Sensors and Materials, Vol. 38, No. 5 (2026) 2381–2401
MYU Tokyo

S & M 4443

Wind Turbine Blades Defect Detection Based on Improved YOLOv8

Wen Wang,[1] Litao Xiao,[1] Lifu He,[1] Ji Jiang,[1] Yang Lyu,[2]
Wen Zou,[2] Haowei Xiong,[3] Baotong Chi,[3] and Wenlong Fu[3*]

[1]Hubei Energy Group Huangshi Wind Power Co., Ltd., Huangshi, Hubei 435000, China
[2]Power China Huadong Engineering Co., Ltd., Hangzhou, Zhejiang 311122, China
[3]College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, Hubei 443002, China

(Received August 12, 2025; accepted December 12, 2025)

Keywords: wind turbine blades, defect detection, improved YOLOv8, mixed local channel attention mechanism, pixels-IoU

Wind turbine blades endure persistent operational stresses, including aerodynamic loads, cyclic fatigue, and environmental corrosion, leading to structural defects such as cracks, perforations, and surface delamination. These defects impair aerodynamic performance, reduce energy output, and may propagate over time, increasing the risk of blade fracture. To find these defects in time, a novel defect detection method for wind turbine blades based on improved YOLOv8 is proposed in this paper. First, standard convolutional layers are systematically replaced with spatial pyramid depthwise convolution modules in the backbone network to improve the recognition of slender and microscale crack defects, and the proposed method enhances defect detection accuracy through the mixed local channel attention mechanism. Additionally, the conventional spatial pyramid pooling (SPP) structure is further redesigned as the Spatial Pyramid Pooling Cross Stage Partial Connections (SPPCSPC) architecture by integrating SPP with cross-stage partial connection (CSPC), thereby enhancing multi-scale feature representation. Finally, the complete intersection over union (IoU) loss is optimized into Pixels-IoU formulation for bounding box regression, which significantly improves detection performance for small targets while ensuring regression accuracy and stability. Experimental results on a wind turbine blade crack defect dataset demonstrate that the proposed method achieves improvements in precision of 9.3% and in mean average precision of 7.3% compared with the baseline YOLOv8 model. These findings validate that the enhanced You Only Look Once v8 (YOLOv8) detection method exhibits superior detection effectiveness, accuracy, and reliability.

1. Introduction

Among various clean energy alternatives, wind power demonstrates exceptional economic viability and sustainability advantages. The effective harnessing of this resource presents a dual

*Corresponding author: e-mail: ctgu_fuwenlong@126.com
https://doi.org/10.18494/SAM5877

ISSN 0914-4935

solution to energy security concerns and climate change mitigation, positioning it as a key enabler for meeting the determined carbon reduction commitments.[1–3] Wind turbine blades, serving as the primary energy conversion mechanism, transform aerodynamic forces into rotational motion. To endure high mechanical stresses during operation, they are commonly manufactured from advanced composite materials, including glass-fiber-reinforced polymers and carbon-fiber-reinforced plastics, which offer an optimal balance between strength and weight. However, as wind power deployment expands into regions with complex terrain, harsh climates, and diverse environmental conditions, the long-term operation of turbine blades often leads to structural defects such as cracks, perforations, and surface delamination.[4] These defects can affect wind turbine blades in two major ways. On the one hand, the progressive accumulation of structural flaws weakens the load-bearing capacity of the blade. When these defects reach a critical threshold, they can lead to structural failure, increased operational noise, and reduced environmental compatibility. Furthermore, compromised blade integrity may induce abnormal vibrations and damage other turbine components. On the other hand, the size and geometry of surface defects can significantly alter the blade's aerodynamic profile, resulting in increased drag and reduced energy conversion efficiency.[5] Therefore, the early identification and evaluation of minor imperfections play a crucial role in optimizing power generation output while prolonging wind turbine service life.

For extended periods, the defects in wind turbine blades have been predominantly detected by manual inspection. While this method does not require costly equipment, it demands substantial time and effort from technicians for preparation and execution. The inspection process is entirely dependent on visual observation, particularly in identifying and classifying various types of defect, making the results highly contingent upon the operator's experience. Although manual inspection offers advantages in terms of low equipment cost, its performance in safety and diagnostic accuracy remains limited and requires significant improvement.[6]

With the advancement of sensor technologies and improvements in detection algorithms, vibration analysis, ultrasonic testing, and infrared thermographic inspection have been widely applied in the detection of wind turbine blade defects. These methods typically involve installing external sensors on the blades to collect operational data, which are then analyzed to identify the type and location of defects. For example, Tcherniak and Mølgaard[7] proposed a semisupervised learning approach for remotely detecting small-scale defects. However, the effectiveness of methods based on vibration analysis is heavily dependent on the layout of the vibration sensor system. Khadka *et al.*[8] introduced a system utilizing stereo cameras to monitor blade motion and reduce manual intervention, but its accuracy and stability decrease significantly at high rotational speeds. Ultrasonic testing, a widely used nondestructive evaluation technique, infers defect characteristics on the basis of differences in acoustic wave reflections across media. Tiwari and Raisutis[9] developed an ultrasonic signal-based defect detection method that enhanced wind turbine blade defect detection performance by employing discrete wavelet transform, variational mode decomposition, and Hilbert transform. Oliveira *et al.*[10] developed a hybrid approach employing wavelet transform denoising followed by principal component decomposition, enabling comprehensive damage assessment via noninvasive testing combined with single-class classification. Infrared thermography leverages infrared sensors to detect

surface radiation and produce thermal images that reveal subtle surface defects. Arora *et al.*[11] employed modulated infrared imaging with multi-frequency, low-power, coded excitation to improve defect detection under noisy conditions. Similarly, Arora *et al.*[12] also adopted the time-domain phase-based digital frequency modulation approach, which improved scanning depth, resolution, and sensitivity. Despite the transition from manual to sensor-based methods, the effectiveness of these approaches is often constrained by sensor performance under varying operational conditions, thus limiting their overall practical applicability.

Recent advances in defect detection have been increasingly dominated by deep-learning-based methodologies. These approaches are generally classified into two-stage and one-stage detection frameworks, with the former category primarily represented by R-CNN and its variants.[13] For example, Mao *et al.*[14] developed a vision-based inspection model leveraging a cascaded R-CNN framework, which automatically and accurately classifies and localizes surface defects, demonstrating strong inspection performance. Diaz and Tittuss[15] employed depth-wise separable convolutions in the backbone of a cascaded Mask R-CNN to reduce computational cost while preserving detection accuracy. While R-CNN-based models offer high detection precision, they rely on a region proposal network to generate candidate regions during inference, resulting in significant computational overhead and reduced processing speed. Consequently, their application in real-time or large-scale industrial scenarios remains limited.

Compared with two-stage detection methods, one-stage detection approaches enable end-to-end processing by simultaneously performing feature extraction, object localization, and confidence score regression on the entire image. One-stage detection methods are primarily implemented using You Only Look Once (YOLO) series models,[16] such as YOLOv3, YOLOv5, and YOLOv8. For example, Zhang and Wen[17] employed foreground extraction combined with Hough transformation as preliminary image processing steps. The enhanced YOLOv5 framework incorporated a dedicated micro-feature detection module, with anchor box optimization achieved through K-means clustering, effectively addressing the challenge of feature loss in detecting minute and geometrically complex defects. Liu *et al.*[18] proposed an improved YOLOX architecture featuring a dual-branch attention module for simultaneous channel and spatial feature extraction, coupled with an adaptive decay-based classification loss to mitigate class distribution skewness in defect detection. Ma *et al.*[19] introduced a multi-kernel block convolution module (M_C2f) and integrated a small object detection head with a bidirectional feature pyramid network into YOLOv8. This design not only enhanced the accuracy of small defect detection but also substantially reduced the number of model parameters. Liu *et al.*[20] improved detection performance while accelerating model inference by adopting the GhostNetv2 lightweight backbone network in YOLOv5, combined with structural pruning and knowledge distillation techniques. He *et al.*[21] achieved the precise localization and classification of multi-scale defects by incorporating a multi-level feature fusion module and an adaptive multi-scale bounding box generation module, demonstrating excellent detection accuracy and localization performance on real-world wind power scene datasets. Liu and Liu[22] proposed the GFD-YOLOX model, which incorporates a dual-frequency fusion bidirectional feature pyramid network (DF-BiFPN) and a hierarchical frequency adaptive fusion (HFAF) module. This architecture effectively enhances multi-scale feature representation and improves

both detection accuracy and robustness for defects of varying sizes in wind turbine blades. Tong *et al.*[23] integrated a multi-head self-attention mechanism via the MHSA-C2f module to capture global contextual information and adopted the Mini-BiFPN structure to improve the retention of small targets in shallow layers. These enhancements effectively improved both the detection accuracy and real-time performance for small-sized defects in wind turbine blades. Although the above defect detection approaches have yielded promising results by effectively balancing detection accuracy and processing speed, their performance remains limited in scenarios involving complex backgrounds and subtle defect features. Therefore, improving detection accuracy under complex scenarios remains a critical challenge in wind turbine blade defect detection.

To improve detection accuracy under complex scenarios, a novel defect detection method for wind turbine blades is proposed on the basis of the improved YOLOv8 framework. This approach enhances the detection accuracy for slender and subtle cracks by incorporating an attention mechanism into the YOLOv8 backbone network and by replacing conventional convolutional modules and loss functions to enhance the performance of the model. Experimental results demonstrate that the proposed improved YOLOv8 defect detection method achieves a higher detection accuracy than the existing YOLO approaches. The main contributions are as follows:

(1) The detection method's sensitivity to subtle defects in complex backgrounds is enhanced through the integration of the mixed local channel attention (MLCA) mechanism combined with the spatial pyramid depthwise convolution (SPD-Conv) module.
(2) The Spatial Pyramid Pooling Cross Stage Partial Connections (SPPCSPC) module is introduced into YOLOv8n. The enhancement of the network's multi-scale defect perception and representational capability is achieved by embedding the cross stage partial connection (CSPC) architecture within the conventional spatial pyramid pooling (SPP) structure.
(3) Detection performance for small targets is optimized through the implementation of a novel Pixels-IoU (PIoU) loss function, which simultaneously improves regression accuracy and stability.

2. YOLOv8 Structure and Improvements

2.1 YOLOv8

Since its initial proposal in 2015, the YOLO series has emerged as one of the most influential architectures in one-stage object detection. Compared with conventional object detection models, superior performance is exhibited by the YOLO model through its implementation of an end-to-end detection framework utilizing single-pass forward propagation. In this model, target bounding boxes and class probabilities are directly predicted through regression, enabling highly efficient detection. Multi-scale feature extraction is performed by convolutional neural networks, complemented by an innovatively designed grid partitioning approach for precise object localization. In addition, compared with previous iterations such as YOLOv3, YOLOv5,[24,25] and YOLOv7,[26] YOLOv8 exhibits outstanding detection speed and enhanced accuracy.[27] The

YOLOv8 framework comprises five distinct configurations: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. These versions are differentiated primarily by the complexity of the model, which is reflected in the depth and width of the network. Specifically, the number of residual modules is progressively increased, thereby enhancing the model's capability for feature extraction and fusion.

As the model version advances, detection accuracy is correspondingly improved by applying these versions. However, this enhancement is achieved at the expense of increased processing time. Among these versions, YOLOv8n, being the most lightweight variant, distinguishes itself through exceptional detection speed, albeit with slightly compromised accuracy. Conversely, the YOLOv8s to YOLOv8x series demonstrates significant accuracy improvements, but the incorporation of excessive residual structures introduces additional computational overhead, resulting in prolonged detection cycles. In practical real-time monitoring scenarios, temporal delays may constitute a detrimental factor, adversely affecting system responsiveness and practical utility. In light of these considerations, we decided to employ YOLOv8n as the backbone network for targeted optimization and propose a detection algorithm that simultaneously balances speed and accuracy. The network architecture of basic YOLOv8n is illustrated in Fig. 1.

As the next-generation real-time detection framework developed by Ultralytics, YOLOv8n adopts a four-stage modular design comprising data preprocessing, feature extraction network, feature fusion network, and prediction output modules. In the data preprocessing stage, mosaic augmentation strategy and dynamic anchor box adjustment are employed to effectively enhance the spatial representation capability of input data. The feature extraction network progressively

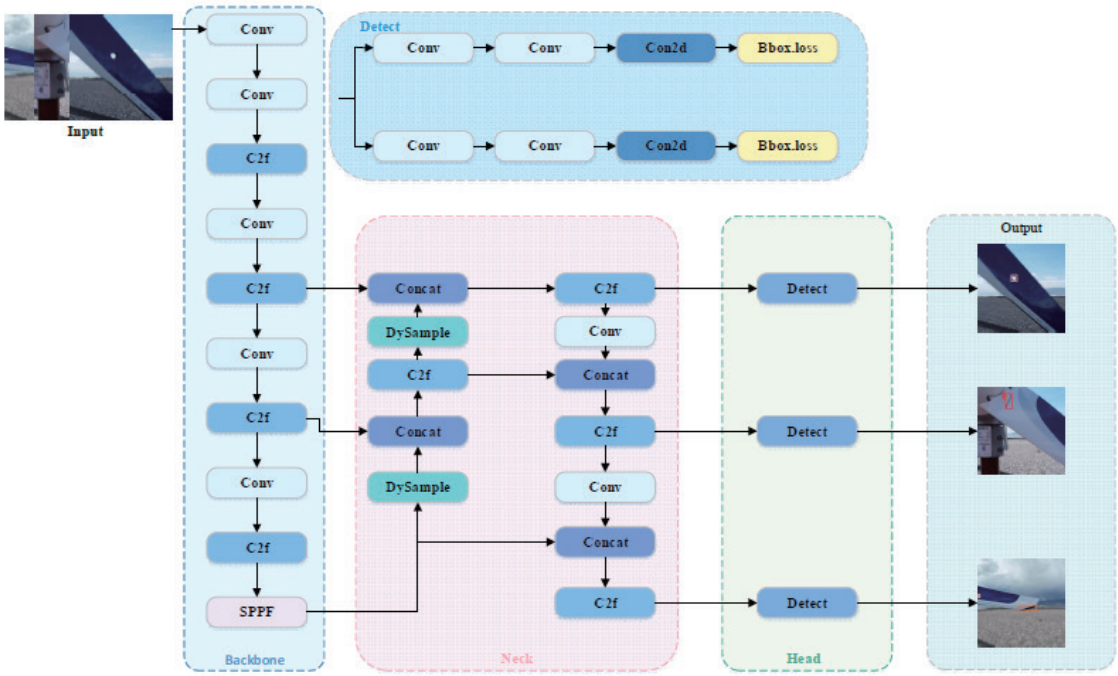

Fig. 1. (Color online) YOLOv8n architecture diagram.

extracts spatial-semantic information from images through a multi-level convolutional neural network, forming a multi-scale feature pyramid that incorporates edge details, texture features, and global context. During the feature fusion stage, a bidirectional feature pyramid structure is utilized, enabling the cross-level interaction of features with different receptive fields through top-down and bottom-up bidirectional feature propagation paths. The prediction output module adopts an anchor-free detection approach based on key points, employing a decoupled dual-branch architecture to separately handle object localization and classification tasks, which eliminates the need for a manual anchor box design based on empirical knowledge, thereby reducing interference caused by the selection of anchor box quantities and dimensions and improving detection accuracy. In terms of loss function design, the YOLOv8n model integrates classification loss and regression loss. The classification loss retains the binary cross-entropy loss function, whereas the regression loss incorporates distribution focal loss and complete intersection over union (CIoU) loss to further enhance localization precision. Additionally, the C2f module in YOLOv8n optimizes the multi-branch connectivity architecture of the efficient layer aggregation network module. By employing streamlined convolutional configurations, model lightweighting is achieved while simultaneously improving the efficiency of gradient information propagation.

2.2 Improved YOLOv8

The proposed improved YOLOv8 network structure is illustrated in Fig. 2. Specifically, the original network is optimized through the following four key innovations.

First, the MLCA mechanism is introduced to strengthen defect feature recognition. Second, standard convolution operations in the backbone network are replaced with SPD-Conv, significantly improving the model's capability to characterize slender cracks and tiny defects. Third, the original spatial pyramid pooling fast (SPPF) module is enhanced by integrating SPP with CSPC, forming an SPPCSPC structure, which effectively strengthens the model's multi-scale feature capturing ability. Finally, the CIoU loss is replaced with the PIoU loss, optimizing the bounding box regression mechanism and significantly improving the accuracy and training stability for small object detection.

2.3 MLCA

Attention mechanisms are often applied in the field of computer vision, aiming to selectively amplify important features while attenuating distracting elements in neural networks. Their integration into deep architectures has proven effective for defect detection tasks, where they help models concentrate on relevant patterns while filtering out background interference. However, most channel attention methods only leverage channel-wise features, neglecting crucial spatial relationships. This limitation often results in degraded model performance. Spatial attention approaches, while addressing this limitation, tend to significantly increase computational complexity and implementation difficulty.

In this paper, we introduce the MLCA mechanism into the improved YOLOv8,[28] aiming to balance performance with computational cost while strengthening object detection networks.

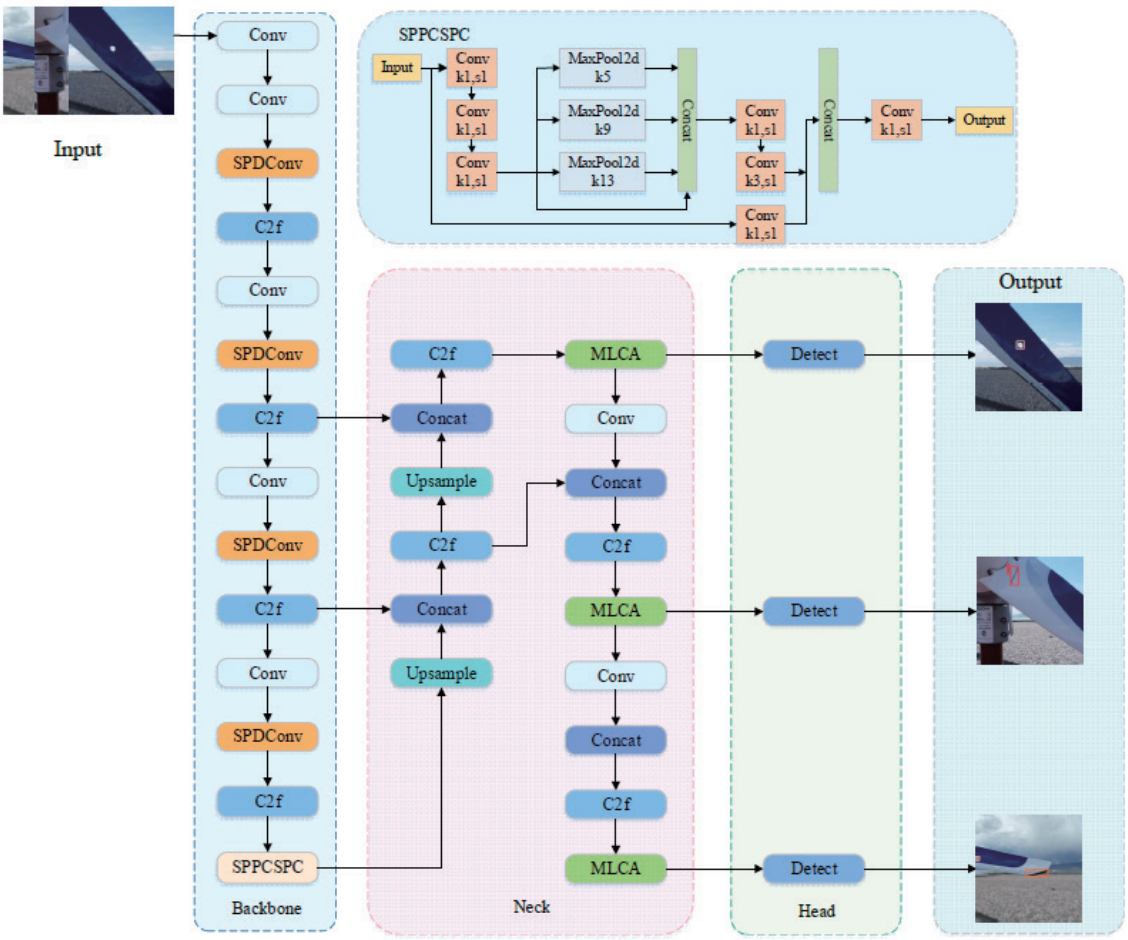

Fig. 2. (Color online) Improved YOLOv8 architecture diagram.

The module integrates channel-wise and spatial features concurrently, making use of both local and global contexts to boost detection accuracy.

Existing attention mechanisms, such as squeeze and excitation mechanisms, and the efficient channel attention mechanism, primarily focus on channel-wise weight allocation. However, this process may lead to imbalanced weight distribution, where certain channels are assigned excessively high weights while others are suppressed, potentially causing critical features in low-weight channels to be overlooked. In contrast, the MLCA mechanism demonstrates outstanding performance in evaluating channel importance, enabling a more effective extraction of discriminative features. The MLCA mechanism not only optimizes the model's performance in image processing tasks but also significantly enhances its overall efficiency.

The structure of MLCA is shown in Fig. 3. For the input feature map, preliminary processing is first performed using local average pooling (LAP) and global average pooling (GAP). The input feature vector is transformed into a $1 \times C \times k \times k$ dimensional vector, where k denotes the convolutional kernel size. The LAP operation aggregates neighboring features to perform downsampling, which effectively enhances the model's focus on essential local patterns. GAP

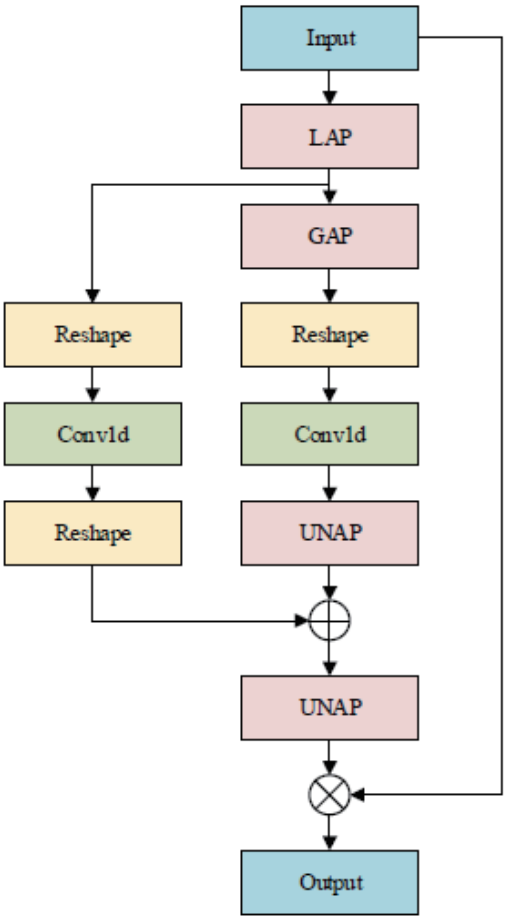

Fig. 3. (Color online) Architectural diagram of MLCA.

summarizes overall feature statistics through spatial averaging, which helps preserve broader contextual relationships. These two branches respectively concentrate on local regional features and global statistical characteristics, providing comprehensive feature capturing.

Following processing by LAP and GAP, the feature information is subjected to sorting operations. Within the MLCA module, the input features are transformed into one-dimensional vectors through two distinct branches: one dedicated to capturing global contextual information, while the other preserves localized spatial features. The globally pooled features are processed through 1 × 1 convolution and reshaped, followed by element-wise addition with locally pooled features. This additive operation, performed position-wise between corresponding elements of both feature maps, effectively integrates global contextual information. The combined features then undergo transformation through Conv1d layers, which compresses the feature channels while maintaining their spatial dimensions. The Conv1d operation in the MLCA architecture aims to capture interaction information between each channel and its k-sized neighborhood channels, where the kernel size k is determined on the basis of the channel dimension C. The kernel size k is derived from the following Eq. (1).

$$k = \Phi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{1}$$

In this equation, C corresponds to the channel count, whereas k represents the kernel width. γ and b are predefined hyperparameters, typically initialized to 2. The subscript "odd" enforces an odd-valued constraint; if the computed k is even, it is automatically incremented by one to satisfy this requirement.

After Conv1d operation, the locally and globally attended feature maps are restored to their original spatial dimensions via un-average pooling (UNAP). These are then multiplied element-wise with the original input features, generating an enhanced feature map that emphasizes relevant characteristics. The feature maps processed by MLCA represent a comprehensive understanding of feature information and an integrated utilization of global contextual knowledge. Serving as input for subsequent defect detection, these enhanced feature maps enable a more intelligent identification of wind turbine blade defects.

2.4 SPD-Conv

In wind turbine blade defect detection scenarios, defect objects are typically small, making it challenging for standard convolutional operations to effectively extract discriminative features, and often leads to frequent missed detections and false alarms, while also increasing susceptibility to background interference, thereby compromising algorithmic robustness. To overcome these limitations, the original stride convolutions with step size two and pooling layers in the YOLOv8n architecture are replaced with SPD-Conv modules in this paper.[29] The SPD-Conv module is specifically designed for small-object and multi-object detection tasks, which is particularly suitable for addressing the issues of fine-grained information loss and low feature learning efficiency that frequently occur in traditional feature extraction methods.

The SPD-Conv module performs downsampling for feature to maintain critical channel information, followed by the execution of nonstride convolution to optimally utilize learnable parameters in the additional convolutional layers through channel refinement. The SPD layer is specifically designed to compress the spatial dimensions of the input and redistribute the encoded information into the channel, thereby preserving critical feature information. The subsequent Conv layer performs conventional convolution operations on each feature map, which serves to mitigate potential oversampling issues that may occur in the SPD layer while maintaining fine-grained details. Through this architecture, spatial dimension reduction is achieved without information loss, with the simultaneous preservation of channel-wise information.

The structure of the SPD-Conv module is shown in Fig. 4, where X refers to the original input, X' denotes an intermediate tensor, and X'' indicates the final output. Components A, B, and C indicate spatial expansion operations, whereas D signifies a standard convolution operation with stride 1. Given an input tensor X with shape (S, S, C_1), a downsampling step generates four distinct partitions ($f_{0,0}, f_{1,0}, f_{0,1}, f_{1,1}$), each of size (S/2, S/2, C_1). These partitions are then merged along the channel axis, forming an intermediate representation denoted as X'.

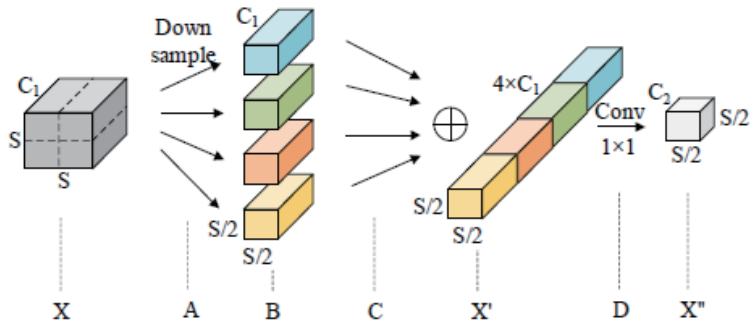

Fig. 4. (Color online) Architectural diagram of SPD-Conv.

This step effectively halves the spatial resolution while quadrupling the channel count. The SPD operation restructures the original tensor into X′ of shape (S/2, S/2, $4C_1$). After this rearrangement, X′ is passed through a convolution layer to obtain the refined output, referred to as X'' with dimensions (S/2, S/2, C_2). This transformation is carefully designed to preserve all discriminative feature information. Notably, the employed convolution operation is nonstride, ensuring that no additional downsampling occurs and that the maximal granularity of the input features is maintained throughout the process.

2.5 SPPCSPC

The SPP module is one of the core components of YOLOv5. This module enables the conversion of feature maps of arbitrary sizes into fixed-size feature vectors, effectively eliminating the necessity for cropping and scaling operations on image regions, thereby mitigating issues such as image distortion. Moreover, the SPP module overcomes the limitations of convolutional neural networks in extracting repetitive image features, significantly enhancing the speed of candidate region generation while reducing computational costs. Building upon the SPP architecture, the YOLOv7 algorithm proposes the SPPCSPC structure, which incorporates the concept of CSPC. The CSPC structure is designed to improve module performance while reducing network parameters and computational complexity. By utilizing partial connectivity, CSPC divides the feature map into two components: one component is directly propagated to the subsequent stage, whereas the other undergoes processing through a series of convolutional layers before being fused with the former. To enhance the feature extraction capability of the network for wind turbine blade defects, the SPPCSPC module is introduced into YOLOv8n in this paper. By integrating the advantages of both the SPP and CSPC structures, this module establishes a cross-stage connection mechanism between low-level and high-level features. This design not only facilitates the effective transmission of multi-level features but also significantly improves the feature representation capacity and receptive field range of the network. Consequently, the algorithm is enabled to better adapt to blade defect detection tasks across varying image resolutions.

The SPPCSPC module is designed to incorporate multi-scale MaxPool operations with varying kernel sizes within the sequential convolutional process. This approach effectively

mitigates image distortion artifacts that may be caused by repeated convolution or resizing operations, while simultaneously addressing the limitations of conventional convolutional networks in extracting redundant features. Through dilated MaxPool operations, diverse receptive fields are obtained from the input feature layers and subsequently fused with features generated by standard convolutional operations, thereby significantly enhancing the network's generalization capability. Consequently, in this study, we replaced the YOLOv8's SPPF module with the SPPCSPC module to enhance multi-scale feature representation.

The structure of the SPPCSPC module is depicted in Fig. 5. Initially, the input is split into two branches using the cross-stage partial mechanism. One branch proceeds through standard convolution operations, whereas the other is fed into a series of convolutions with kernels of sizes 1 × 1, 3 × 3, and again 1 × 1, after which SPP is applied. In this stage, features are extracted via max pooling layers with increasingly large kernels of 5 × 5, 9 × 9, and 13 × 13, enabling the capture of multi-scale spatial information. These outputs are concatenated across the channel axis, then passed through additional 1 × 1 and 3 × 3 convolutions for refinement. The final feature representations from both paths are fused and further refined using a concluding 1 × 1 convolution. This architecture reduces computational overhead by 50%, while simultaneously improving inference speed and detection accuracy.

2.6 PIoU

The primary function of the bounding box loss function is to quantify the discrepancy between predicted and ground-truth bounding box parameters, thereby enhancing object localization accuracy. In YOLOv8n, the CIoU loss function is employed, which comprehensively evaluates localization regression by considering aspect ratio differences,[30] overlapping regions, and centroid distances between predicted and ground-truth boxes. The mathematical formulation of CIoU is presented in Eqs. (2)–(4).

$$L_{CIoU} = 1 - IoU + \frac{d^2}{c^2} + \alpha v \tag{2}$$

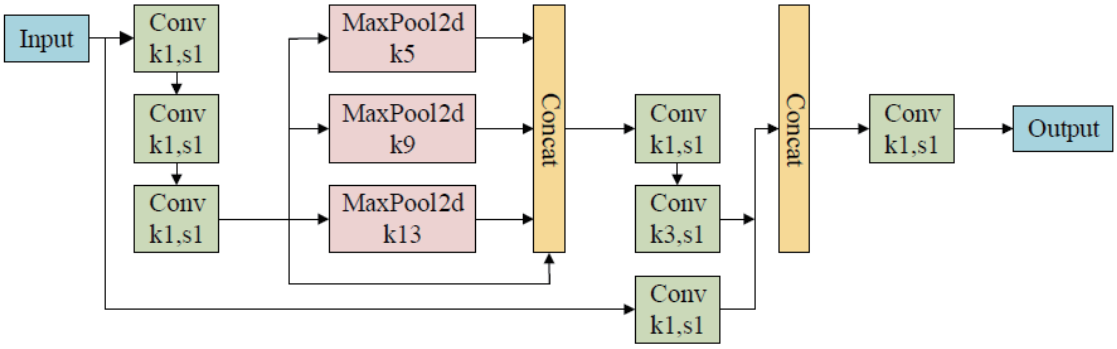

Fig. 5. (Color online) Architectural diagram of SPPCSPC.

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w_{gt}}{h_{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{4}$$

Here, the IoU metric measures the degree of overlap between predicted bounding boxes and their ground-truth counterparts. The Euclidean distance d quantifies the spatial separation between the centroids of the predicted and ground-truth boxes, whereas c denotes the diagonal length of the minimal enclosing rectangle that encompasses both boxes. To assess shape alignment, the term v captures the aspect ratio discrepancy between the boxes. Additionally, α acts as an adaptive parameter to harmonize the effects of overlap area and geometric dissimilarity in the CloU computation. This design guarantees that CloU integrates both spatial coverage and shape fidelity. In this context, w_{gt} and h_{gt} represent the ground-truth box's width and height, respectively, whereas w and h are the corresponding dimensions of the predicted box.

Although the CIoU loss function is designed to enhance bounding box regression performance, certain limitations have been identified regarding box expansion handling and size adaptability. First, the CIoU metric demonstrates suboptimal performance when processing targets with significantly divergent aspect ratios, particularly in scenarios involving complex shapes or substantial target variations, potentially leading to localization inaccuracies. Second, owing to the inherent complexity of the CIoU loss function, gradient updates during training may exhibit instability, increasing optimization difficulty. Third, when bounding boxes are distantly separated or exhibit minimal overlap, the localization effectiveness of CIoU remains relatively weak, failing to provide substantial precision improvement. Furthermore, its limited adaptability to varying target sizes restricts flexible loss computation across different object scales.

In wind turbine blade defect detection tasks, additional challenges are introduced by complex backgrounds and suboptimal imaging conditions, often resulting in degraded dataset samples. Moreover, substantial variations in defect morphologies lead to highly diverse bounding box shapes with unstable aspect ratios. Consequently, the penalty mechanism of the original YOLOv8n's CIoU loss is demonstrated to be insufficient, exhibiting an overly smooth loss curvature that ultimately compromises model generalizability and detection performance for wind turbine blade defects. To address these limitations, the conventional CIoU loss function in the original network is replaced with an enhanced PIoU loss function in this paper. The PIoU loss is specifically designed to mitigate the adverse effects of inappropriate penalty factors inherent in existing IoU-based loss functions, which frequently lead to undesired bounding box expansion and significantly slower convergence during regression. This is achieved through the integration of a target-size adaptive penalty factor with a gradient adjustment mechanism based on bounding box quality metrics.

The PIoU loss function guides bounding boxes along more efficient regression paths, thereby achieving faster convergence than conventional approaches. To resolve both the bounding box expansion issue and fundamental limitations of IoU-based losses, PIoU introduces an adaptive

target-size penalty factor *P*. This factor is uniquely determined by the target box dimensions and remains invariant to the size of the minimum enclosing box covering both the predicted and target boxes. Consequently, the penalty factor maintains its stability even when predicted boxes undergo expansion, significantly improving adaptability to varying target sizes. The mathematical formulation is presented in Eq. (5).

$$P = \frac{1}{4}\left(\frac{dw_1}{w_{gt}} + \frac{dw_2}{w_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}}\right) \tag{5}$$

In this equation, the absolute distances between the corresponding edges of the predicted and ground-truth bounding boxes are denoted as dw_1, dw_2, dh_1, and dh_2, whereas the width and height of the ground-truth box are represented by w_{gt} and h_{gt}, respectively, as illustrated in Fig. 6.

To further optimize the loss function, an adaptive penalty function $f(P)$ based on predicted bounding box quality is introduced in PIoU, whose mathematical formulation is presented in Eq. (6).

$$f(P) = 1 - e^{-P^2} \tag{6}$$

A higher P value reflects greater prediction errors in bounding box localization, triggering a reduced $f(P)$ output to mitigate the impact of inaccurate predictions. When P approaches 1, suggesting close proximity between the predicted and target boxes, an increased $f(P)$ value is generated to accelerate loss function regression. As P tends toward 0, the $f(P)$ value progressively decreases with improving prediction quality, continuously optimizing the predicted boxes until full alignment with targets is achieved. The penalty function $f(P)$ is designed to assign maximal gradients to moderately accurate predictions, enabling their rapid regression toward target boxes and consequent quality enhancement. This mechanism directs the detection network's focus

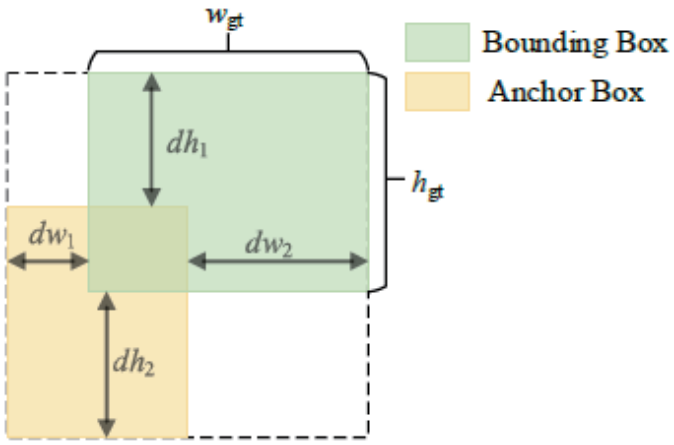

Fig. 6. (Color online) Architectural diagram of PIoU.

toward intermediate-quality anchor boxes. The computational formulation of the PIoU loss function is presented in Eq. (7).

$$PIoU = IoU - f(P), \quad -1 \le PIoU \le 1 \tag{7}$$

$$L_{PIoU} = 1 - PIoU = L_{IoU} + f(P), \quad 0 \le L_{PIoU} \le 2 \tag{8}$$

3. Experiment and Results Analysis

A dataset consisting of 2362 high-resolution images of wind turbine blade defects was utilized to validate the effectiveness of the method proposed in this paper. Through manual screening and defect characterization, a preliminary defect image dataset was constructed, comprising 1305 images containing three typical defect types: cracks, perforations, and surface delamination. To improve the robustness and generalization capability of the target detection model across diverse application scenarios, the dataset was expanded using the Mosaic data augmentation technique, increasing the total number of defect images to 3272. The augmented dataset was then partitioned into training, validation, and test sets at a ratio of 7:2:1 to ensure a scientifically rigorous model training and evaluation process. Furthermore, all defect targets within the images were manually annotated using the LabelImg tool to ensure annotation accuracy and consistency.

3.1 Computational configuration and hyperparameter selection

The computational infrastructure in this paper was configured using Ubuntu 22.04, with Pytorch 2.0.1 serving as the primary deep learning framework and Python 3.10 providing script interpretation. To accelerate the model training process and ensure the efficient processing of large-scale image data, CUDA 11.8 with GPU acceleration support was utilized. The hardware platform consisted of an Intel Xeon E5-2686 v4 processor operating at a base frequency of 2.30 GHz, complemented by an NVIDIA RTX 3080Ti graphics processing unit with 12 GB of GDDR6X memory. Additionally, the specific hyperparameter settings adopted during model training are detailed in Table 1.

Table 1
Experimental parameter settings.

Parameter	Value
Image size	640 × 640
Batch size	16
Number of iterations	200
Momentum coefficient	0.937
Initial learning rate	0.01
Weight decay coefficient	0.0005

3.2 Evaluation metrics

Model detection accuracy was evaluated using three key metrics in this paper: precision (*P*), recall (*R*), and mean average precision (*mAP*). The mathematical formulations of these evaluation metrics are provided in Eqs. (9)–(12).

$$P = \frac{TP}{TP + FP} \times 100\% \tag{9}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{10}$$

$$AP = \int_0^1 P(R)\, dR \times 100\% \tag{11}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \times 100\% \tag{12}$$

Here, the key evaluation metrics include *TP* representing correctly identified defects, *FN* indicating missed defect detections, and *FP* denoting incorrect defect classifications. The detection performance is quantified through average precision (*AP*), derived from the area under the *P*–*R* curve. For comprehensive assessment, *mAP* is calculated by averaging *AP* values across all three defect categories, providing an overall measure of model accuracy.

3.3 Ablation experiments

To systematically evaluate the individual contributions and mutual effects of modules, the experiments are designed to quantitatively analyze the impact of the different modules on key detection metrics, including *P*, *R*, and mAP. The specific improvements implemented are as follows: the standard convolutional layers in the backbone network are replaced with SPD-Conv to enhance shallow feature extraction; the original SPPF module is substituted with the SPPCSPC structure to strengthen multi-scale feature fusion; the MLCA mechanism is introduced into the neck network to optimize cross-scale feature interaction; and the CIoU loss function is replaced with the PIoU loss function, which incorporates stronger geometric constraints, to improve bounding box regression accuracy. Detailed configurations and results of these experiments are summarized in Table 2, where '√' indicates the adoption of the specific module in the experiment, and its absence denotes otherwise.

As can be observed from the experimental results presented in Table 2, the proposed method, which integrates multiple improvement strategies, achieves the best overall detection performance. The results from Experiments 1–5 in Table 2 demonstrate that the improved YOLOv8 detection model, upon the individual introduction of each optimization module, exhibits superior overall performance across multiple key metrics. This validates the effectiveness and practical value of these modules for the wind turbine blade defect detection task. Specifically, upon replacing the standard convolutional layers (Conv) in the backbone

Table 2
Results of ablation experiments.

Number	SPD-Conv	SPPCSPC	MLCA	PIoU	*P* (%)	*R* (%)	*mAP*50 (%)
1					81.8	85.3	87.6
2	√				86.5	85.6	93.9
3		√			90.1	87.0	91.8
4			√		86.1	86.8	92.6
5				√	90.6	87.9	92.4
6	√	√			87.0	86.4	92.8
7	√		√		86.4	89.8	93.4
8	√			√	88.2	89.3	93.6
9		√	√		89.8	87.3	92.4
10		√		√	90.4	88.3	92.5
11			√	√	89.3	88.2	93.2
12	√	√	√	√	91.1	89.1	94.9

network with spatial pyramid depthwise convolution (SPD-Conv) modules, the model exhibited improvements of 4.7% in *P*, 0.3% in *R*, and 6.3% in mAP. This demonstrates that the model effectively enhances feature extraction capabilities for defects without incurring a loss in recall rate. Furthermore, the original SPPF structure was substituted with the SPPCSPC module. Owing to its robust multi-scale feature representation ability, the SPPCSPC module equips the model with superior recognition and localization performance for crack defects, consequently leading to increased *P* and *R*. However, when detecting regions with complex backgrounds, the SPPCSPC module inevitably introduces minor false alarms, resulting in a slight degradation in mAP. Nonetheless, the integration of the SPPCSPC module contributes to achieving an optimal balance among detection *P*, *R*, and mAP. In addition, incorporating the MLCA mechanism into the neck structure of the model yielded more balanced enhancements in *P*, *R*, and mAP compared with the baseline model. This indicates that MLCA strengthens the model's focus on complex structural defects by enhancing cross-scale feature interactions. Finally, the CIoU loss function employed in bounding box regression was replaced with the PIoU loss. The improvements in *P* and *R* significantly surpass those achieved by other modules individually, while the gain in mean average precision is comparable to that of the MLCA module. This validates that the PIoU loss function effectively improves defect localization accuracy by optimizing shape alignment between the predicted and ground-truth bounding boxes.

Experiments 6–11 in Table 2 are conducted to evaluate the interactions of each module. In Experiment 6, both the SPD-Conv and SPPCSPC modules were introduced into the baseline model. The results indicate that the feature retention capability of SPD-Conv may introduce redundant background information that SPPCSPC struggles to fully process, leading to a slight decrease in *P*. However, the multi-scale feature fusion ability of SPPCSPC assists SPD-Conv in detecting more defects, thereby increasing *R* and resulting in an mAP50 value that approximates the intermediate performance of the two modules used individually. In Experiment 7, the SPD-Conv and MLCA modules were combined. SPD-Conv provides feature maps that are rich and well preserved in spatial information, whereas the MLCA mechanism guides the model to focus on key regions within these maps, significantly reducing missed detections and thus markedly

improving *R*. At the same time, mAP50 also achieves favorable results. In Experiment 8, the superior localization capability of PIoU compensates for the precision drop caused by SPD-Conv, achieving a better balance between *P* and *R*. In Experiments 9 and 10, which involve the SPPCSPC module, the models maintain high *P*, but the improvement in *R* is relatively limited. This can be attributed to the MLCA introducing some background interference during cross-scale feature interaction in Experiment 9 and the PIoU loss function exhibiting certain localization inaccuracies when combined with the SPPCSPC module in Experiment 10. Finally, the combination of MLCA and PIoU in Experiment 11 results in all evaluation metrics outperforming the baseline model with a relatively balanced profile, validating the effectiveness of integrating an attention mechanism that focuses on key features with a loss function designed for precise localization regression.

In summary, every newly introduced enhancement component contributed to measurable performance gains across multiple evaluation metrics. The optimized model incorporating all modules demonstrated comprehensive superiority over the baseline YOLOv8n across all evaluation metrics, conclusively validating the effectiveness and feasibility of the proposed methodology for intelligent wind turbine blade defect detection.

3.4 Comparative experiments of different defect detection methods

To further evaluate the efficacy and superiority of the proposed improved object detection algorithm for wind turbine blade defect detection, comprehensive validation experiments were performed. Several YOLO-series detection models, including YOLOv3, YOLOv5, YOLOv6, and YOLOv8, were selected as benchmarks. Systematic comparative experiments were conducted under identical datasets and experimental configurations, with the results summarized in Table 3. The experimental findings from Table 3 reveal that the enhanced model consistently outperforms the original YOLOv8 across all key performance metrics. Specifically, improvements of 9.3% in *P*, 3.8% in *R*, and 7.3% in mAP were achieved. These results confirm that the proposed architectural optimizations and module integration strategies significantly enhance overall detection performance, particularly in recognizing fine-grained defects such as cracks, perforations, and surface delamination, where strong robustness and accuracy were demonstrated. Furthermore, when compared with other detection models including YOLOv3, YOLOv5, and YOLOv6, the improved model exhibits consistently better performance across all evaluation metrics. This verifies its broad adaptability and practical value for wind turbine blade defect detection tasks.

Table 3
Comparative analysis of different model performances characteristics.

Model	*P* (%)	*R* (%)	*mAP*50 (%)	*mAP*50–95 (%)
Yolov3	69.1	71.4	71.0	41.2
Yolov5	81.0	75.5	84.3	52.5
Yolov6	81.2	74.8	83.4	54.8
Yolov8	81.8	85.3	87.6	57.1
Proposed Model	91.1	89.1	94.9	63.5

To provide a more intuitive demonstration of the detection performance of the proposed improved model for wind turbine blade defect detection, comparative visualizations of detection results between the enhanced model and mainstream YOLO-series models are presented. Representative visual samples are illustrated in Fig. 7, where the detection outputs of different models on identical defect images are displayed, including the defect type identification and confidence score distribution of bounding boxes.

From the visualization results, it can be clearly observed that YOLOv5 and YOLOv6 exhibit missed detections when identifying fine cracks, particularly in regions with high fineness or blurred edges, where bounding boxes are either absent or associated with low confidence scores, adversely affecting overall recognition performance. In contrast, the proposed improved model demonstrates the accurate identification of various defect types, including cracks and perforations, in the same test samples. Not only are the detection results more complete, but the bounding boxes also exhibit precise localization and higher confidence scores, indicating defect localization capability and feature representation ability. These findings further validate the synergistic advantages of the introduced modules including SPD-Conv, SPPCSPC, MLCA, and PIoU in feature extraction, scale perception, and bounding box regression. Significant improvements in target recognition accuracy and stability are achieved under the complex surface textures of wind turbine blades. The experimental results demonstrate that the improved YOLOv8 model outperforms traditional approaches in wind turbine blade defect detection. Comprehensive quantitative assessments and visual analysis further confirm the model's higher accuracy and robust practical applicability, highlighting its value for intelligent wind power system applications.

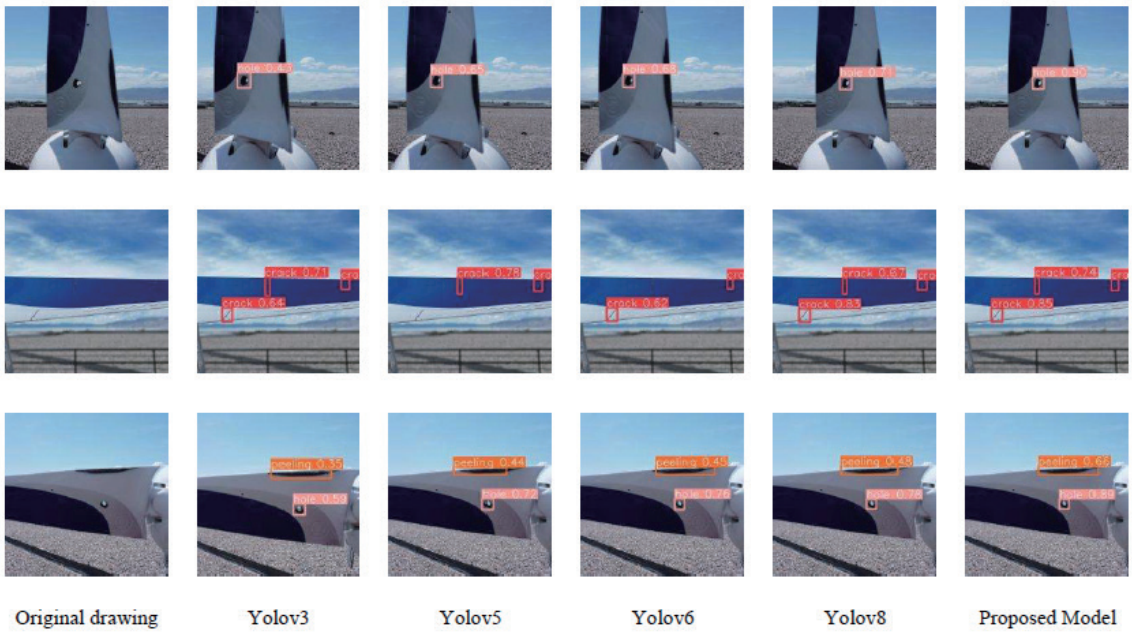

Fig. 7. (Color online) Visualized comparison results diagram.

4. Conclusions

To address the challenges of wind turbine maintenance and blade defect detection, we developed an improved YOLOv8-based detection framework incorporating several key enhancements. The proposed method integrates the MLCA mechanism to significantly reduce both missed detections and false alarms while improving defect recognition accuracy. By replacing conventional convolution blocks with SPD-Conv modules in the backbone network, the system achieves more effective feature preservation and transmission. The architecture further enhances multi-scale analysis capabilities through the CSPC-upgraded SPPCSPC module for feature fusion across different scales. Additionally, the introduction of the novel PIoU loss function substantially improves bounding box localization precision. Experimental evaluations demonstrate notable performance gains, with the enhanced model achieving 9.3%, 3.8%, and 7.3% improvements in *P*, *R*, and mAP, respectively, compared with the baseline YOLOv8, confirming its effectiveness for practical wind turbine inspection applications. Given the wide range of wind turbine blade defects found in real-world settings, subsequent studies should prioritize expanding both the diversity and quantity of flaw samples in current datasets, refining the model's design, and boosting its ability to generalize.

Acknowledgments

Not applicable.

Authors' contributions

Wen Wang conceived the research, Litao Xiao performed the experiments and was responsible for paper writing, Haowei Xiong and Baotong Chi provided guidance for experiments and contributed to revise the paper, Lifu He helped analyze results, Yang Lyu and Wen Zou provided recommendations for this research, and Ji Jiang and Wenlong Fu provided an overall review of the article.

Funding

This work was supported by Hubei Energy Group New Energy Development Co., Ltd. (ENXN-HSF-GC-2025003).

Availability of data and materials

The data cannot be made publicly available upon publication because they contain commercially sensitive information. The data that support the findings of this study are available upon reasonable request from the authors.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1 Y. Du, S. Zhou, X. Jing, and Y. Li: Mech. Syst. Signal Process. **141** (2020) 106445. https://doi.org/10.1016/j.ymssp.2019.106445
2 M. Memari, P. Shakya, M. Shekaramiz, and T. Mohsenin: IEEE Access **12** (2024) 33236. https://doi.org/10.1109/ACCESS.2024.3371493
3 P. Stałowska, C. Suchocki, and A. Zagubień: Measurement **246** (2025) 116706. https://doi.org/10.1016/j.measurement.2025.116706
4 H. Zhao, Y. Gao, and W. Deng: IEEE Internet Things J. **11** (2024) 32804. https://doi.org/10.1109/JIOT.2024.3409823
5 W. Nsengiyumva, S. Zhong, J. Lin, and S. Li: Compos. Struct. **256** (2021) 112951. https://doi.org/10.1016/j.compstruct.2020.112951
6 Y. Zhang, L. Wang, C. Huang, and J. Liu: IEEE Trans. Ind. Appl. **61** (2024) 653. https://doi.org/10.1109/TIA.2024.3481190
7 D. Tcherniak and L. L. Mølgaard: Struct. Health Monit. **16** (2017) 536. https://doi.org/10.1177/1475921717722725
8 A. Khadka, B. Fick, A. Afshar, and J. N. Breunig: Mech. Syst. Signal Process. **138** (2020) 106446. https://doi.org/10.1016/j.ymssp.2019.106446
9 K. A. Tiwari and R. Raisutis: J. Strain Anal. Eng. Des. **53** (2018) 546. https://doi.org/10.1177/0309324718772668
10 M. A. Oliveira, E. F. Simas Filho, M. C. S. Albuquerque, and I. C. Silva: Ultrasonics **108** (2020) 106166. https://doi.org/10.1016/j.ultras.2020.106166
11 V. Arora, J. A. Siddiqui, R. Mulaveesala, and A. Rani: IEEE Sens. J. **15** (2015) 663. https://doi.org/10.1109/JSEN.2014.2361391
12 V. Arora, R. Mulaveesala, A. Rani, and S. Dua: Nondestruct. Test. Eval. **34** (2019) 23. https://doi.org/10.1080/10589759.2018.1546304
13 M. Davis, E. N. Dejesus, M. Shekaramiz, and T. Mohsenin: Appl. Sci. **14** (2024) 6319. https://doi.org/10.3390/app14146319
14 Y. Mao, S. Wang, D. Yu, and L. Zhang: Intell. Data Anal. **25** (2021) 463. https://doi.org/10.3233/IDA-205143
15 P. M. Diaz and P. Tittuss: Signal Image Video Process. **17** (2023) 2333. https://doi.org/10.1007/s11760-022-02450-6
16 J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (IEEE, 2016) 779–788. https://doi.org/10.1109/CVPR.2016.91
17 R. Zhang and C. Wen: Adv. Theory Simul. **5** (2022) 2100631. https://doi.org/10.1002/adts.202100631
18 Y. Liu, Y. Zheng, Z. Shao, and Y. Li: Adv. Eng. Inf. **59** (2024) 102292. https://doi.org/10.1016/j.aei.2023.102292
19 L. Ma, X. Jiang, Z. Tang, M. Li, and Z. Wang: IEEE Sens. J. **24** (2024) 28409. https://doi.org/10.1109/JSEN.2024.3430351
20 Y. Liu, Y. Zheng, T. Wei, and Y. Li: Eng. Appl. Artif. Intell. **138** (2024) 109422. https://doi.org/10.1016/j.engappai.2024.109422
21 Y. He, X. Niu, C. Hao, and Y. Liu: Mech. Syst. Signal Process. **219** (2024) 111592. https://doi.org/10.1016/j.ymssp.2024.111592
22 D. Liu and M. Liu: Appl. Soft Comput. (2025) 113914. https://doi.org/10.1016/j.asoc.2025.113914
23 L. Tong, C. Fan, Z. Peng, and J. Wang: Sustainability **16** (2024) 4467. https://doi.org/10.3390/su16114467

24 M. Lei, X. Wang, M. Wang, and Y. Zhang: Energies **17** (2024) 1796. https://doi.org/10.3390/en17081796
25 Y. Zhang, Y. Yang, J. Sun, and L. Wang: Measurement **220** (2023) 113222. https://doi.org/10.1016/j.measurement.2023.113222
26 S. Z. Rizvi, M. Jamil, and W. Huang: Comput. Electr. Eng. **120** (2024) 109615. https://doi.org/10.1016/j.compeleceng.2024.109615
27 W. Zhao, J. Cai, Y. Ma, and Z. Wang: Shandong Electr. Power Technol. **52** (2025) 1. https://doi.org/10.20097/j.cnki.issn1007-9904.2025.05.001
28 D. Wan, R. Lu, S. Shen, and G. Zhang: Eng. Appl. Artif. Intell. **123** (2023) 106442. https://doi.org/10.1016/j.engappai.2023.106442
29 Y. Zhou, H. Qian, and P. Ding: J. Real-Time Image Process. **20** (2023) 103. https://doi.org/10.1007/s11554-023-01358-9
30 W. Deng, G. Liu, and J. Meng: Sensors **25** (2025) 2219. https://doi.org/10.3390/s25072219