Table A.1
Ablation comparison experiment results. ✓ represents the use of the corresponding improvement method.

MEIEM	Pruning	Knowledge Distillation	P%	R%	mAP@0.5%	mAP@0.5:0.95%	Params	GFLOPs	Memory
–	–	–	92.0	89.8	92.7	64.4	2.58	6.3	5.5
✓	–	–	92.3	88.9	93.8	63.1	2.53	6.3	5.6
✓	✓	–	92.6	88.3	93.3	62.7	0.55	2.1	1.6
✓	✓	✓	**95.0**	**90.8**	**95.6**	**64.6**	0.55	2.1	1.6

Table A.2
Efficient Structured Pruning comparison experiment.

Speedup	P%	Params	GFLOPs	Memory	FPS
MEIEM	92.3	2.53	6.3	5.6	186.8
2.0X	89.8	0.89	3.1	2.3	185.4
2.2X	91.2	0.77	2.8	2.1	185.8
2.4X	**92.8**	0.70	2.6	1.9	182.3
2.6X	88.9	0.65	2.4	1.8	189.8
2.8X	92.3	0.58	2.2	1.7	188.2
3.0X	92.6	**0.54**	**2.1**	**1.6**	**190.6**

Table A.3
Precision and Recall of Obscured Coccinellidae Detection.

Models	P%	R%	Models	P%	R%
RT-DETR	91.4	77.6	Yolov5	93.9	80.1
Yolov6	91.4	81.5	Yolov8	92.5	81.4
Yolov11	93.8	79.9	Ours	**93.9**	**82.3**

A. Additional experiments

A.1. Ablation studies

Table A.1 shows ablation studies of MEIEM, Efficient structured pruning and Multiple Equilibrium Knowledge Distillation Strategy. The experimental results show that MEIEM can independently enhance the detection performance of the entire network, increasing mAP@0.5 by 1.1 percentage points and increasing precision by 0.3 percentage points. Efficient structured pruning can reduce the number of parameters by 2.08M, GFLOPs by 4.2, and memory by 3.9MB, but it will slightly deteriorate mAP@0.5:0.95 and recall. A slight decrease in accuracy is acceptable in the lightweighting process. Knowledge distillation compensates for the accuracy loss caused by model channel pruning. Our knowledge distillation strategy improves the precision to 95.0% and mAP@0.5 to 95.6%, indicating that multiple equilibrium knowledge distillation, a lightweight method, is helpful to ensure the accuracy of the detector. Therefore, the ablation experiments verify the effectiveness of the three innovations in improving the accuracy and lightweightness of the baseline model.

A.2. Comparison of feature maps between C3K2 and C3K2-MEIEM

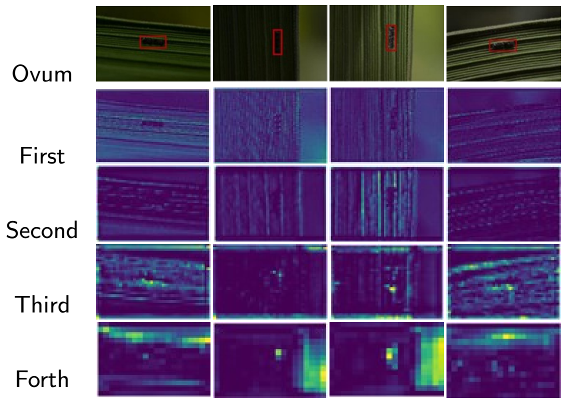
As shown in Figure A.1, we present the feature maps output by the C3k2 modules in the YOLOv11 backbone and those output by the C3k2-MEIEM modules in our backbone. Yellow represents the pixel position of the target ovum, and blue represents the background pixel. First C3k2 and First C3k2-MEIEM are located in the shallow position of the backbone, and their feature maps basically express the edge and texture of ovum. As the convolutional layers in the backbone are deepened, the model gradually determines the ovum position. Forth C3k2 and Forth C3k2-MEIEM are located in the deep layer of the backbone, and their feature maps express pixel semantic information. As shown in Figure A.1a, the yellow part of the feature map output by Forth C3k2 does not overlap with ovum location, and even background pixels are identified as ovum pixels. This shows that C3k2 is not suitable for ovum recognition under low illumination. As shown in Figure A.1b, the yellow area of the feature map output by Forth C3k2-MEIEM basically converges to ovum location. This shows that MEIEM we proposed is basically suitable for pest detection under low illumination.

A.3. Comparative experiments of pruning rate

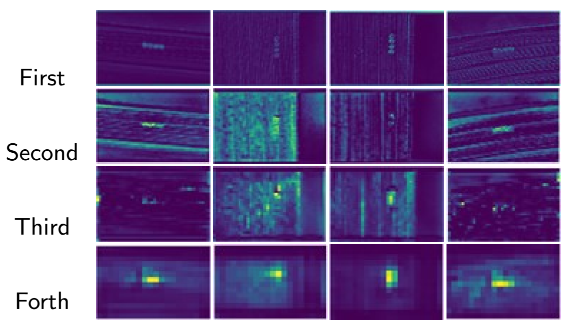
As shown in Table A.2, we conducted comparative experiments of pruning rate to demonstrate the effect of model compression. The experiment found that pruning did not actually reduce the model precision. When Speedup was 3.0, the amount of calculation was reduced to 1/3 of Baseline-MEIEM, greatly reducing the computational cost, and the precision increased by 0.3 percentage points. Other Speedups cannot compress models to such an extreme. When Batch Size is set to 4, the detection rate of NVIDIA GeForce RTX 3090 reaches 190.6 FPS at 3.0X Speedup, which is 3.8 frames faster than Baseline-MEIEM. The results show that efficient structured pruning can retain the basic performance of the model and achieve lightweight, and at the same time illustrate the robustness and generalizability of model pruning in rice pest detection.

A.4. Performance Comparison of Obscured Coccinellidae Detection

Coccinellidae has a wide range of activities, small size, and large number. Sometimes they are blocked by rice ears, so they are easily missed. Recall is an important indicator to measure missed detection. The lower the recall rate, the more serious the missed detection. As shown in Table A.3, on coccinellidae, MPD-Net can achieve the highest recall while maintaining a high precision, indicating that our architecture can maintain the accuracy of detection without missing coccinellidae. Figure A.2 shows detection results of

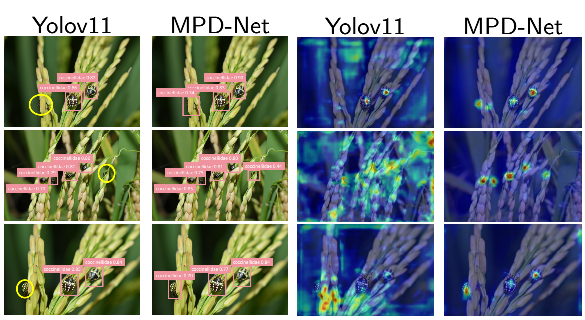


(a) C3k2 of Yolov11 Backbone



(b) C3k2-MEIEM of Our Backbone

Figure A.1: Feature Map of C3k2 and C3k2-MEIEM



(a) Prediction Box and Classification Confidence

(b) Heat map of Coccinellidae Detection

Figure A.2: Detection results of Obscured Coccinellidae. The pink boxes indicate the detection results, and the yellow circles indicate the missed pests.

obscured coccinellidae. As shown in Figure A.2a, there are missed detections in each detection result graph of Yolov11, and the missed coccinellidae are marked with yellow circles. However, MPD-Net can identify coccinellidae in all positions, including coccinellidae hiding behind rice ears. Figure A.2b shows the interpretability analysis of KPCA-CAM[47]. Red represents high model attention, and blue represents low model attention. Yolov11 does not pay attention to the occluded coccinellidae and also allocates attention to the complex background. However, our proposed framework can pay attention to the occluded coccinellidae. This shows the robustness of MPD-Net for occluded pest detection.