wind

MDPI

Article

Improved YOLOv8n-Based Model for Wind Turbine Blade Defect Detection

Hao Ren and Zhanjun Tang *

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 20242204241@stu.kust.edu.cn
* Correspondence: tzj504@kust.edu.cn

Abstract

Wind turbine blade defect detection in complex environments is challenged by weak defect features, missed crack detections, and false detections caused by background interference. To address these problems, this study proposes an improved YOLOv8n-based detection model. First, MS-CBAM is introduced before the SPPF module to enhance channel-wise feature refinement and multi-scale spatial feature extraction. Second, four stride-2 down-sampling convolutional layers in the backbone are replaced with EPConv, which combines efficient multi-scale channel attention with directional pinwheel-shaped convolution to strengthen the representation of weak and elongated defects. Finally, the original CIoU loss is replaced with PIoU v2 to improve bounding-box regression. Experiments on a self-constructed wind turbine blade defect dataset show that the proposed model achieves a precision of 92.1%, a recall of 85.1%, an $mAP_{0.5}$ of 90.7%, and an $mAP_{0.5:0.95}$ of 68.0%. Compared with the original YOLOv8n, these values represent improvements of 0.4, 5.9, 3.9, and 5.2 percentage points, respectively. The model contains approximately 3.0 M parameters, requires 8.9 GFLOPs, and achieves a network-forward inference speed of 77.1 FPS on an NVIDIA GeForce RTX 5060 Laptop GPU. Class-wise evaluation further shows that crack $AP_{0.5}$ increases from 78.4% to 85.0%, while crack $AP_{0.5:0.95}$ increases from 50.7% to 55.9%. These results demonstrate that the proposed modifications improve the detection and localization of weak and elongated defects while maintaining real-time inference capability on the tested GPU platform.

Keywords: wind turbine blade; defect detection; attention mechanism; YOLOv8n

1. Introduction

With the growing imbalance between global energy supply and demand and the widespread implementation of low-carbon development strategies, wind power, as a clean and renewable form of energy generation, has become an important means of alleviating the energy crisis and promoting sustainable social development [1]. Wind turbine blades are critical components that directly affect power generation efficiency, operational safety, and service life of wind turbines [2]. Since wind turbine blades are commonly deployed in remote areas such as mountainous regions, deserts, and offshore environments, they are continuously exposed to complex and harsh natural conditions. Consequently, they are highly susceptible to various defects, including surface peeling, coating damage, pitting, and cracking. Relevant statistics indicate that blade failures account for more than 30% of all wind turbine accidents, making blade defects one of the most prominent and direct bottlenecks restricting the efficient and stable operation of wind farms [3]. Therefore,

Check for updates

Academic Editor: Giuseppe Failla

Received: 9 July 2026
Revised: 9 August 2026
Accepted: 13 August 2026
Published: 20 August 2026

Copyright: © 2026 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.

regular condition inspection of wind turbine blades and timely mitigation of potential hazards are essential for ensuring the continuous and stable operation of wind turbines and reducing operation and maintenance risks.

Traditional wind turbine blade defect detection based on manual inspection is often unable to meet the requirements of efficient detection in practical applications [4]. Arcos-Jiménez [5] developed an ultrasonic nondestructive testing platform and employed wavelet transform algorithms for signal filtering, denoising, and feature extraction, thereby addressing the difficulty of identifying mud contamination on blade surfaces. Tcherniak [6] used electromagnetic actuators to excite wind turbine blades, acquired vibration signals through accelerometers, extracted vibration features within the medium-frequency range, and combined them with a semi-supervised learning algorithm to achieve accurate blade fault detection. Liu [7] analyzed the strain transfer efficiency of fiber Bragg grating (FBG) sensors under various loading conditions and established a corresponding computational model for identifying wind turbine blade damage based on strain variation characteristics. However, these methods generally involve relatively complex data processing procedures and high long-term sensor maintenance costs. Moreover, ultrasonic testing imposes stringent requirements on operating environments and often encounters difficulties in achieving precise defect localization.

Deep-learning-based object detection methods can generally be classified into two-stage and one-stage detection algorithms [8]. Two-stage detectors typically follow a process in which candidate regions are first generated and subsequently subjected to object classification and bounding-box refinement, as exemplified by R-CNN [9] and Faster R-CNN [10]. Chen Liqiong [11] proposed a weld defect detection method based on Faster R-CNN by introducing a channel attention mechanism into the backbone network to modify the residual block structure and replacing conventional ROI Pooling with ROI Align. The proposed method achieved a mean average precision ($mAP_{0.5}$) of 92.34%. Yang [12] proposed a coal–rock recognition method based on the fusion of laser point clouds and images, termed Multi-Modal Frustum PointNet (MMFP). In this method, MobileNetV3 was adopted as the backbone of Mask R-CNN, while a dilated convolutional block attention mechanism and an Inception structure were incorporated. A self-attention-based PointNet was further constructed to segment point clouds within the frustum region, achieving an accuracy of 93.6%. Although such two-stage methods generally provide high detection accuracy and strong generalization capability, they are often associated with slow inference speed and large model sizes.

In contrast, one-stage detection algorithms such as YOLO and SSD eliminate the independent region proposal stage and perform object classification and localization regression through a single forward pass, thereby offering substantial advantages in detection efficiency and better meeting the engineering requirements of real-time online wind turbine blade defect detection. Lv [13] combined an improved SSD detection framework with a ResNet feature extraction network to develop the EADD damage detection model, which achieved improved performance in wind turbine blade damage detection; however, its detection accuracy still leaves room for further improvement. Sun [14] proposed a lightweight RGS-YOLO model based on YOLOv8n by introducing the C2f-RFAConv module, incorporating GSConv and VoV-GSCSP modules into the neck network, and designing an SCConv Head for wind turbine surface defect detection. The results demonstrated that the model substantially reduced both the number of parameters and computational cost while improving detection accuracy. Nevertheless, the relatively small dataset used in that study means that its generalization capability across broader real-world scenarios remains to be further validated. Tang Zhanjun [15] proposed EPW-YOLOv7 for real-time wind turbine blade defect detection. By enhancing feature extraction capability while reducing network param-

eters and computational complexity, the method markedly improved both the recognition accuracy and detection speed of multi-scale defects. However, its performance remained noticeably limited for crack-type defects. Zhang [16] proposed a lightweight GCB-YOLO algorithm based on YOLOv5s, in which GhostNet was used to replace part of the original backbone network, a coordinate attention (CA) mechanism was subsequently introduced, and a bidirectional feature pyramid network (BiFPN) was employed to optimize the neck network. This method achieved a favorable balance between wind turbine blade defect detection accuracy and inference speed, facilitating deployment on edge devices. However, because the method currently relies solely on RGB images as input, it remains sensitive to environmental interference caused by complex illumination, rain, snow, and fog, and cannot directly estimate defect depth. Although the methods reported in existing studies have improved the efficiency of wind turbine blade defect detection, some models still exhibit insufficient defect recognition accuracy under complex background interference.

2. Improved YOLOv8n Model

YOLOv8n is a lightweight object detection model with a compact architecture, mature implementation, and a favorable balance between detection accuracy and inference efficiency. These characteristics make it suitable for real-time inspection tasks with limited computational resources. In addition, its modular network structure facilitates controlled architectural modification and ablation analysis. Therefore, YOLOv8n was selected as the baseline model in this study to investigate whether targeted feature enhancement and bounding-box regression optimization could improve the detection of small, elongated, and weak-feature defects on wind turbine blades. The selection of YOLOv8n as the baseline does not imply that it universally outperforms all later YOLO variants; rather, it provides a stable and reproducible framework for evaluating the proposed modifications. The architecture of the improved network is illustrated in Figure 1.

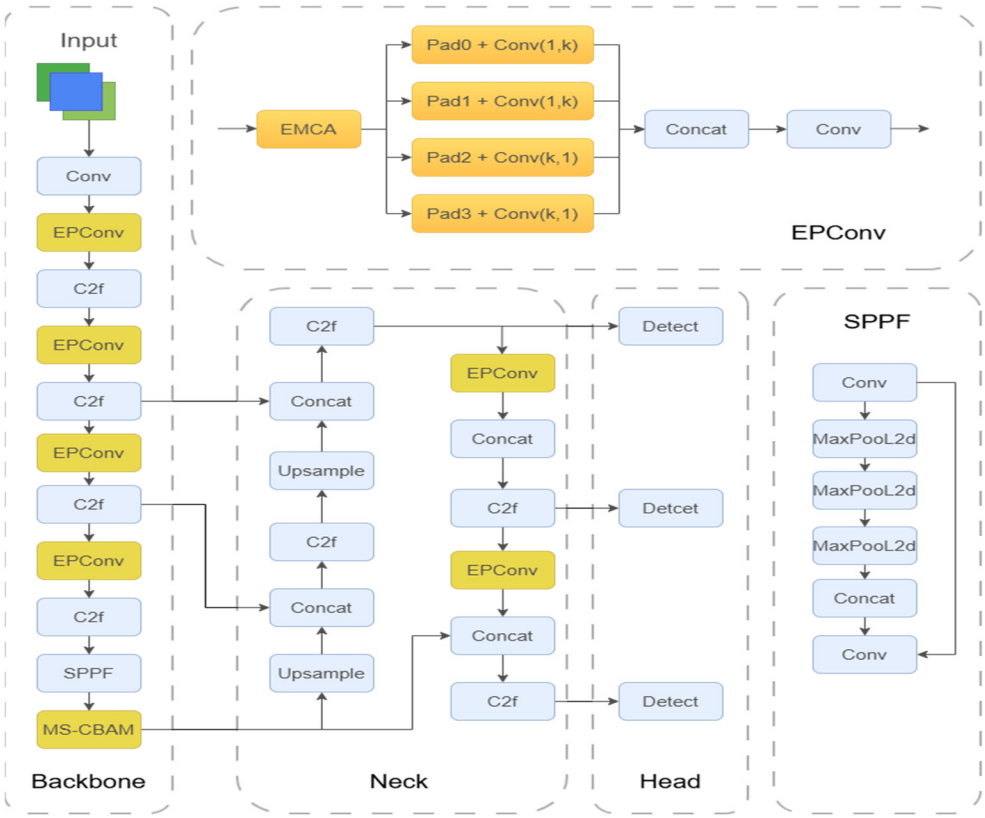


Figure 1. Architecture of the improved YOLOv8n network.

2.1. EPConv Convolutional Module

Under complex background conditions, discriminative defect features often exhibit relatively weak responses and are severely disturbed by redundant background information. Meanwhile, the directional characteristics of elongated cracks may not be sufficiently captured. To address these issues, the four stride-2 downsampling convolutional layers in the YOLOv8n backbone are selectively replaced with the efficient EPConv module. The EPConv module consists of the Efficient Multi-scale Channel Attention (EMCA) module [17] and Pinwheel-shaped Convolution (PConv) [18]. The structure of the EMCA module is illustrated in Figure 2.

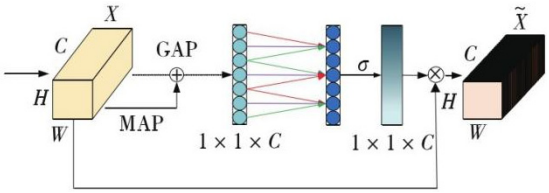


Figure 2. Structure of EMCA.

EMCA is developed based on the ECA attention mechanism [19]. By simultaneously introducing global average pooling and global max pooling, it enhances the extraction of channel-wise statistical information, enabling the model to capture the overall response distribution of features while highlighting locally salient characteristics. EMCA employs one-dimensional convolution with an adaptive kernel size to model inter-channel dependencies and uses a Sigmoid function to generate channel attention weights. Subsequently, these weights are applied to the original feature maps to recalibrate features along the channel dimension, thereby enhancing the representation capability of important defect regions and reducing the adverse influence of ineffective and redundant features on detection results. The corresponding calculation formulas are given as follows:

$$z_{gap,c} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_C(i,j) \tag{1}$$

$$z_{gmp,c} = \max_{1 \le i \le H, 1 \le j \le W} X_c(i,j) \tag{2}$$

$$z_c = z_{gap,c} + z_{gmp,c} \tag{3}$$

$$k = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{4}$$

$$y = \mathrm{Conv1D}(z,k) \tag{5}$$

$$\omega = \sigma(y) \tag{6}$$

$$\widetilde{X}_C = \omega_C \cdot X_C, c = 1,2,\ldots,C \tag{7}$$

where $X_c(i,j)$ denotes the feature value of the input feature map X, at spatial position (i,j) in the c-th channel. H and W denote the height and width of the feature map, respectively, while C represents the number of channels. $Z_{gap,c}$ and $Z_{gmp,c}$ denote the channel descriptors obtained from the c-th channel through global average pooling and global max pooling, respectively. Z_C denotes the fused channel feature. k represents the kernel size of the one-dimensional convolution. gamma is the linear scaling coefficient, and b denotes the bias term. $|\cdot|_{odd}$ denotes the operation of selecting the nearest odd integer. Conv1D(·) denotes the one-dimensional convolution operation. $\sigma(\cdot)$ denotes the Sigmoid activation function. ω_C denotes the attention weight of the (c)-th channel. $\widetilde{X}_C$ denotes the output feature after channel-wise recalibration.

The PConv module is illustrated in Figure 3. By employing asymmetric padding in different directions together with direction-oriented convolutions along the horizontal and vertical axes, the module enhances the network's capability to extract elongated structures and directional edge features. The corresponding computational formulations are given as follows:

$$X_i = SiLU(BN(Conv_i(P_i(X)))), i = 1, 2, 3, 4 \tag{8}$$

where $P_i = (left_i, right_i, top_i, bottom_i)$ denotes the asymmetric padding operation corresponding to the (i)-th branch, and $Conv_i(\cdot)$ denotes the directional convolution operation in the (i)-th branch, with a kernel size of either 1 × k or k × 1. Subsequently, the outputs of the four branches are concatenated along the channel dimension, followed by a 2 × 2 convolution to fuse the concatenated features.

$$X' = Cat(X_1, X_2, X_3, X_4) \tag{9}$$

$$Y = SiLU(BN(Conv_{2\times2}(X'))) \tag{10}$$

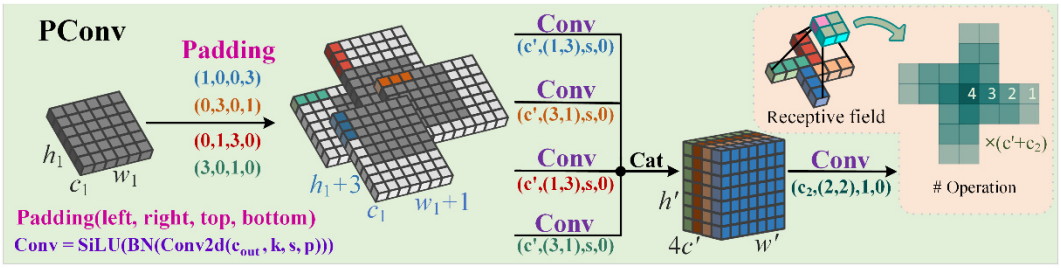


Figure 3. Structure of PConv.

To improve the robustness of the model in detecting small cracks and weak-texture defects, an EPConv structure is proposed in this study, as illustrated in Figure 4. The structure integrates EMCA and PConv in a sequential manner, enabling the feature representations generated by the two modules to complement and fuse with each other. Consequently, an output feature map with enhanced multi-scale feature representation and reduced computational overhead is obtained.

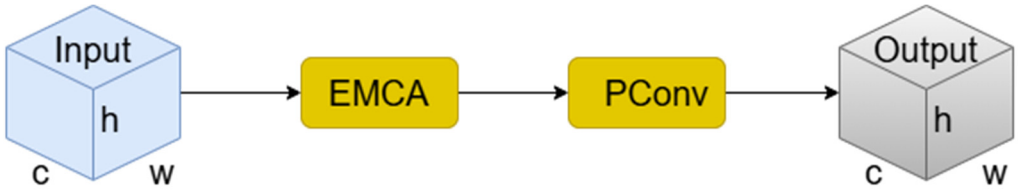


Figure 4. Structure of EPConv.

By integrating EMCA with PConv to construct the EPConv module, the model can not only emphasize effective channel-wise responses to defects such as cracks, breakages, and corrosion, but also enhance the directional continuity and boundary structural features of elongated cracks in the spatial dimension. Moreover, this integration compensates for the limited spatial directional modeling capability of channel attention alone and the insufficient channel-selection capability of standalone directional convolution. As a result, the feature representation ability of the model for wind turbine blade defects under complex backgrounds is further strengthened, while false detections caused by background interference are effectively reduced.

In the constructed network, EPConv is employed as a selective replacement rather than being used to replace all convolutional layers of the original YOLOv8n. Specifically, the four 3 × 3 convolutional layers with a stride of 2 in the backbone are replaced with EPConv.

These layers correspond to the P2/4, P3/8, P4/16, and P5/32 feature extraction stages, respectively. The initial stem convolution, the internal convolutional layers of the C2f modules, and the convolutional layers in the neck are retained as standard convolutions.

This selective deployment enables channel-wise feature refinement and directional feature extraction during hierarchical downsampling while avoiding excessive computational overhead. EPConv at the shallow P2/4 and P3/8 stages helps preserve weak textures, edge details, and small-crack features under relatively high spatial resolution. At the deeper P4/16 and P5/32 stages, EPConv strengthens the directional continuity and high-level semantic representation of elongated defects. Meanwhile, retaining standard convolutions in the stem and neck helps maintain stable low-level feature extraction and efficient multi-scale feature fusion. In the PConv component, the directional convolution kernel size was set to k = 3. This setting provides a directional receptive field for extracting elongated defect boundaries while limiting the additional computational cost. The kernel size was kept identical in all corresponding EPConv layers to ensure a controlled experimental configuration.

2.2. MS-CBAM

Small cracks in wind turbine blade defect images often exhibit weak feature responses, while defects at different scales show substantial variations in appearance. Moreover, under complex background conditions, it is difficult for the model to accurately focus on defect-relevant regions. To address these issues, the MS-CBAM is incorporated into the YOLOv8n model in this study. The module consists of the Convolutional Block Attention Module (CBAM) [20] and a dual-branch parallel Multi-Scale Convolutional Attention (MSCA) module [21]. The architecture of MSCA is illustrated in Figure 5.

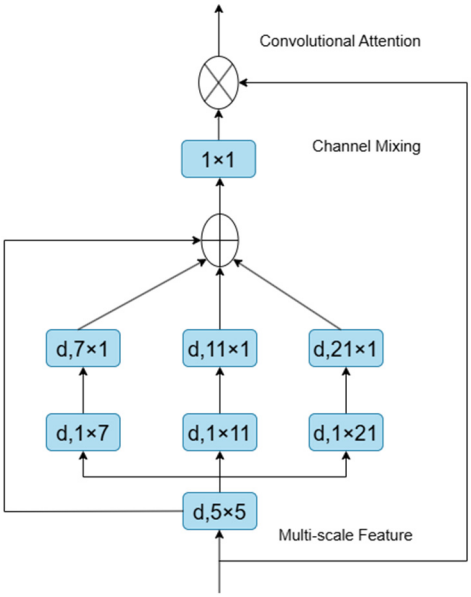


Figure 5. Structure of MSCA.

The MSCA module first employs a 5×5 depthwise convolution to extract local spatial features from the input feature map. Subsequently, three parallel strip depthwise convolution branches are constructed, with kernel sizes of $1 \times 7/7 \times 1$, $1 \times 11/11 \times 1$ and $1 \times 21/21 \times 1$, respectively, to capture long-range spatial contextual information at different scales. The local branch is then integrated with the multi-scale strip convolution branches, while a 1×1 convolution is employed to facilitate cross-channel information interaction. Finally, spatial attention weights are generated through the Sigmoid function and applied element-wise to the input feature map, thereby enhancing feature representa-

tions in the spatial dimension and producing the final output features. The corresponding computation is formulated as follows:

$$F_0 = DWConv_{5\times5}(F) \tag{11}$$

$$F_i = DWConv_{K_i\times1}(DWConv_{1\times K_i}(F_0)), K_i \in \{7, 11, 21\}, i = 1, 2, 3 \tag{12}$$

$$Att = \sigma(Conv_{1\times1}(F_0 + \sum_{i=1}^{3} F_i)) \tag{13}$$

$$Out = Att \otimes F \tag{14}$$

where (F) denotes the input feature map. $DWConv_{5\times5}$ denotes the 5×5 depthwise convolution operation used to extract local spatial features. $DWConv_{1\times K_i}$ and $DWConv_{K_i\times1}$ denotes the strip depthwise convolution operation used to capture long-range spatial dependencies at different scales. K_i takes values of 7, 11, and 21, respectively. $Conv_{1\times1}$ is employed to fuse the multi-scale branch features and facilitate cross-channel information interaction. $\sigma(\cdot)$ denotes the Sigmoid activation function. Att represents the generated multi-scale convolutional attention weights. $\otimes$ denotes element-wise multiplication. Out represents the output feature of the MSCA module.

Traditional CBAM enhances feature representations by jointly employing channel attention and spatial attention. However, during spatial attention modeling, it typically relies on single-scale convolution to capture local spatial responses, which limits its ability to adequately characterize the spatial features of defects at different scales. In particular, for elongated targets such as cracks, CBAM provides insufficient modeling of directional continuity and multi-scale contextual information, making it difficult to simultaneously preserve the fine-grained details of small cracks and the regional characteristics of large-scale defects. In contrast, MSCA employs strip convolution branches with different receptive-field scales to capture long-range spatial dependencies, which is beneficial for extracting the linear contours and directional extension patterns of cracks. Nevertheless, when used independently, its capability to distinguish defect-related semantic information from background-redundant information across the channel dimension remains limited. Therefore, this study integrates the attention-enhancement mechanism of CBAM with the multi-scale spatial modeling capability of MSCA to construct the MS-CBAM. The architecture of the proposed module is illustrated in Figure 6.

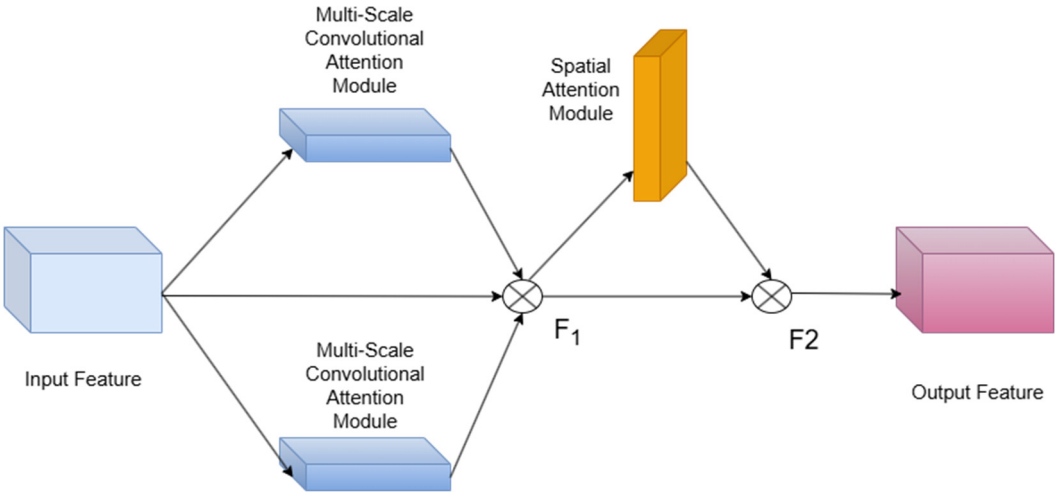


Figure 6. Structure of MS-CBAM.

The MS-CBAM compensates for the limitations of conventional CBAM in single-scale spatial perception while addressing the lack of effective channel-wise feature selection in standalone multi-scale spatial convolution. By jointly exploiting channel-wise feature refinement and multi-scale spatial modeling, the model can more accurately focus on defect-relevant regions, including crack edges, crack endpoints, discontinuous textures, coating spalling, and corrosion. Meanwhile, the module enhances the representation capability of the model for both small and multi-scale wind turbine blade defects and improves feature extraction robustness under complex background interference.

In the improved YOLOv8n architecture, MS-CBAM is inserted after the SPPF module at the end of the backbone and before the feature maps are transmitted to the neck. This position was selected because the output of SPPF contains relatively rich high-level semantic information and aggregated multi-scale context. Applying MS-CBAM at this stage allows the model to refine defect-relevant channel responses and multi-scale spatial structures before subsequent feature fusion in the neck. Compared with repeatedly inserting attention modules at multiple stages, a single deployment after SPPF limits the increase in parameters and computational complexity.

The MSCA branch first employs a 5×5 depthwise convolution to capture local spatial information. Strip convolution kernels of 7, 11, and 21 are then used to model spatial dependencies at progressively larger receptive-field scales. The smaller strip kernel preserves local crack details, whereas the larger kernels capture long-range directional continuity and contextual information associated with elongated defects. These kernel sizes follow the multi-scale configuration of MSCA and were fixed for all experiments to ensure fair comparison.

2.3. PIoU V2 Loss Function

Cracks are typically characterized by large aspect ratios, narrow widths, and indistinct boundaries. Consequently, even slight localization deviations may lead to missed detections or low-confidence predictions. The CIoU loss function adopted in YOLOv8n primarily considers the overlap area, center-point distance, and aspect ratio between the predicted and ground-truth bounding boxes. However, for elongated bounding boxes, endpoint displacement between the predicted and ground-truth boxes may not always be adequately reflected by changes in IoU, resulting in insufficient characterization of localization errors. To address this limitation, the CIoU loss function is replaced with the PIoU v2 loss function in this study. Its specific formulation is defined as follows:

$$P = \frac{1}{4}\left(\frac{d_{w1} + d_{w2}}{|w^{gt}|} + \frac{d_{h1} + d_{h2}}{|h^{gt}|}\right) \tag{15}$$

$$L_{PIoU} = 1 - IoU - e^{-P^2} + 1 \tag{16}$$

$$q = e^{-P}, x = \lambda q \tag{17}$$

$$L_{PIoUv2} = 3xe^{-x^2} L_{PIoU} \tag{18}$$

where IoU denotes the intersection over union between the predicted and ground-truth bounding boxes. P represents the boundary offset penalty factor. d_{w1} and d_{w2} the two horizontal offset terms denote the displacement distances between the left and right boundaries of the predicted and ground-truth boxes, respectively. d_{h1} and d_{h2} the two vertical offset terms denote the displacement distances between the upper and lower boundaries of the predicted and ground-truth boxes, respectively. w^{gt} and h^{gt} represent the width and height of the ground-truth bounding box, respectively. Q denotes the quality modulation factor derived from the boundary offset penalty factor. Λ is a balancing hyperparameter. X

represents the non-monotonic focusing variable. L_{PIoUv2} denotes the final bounding-box regression loss.

The balancing hyperparameter λ was set to the default value provided in the original PIoU v2 implementation and was kept unchanged in all experiments. This value was selected according to the recommended configuration of PIoU v2 and was kept unchanged throughout the ablation and comparison experiments. No category-specific adjustment of the loss hyperparameters was performed.

PIoU v2 introduces a target-size-adaptive penalty factor into the original PIoU loss and adopts a non-monotonic focusing mechanism. Specifically, the loss function evaluates the boundary offsets between the predicted and ground-truth bounding boxes along both the horizontal and vertical directions and normalizes these offsets with respect to the width and height of the ground-truth box. This design enables the model to perceive boundary localization errors of elongated targets more sensitively. For defect types such as cracks and scratches, which are typically characterized by large aspect ratios, narrow widths, and indistinct edges, even slight boundary deviations can cause substantial localization errors. By imposing more effective constraints on such boundary deviations, PIoU v2 facilitates more accurate regression of the endpoints and boundary positions of elongated defects, thereby improving bounding-box localization accuracy.

3. Experiments and Analysis

3.1. Wind Turbine Blade Defect Dataset

This study focuses on wind turbine blades from a wind farm in Yunnan Province, China. Images were collected using a DJI unmanned aerial vehicle under different viewing angles and illumination conditions. After image screening and preprocessing, a total of 738 original images with a resolution of 640 × 640 pixels were retained. All images were manually annotated using LabelImg and classified into seven common wind turbine blade defect categories: burn, crack, deformation, dirt, oil stain, coating spalling, and corrosion.

To reduce the risk of data leakage, the 738 original images were divided into training, validation, and test subsets at a ratio of 8:1:1 before any data augmentation was performed. MixUp and Mosaic were applied exclusively to the training subset, whereas the validation and test subsets contained only non-augmented original images. Consequently, augmented variants derived from a training image could not appear in the validation or test subsets. After augmentation, the complete dataset contained 3573 images. The category distribution consisted of 230 burn samples, 362 crack samples, 536 deformation samples, 824 dirt samples, 465 oil-stain samples, 909 coating-spalling samples, and 247 corrosion samples, as illustrated in Figure 7.

The current partitioning strategy guarantees separation at the original-image level and prevents direct leakage of augmented variants. However, complete metadata identifying the physical blade, turbine, and UAV inspection sequence were not available for all images. Therefore, strict independence at the blade-, turbine-, or inspection-sequence level could not be verified. In addition, viewing angle and illumination were not recorded as structured categorical metadata during image acquisition, and quantitative condition-stratified statistics cannot be reported retrospectively. The present dataset covers the viewing-angle and illumination variations observed during the inspection campaign, but it should not be interpreted as representing all wind-farm seasons, weather conditions, turbine types, or UAV acquisition environments. These limitations are considered when interpreting the reported generalization and robustness results.



Figure 7. Dataset of Wind Turbine Blade Defects.

3.2. Evaluation Metrics

Considering the practical requirements of wind turbine blade defect detection, this study employs FPS, GFLOPs, P, Params, Size, mAP0.5:0.95, R, and $mAP_{0.5}$ as evaluation metrics to comprehensively assess the detection performance and computational efficiency of the model. The corresponding calculation formulas are given as follows:

$$P = \frac{TP}{TP + FP} \tag{19}$$

$$R = \frac{TP}{TP + FN} \tag{20}$$

$$AP = \int_0^1 P(R)dR \tag{21}$$

$$mAP_{0.5} = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{22}$$

where TP denotes the number of correctly detected positive samples; FP represents the number of negative samples incorrectly classified as positive; FN denotes the number of positive samples that are missed by the detector; AP represents the average precision; P(R) denotes the precision–recall curve; and N represents the total number of defect categories.

3.3. Experimental Setup and Network Training

The experiments were conducted under Windows 11 on a hardware platform equipped with an AMD Ryzen 9 8945HX with Radeon Graphics CPU and an NVIDIA GeForce RTX 5060 Laptop GPU. The software environment consisted of Python 3.10.14, PyTorch 2.9.0, and CUDA 13.0. All models used an input resolution of 640 × 640 pixels, a batch size of 32, an initial learning rate of 0.01, stochastic gradient descent as the optimizer, and a total of 300 training epochs.

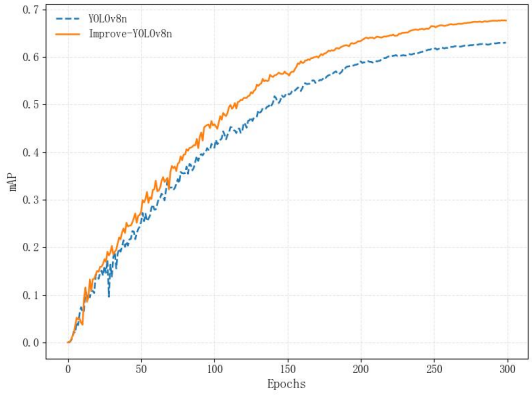
To reduce the influence of training strategy on the comparison results, all baseline, ablation, attention-module, convolution-module, and model-comparison experiments used the same dataset partition, input resolution, data-augmentation configuration, optimizer,

learning rate, batch size, number of epochs, and fixed random-seed configuration. No module-specific adjustment of the training strategy was performed. Therefore, the differences among the evaluated configurations primarily reflect changes in network structure or bounding-box regression loss.

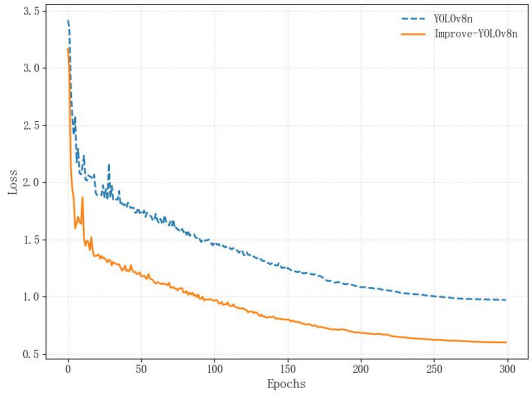
To ensure reproducibility of the inference-speed evaluation, FPS was measured on the NVIDIA GeForce RTX 5060 Laptop GPU using a batch size of 1 and FP32 numerical precision. Before timing, each model was executed for 200 warm-up iterations to stabilize GPU computation. Subsequently, 1000 forward-inference iterations were performed using randomly generated input tensors with a resolution of 640 × 640 pixels. CUDA synchronization was performed immediately before and after every timed iteration.

The reported FPS represents the reciprocal of the average single-image network-forward latency. Image loading, image decoding, preprocessing, non-maximum suppression, visualization, and result storage were not included. All models were measured using the same hardware platform, input resolution, batch size, numerical precision, warm-up procedure, and timing method. Therefore, the reported FPS values are suitable for controlled model-level comparison but do not represent the end-to-end throughput of a complete UAV inspection system.

Under identical experimental conditions, the training curves of the improved YOLOv8n model and the original YOLOv8n model are compared in Figure 8. As shown in Figure 8, the improved YOLOv8n model replaces the CIoU loss with PIoU v2 as the bounding-box regression loss. Compared with the original YOLOv8n model, the improved model exhibits a faster convergence rate. In addition, the comparison of the mean average precision curves demonstrates a substantial improvement in detection accuracy, thereby confirming the effectiveness and feasibility of the proposed method.



(**a**) mAP-mean curve



(**b**) Loss curve

Figure 8. Training curves of the original YOLOv8n and improved YOLOv8n. (**a**) mAP-mean curve; (**b**) Loss curve.

3.4. Experimental Results and Analysis

3.4.1. Ablation Experiments

To independently evaluate the contribution of each proposed modification and examine the interaction among the modules, six controlled ablation configurations were constructed using YOLOv8n as the baseline. The first configuration was the original YOLOv8n without modification. The second, third, and fourth configurations incorporated MS-CBAM, EPConv, and PIoU v2 individually, respectively. The fifth configuration combined MS-CBAM and EPConv, while retaining the original CIoU loss. The sixth configuration represented the complete proposed model, in which MS-CBAM, EPConv, and PIoU v2 were jointly employed. All configurations were trained and evaluated using identical dataset partitions, input resolution, training epochs, optimizer settings, and random-seed configuration. In addition to P, R, and mAP0.5, mAP0.5:0.95, FPS, GFLOPs, parameter count, and model size were reported to provide a more comprehensive evaluation of accuracy, localization quality, and computational complexity. The corresponding experimental results are presented in Table 1.

Table 1. Ablation experiment.

Number	MS-CBAM	EPConv	PIoU v2	P (%)	R (%)	$mAP_{0.5}$ (%)	$mAP_{0.5–0.95}$ (%)	FPS	GFLOPs	Params (M)	Size (MB)
1				91.7	79.2	86.8	62.8	161.3	8.1	3.0	6.0
2	√			90.7	83.6	88.6	65.8	130.1	8.2	3.1	6.2
3		√		91.5	81.4	88.7	64.2	95.1	8.7	3.0	6.1
4			√	89.5	84.5	88.8	63.6	142.3	8.1	3.0	6.0
5	√	√		92.4	83.6	90.1	67.0	85.7	8.9	3.0	6.1
6	√	√	√	92.1	85.1	90.7	68.0	77.1	8.9	3.0	6.1

1. The individual-module experiments show that the three modifications improve different aspects of the detector. After introducing MS-CBAM alone, the recall and $mAP_{0.5}$ increase noticeably, indicating that joint channel refinement and multi-scale spatial modeling help the network retain weak responses from small and multi-scale defects. The improvement in recall suggests that MS-CBAM mainly contributes to reducing missed detections.
2. EPConv used independently improves the representation of elongated and direction-sensitive defect features. Compared with the baseline, the improvement in $mAP_{0.5}$ and $mAP_{0.5:0.95}$ indicates that the combination of EMCA and directional PConv provides more discriminative feature representations than standard stride-2 convolution. The directional branches help preserve crack continuity and boundary information, whereas EMCA suppresses redundant channel responses produced by complex backgrounds.
3. Replacing CIoU with PIoU v2 alone improves bounding-box regression without changing the feature-extraction architecture. Its improvement is more evident in $mAP_{0.5:0.95}$ than would be reflected by $mAP_{0.5}$ alone, showing that the boundary-offset penalty and non-monotonic focusing mechanism provide stricter constraints on localization errors.
4. Combining MS-CBAM and EPConv produces a larger improvement than either architectural module alone. This result indicates a complementary relationship between the two modules. MS-CBAM refines defect-relevant regions at the high-level semantic stage, while EPConv improves directional feature extraction during hierarchical downsampling. Their combination simultaneously enhances multi-scale region attention and directional boundary representation.
5. The complete model achieves the best overall balance after PIoU v2 is further introduced. Compared with the original YOLOv8n, P, R, and $mAP_{0.5}$ increase from 91.7%, 79.2%, and 86.8% to 92.1%, 85.1%, and 90.7%, respectively. The corresponding im-

provements are 0.4, 5.9, and 3.9 percentage points. The complete model also improves mAP0.5:0.95, confirming that its advantage is not limited to the relatively tolerant IoU threshold of 0.5.

Although the complete model introduces a moderate increase in computational complexity and a reduction in FPS, it contains approximately 3.0 M parameters and requires 8.9 GFLOPs. Therefore, the value of 8.9 represents computational complexity rather than the number of parameters. The model still reaches 77.1 FPS under the specified measurement protocol, demonstrating that the improvement in detection performance is achieved while retaining real-time inference capability.

3.4.2. Comparative Experiments

Table 2 presents the class-wise detection performance of the baseline and the improved model. Performance improvements are observed for all seven defect categories, although the magnitude of improvement varies among categories.

Table 2. AP performance of each wind turbine blade defect category for different models.

Defect Category	Baseline_$AP_{0.5}$ (%)	Ours_$AP_{0.5}$ (%)	Baseline_$AP_{0.5–0.95}$ (%)	Ours_$AP_{0.5–0.95}$ (%)	Baseline_R	Ours_R
burning	69.1	79.4	41.7	50.0	55.5	72.2
crack	78.4	85.0	50.7	55.9	71.8	73.3
deformity	92.1	94.1	72.6	75.1	87.9	92.5
dirt	91.2	92.4	72.8	76.8	86.9	90.2
oil	91.5	94.4	70.9	75.3	83.0	89.6
peeling	91.0	94.1	63.6	69.1	86.2	87.5
rusty	93.9	95.1	67.5	71.5	82.6	89.7

For the crack category, $AP_{0.5}$ increases from 78.4% to 85.0%, while $AP_{0.5:0.95}$ increases from 50.7% to 55.9%. Recall also increases from 71.8% to 73.3%. Crack defects generally have narrow widths, large aspect ratios, discontinuous textures, and weak contrast with the blade surface. Therefore, their detection depends not only on semantic recognition but also on the preservation of directional continuity and accurate endpoint localization. The improvement in both $AP_{0.5}$ and $AP_{0.5:0.95}$ supports the effectiveness of EPConv and PIoU v2 for elongated defect representation and stricter boundary regression.

The burn category exhibits an $AP_{0.5}$ improvement from 69.1% to 79.4%, an $AP_{0.5:0.95}$ improvement from 41.7% to 50.0%, and a recall improvement from 55.5% to 72.2%. Burn defects may present irregular boundaries and weak local contrast under uneven illumination. The substantial recall improvement indicates that MS-CBAM helps the model focus on weak defect-related regions and reduces missed detections.

For deformation, dirt, oil stains, coating spalling, and corrosion, the improved model also achieves higher AP and recall. In particular, the improvement in $AP_{0.5:0.95}$ across multiple categories indicates that the proposed method improves not only category recognition but also bounding-box localization under stricter IoU criteria.

Overall, the class-wise results demonstrate that the performance improvement is not produced by only one dominant defect category. The proposed modules provide consistent benefits for defects with different textures, shapes, and scales, with particularly meaningful improvements for weak-feature and elongated categories.

To determine whether the improvement results merely from introducing an additional attention mechanism, MS-CBAM was compared with conventional CBAM, CoordAttention, and SimAM under the same experimental conditions, as shown in Table 3.

Table 3. Comparison of different attention modules.

Model	P	R	$mAP_{0.5}$ (%)	$mAP_{0.5–0.95}$ (%)	Params (M)
Baseline	91.7	79.2	86.8	62.8	3.0
CBAM	89.9	81.1	88.5	62.1	3.0
MS-CBAM	90.7	83.6	88.6	65.8	3.1
Attention	85.9	84.1	87.5	62.0	3.0
SimAN	85.9	82.7	87.1	61.7	3.0

Compared with the baseline, conventional CBAM improves recall and mAP0.5, demonstrating that channel and spatial attention are beneficial for suppressing background interference. However, its spatial attention branch primarily relies on a single convolutional receptive field, limiting its ability to simultaneously model small local defects and long-range directional structures.

MS-CBAM achieves an $mAP_{0.5}$ of 88.6% and an $mAP_{0.5:0.95}$ of 65.8%, outperforming conventional CBAM in both metrics. In particular, its $mAP_{0.5:0.95}$ is 3.0 percentage points higher than that of the baseline and 3.7 percentage points higher than that of CBAM. This result indicates that integrating multi-scale strip convolution into the attention mechanism improves spatial localization under stricter IoU thresholds rather than merely increasing architectural complexity.

CoordAttention and SimAM improve recall to different degrees but do not achieve the same overall AP performance as MS-CBAM. This suggests that attention mechanisms emphasizing coordinate information or parameter-free energy functions cannot fully replace the joint channel refinement and multi-scale spatial modeling required for wind turbine blade defects.

To evaluate whether the performance improvement is specifically related to the proposed combination of EMCA and PConv, EPConv was compared with standalone PConv, GhostConv, and GSConv, as illustrated in Table 4.

Table 4. Comparison of different convolution modules.

Model	P	R	$mAP_{0.5}$ (%)	$mAP_{0.5–0.95}$ (%)	Params (M)
Baseline	91.7	79.2	86.8	62.8	3.0
PConv	93.2	80.8	88.1	63.4	2.9
EPConv	91.5	81.4	88.7	64.2	3.0
GhostConv	89.2	82.2	86.9	62.2	2.7
GSConv	88.5	83.3	87.8	62.3	2.7

PConv achieves relatively high precision because its directional convolution branches are effective for extracting elongated edges and crack-like structures. However, standalone PConv lacks explicit channel-wise feature selection and may retain redundant responses produced by complex backgrounds.

EPConv combines directional spatial modeling with efficient multi-scale channel attention. Compared with standalone PConv, EPConv achieves higher recall, $mAP_{0.5}$, and mAP0.5:0.95. This result indicates that EMCA and PConv provide complementary functions: EMCA emphasizes defect-relevant channels and suppresses redundant

background responses, whereas PConv captures directional continuity and asymmetric boundary structures.

GhostConv and GSConv reduce the number of parameters, but their $mAP_{0.5:0.95}$ values are lower than those of EPConv. These lightweight convolution methods mainly focus on reducing redundant computation, whereas EPConv is specifically designed to strengthen the representation of weak and elongated defect features. Therefore, the improvement achieved by EPConv cannot be attributed solely to replacing standard convolution or changing model complexity.

The table further verifies the performance advantages of the improved YOLOv8n model in wind turbine blade defect detection, comparative experiments were conducted against other object detection models under the same experimental configuration and unified initial training parameters. Specifically, the proposed model was compared with Faster R-CNN, SSD, Hyper-YOLO [22], and several YOLO-series models. The comparative results are presented in Table 5.

Table 5. Model comparison experiment.

Model	P (%)	R (%)	$mAP_{0.5}$ (%)	$mAP_{0.5–0.95}$ (%)	FPS	GFLOPs	Params (M)	Size (MB)
Faster-RCNN	96.6	87.7	86.5	83.1	3.5	182.2	41.3	316.1
SSD	93.0	83.9	83.6	74.1	33.72	61.5	24.5	187.3
Hyper-YOLO	90.4	83.9	89.7	68.4	76.3	10.8	3.9	7.9
SLF-YOLO-*n*	88.2	80.0	85.7	62.5	249.88	7.3	2.6	5.5
GCB-YOLO	92.7	86.8	89.8	67.6	175.44	8.1	3.7	14.1
YOLOv8n	91.7	79.2	86.8	62.8	161.3	8.1	3.0	6.0
YOLOv11n	87.8	80.7	87.1	61.5	132.9	6.3	2.5	5.3
YOLO26n	78.5	71.2	78.2	44.7	54.5	5.2	2.4	5.2
Ours	92.1	85.1	90.7	68.0	77.1	8.9	3.0	6.1

Faster R-CNN achieved the highest precision of 96.6%, recall of 87.7%, and $mAP_{0.5:0.95}$ of 83.1%, demonstrating the localization capability of the two-stage detector. However, it required 182.2 GFLOPs, contained 41.3 M parameters, had a model size of 316.1 MB, and achieved only 3.5 FPS. These computational and storage requirements limit its suitability for real-time wind turbine inspection. SSD achieved a precision of 93.0% and a recall of 83.9%, but its $mAP_{0.5}$ was only 83.6%, while its computational complexity and model size reached 61.5 GFLOPs and 187.3 MB, respectively.

Among the evaluated lightweight one-stage detectors, the proposed model achieved the highest $mAP_{0.5}$ of 90.7%. Its precision and recall reached 92.1% and 85.1%, respectively, while its $mAP_{0.5:0.95}$ reached 68.0%. The model required 8.9 GFLOPs, contained approximately 3.0 M parameters, and had a model size of 6.1 MB. Compared with Hyper-YOLO, the proposed model improved $mAP_{0.5}$ from 89.7% to 90.7%. Compared with GCB-YOLO, $mAP_{0.5}$ increased from 89.8% to 90.7%, although GCB-YOLO achieved a higher recall of 86.8%. These results indicate that the proposed model provides a favorable balance between detection accuracy and computational complexity rather than outperforming every comparison method on every metric.

Compared with the original YOLOv8n, the proposed model increased precision by 0.4 percentage points, recall by 5.9 percentage points, $mAP_{0.5}$ by 3.9 percentage points, and $mAP_{0.5:0.95}$ by 5.2 percentage points. Meanwhile, FPS decreased from 161.3 to 77.1, corresponding to a throughput reduction of approximately 52.2%. The average network-forward latency consequently increased from approximately 6.20 ms to 12.97 ms per image. This result reflects a clear accuracy–efficiency trade-off: the proposed modules substantially reduce missed detections and improve localization, but they also introduce additional

inference latency. Nevertheless, the model-forward speed remains above the frame rate of typical 25–30 FPS inspection video on the tested GPU.

3.4.3. Visualization Analysis

Figure 9 presents representative detection results for multiple wind turbine blade defect categories under different background conditions. The confidence scores and predicted bounding boxes are retained to enable direct comparison between the original YOLOv8n and the improved model.

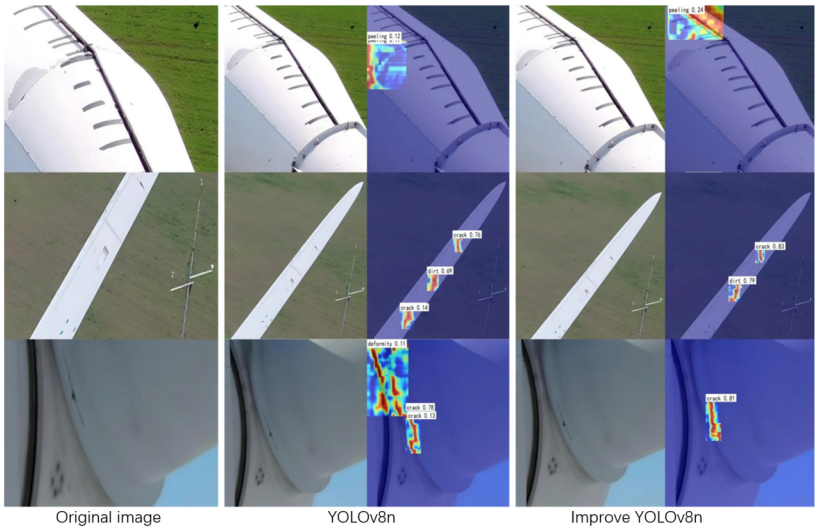


Figure 9. Representative successful and failure cases of the original YOLOv8n and the improved model. Confidence scores, missed detections, false positives, and duplicate detections are retained for comparison.

For coating-spalling and dirt defects with relatively clear regional structures, both models can identify the principal targets. However, the improved model produces more complete bounding boxes and generally higher confidence scores. For narrow crack defects, the baseline model is more likely to generate low-confidence predictions, incomplete localization, or missed detections because the crack texture is weak and spatially discontinuous. In contrast, the improved model better preserves the directional continuity of the crack and provides more accurate localization.

The improved model also reduces false-positive and duplicate detections in regions containing blade edges, shadows, stains, and complex background textures. This improvement can be attributed to the channel-refinement capability of MS-CBAM and EPConv, which suppresses irrelevant background responses while emphasizing defect-related features.

Nevertheless, failure cases remain. For example, when coating spalling occupies an extremely small region or has very low contrast with the blade surface, the improved model may still produce low-confidence predictions or fail to detect the defect. These cases indicate that the current method remains sensitive to extremely weak visual features and severe scale variation. Therefore, the visualization results demonstrate both the effectiveness and the remaining limitations of the proposed method.

3.4.4. Generalization Experiments

To further evaluate the generalization performance of the improved YOLOv8n model across different defect detection scenarios, generalization experiments were conducted using the same experimental parameter settings as those described in Section 3.2. The publicly available PCB defect dataset from Peking University [23] was selected for evalua-

tion. The dataset contains 693 images covering six defect categories: missing hole, mouse bite, open circuit, short circuit, spur, and spurious copper. The dataset was subsequently divided into training, validation, and test sets at a ratio of 8:1:1. The PCB dataset was selected because several PCB defects share morphological characteristics with wind turbine blade defects despite the substantial difference in visual domain. Open-circuit, short-circuit, spur, and spurious-copper defects frequently exhibit narrow structures, small local regions, weak contrast, or irregular boundaries. Therefore, this dataset can be used to examine whether the multi-scale attention, directional feature extraction, and bounding-box regression mechanisms introduced in the proposed model are applicable to another industrial surface-inspection task. The corresponding experimental results are presented in Table 6.

Table 6. Class-wise $AP_{0.5}$ results on the PCB defect dataset.

Model	$AP_{0.5}$ (%)						$mAP_{0.5}$ (%)
	Missing Hole	Mouse Bite	Open Circuit	Short Circuit	Spur	Spurious Copper	
YOLOv8n	98.6	84.4	94.0	97.4	84.3	91.5	90.7
Ours	99.5	85.2	94.8	97.7	87.2	91.7	92.7

As shown in Table 6, the original YOLOv8n achieved $AP_{0.5}$ values of 98.6%, 84.4%, 94.0%, 97.4%, 84.3%, and 91.5% for missing hole, mouse bite, open circuit, short circuit, spur, and spurious copper, respectively. The corresponding values obtained by the proposed model were 99.5%, 85.2%, 94.8%, 97.7%, 87.2%, and 91.7%. Based on the arithmetic mean of the six class-wise AP values, $mAP_{0.5}$ increased from 91.7% to 92.7%, corresponding to an improvement of 1.0 percentage point. The largest improvement was observed for the spur category, for which $AP_{0.5}$ increased by 2.9 percentage points. This result is consistent with the intended capability of EPConv and PIoU v2 to improve the representation and localization of narrow and irregular defect structures.

The PCB experiment evaluates cross-dataset architectural applicability rather than zero-shot domain adaptation. Both models were retrained on the PCB training set before evaluation, and PCB images differ substantially from UAV-acquired blade images in background texture, imaging distance, environmental interference, and defect appearance. Therefore, the PCB results do not demonstrate generalization across different wind farms or UAV acquisition conditions. Instead, they provide supporting evidence that the proposed architectural modifications can also benefit another industrial surface-defect detection task.

4. Conclusions

Wind turbine blade defect detection still faces several challenges, including low detection accuracy, frequent missed detections of small defects, and a high false detection rate under complex backgrounds. To address these issues, this study improves the YOLOv8n model. First, the MS-CBAM multi-scale convolutional attention module is introduced into the backbone network to enhance feature extraction and localization for small defects and defects with complex shapes. Subsequently, four stride-2 downsampling convolutional layers in the backbone are selectively replaced with EPConv to enhance channel-wise feature selection and directional feature extraction while controlling computational overhead. Finally, the original CIoU loss function is replaced with PIoU v2 to optimize the bounding box regression process, thereby improving defect localization accuracy and accelerating model convergence during training. The improved model achieves P, R, and $mAP_{0.5}$ values of 92.1%, 85.1%, and 90.7%, respectively. computational complexity of only 8.9 GFLOPs and an inference speed of 77.1 frames·s^{-1}.

Nevertheless, the improved model introduces a moderate increase in computational complexity and a reduction in inference speed compared with the baseline. In addition, practical deployment should incorporate secure image transmission and storage, restricted data access, model-integrity verification, confidence-based warnings, and human verification of safety-critical detections. Future work will establish a larger dataset with blade-, turbine-, and sequence-level identifiers and will conduct size-stratified evaluation and bounding-box geometry analysis. Model pruning, quantization, and knowledge distillation will also be investigated to improve deployment efficiency while maintaining detection accuracy.

Author Contributions: Conceptualization, H.R. and Z.T.; methodology, H.R.; software, H.R.; validation, H.R. and Z.T.; formal analysis, H.R.; investigation, H.R.; resources, Z.T.; data curation, H.R.; writing—original draft preparation, H.R.; writing—review and editing, H.R. and Z.T.; visualization, H.R.; supervision, Z.T.; project administration, Z.T.; funding acquisition, Z.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw UAV inspection images of wind turbine blades supporting the conclusions of this study are available on request from the corresponding author. The dataset involves real commercial wind farm operation data and is restricted by the confidentiality agreement signed with the wind power enterprise; thus, full public release is not allowed. Non-commercial academic researchers can contact the corresponding author and sign a confidentiality agreement to obtain the complete dataset for research purposes.

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

References

1. Song, Y.; Wu, Y. Review of surface defect detection of wind turbine blades based on UAV aerial photography. *Chin. J. Sci. Instrum.* **2024**, *45*, 1–25.
2. Chatterjee, J.; Dethlefs, N. Scientometric review of artificial intelligence for operations & maintenance of wind turbines: The past, present and future. *Renew. Sustain. Energy Rev.* **2021**, *144*, 111051. [CrossRef]
3. Wang, Y.; Han, Q.; Zhang, X.; Wu, C.; Yang, M. Wind turbine blade damage detection system based on machine vision. *Sci. Technol. Eng.* **2022**, *22*, 4879–4886.
4. He, B.; Jia, J.H.; Zhao, F.; Song, X. Application of UAV in wind turbine blade detection. *Electr. Eng.* **2019**, *40*, 64–65.
5. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers. *Reliab. Eng. Syst. Saf.* **2019**, *184*, 2–12. [CrossRef]
6. Tcherniak, D.; Mølgaard, L.L. Active vibration-based structural health monitoring system for wind turbine blade: Demonstration on an operating Vestas V27 wind turbine. *Struct. Health Monit.* **2017**, *16*, 536–550. [CrossRef]
7. Liu, Z.; Liu, X.; Zhu, S.-P.; Zhu, P.; Liu, W.; Correia, J.A.; De Jesus, A.M. Reliability assessment of measurement accuracy for FBG sensors used in structural tests of the wind turbine blades based on strain transfer laws. *Eng. Fail. Anal.* **2020**, *112*, 104506. [CrossRef]
8. Ma, M.; Liu, H.; Gao, Y.; Li, S.; Ma, C.; Hu, D. Fully autonomous inspection of wind turbine blades based on UAV and artificial intelligence. In *Proceedings of the ISIA 2023: 2023 International Conference on Intelligent Sensing and Industrial Automation, Virtual Event, China, 9–10 December 2023*; ACM: New York, NY, USA, 2023; pp. 44–48.
9. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
10. Shi, L.; Long, Y.; Wang, Y.; Chen, X.; Zhao, Q. Evaluation of internal cracks in turbine blade thermal barrier coating using enhanced multi-scale Faster R-CNN model. *Appl. Sci.* **2022**, *12*, 6446. [CrossRef]
11. Chen, L.; Mei, H.; Hu, H.; Zhao, K. Weld defect detection method based on improved Faster R-CNN. *Sci. Technol. Eng.* **2025**, *25*, 2027–2033.

12. Liu, Y.; Si, L.; Wang, Z.; Chen, M.; Li, X.; Wei, D.; Gu, J. A novel coal-rock recognition method in coal mining face based on fusing laser point cloud and images. *Int. J. Min. Sci. Technol.* **2025**, *35*, 1057–1071. [CrossRef]
13. Lv, L.; Yao, Z.; Wang, E.; Ren, X.; Pang, R.; Wang, H.; Zhang, Y.; Wu, H. Efficient and accurate damage detector for wind turbine blade images. *IEEE Access* **2022**, *10*, 123378–123386. [CrossRef]
14. Sun, Z.; Wang, X. RGS-YOLO: A lightweight solution for surface defect detection in wind turbines. *PLoS ONE* **2026**, *21*, e0343966. [CrossRef] [PubMed]
15. Tang, Z.; Zhang, C.; Wang, J.; Lu, P.; Liu, H.; Jian, H. Research on wind turbine blade defect detection based on improved YOLOv7. *J. Beijing Univ. Aeronaut. Astronaut.* **2026**, *52*, 1903–1914.
16. Zhang, Z.; Dong, C.; Wei, Z.; Chen, X.; Zan, W.; Xue, Y. GCB-YOLO: A lightweight algorithm for wind turbine blade defect detection. *Wind Energy* **2025**, *28*, e70029. [CrossRef]
17. Bakr, E.M.; El-sallab, A.; Rashwan, M. EMCA: Efficient multiscale channel attention module. *IEEE Access* **2022**, *10*, 103447–103461. [CrossRef]
18. Yang, J.; Liu, S.; Wu, J.; Su, X.; Hai, N.; Huang, X. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2025; Volume 39, pp. 9202–9210.
19. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
20. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
21. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; Hu, S.-M. SegNeXt: Rethinking convolutional attention design for semantic segmentation. In *Proceedings of the Advances in Neural Information Processing Systems 35*; NeurIPS: New Orleans, LA, USA, 2022; Volume 35, pp. 1140–1156.
22. Feng, Y.; Huang, J.; Du, S.; Ying, S.; Yong, J.-H.; Li, Y.; Ding, G.; Ji, R.; Gao, Y. Hyper-YOLO: When visual object detection meets hypergraph computation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *47*, 2388–2401. [CrossRef] [PubMed]
23. Cai, X.; Wu, J. PCB bare board defect detection based on improved YOLOv7-tiny. In *Proceedings of the 2023 China Automation Congress*; IEEE: New York, NY, USA, 2023; pp. 5768–5773.

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.