLE I: Schematic Diagram of Partial Dataset

Dataset	Function	Number of images	Image resolution(pixels)
Trainingset	Train the model	5000	2 000×1 500
Validationset	Select a model	725	2 000×1 500
Testset	Evaluate the model	726	2 000×1 500

B. Experimental Details

The computational resource used for the experiment was a Tesla T4 GPU, equipped with an Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz. The deep learning framework was Pytorch 2.10, with Python version 3.10 and CUDA version 11.8. The experimental parameters were set as follows: the input image size was 640×640, the batch size was 16, and the initial learning rate was set to 0.01. All other experiments were trained for 300 epochs.

C. Comparative Experiments

We conducted a comprehensive evaluation of MADN across three benchmark tests, with a particular focus on lightweight metrics, including the number of parameters (Params), floating-point operations (GFLOPs), inference speed (FPS), and model size (Model Size). Simultaneously, we also paid attention to accuracy metrics, such as precision, recall, and the mean Average Precision at 50% Intersection over Union (mAP50).

To validate the practicality and generalization capability of our improved algorithm, we conducted tests using three different public datasets. The first is the Plant-Village [11] dataset provided by Hughes et al., from which we selected 3,782 images specifically tailored for cotton pest and disease detection. The second dataset, IDADP [12], is offered by institutions including the Hefei Institutes of Physical Science (a subsidiary of the Chinese Academy of Sciences). We utilized 1,522 images from this dataset that are relevant to cotton pest and disease detection. The third dataset, IP102 [13], is a crop pest and disease dataset designed for object classification and detection tasks. We employed 2,800 images from it for testing purposes.

As demonstrated in Table 2, both the accuracy and mAP50 metrics have improved across all three public datasets. This indicates that our MADN model not only maintains a high level of accuracy but also exhibits remarkable generalization capability.

TABLE II: Comparative experiments with public datasets

Dataset	Model	Precision	mAP50
Plant-Village	YOLOv10n	0.903	0.928
	MADN	**0.927**	**0.954**
IDADP	YOLOv10n	0.897	0.921
	MADN	**0.919**	**0.943**
IP102	YOLOv10n	0.904	0.935
	MADN	**0.926**	**0.951**

To further substantiate the performance advantages of the proposed method and fulfill the accuracy and speed requirements for critical component detection in real-world application scenarios, while also taking into account the computational constraints of hardware devices, we conducted comparative experiments under identical experimental conditions and parameter configurations. Specifically, we benchmarked the MADN deep neural network against current mainstream object detection algorithms as well as several lightweight models improved upon YOLOv10n. The training duration for all models was uniformly set to 300 epochs.

As illustrated in Table 3, when juxtaposed with other network architectures, our proposed model excels by exhibiting the shortest inference time and the smallest parameter count. This characteristic renders it exceptionally well-suited for hardware platforms with constrained resources. Concurrently, the model attains the highest level of detection accuracy. Through these rigorous experiments, we have compellingly demonstrated both the efficiency and the applicability of our enhanced model.

To verify the reliability of our proposed method, we conducted extensive ablation experiments on the CottonInsect dataset.

As shown in Table 4, the detection accuracy of the unimproved YOLOv10 network is 90.2%, and the detection time per image is 1.3 milliseconds.

After introducing our proposed MDFA module, the network's detection accuracy is further improved by 2.1%, and the detection time is reduced by 0.1 milliseconds. The MDFA enhances the representation of key features of cotton pests and diseases by utilizing multiple dilation rates and integrating channel and spatial attention mechanisms. This design meets the need to simultaneously capture detailed information and extensive contextual information in images, which is crucial for the detection and recognition of cotton pest and disease targets in complex scenarios.

After adopting DualConv, although the description initially had a slight confusion (corrected here), the average detection accuracy of the network is improved by 1.5%, and the detection time is astonishingly reduced by 0.5 milliseconds. This is because DualConv replaces the original convolutional modules in the YOLOv10 backbone network, which to some extent reduces the model's complexity, thereby significantly shortening the detection time. This enables the model to be better suited for farmland pest and disease detection scenarios, still performing well on embedded systems with limited hardware resources.

Finally, we replaced CIoU with MPDIoU in the model, effectively mitigating bounding box distortions caused by large sample variability, improving the model's robustness, and reducing computational load and memory consumption. To balance the relationship between computational cost and detection performance, we further reduced the number of network model parameters.

Compared to the unimproved version, our proposed MADN network achieves a cumulative improvement in detection accuracy of 2.8%, with improved detection performance for small targets in key areas. Additionally, the detection time is reduced by 0.4 milliseconds, further enhancing the network's detection rate. Moreover, our MADN model excels in other metrics, exhibiting the shortest inference time and the fewest parameters. This characteristic makes it particularly suitable for hardware platforms with constrained resources. Simultaneously, the model attains the highest level of detection accuracy. Through these rigorous experiments, we have compellingly demonstrated both the efficiency and the applicability of our enhanced model.

V. Conclusion

This paper proposes a lightweight cotton field pest and disease detection network, MADN, which reduces the model's computational complexity and the number of parameters while enhancing the algorithm's robustness and detection accuracy, thereby improving the detection performance for cotton pests and diseases in complex scenarios. The designed MDFA module not only strengthens the feature representation of key characteristics of cotton field pests and diseases but also meets the need to simultaneously capture detailed information and extensive contextual information in images, which is crucial for the detection and recognition of cotton pest and disease targets in complex scenarios. Additionally, we introduce the DualConv module to reduce the model's complexity, thereby significantly shortening the detection time. This enables the model to be better suited for farmland pest and disease detection scenarios, still performing well on embedded systems with limited hardware resources. Finally, we replace CIoU with MPDIoU in the model, effectively mitigating bounding box distortions caused by large sample variability, improving the model's robustness, and reducing computational load and memory consumption. To balance the relationship between computational cost and detection performance, we further reduce the number of network model parameters. Experimental results demonstrate that MADN exhibits outstanding performance in cotton field pest and disease detection tasks.

References

[1] L. Yan-Hui, L. Ge-Mei, Z. Yong-Jun, and Y. Xian-Ming, "Advances in the management of insect pests of cotton in china since the 21 st century.," *Chinese Journal of Applied Entomology*, vol. 57, no. 3, 2020.

[2] Z. Liang, C. Bing, and X. Chunhua, "Remote sensing monitoring technique on cotton leaves infected by spider mites," *Journal of Gansu Agricultural University*, vol. 50, no. 5, pp. 94–99, 2020.

[3] D. Yimamu, Z. Jianping, and X. Yan, "Cotton pest monitoring based on logistic algorithm and remote sensing image," *Journal of South China Agricultural University*, vol. 43, no. 2, pp. 87–95, 2022.

[4] W. Zenglong, J. Yongyong, and P. Haifeng, "A method for cotton pests identification and statistics based on faster-rcnn," *Popular Technology*, vol. 25, no. 5, pp. 5–7, 12, 2023.

[5] N. Zhang, X. Zhang, and T. Bai, "A method for identifying cotton leaf diseases and pests in natural environments based on cbam-yolo v7," *Transactions of the Chinese Society for Agricultural Machinery (English Edition)*.

[6] R. Liu, "Cotton pest detection algorithm based on improved yolov8," *Agricultural Engineering*, vol. 14, no. 7, pp. 42–47, 2024.

[7] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458*, 2024.

[8] J. Zhong, J. Chen, and A. Mian, "Dualconv: Dual convolutional kernels for lightweight deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 9528–9535, 2022.

[9] S. Ma and Y. Xu, "Mpdiou: a loss for efficient and accurate bounding box regression," *arXiv preprint arXiv:2307.07662*, 2023.

[10] M. Yang, Y. Chen, and Y. Li, "Cottoninsect: An image dataset for cotton field insect recognition research." DS/OL.

[11] D. Hughes, M. Salathé, *et al.*, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2017.

[12] C. Lei and Y. Yuan, "Image dataset for crop disease recognition in the field." [DB/OL], 2019.

[13] X. Wu, C. Zhan, Y.-K. Lai, M.-M. Cheng, and J. Yang, "Ip102: A large-scale benchmark dataset for insect pest recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8787–8796, 2019.

[14] Y. Fang, X. Guo, K. Chen, Z. Zhou, and Q. Ye, "Accurate and automated detection of surface knots on sawn timbers using yolo-v5 model," *BioResources*, vol. 16, no. 3, p. 5390, 2021.

[15] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, and W. e. a. Nie, "Yolov6: A single-stage object detection framework for industrial applications," *arXiv preprint arXiv:2209.02976*, 2022.

[16] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7464–7475, 2023.

[17] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 116–131, 2018.

[18] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314–1324, 2019.

[19] K. Zhou, J.-L. Shi, J.-Y. Fu, S.-X. Zhang, T. Liao, C.-Q. Yang, J.-R. Wu, and Y.-C. He, "An improved yolov10 algorithm for automated damage detection of glass curtain-walls in high-rise buildings," *Journal of Building Engineering*, p. 111812, 2025.

[20] L. Aktouf, Y. Shivanna, and M. Dhimish, "High-precision defect detection in solar cells using yolov10 deep learning model," in *Solar*, vol. 4, pp. 639–659, MDPI, 2024.

[21] Y. Li, W. Leong, and H. Zhang, "Yolov10-based real-time pedestrian detection for autonomous vehicles," in *2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 1–6, IEEE, 2024.

[22] X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan, "Efficientvit: Memory efficient vision transformer with cascaded group attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14420–14430, 2023.

[23] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1580–1589, 2020.

[24] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16133–16142, 2023.