Drone-Based Aerial Inspection of Solar Panel Defects using an Improved YOLO Model

Tao Chen, Mengmei Wang*, Tianxiang Hou, Yang Cao
Jiangsu Normal University KeWen College, Xuzhou, Jiangsu, China
**Corresponding Author*

Abstract: As the scale of photovoltaic power plants continues to expand, traditional manual inspection methods are inefficient, costly, and pose significant safety risks, making them inadequate for actual operation and maintenance needs. The integration of drone aerial photography and computer vision technology provides an efficient and safe solution for detecting defects in photovoltaic panels. To address the problem of insufficient detection accuracy caused by the small size of defect targets, complex backgrounds, and unclear features in aerial images, this paper proposes a defect detection model based on an improved YOLOv8. The model incorporates a coordinate attention mechanism into the backbone network to enhance the localisation of tiny defects; it replaces the original structure with a weighted bidirectional feature pyramid network to improve multi-scale feature fusion; and it uses a SIoU loss function to optimise bounding box regression accuracy. Experimental results show that the improved model achieves an average precision of 92.7% on a self-built dataset, an increase of 4.3 percentage points over the baseline model, and its detection speed meets real-time requirements, providing effective technical support for intelligent operation and maintenance of photovoltaic power plants.

Keywords: UAV Aerial Photography; PV Panel Defect Detection; YOLOv8; Attention Mechanism; Feature Fusion

1. Introduction

1.1 Research Background and Significance

Amid the global trend toward clean and low-carbon energy transition, coupled with the strong impetus of the “dual carbon” goals, China’s solar photovoltaic (PV) power generation industry has experienced explosive growth. By the end of 2025, China’s cumulative PV installed capacity will exceed 800 GW, securing the top spot in the global PV market and powerfully driving the optimization of the energy structure and green development. However, PV panels—the core components of PV power plants—are constantly exposed to complex outdoor environments, where they are subject to various factors such as wind and sand erosion, snow accumulation, hot-spot effects, and obstruction by bird droppings. These conditions can lead to defects such as cracks, hot spots, glass shards, and surface obstructions. These defects not only reduce the power generation efficiency of solar panels by 10%–30%, resulting in significant economic losses, but in severe cases, they may also trigger safety incidents such as short circuits and fires, threatening the safe and stable operation of power plants. Therefore, conducting research on efficient and precise solar panel defect detection is of great significance for ensuring the safe operation and maintenance of power plants and enhancing power generation efficiency.

Traditional PV panel inspections[1] primarily rely on manual visual inspections or handheld thermal imaging cameras, which suffer from low efficiency, high rates of missed defects, high labor intensity, and numerous safety hazards associated with working at heights, making it difficult to meet the operational and maintenance demands of large-scale PV power plants. Today, with the deep integration of drone technology and artificial intelligence, automated inspections using drones equipped with visible-light or infrared cameras—combined with deep learning algorithms for automatic defect identification—are becoming the industry standard. Among these, object detection algorithms based on convolutional neural networks provide reliable technical support for the efficient inspection of photovoltaic panels.

1.2 Current State of Research at Home and

Abroad

In recent years, defect detection in photovoltaic panels has become a hot research topic among scholars both domestically and internationally, with a surge in related studies. Early studies mainly relied on traditional image processing methods such as edge detection and threshold segmentation, but their limitations were obvious: they were sensitive to illumination changes, detection performance could be easily affected by environmental interference, and they lacked generalization ability, making it difficult to adapt to complex and variable real-world scenarios. Since 2018, deep learning has gradually become the mainstream technology for photovoltaic panel defect detection. Guo et al. implemented photovoltaic hot spot detection based on Faster R-CNN [2], but such two-stage methods have slow inference speed and struggle to meet real-time requirements. Therefore, single-stage detectors such as SSD and the YOLO series have received widespread attention.

The YOLO series of models has performed exceptionally well in the field of object detection, thanks to their end-to-end detection architecture and excellent balance between accuracy and speed. For PV detection scenarios, researchers have continuously refined their approaches. For instance, YOLOv5 introduced Transformer modules to enhance global feature extraction, adopted lightweight designs to adapt to edge computing devices, and employed data augmentation strategies to improve the model's robustness against complex backgrounds. However, due to the variable perspectives of drone aerial images, the wide variation in target sizes, and the small proportion of defect areas, existing models still suffer from issues such as missed detection of small targets and inaccurate localization.

1.3 Main Research Focus

This paper uses YOLOv8 [3] as the baseline model and focuses on the practical challenges faced in detecting defects on solar panels in drone aerial photography scenarios, proposing three targeted improvements. First, a Coordinate Attention (CA) module is embedded at the end of the backbone network to enhance the model's sensitivity to spatial positions, thereby improving its ability to locate small defects; Second, a Weighted Bidirectional Feature Pyramid Network (BiFPN) [4] is adopted to replace the original Path Aggregation Network (PANet), enabling efficient weighted fusion of multi-scale features; third, the SIoU loss function is used to replace the CIoU loss function, accelerating model convergence and improving bounding box regression accuracy. Through these improvements, we have successfully developed a high-precision, lightweight solar panel defect detection model suitable for UAV aerial photography scenarios, and its effectiveness has been thoroughly validated through comparative experiments.

2. Technical Background

2.1 YOLOv8 Object Detection Algorithm

As the latest version of the YOLO series launched by Ultralytics in 2023, YOLOv8 significantly surpasses previous models in both detection accuracy and inference speed. Its network architecture mainly consists of three parts: the Backbone is based on the improved CSPDarknet53, replacing the original C3 module with the C2f module to optimize gradient flow while maintaining lightweight characteristics; the Neck uses a PANet structure, achieving efficient multi-scale feature fusion through top-down and bottom-up bidirectional paths; the Head adopts a decoupled design, separating classification and regression tasks and introducing an anchor-free detection mechanism to simplify the post-processing workflow. In addition, YOLOv8 offers five variants: n, s, m, l, and x. Among them, YOLOv8n has the smallest number of parameters, suitable for deployment on embedded devices; YOLOv8m achieves a good balance between accuracy and speed, meeting the needs of most application scenarios.

2.2 Attention Mechanism

Attention mechanisms are widely used in object detection by assigning weights to different regions of feature maps to guide the model to focus on key information. Common attention modules each have their own characteristics: the SE module [5] focuses on the channel dimension, enhancing feature representation through compression and activation, but ignores spatial location information; The CBAM module combines channel and spatial attention, outperforming the SE module but incurring higher computational costs due to its complex structure; the CA module innovatively embeds spatial information into channel attention, balancing spatial accuracy with computational efficiency, making it particularly suitable for detection scenarios involving significant variations in object size.

2.3 Feature Pyramids and Loss Functions

Feature Pyramid Networks effectively address the challenge of information loss for small objects in deep neural networks. Traditional FPNs rely solely on simple element-wise addition or concatenation operations, without accounting for differences in the importance of features across different levels. In contrast, BiFPN introduces a weighted feature fusion mechanism, learning unique weights for each input feature and enhancing information flow efficiency through bidirectional cross-scale connections. Regarding the loss function, while YOLOv8's default CIoU accounts for area of overlap, center-to-center distance, and aspect ratio, it does not consider the directional discrepancy between predicted and ground-truth bounding boxes. SIoU introduces an angular penalty term, further improving regression accuracy and convergence speed.

3. An Improved YOLO Model for Solar Panel Defect Detection

3.1 Coordinate Attention Mechanism Module

In aerial drone images, defects in photovoltaic panels often appear as localized anomalies, such as linear structures resembling cracks or circular high-temperature regions resembling hotspots. These defects typically account for less than 0.1% of the pixels in the original image, making them a classic small-object detection problem. To enhance the model's sensitivity to tiny targets, this paper introduces the Coordinate Attention (CA) mechanism after the C2f module at the end of the backbone network to strengthen the representation capability of small-scale features.

The core idea of the CA module is to decompose two-dimensional global pooling into two one-dimensional encoding processes. For an input feature map $X \in RC \times H \times W$, average pooling is first performed along the horizontal and vertical directions to obtain feature vectors $zh \in RC \times H$ and $zw \in RC \times W$. Then, these two vectors are concatenated, convolved, normalized, and processed with a non-linear activation to generate an intermediate feature map $f \in RC/r \times (H\ W)$. Next, f is decomposed along the spatial dimension into two tensors, which, after Sigmoid [6] activation, are multiplied with the original features as attention weights. Unlike the global compression method of the SE module, the CA module can capture long-range dependencies while preserving spatial structure, enhancing the model's spatial awareness for targets such as slits or small hotspots.

3.2 Weighted Bidirectional Feature Pyramid Network

The original YOLOv8 PANet [7] performs feature fusion through both top-down and bottom-up paths; however, during the fusion process, all layers are assigned the same weight, which fails to reflect the varying contributions of features at different resolutions to the detection task, thereby limiting the utilization of multi-scale features. To address this issue, this paper introduces BiFPN for optimization. BiFPN features four main improvements: First, it employs weighted feature fusion, assigning learnable weights to each input feature and utilizing a fast normalized fusion method to allow the network to adaptively adjust the contribution of features at each layer; second, it adds cross-scale connections by introducing skip connections between features at the same level to preserve original information; third, it retains PANet's bidirectional structure while removing nodes with only one input edge to simplify computation. In solar panel detection, shallow-layer features are rich in edge texture information, which aids in identifying fine defects such as cracks, while deep-layer features have a large receptive field, making them suitable for locating regional anomalies such as hotspots. By leveraging the weighted fusion mechanism, BiFPN enables the model to dynamically select the optimal feature scale based on defect type, thereby enhancing the consistency of multi-scale object detection.

3.3 Improvements to the Loss Function

The loss function of YOLOv8 consists of classification loss and regression loss. The classification loss uses binary [8] cross-entropy, while the regression loss uses CIoU loss. The calculation formula of CIoU loss is $LCIoU = 1 - IoU + \rho^2 (b, bgt)/c^2 + \alpha v$, where ρ represents the Euclidean distance between the centers of the predicted box and the ground truth box, c is the diagonal length of the smallest enclosing rectangle of the two, α is the weighting coefficient, and v is used to measure the consistency of the aspect ratio. However, CIoU does not consider the directional misalignment between the predicted box and the ground truth box. In the scenario of drone oblique shooting, solar panels are prone to perspective distortion, and angular deviation will further affect positioning accuracy. In contrast, SIoU introduces an angle penalty term on the basis of CIoU and redesigns the distance loss and shape loss. Its loss function is defined as $L_{SIoU} = 1 - IoU + \Delta + \Omega$. Experimental results show that SIoU can accelerate model convergence and improve the positioning accuracy for slender and small-scale targets.

3.4 Model Lightweighting Optimization

Given the stringent real-time requirements of drone inspections, this paper focuses on lightweight design while ensuring detection accuracy. Specifically, the number of channels in the C2f module [9] of the backbone network was reduced to 75% of the original version, and the number of feature layers in the BiFPN was limited to three to control computational load. Ultimately, the model has approximately 23.7 million parameters and requires 48.2 billion floating-point operations. On edge devices such as the Jetson Orin NX, the inference speed reaches 45 FPS, meeting real-time detection requirements.

4. Experimental Design and Analysis of Results

4.1 Dataset Construction

The dataset used in this paper was obtained from a drone inspection project at a 100 MW photovoltaic power plant in Qinghai Province. In this project, the drone flew at an altitude of 30 meters and was equipped with a 20-megapixel visible-light camera to capture images, resulting in a total of 5,000 images with a resolution of 5472×3648. These images underwent preprocessing, including cropping and downsampling, and were ultimately standardized to a resolution of 640×640 pixels. Defect annotation was performed by professional O&M personnel, with defects categorized into four types: first, cracks, defined as linear fractures on the glass surface or solar cells; second, hotspots, manifested as areas of abnormally elevated temperature in infrared images; third, stains, caused by obstructions such as bird droppings or dust accumulation; and fourth, fragments, referring to broken corners or large-scale damage. The dataset is divided into training, validation, and test sets in a 7:2:1 ratio, containing a total of 18,762 valid annotated defect instances. To enhance the model’s generalization ability, various strategies such as Mosaic data augmentation, random flipping, and color jittering were employed during training.

4.2 Experimental Environment and Parameter Settings

The experiment is based on the Ubuntu 20.04 operating system, with hardware using Intel Xeon Silver 4210 CPU and NVIDIA RTX 3090 GPU (24GB memory). The deep learning framework is PyTorch 2.0, with CUDA version 11.8. During the training phase, input images are uniformly resized to 640×640 resolution, with a batch size set to 16. The optimizer used is SGD, with an initial learning rate of 0.01, momentum and weight decay parameters set to 0.937 and 0.0005 respectively. The model is trained for a total of 200 epochs, with the learning rate decayed using a cosine annealing strategy. YOLOv8m is chosen as the baseline model for the experiment, and the improved model is trained under the same parameter configuration.

4.3 Evaluation Criteria

In the field of object detection, multiple metrics are commonly used to evaluate model performance. Precision measures the accuracy of the model's predictions, that is, the proportion of predicted positive samples that are actually positive; recall reflects the model's ability to identify true positive samples, that is, the proportion of correctly detected positive samples among all true positive samples. Average Precision (AP) uses an IoU threshold of 0.5 to calculate the mean of precision scores for each category. Parameter count refers to the total number of trainable parameters in the model, while FPS (frames per second) intuitively reflects the model's inference speed, i.e., the number of image frames that can be processed per second.

4.4 Ablation Experiment

To validate the effectiveness of each improved module, we designed ablation experiments using YOLOv8m as the baseline and progressively incorporating CA, BiFPN, and SIoU. The experimental results are shown in Table 1.

Table 1. Ablation Experiment

Model	CA	BiFPN	SIoU	mAP@0.5 (%)	Number of participants (M)	FPS
YOLOv8m	×	×	×	88.4	25.9	52
+ CA	√	×	×	90.1	26.2	49
+ BiFPN	×	√	×	90.8	23.8	48
+ SIoU	×	×	√	89.6	25.9	52
Improved Model	√	√	√	92.7	23.7	45

As shown in the table, introducing the CA module alone increased the mAP by 1.7 percentage points, indicating that the attention mechanism effectively enhances the model's ability to capture features of small defects; the addition of the BiFPN module increased the mAP by 2.4 percentage points, fully validating the significant advantages of weighted feature fusion in PV scenarios; The SIoU loss function further increased mAP by 1.2 percentage points. When all three components work in concert, the improved model achieves an mAP of 92.7%, representing a 4.3-percentage-point increase over the baseline model. Meanwhile, the number of parameters has decreased slightly, and the inference speed fully meets the requirements for real-time detection.

4.5 Comparative Experiments

We compared the improved model with mainstream object detection algorithms; the results are shown in Table 2.

Table 2. Comparative Experiments

Model	mAP@0.5 (%)	Number of participants (M)	FPS
Faster R-CNN	85.2	137.2	18
SSD512	80.7	26.3	35
YOLOv5m	86.9	21.2	56
YOLOv7	88.1	37.2	48
YOLOv8m	88.4	25.9	52
Improved Model	92.7	23.7	45

Experimental results show that the improved model significantly outperforms other comparison algorithms in terms of mAP, achieving a 4.3 percentage point improvement over YOLOv8m [10]. Although its frame rate (FPS) is slightly lower than that of YOLOv5m, it still meets the real-time requirements (>30 FPS) for drone inspections. Compared to Faster R-CNN, the improved model has 82% fewer parameters and is 2.5 times faster, making it more suitable for edge deployment.

4.6 Visualization of Test Results

To visually demonstrate the effectiveness of the improvements, we selected three typical scenarios—cracks, hotspots, and object occlusion—to compare detection results. In crack detection, the baseline model failed to detect some fine cracks, while the improved model enhanced the response in crack regions using the CA module; In hotspot detection, BiFPN integrates multi-scale features to prevent deep-layer features from obscuring small hotspots; in the stain-occlusion scenario, the SIoU loss ensures that predicted bounding boxes align more closely with ground-truth boxes, reducing localization errors. Visualization demonstrates that the improved model exhibits superior detection capabilities in complex scenarios.

5. Conclusions and Outlook

5.1 Summary of the Study

This paper addresses challenges in drone-based aerial inspection of solar panel defects, such as the difficulty in detecting small objects, complex backgrounds, and inaccurate localization, and proposes a detection model based on an improved YOLOv8 architecture. The model introduces a coordinate attention mechanism into the backbone network to enhance the utilization of spatial position information, thereby significantly improving the detection accuracy of small objects. By replacing PANet with BiFPN, the model achieves weighted fusion of multi-scale features, improving the consistency of defect detection across different scales; It employs the SIoU loss function to optimize bounding box regression, accelerating convergence and improving localization accuracy. In experiments on a custom dataset, the improved model achieved an mAP of 92.7%, a 4.3% improvement over the baseline, with inference speeds meeting real-time requirements, providing reliable technical support for the intelligent operation and maintenance of photovoltaic power plants.

5.2 Outlook

Although the method described in this paper has demonstrated significant effectiveness, there remain many areas worthy of further exploration. Regarding detection methods, the current approach relies solely on visible-light images; future work could integrate infrared and electroluminescence (EL) images to achieve complementary multimodal information and enhance the comprehensiveness of defect identification. In terms of the model, further size reduction is needed, and efficient deployment solutions on edge computing platforms such as Jetson and Huawei Ascend should be explored. Given the significant geographical and climatic variations among PV power plants, which can lead to performance degradation when the model is applied to new scenarios, research into incremental learning and domain adaptation techniques is warranted to endow the model with continuous learning capabilities. Furthermore, it is essential not only to detect the presence of defects but also to quantitatively assess their severity, thereby providing precise data to support O&M decision-making. With the advancement of UAV, 5G, and AI technologies, the trend toward unmanned and intelligent O&M of PV power plants is gaining momentum, and this paper provides a feasible technical pathway for this purpose.

Acknowledgments

This work was supported by the Research outcomes from the Jiangsu Provincial Social Sciences Applied Research Excellence Program; the 2025 Jiangsu Provincial Universities “Qinglan Project” Outstanding Young Faculty Development Grant (Recipient: Wang Mengmei); funding from Kewen College “ Talent Program ” ; and the 2025 Jiangsu Undergraduate Universities Special Research Project on “ Teaching Reform in General Education Courses on Artificial Intelligence ” (2025ZNT-20), "Jiangsu Provincial College Students' Innovation and Entrepreneurship Training Program (Project No.: S202513988022)".

References

[1] Cui Jianwei, Wang Yueming. A Defect Detection and Localization System for Photovoltaic Panels Based on YOLOv11. Electronic Measurement Technology, 1–9 [2026-03-23].

[2] Ji Ruirui, Mei Yuan, Yang Sifan, et al. An Infrared Hotspot Detection Algorithm for Photovoltaic Modules Based on Improved Faster R-CNN. Laser & Infrared, 2024, 54(04): 584–592.

[3] Sun Qiyao, Sun Junhai, Zhu Xianyuan, et al. A Method for Defect Detection in Infrared Photovoltaic Modules Based on an Improved YOLOv8. Journal of Solar Energy, 2026, 47(02): 148–154. DOI: 10.19912/j.0254-0096.tynxb.2024-1871.

[4] Li Zhe, Yang Jie, Zhang Yi, et al. Precise Identification of Aircraft Arresting Gear Towing Status Based on Coordinate Attention and Weighted Bidirectional Feature Pyramid Network . Chinese Ship Research, 2025, 20(04): 124-133. DOI:10.19693/j.issn.1673-3185.04005.

[5] Xiang Linying, Shao Shao, Ma Lepeng, et al. An Image Restoration Algorithm Based on SE Modules and Hollow Convolution. Journal of Command and Control, 2025, 11(05): 649–654.

[6] Wu Yifan, Ma Lingkun, Yan Meiyi. Research on a Novel Variable Step Size Algorithm Based on the Sigmoid Function. Applied Electronics, 2026, 52(02): 62–65. DOI:10.16157/j.issn.0258-7998.256591.

[7] Sang Zhi-lei, Qi Long, Zhou Xun-rong. Research on a PANet-Based Defect Detection Model for Toy Parts. Toy World, 2025, (06): 7–9.

[8] Zhang, Jing. A Study on Document Image Binarization Methods Based on Metric Loss and Attention. Taiyuan University of Technology, 2024. DOI: 10.27352/d.cnki.gylgu.2024.002711.

[9] Liu Tianbo. Design and Implementation of a Blockchain-Based Precision Poverty Alleviation System under the C2F Model. South-Central University for Nationalities, 2021.

[10] Gu Ming, Zhang Zhenwei, Chen Mingming, et al. Defect Detection in Photovoltaic Modules Based on an Improved YOLOv5-RCD. Journal of Xi'an University of Engineering, 1–9 [2026-03-23].