Defect Detection of Photovoltaic Cells Based on an Improved YOLOv8

Zhihui LI[1], Liqiang WANG[2]*
Tianjin University of Technology and Education, Tianjin 300222, China[1, 2]
Tianjin Engineering Research Center of Fieldbus Control Technology, Tianjin 300202, China[2]

***Abstract*—Currently, defect detection in photovoltaic (PV) cells faces challenges such as limited training data, data imbalance, and high background complexity, which can result in both false positives and false negatives during the detection process. To address these challenges, a defect detection network based on an improved YOLOv8 model is proposed. Firstly, to tackle the data imbalance problem, five data augmentation techniques—Mosaic, Mixup, HSV transformation, scale transformation, and flip—are applied to improve the model's generalization ability and reduce the risk of overfitting. Secondly, SPD-Conv is used instead of Conv in the backbone network, enabling the model to better detect small objects and defects in low-resolution images, thereby enhancing its performance and robustness in complex backgrounds. Next, the GAM attention mechanism is applied in the detection head to strengthen global channel interactions, reduce information dispersion, and enhance global dependencies, thereby improving network performance. Lastly, the CIoU loss function in YOLOv8 is replaced with the Focal-EIoU loss function, which accelerates model convergence and improves bbox regression accuracy. Experimental results show that the optimized model achieves a mAP of 86.6% on the augmented EL2021 dataset, representing a 5.1% improvement over the original YOLOv8 model, which has 11.24 × 10^6 parameters. The improved algorithm outperforms other widely used methods in photovoltaic cell defect detection.**

***Keywords*—Photovoltaic cells; defect detection; YOLOv8; loss function**

I. Introduction

With the clear goals of "carbon peak" and "carbon neutrality", the development and utilization of clean energy have garnered increasing attention. Solar energy is particularly favored for its safety, stability, low cost, and wide applicability. Currently, silicon cells are primarily used to convert solar energy into electricity. However, silicon cells are often prone to defects such as cracks, short circuits, and black cores due to material properties during production. Therefore, efficient detection techniques are essential for promptly identifying and addressing these issues, ultimately improving the yield and conversion efficiency of solar cells.

Traditionally, surface defects in solar cells were detected through manual inspection and machine vision technology using industrial cameras [1]. However, these methods are not only labor-intensive and inefficient, but also susceptible to human error, which can lead to missed or misidentified defects. Since the rise of deep learning theory in the 1950s, deep learning models based on convolutional neural networks have been widely used in image recognition [2], [3], and natural language processing [4], [5], among other fields. However, directly applying deep learning to defect detection in solar cells remains challenging.

Currently, machine vision-driven defect detection technology, with its efficiency and low-cost advantages, is gradually replacing traditional image processing methods and manual inspection. The YOLO series, with its excellent overall performance, has become a widely adopted framework in object detection [6]. Su et al. [7] also integrated channel and spatial attention mechanisms into the Faster R-CNN framework to effectively detect three types of defects in EL images. Although Faster R-CNN detectors provide high-precision results, they suffer from slow speed, high memory usage, and high computational resource demands. In comparison, SSD has some applications in small object detection, but its performance has not yet reached YOLO's real-time detection level and needs further improvement. In [8], researchers integrated an innovative spatial pyramid pooling technique and channel attention mechanism into the YOLOv5 model to accurately detect and locate crack and fractures in battery electrochemical luminescence (EL) images. In [9], researchers integrated a branch attention module into the YOLOX model to improve small object detection accuracy. The module captures key spatial and channel-level information, optimizing classification and localization tasks, leading to a significant increase in detection accuracy. Li et al. [10] incorporated the GCSC global self-attention mechanism into the YOLOv7 algorithm, enabling effective recognition of four specific defect types in EL images, yielding significant results.

YOLOv8 builds on the YOLO series' strengths, adding new features and optimizations for greater flexibility and improved detection accuracy. The author in [11] focused on small object detection in specific scenarios and based on the YOLOv8 framework. The author in [12] proposed a novel down sampling technique and feature fusion network to retain background features while effectively integrating shallow and deep features. Moreover, a data augmentation strategy based on original samples was used to generate new ones, alleviating the class imbalance in the dataset. The author in [13] developed a tailored algorithm for PV cell EL image defect detection, designed to enhance YOLOv8's performance by optimizing the learning rate and model parameters. However, the current model's accuracy still leaves room for improvement in detecting defects of varying categories and sizes.

From the above analysis, YOLOv8 plays a key role in object detection, providing excellent real-time performance

with relatively low hardware demands, facilitating efficient real-time detection. The main contribution of this paper is: 1) By integrating five data augmentation techniques—Mosaic, Mixup, HSV adjustment, scale transformation, and flipping—the dataset was effectively expanded while preserving the original feature information. 2) SPD-Conv [14] is used in the backbone network to replace the standard convolution in the original network to enhance feature extraction capability. 3) The GAM [15] attention mechanism is incorporated between the model's neck and head to strengthen global channel interactions. 4) The CIoU loss function in the model is replaced by Focal-EIoU [16] to accelerate convergence and improve bbox regression precision.

II. Related Work

At present, there are two publicly available electro-luminescence (EL) image datasets globally. One, presented by Buerhop-Lutz et al. [17], originates from the ELPV dataset and mainly focuses on identifying photoluminescence errors using optical methods. The other, called the PVEL-AD-2021 dataset, was proposed by Su et al. [18] and aims to detect anomalies in the brightness images of photovoltaic (PV) cells. This dataset is regarded as a valuable asset in the field of open-world industrial anomaly detection. Developed over two years by a dedicated research team, the PVEL-AD dataset has evolved from the initial PVEL-AD-2019 version to the latest PVEL-AD-2021 version, featuring substantial improvements. The attention module [19] adaptively adjusts the weight of feature pixels in the input image, boosting focus on crucial information and minimizing distractions from unrelated details. As a result, many researchers have adopted defect detection methods that combine attention modules with convolutional neural networks (CNNs). This study uses the publicly available PVEL-AD-2021 dataset.

In recent years, detection methods based on machine vision and computer vision have been widely applied to the detection of surface defects in solar cells. The author in [20] presents a hybrid fuzzy convolutional neural network (HFCNN), which effectively integrates traditional fuzzy theory with convolutional neural network (CNN) technology, achieving notable success in electro-luminescence (EL) image processing. However, it is important to note that the application of these studies is currently limited to defect recognition in simple EL images. Su et al. [21] performed an extensive evaluation of the PVEL-AD dataset to compare the performance of various defect image recognition models. The models include Faster RPAN-CNN, BAF detector, EfficientDet-D0, EfficientDet-D1, EfficientDet-D2, and three different variants of the YOLOv5 network architecture. To address the challenges posed by complex defect patterns and uneven background structures, Acikgoz et al. [22] proposed an advanced solution using a deep evolutionary neural network model. To tackle the photovoltaic cell defect classification issue, they proposed an innovative classification method that combines Spatial Pyramid Pooling (SPP) with residual connections. Wang et al. [23] proposed a technique that integrates the attention mechanism (CA) into feature maps and uses ResNet152-Xception for feature fusion, enhancing the feature extraction ability of the existing model. To improve the recognition accuracy of defects at various scales in EL images, Fu and Cheng [24] introduced a new component called ELCN and integrated it into the YOLOv7 algorithm. Lu et al. [25] incorporated a coordinated attention (CA) mechanism and HEAD into YOLOv5 to enhance the model's detection precision.

The EL images contain numerous subtle defects, often accompanied by strong background noise, resulting in an imbalanced or skewed defect dataset. Therefore, most previous researchers focused on three common categories (crack, finger defects, and black_core) or four defect types (crack, finger defects, black_core, and thick_line). Su et al. expanded upon this research, incorporating eight different defect types for analysis, including black_core, corner defects, crack, finger defects, fragment, scratch, star_crack, and thick_line. Lu et al. further expanded this scope, covering nine different defect categories, including but not limited to: black_core, corner defects, crack, finger defects, horizontal_dislocation, short_ circuit, star_crack, thick_line, and vertical_dislocation.

Despite the aforementioned research achievements, several challenges persist in the field of solar cell defect detection: 1) The difficulty in acquiring solar cell defect images results in a limited dataset, often leading to insufficient training and poor accuracy. 2) Solar cell defects are diverse and vary in shape even within the same type, and the existing models' insufficient accuracy in identifying these specific defects increases the risk of both false positives and false negatives. 3) Current detection models need further improvement in recognizing target defects and handling complex feature variations. If these issues are not addressed effectively, they could severely limit the reliability of industrial production. Therefore, further in-depth research is urgently needed.

III. Improved YOLOv8 Model

A. YOLOv8 Model

YOLOv8 has four versions: YOLOv8n, YOLOv8s, YOLOv8l, and YOLOv8x [26]. These models have different depth and width parameters. The smaller the network model, the lower the hardware requirements, making deployment easier. To ensure detection accuracy, YOLOv8s is used in this study. YOLOv8 can be roughly divided into three components: the backbone, the neck, and the head. The Backbone adopts the CSPDarknet53 architecture. Unlike previous versions such as YOLOv5, YOLOv8 uses the C2f (CSPLayer_2Conv) module instead of the C3 module. The C2f module has fewer parameters and superior feature extraction ability, which contributes to the light-weighting of the network while enhancing the model's detection accuracy and speed. The Neck consists of the Feature Pyramid Network (FPN) [27] and Path Aggregation Network (PAN) [28]. The Head has three detection heads and adopts the current mainstream decoupled head structure, which separates the classification and detection heads. Additionally, it switches from Anchor-Based to Anchor-Free.

B. An Enhanced Approach based on the SPD-Conv Convolution Module

Given the presence of many small targets and low-resolution defects in the dataset, the SPD-Conv convolution module is incorporated into the backbone to improve

YOLOv8's feature extraction ability. After integrating the SPD-Conv convolution module into the YOLOv8 model, it not only enhances the feature representation capability but also preserves the original architecture of the model, thereby reducing the demand for high-quality input data. The SPD-Conv module replaces the stride convolution layers and pooling layers in the traditional CNN architecture. The input image is then divided into a series of smaller blocks, each representing different feature regions of the image. These blocks' various feature regions are then converted into the number of channels. As shown in Fig. 1, the number of channels is four times the input channels at this stage, which mainly reduces the spatial dimension while increasing the channel dimension. Lastly, the convolution operation is carried out using non-stride convolution layers, meaning the convolution moves pixel by pixel, preserving as much information as possible. This approach cleverly reduces the spatial dimension while ensuring the integrity of the information and preserving the richness of the channel information.

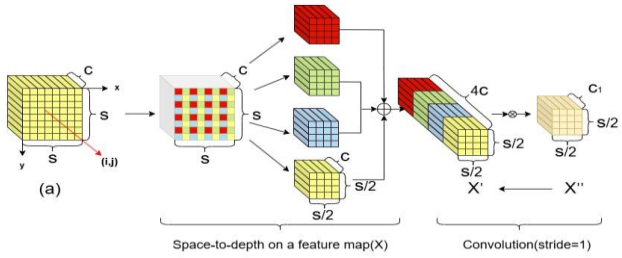

Fig. 1. SPD Transformation.

C. GAM Attention Module

In the precise and intricate task of photovoltaic cell defect detection, the conventional YOLOv8s model encounters notable challenges when handling large amounts of defect features, complex background details, and the coexistence of both large and small-scale targets. The complex background information, after undergoing convolution in YOLO, produces substantial interference, resulting in false detections and misclassifications in defect detection for the corresponding categories. To minimize the interference from complex background information and improve defect feature extraction, introducing an attention mechanism is a good choice. Today, the significance of attention mechanisms in enhancing feature representation is widely acknowledged. Like most lightweight networks, SE [29] modules are often used as the core of their attention mechanism. However, a limitation of the SE module is that it focuses on information interaction between channels but neglects crucial positional information. CAM [30] and CBAM [31] try to capture spatial attention information through convolutional operations, but this method is constrained by the local receptive field of convolution, which can only extract relationships within a local scope and fails to effectively capture long-range or global relationship information. GAM can reduce information loss and amplify the global dimensional interactions， mainly using channel attention and spatial attention mechanisms to expand the global receptive field. This mechanism improves defect classification and can be easily inserted into the core structure of mobile networks. Therefore, this paper integrates the Global Attention Mechanism (GAM) between the neck and head of YOLOv8 to enhance the network's ability to retain information and amplify global cross-dimensional interactions. This improves the ability to accurately identify various defects and reduces the interference from complex backgrounds.

The GAM module begins with the channel-space attention mechanism, and the entire process is illustrated in Fig. 2. Given the input feature map, the intermediate states and outputs are defined by Eq. (1) and Eq. (2):

$$F_2 = M_c(F_1) \otimes F_1 \tag{1}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{2}$$

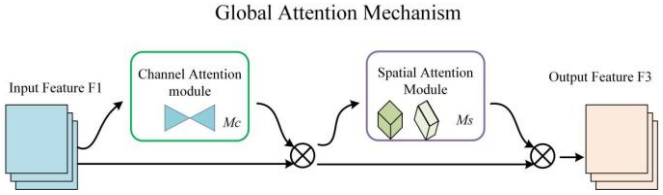

Fig. 2. GAM Module.

In Fig. 2, Mc and Ms denote the channel attention module and the spatial attention module, respectively. The channel attention submodule uses a three-dimensional arrangement to preserve the integrity of information along all three dimensions. It then strengthens the channel-space correlation across dimensions using a multi-layer perception (MLP) with two levels, which has an encoder-decoder structure like BAM, and applies a compression ratio of r. The detailed structure of the channel attention submodule is shown in Fig. 3.

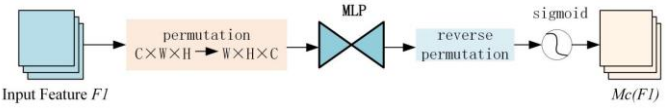

Fig. 3. Channel attention module.

In the design of the spatial attention sub-module, two convolutional layers are used to integrate spatial features, and the same reduction ratio r as in BAM is applied, which also originates from the channel attention submodule. At the same time, since max pooling may cause information loss and have adverse effects, we chose to omit this step to better preserve the details of the feature map. The spatial attention sub-module, without group convolution, is shown in Fig. 4.

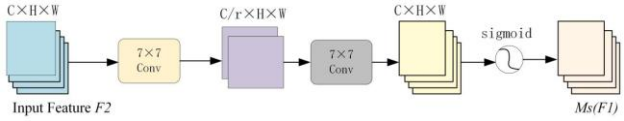

Fig. 4. Spatial attention submodule.

D. FOCAL_EIOU Loss Function

To solve the issue of the imbalance between positive and negative examples in object detection tasks in photovoltaic defect detection, a focal loss function is introduced as the optimization objective. The focal loss function adjusts the weights of positive and negative samples to handle difficult-to-

classify examples, improving the model's ability to recognize complex instances. It takes into account not only the overlap between the predicted bounding box and the ground truth but also additional metrics, such as the distance between the centers of the boxes and the relative differences in width and height. This ensures that effective gradient information is provided even when the overlap is minimal or absent, contributing to the model's training and convergence. Thus, the YOLOv8 model incorporates both Focal Loss and EIoU, two advanced enhancement strategies, which not only help the network focus more on each target instance but also greatly improve the model's detection accuracy and reliability.

The Focal Loss function addresses the issue of class imbalance by adjusting the weight assigned to each sample through a modulation factor. The modulation coefficient is determined by the following formula:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \tag{3}$$

The α_t parameter is designed to reduce the weight of easily classified samples, allowing the model to focus on training difficult samples and thus better handle class imbalance. The adjustment factor is shown in equation (3) $(1-p_t)^{\gamma} .. \gamma \geq 0$ is tunable focusing parameter. To balance the number of easily classifiable and hard-to-classify samples, an appropriate weight ratio for the samples needs to be carefully set, which typically depends on practical experience and fine-tuning. By systematically trying different weight ratios and evaluating the model's performance on the validation set using cross-validation, the ideal weight ratio that best promotes the balance between the two can be selected.

The common loss function, GIoU, considers only the IoU value when the predicted box and the ground truth box intersect. If the IoU value is 0, this will result in a loss function with no gradient over a large area. CIoU only considers the distance between the center points and the overlap area, but does not take the aspect ratio into account, which leads to slower model convergence. The calculation of the EIoU loss function is represented by the following formula:

$$\begin{aligned} L_{EIOU} &= L_{IOU} + L_{dis} + L_{asp} \\ &= 1 - IOU + \frac{\rho^2(b,b^{gt})}{c^2} + \frac{\rho^2(\omega,\omega^{gt})}{C_{\omega}^2} + \frac{\rho^2(h,h^{gt})}{C_h^2} \end{aligned} \tag{4}$$

Here, IoU stands for Intersection over Union, which is the ratio of the intersection area between the predicted and ground truth boxes to the area of their union. b and b^{gt} represent the predicted box and the ground truth box, respectively.

$\rho(b,b^{gt})$ denotes the Euclidean distance between the centers of the predicted and ground truth boxes. c represents the diagonal distance of the minimum enclosing region of the predicted and ground truth boxes. From formula (4), it can be observed that this loss function comprises three components: overlap loss, center loss, and width-height loss. The difference from the original CIoU loss function used in the YOLOv8 model is the inclusion of the width-height loss, which accelerates the model's convergence. This paper combines Focal Loss with EIoU, starting from the gradient perspective, to separate high-quality anchor boxes from low-quality anchor boxes. The formula is as follows:

$$L_{Foccal-EIOU} = IOU^{\gamma} L_{EIOU} \tag{5}$$

In this case, $IOU = |A \cap B| / |A \cup B|$. γ is the parameter used to control the extent of outlier suppression. The Focal loss discussed here dynamically adjusts the loss according to the IOU (Intersection over Union) value: as the IOU increases, indicating better overlap between the predicted and target boxes, the loss value becomes larger. This design resembles a weighting strategy, assigning a higher penalty to high-quality regression targets, with the goal of further enhancing regression accuracy.

E. Network Structure

An improved network structure is proposed to tackle the high background complexity and the class imbalance. The architecture described in Fig. 5 shows significant effectiveness in addressing this issue. Specifically, in the YOLOv8 backbone network, the original convolution layer with a stride of 2 is replaced with SPD-Conv, enabling the model to more precisely capture detailed features in medium and small-scale targets. Furthermore, to reduce the interference of complex backgrounds in photovoltaic defect detection, this paper introduces the GAM attention mechanism after each C2f block in the model's downsampling stage. This improvement helps enhance the ability to extract effective feature information when detecting different defect categories, thereby improving detection accuracy. Lastly, this paper substitutes the C-IoU used in YOLOv8 with the Focal-EIoU loss function. This enhancement dynamically adjusts the loss based on the IOU and separates high-quality anchor boxes from low-quality ones. This modification improves the gradient behavior of the model's loss function without adding extra parameters, facilitating better training and convergence.

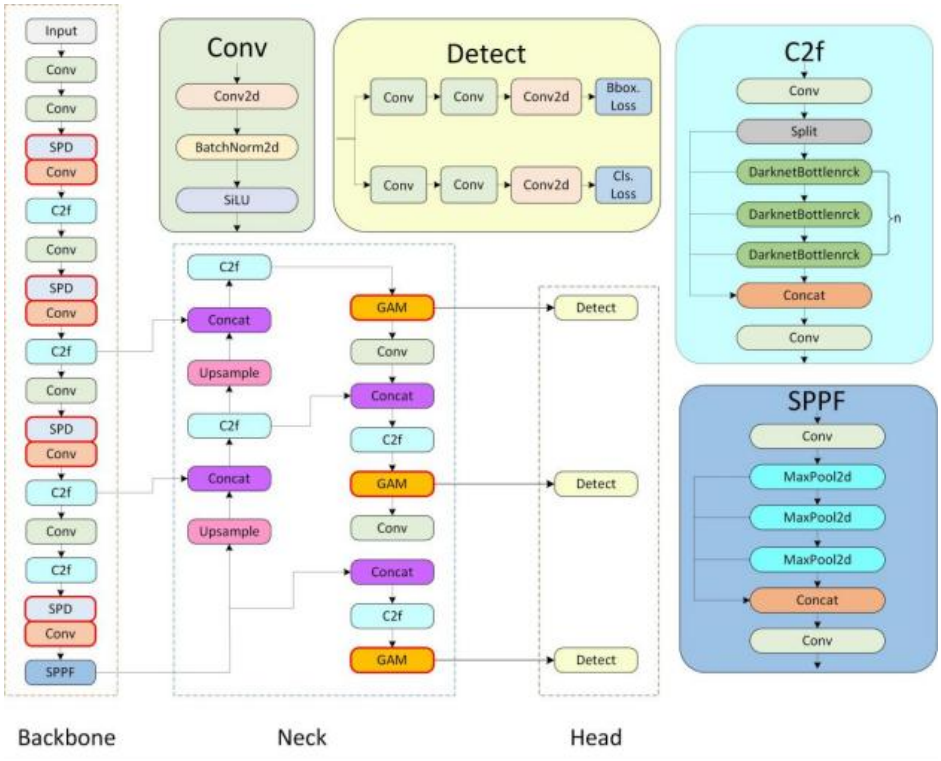

Fig. 5. Improved YOLOv8 network architecture.

IV. Experiments and Results

A. Dataset

This study uses two internationally recognized public datasets for experimentation and evaluates the results of the proposed object detection method. The first dataset is the PVEL-AD-2021 dataset, released through a collaboration between Hebei University of Technology and Beijing University [18]. It includes a category of normal images and features 12 distinct types of abnormal defect images, such as crack, finger, fragment, black_core, star_crack, thick_line, printing_error, corner, scratch, short_circuit, horizontal_dislocation and vertical_dislocation. Samples of these defect types are shown in Fig. 6.

The second dataset is the COCO2017 dataset. It contains over 118,000 training images, 5,000 validation images, and 40,000 test images. The dataset includes 80 common object categories, such as people, vehicles, furniture, animals, etc. covering various scenes from daily life. The COCO2017 dataset has become a critical benchmark for assessing and enhancing the performance of modern computer vision algorithms due to its extensive coverage, detailed annotations, and diverse application scenarios.

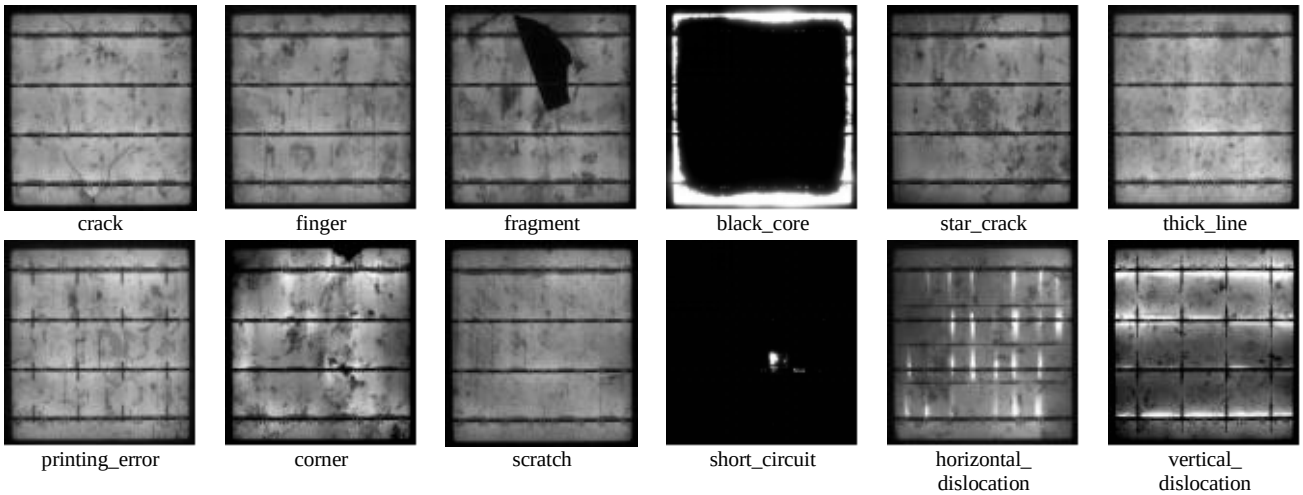

Fig. 6. An illustration of samples from 12 different types of defect images in the defect dataset.

B. Experimental Conditions

The experiment was conducted on a Windows 10 platform with an NVIDIA RTX A4000 GPU, Intel(R) Xeon(R) Gold 6248R CPU @ 3.0GHz 2.99GHz (2 processors), 768 GB RAM, using Torch 2.4.0+ cu11.8 and Python 3.8.0 as the programming environment. The training parameters are listed in Table I below.

TABLE I EXPERIMENTAL CONDITIONS

Parameters	Number
Training Epochs	300
Batch Size	16
Momentum	0.937
Cosine Annealing Hyperparameters	0.01
Initial Learning Rate	0.01
Weight Decay	0.0005

C. Evaluation Metrics

The model's performance was evaluated using multiple metrics in the experiment, including Precision, Recall, and mean Average Precision (mAP). The formulas for each metric are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$AP = \int_0^1 PdR \tag{8}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{9}$$

In the formula, True Positive (TP) refers to the case where both the detection result and the actual situation are positive, meaning that the model correctly identifies and labels the existing objects. False Positive (FP) refers to the situation where the detection result is positive, but the actual situation is negative, meaning the model falsely identifies an object that does not exist. False Negative (FN) refers to the case where the detection result is negative, but the actual situation is positive, meaning the model fails to identify an existing object. Accuracy is the ratio of correctly detected results to all samples identified as having defects (or target objects), reflecting the model's precision in defect (or object) detection. Recall indicates the proportion of correctly detected results among all actual defect (or target object) samples. It reflects the model's effectiveness in identifying true defects (or target object) samples. “AP” (Average Precision) refers to the average accuracy for a specific category, which is equivalent to the area under the P-R (Precision-Recall) curve, used to evaluate the model's performance for that category. “mAP” (mean Average Precision) represents the average of average precision values across multiple classes. It quantifies the model's average performance across all classes and is a critical metric for assessing object detection accuracy.

D. Results and Analysis of Experiments

1) Comparative experiments: Based on the research by Su et al. on the PVEL-AD-2021 dataset, YOLOv5 outperformed other models, such as Faster RPAN-CNN, BAF-Detector, EfficientDet-D0, EfficientDet-D1, and EfficientDet-D2, in PVEL image defect detection. In light of this, this study further compares the defect detection performance of the proposed model with the YOLO series (including YOLOv5, YOLOv7-tiny, YOLOv7, YOLOv8, YOLOv9, and YOLOv11) on the PVEL-AD-2021 dataset, and summarizes the comparison results in Table II.

TABLE II THE OUTCOMES OF DIVERSE DETECTIONS CONDUCTED ON THE PVEL-AD-2021 DATASET

Model	mAP50	mAP50-95	Parameters	FPS
YOLOv5	0.635	0.483	2704372	9.1
YOLOv7-tiny	0.541	0.385	6044754	13.3
YOLOv7	0.517	0.36	37255890	105.3
YOLOv8s	0.815	0.545	9140774	73.9
YOLOv9s	0.809	0.531	9751848	70.4
YOLOv11s	0.816	0.535	9432420	63.5
ours	0.866	0.558	11245812	71.5

In contrast to the rapid decline in training loss, the improved model reaches the performance improvement inflection point earlier than YOLOv8s, after producing more stable results progressively. Fig. 7 shows the comparison results between YOLOv8 and our proposed detection model in terms of mAP 50. Clearly, in terms of detection accuracy for PVEL image defects, the proposed model in this paper shows better performance than YOLOv8s.

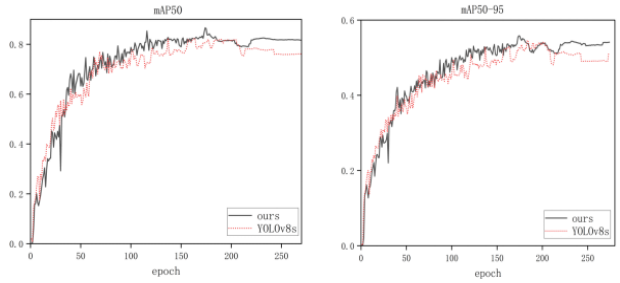

Fig. 7. The comparison results of YOLOv8, and our proposed detection model on mAP50 and mAP50-95.

The algorithm presented in this study successfully enhances object detection accuracy while preserving the model's complexity. After comparing the performance of the enhanced algorithm with the original YOLOv8s model, it is clear that this method achieves significant improvement, specifically a 5.1% increase in the mAP@0.5 metric.

2) Ablation experiment: To evaluate the performance enhancement achieved by adding the SPD-Conv module,

GAM module, and Focal-EIoU loss function to YOLOv8s, a series of ablation experiments were conducted. According to Table III (I represents SPD-Conv, II represents GAM, III represents Focal-EIoU.), YOLOv8s does not fully recognize all 12 types of defects in the PVEL dataset. However, it can be seen that integrating SPD-Conv into the original YOLOv8s increased the mAP by 4.4%, improving the recognition accuracy of finger, black_core, corner, horizontal_ dislocation, and vertical_dislocation defects. Furthermore, when the GAM attention mechanism was introduced, the mAP increased by 1.7%, and the recognition accuracy of finger, black_core, corner, and scratch defects also showed slight improvement. Finally, after replacing the loss function with Focal-EIoU, mAP increased by 0.7%, with improved accuracy in detecting finger, black_core, star_crack, and corner defects. The overall model proposed in this paper improved mAP by 5.1%, with a slight increase in model size and number of parameters. The most important finding from the ablation experiments is that the improved model accurately identifies and locates defects across 12 categories, offering a viable and efficient solution for PVEL image defect detection.

TABLE III THE STATISTICAL RESULTS OF THE ABLATION EXPERIMENTS

Model	Detected types of defects(mAP50)												
	crack	finger	black_ core	thick_ line	star_ crack	corner	fragment	scratch	horizontal_ dislocation	vertical_ dislocation	printing_ error	short_ circuit	mAP50
YOLOv8s	0.82	0.925	0.99	0.916	0.76	0.497	0.995	0	0.94	0.946	0.995	0.995	0.815
YOLOv8s +I	0.79	0.93	0.994	0.905	0.746	0.995	0.995	0.0058	0.969	0.989	0.995	0.995	0.859
YOLOv8s +II	0.778	0.93	0.995	0.912	0.77	0.606	0.995	0.224	0.922	0.866	0.995	0.995	0.832
YOLOv8s +III	0.808	0.938	0.994	0.909	0.773	0.543	0.995	0.0234	0.934	0.955	0.995	0.995	0.822
YOLOv8s +I+II	0.792	0.925	0.994	0.897	0.838	0.695	0.995	0.0995	0.873	0.93	0.995	0.995	0.836
YOLOv8s +II+III	0.805	0.939	0.994	0.906	0.768	0.662	0.995	0.0392	0.981	0.938	0.995	0.99	0.835
YOLOv8s +I+II	0.827	0.929	0.993	0.897	0.756	0.995	0.995	0	0.971	0.994	0.995	0.995	0.862
YOLOv8s +I+II+III	0.782	0.931	0.993	0.901	0.747	0.745	0.995	0.398	0.932	0.981	0.995	0.995	0.866

The confusion matrix shown in Fig. 8 is used to evaluate the classification performance of PV cells for 12 types of defects: black_core, corner, crack, finger defects, fragment, horizontal_dislocation, printing_error, scratch, short_circuit, star_crack, thick_line, and vertical_dislocation. In the confusion matrix, 54 crack defects were misclassified, making up 19.92% of the total crack defects. 56 finger defects were misclassified, accounting for 9.98% of the total number of finger defects. 3 black_core defects were misclassified, constituting 1.38% of the total black_core defects.22 thick_line defects were misclassified, constituting 12.64% of the total thick_line defects. In addition, 6 star_crack were misclassified, accounting for 21.42% of the total star_crack. 1 corner defect, representing 50% of their respective totals. The misclassification rate for scratch defects is 100%.10 horizontal_dislocation defects were misclassified, representing 5.84% of the total. 1 vertical_dislocation defect was misclassified, representing 2.77% of the total. There were no misclassifications for fragment and printing_error. 1 short_circuit defect was misclassified, representing 0.98% of the total. This shows that the improved model in this study minimizes prediction errors and exhibits strong classification performance for PVEL defects.

Based on the observations shown in Fig. 9, the model exhibits outstanding anomaly detection capabilities across a range of common defects (such as finger defects, black_ core, fragment, horizontal_dislocation, printing_error, short_circuit, thick_line, and vertical_dislocation), with an mAP50 metric close to 100%. However, when distinguishing between crack, corner, and star_crack, which are similar and challenging to differentiate, the mAP50 value indicates a moderate level. Additionally, the mAP50 value for scratch dropped significantly. This is mainly due to the high background complexity, which causes frequent misdetections and misclassifications for small defects. It is recommended to consider adding supplementary features or using alternative algorithms to enhance feature extraction for these specific defects.

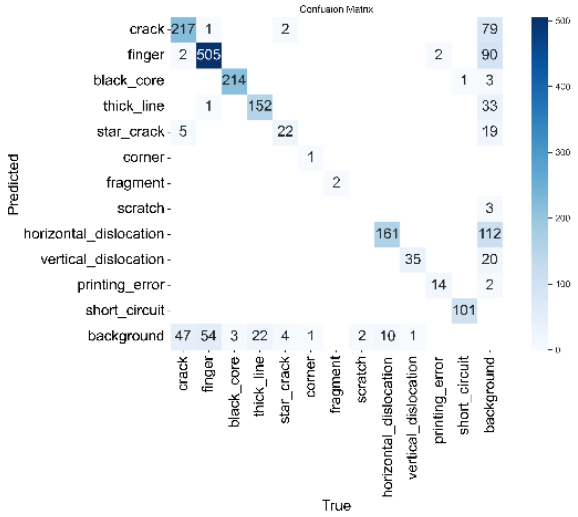

Fig. 8. Confusion matrix of the improved YOLOv8s model.

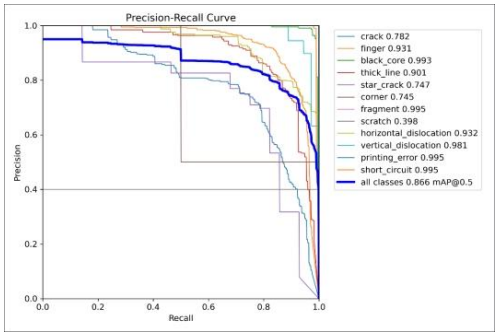

Fig. 9. Precision-recall curve for the enhanced YOLOv8s.

Lastly, Fig. 10 illustrates the predicted bounding boxes, identified upon completion of training, demonstrating accurate alignment with the ground truth boxes in height.

3) Network generalizability: In order to further assess the effectiveness of the improved YOLOv8, this research performed comparative experiments using YOLOv8 on the publicly available COCO2017 dataset. The results presented in Table IV clearly show that, compared to YOLOv8 on the COCO 2017 dataset, the mAP 50 value significantly increased by 4.1%. This result strongly demonstrates the accuracy and performance advantages of the improved model in object detection tasks.

TABLE IV THE OUTCOMES OF DIVERSE DETECTIONS CONDUCTED ON COCO2017

Model	Categories of detected defects	mAP50
YOLOV8	80	0.623
ours	80	0.664

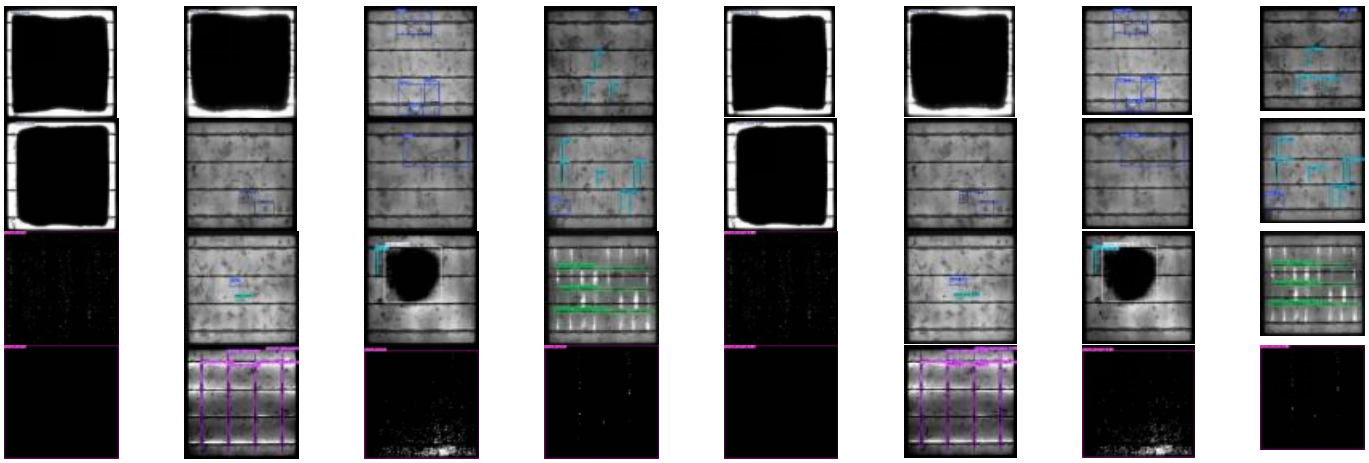

Fig. 10. Comparing the outcomes of randomly sampled ground truth bounding boxes with predicted bounding boxes.

V. CONCLUSION

This study addresses the challenges of difficult data collection, complex defect classification, and high background complexity, which lead to missed detections in photovoltaic cell surface defect detection, by proposing an optimized YOLOv8s model. The results from the comparison and ablation experiments indicate that the optimized YOLOv8s model improves the mAP by 5.1% compared to the original model, exhibiting significant detection adaptability for 12 defect types in solar cells. This indicates that the model has great potential for application in photovoltaic cell defect detection. The future direction of work will involve further optimizing the model to achieve higher identification accuracy to meet industrial needs. The existing public dataset contains class imbalance and issues with detecting certain defect types, necessitating further dataset optimization. Moreover, it is important to consider improving detection accuracy and speed while reducing model parameters, thereby enhancing the model's practicality.

REFERENCES

[1] Jha SB, Babiceanu RF. Deep CNN-based visual defect detection: Survey of current literature. Computers in Industry. 2023 Jun 1; 148:103911

[2] Liu L, Shen C, van den Hengel A. Cross-convolutional-layer pooling for image recognition. IEEE transactions on pattern analysis and machine intelligence. 2016 Dec 9;39(11):2305-13.

[3] Liu N, Wan L, Zhang Y, Zhou T, Huo H, Fang T. Exploiting convolutional neural networks with deeply local description for remote sensing image classification. IEEE access. 2018 Jan 26; 6:11215-28.

[4] Hassan A, Mahmood A. Convolutional recurrent deep learning model for sentence classification. Ieee Access. 2018 Mar 12; 6:13949-57.

[5] Ren X, Zhou Y, Huang Z, Sun J, Yang X, Chen K. A novel text structure feature extractor for Chinese scene text detection and recognition. IEEE Access. 2017 Mar 3; 5:3193-204.

[6] Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. Machine Learning and Knowledge Extraction. 2023 Nov 20;5(4):1680-716.

[7] Su B, Chen H, Chen P, Bian G, Liu K, Liu W. Deep learning-based solar-cell manufacturing defect detection with complementary attention network. IEEE Transactions on Industrial informatics. 2020 Jul 8;17(6):4084-95.

[8] Xu S, Qian H, Shen W, Wang F, Liu X, Xu Z. Defect detection for PV Modules based on the improved YOLOv5s. In2022 China Automation Congress (CAC) 2022 Nov 25 (pp. 1431-1436). IEEE.

[9] Lü Zhixuan, Wei Xia, Ma Zhigang. Improved YOLOX Lightweight Helmet Detection Method. Journal of Computer Engineering & Applications. 2023 Jan 1;59(1).

[10] Li J, Wu W, Chen H. GCSC-Detector: A Detector for Photovoltaic Cell Defect Based on Deep Learning. In2023 42nd Chinese Control Conference (CCC) 2023 Jul 24 (pp. 6913-6917). IEEE.

[11] Lou H, Duan X, Guo J, Liu H, Gu J, Bi L, Chen H. DC-YOLOv8: small-size object detection algorithm based on camera sensor. Electronics. 2023 May 21;12(10):2323.

[12] Ruiz-Ponce P, Ortiz-Perez D, Garcia-Rodriguez J, Kiefer B. Poseidon: A data augmentation tool for small object detection datasets in maritime environments. Sensors. 2023 Apr 2;23(7):3691.

[13] Phan QB, Nguyen TT. A Novel Approach for PV Cell Fault Detection using YOLOv8 and Particle Swarm Optimization. In2023 IEEE 66th International Midwest Symposium on Circuits and Systems (MWSCAS) 2023 Aug 6 (pp. 634-638). IEEE.

[14] Sunkara R, Luo T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. InJoint European conference on machine learning and knowledge discovery in databases 2022 Sep 19 (pp. 443-459). Cham: Springer Nature Switzerland.

[15] Shen S, Yang J. Better YOLO with Attention-Augmented Network and Enhanced Generalization Performance for Safety Helmet Detection. arxiv preprint arxiv:2405.02591. 2024 May 4.

[16] Zhang YF, Ren W, Zhang Z, Jia Z, Wang L, Tan T. Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing. 2022 Sep 28; 506:146-57.

[17] Buerhop-Lutz C, Deitsch S, Maier A, Gallwitz F, Berger S, Doll B, Hauch J, Camus C, Brabec CJ. A benchmark for visual identification of defective solar cells in electroluminescence imagery. In35th European PV Solar Energy Conference and Exhibition 2018 Sep 24 (Vol. 12871289, pp. 1287-1289).

[18] Su B, Zhou Z, Chen H. PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection. IEEE Transactions on Industrial Informatics. 2022 Mar 29;19(1):404-13.

[19] Li M, Wei M, He X, Shen F. Enhancing part features via contrastive attention module for vehicle re-identification. In2022 IEEE International Conference on Image Processing (ICIP) 2022 Oct 16 (pp. 1816-1820). IEEE.

[20] Ge C, Liu Z, Fang L, Ling H, Zhang A, Yin C. A hybrid fuzzy convolutional neural network-based mechanism for photovoltaic cell defect detection with electroluminescence images. IEEE Transactions on Parallel and Distributed Systems. 2020 Dec 21;32(7):1653-64.

[21] Su B, Chen H, Zhou Z. BAF-detector: An efficient CNN-based detector for photovoltaic cell defect detection. IEEE Transactions on Industrial Electronics. 2021 Apr 7;69(3):3161-71.

[22] Acikgoz H, Korkmaz D, Budak U. Photovoltaic cell defect classification based on integration of residual-inception network and spatial pyramid pooling in electroluminescence images. Expert Systems with Applications. 2023 Nov 1; 229:120546.

[23] Wang J, Bi L, Sun P, Jiao X, Ma X, Lei X, Luo Y. Deep-learning-based automatic detection of photovoltaic cell defects in electroluminescence images. Sensors. 2022 Dec 27;23(1):297.

[24] Fu H, Cheng G. Convolutional neural network based efficient detector for multicrystalline photovoltaic cells defect detection. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. 2023 Aug 1;45(3):8686-702.

[25] Lu S, Wu K, Chen J. Solar cell surface defect detection based on optimized YOLOv5. IEEE Access. 2023 Jul 11.

[26] Sapkota R, Meng Z, Ahmed D, Churuvija M, Du X, Ma Z, Karkee M. Comprehensive performance evaluation of yolov10, yolov9 and yolov8 on detecting and counting fruitlet in complex orchard environments. Authorea Preprints. 2024 Jul 9.

[27] Zhu L, Lee F, Cai J, Yu H, Chen Q. An improved feature pyramid network for object detection. Neurocomputing. 2022 Apr 28; 483:127-39.

[28] Yu H, Li X, Feng Y, Han S. Multiple attentional path aggregation network for marine object detection. Applied intelligence. 2023 Jan;53(2):2434-51.

[29] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 7132-7141).

[30] Park J. Bam: Bottleneck attention module. arxiv preprint arxiv:1807.06514. 2018.

[31] Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. InProceedings of the European conference on computer vision (ECCV) 2018 (pp. 3-19).