China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 11, No. 2, June 2024 https://doi.org/10.15878/j.instr.202400166

Article

An Efficient YOLOX-Based Method for Photovoltaic Cell Defect Detection

Junjie Wang[1,2,3,†], Li Bi[1,*], Xunde Ma[1,†], Pengxiang Sun[4]

[1] College of Information Engineering, Ningxia University, Yinchuan 750021, China

[2] Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

[3] University of Science and Technology of China, Hefei 230026, China

[4] College of Economics and Management, Taiyuan University of Technology, Taiyuan 030024, China

[*] Corresponding author email: billy68@nxu.edu.cn

[†] These authors contributed equally to this work

Abstract: Defect detection technology is crucial for the efficient operation and maintenance of photovoltaic systems. However, the diversity of defect types, scale inconsistencies, and background interference significantly complicate the detection task. Therefore, this paper employs the YOLOX model as the backbone network structure and optimizes various modules to address these issues. First, we adopt a transfer learning strategy to accelerate model convergence and avoid insufficient accuracy due to the limited number of defect samples. Second, we introduce the SENet module into the feature extraction process to enhance the contrast between defects and their background. Then, we incorporate the ASFF strategy at the end of the PAFPN network to adaptively learn and emphasize both high- and low-level semantic features of defects. Furthermore, model accuracy is enhanced by refining the loss functions for positioning, classification, and confidence. Finally, the proposed method achieved excellent results on the Photovoltaic Electroluminescence Anomaly Detection dataset (PVEL-AD), with a mAP of 96.7% and a detection speed of 71.47FPS. Specifically, the detection of small target defects showed significant improvement.

Keywords: photovoltaic cell; defect detection; deep learning; YOLOX; electroluminescence

Citation: JunjieWang, Li Bi, Xunde Ma, and Pengxiang Sun. "An Efficient YOLOX-Based Method for Photovoltaic Cell Defect Detection." Instrumentation 11, no. 2 (June 2024). https://doi.org/10.15878/j.instr.202400166.

1 Introduction

Photovoltaic power generation has achieved significant success as a crucial renewable energy technology, with its proportion in the power generation sector steadily increasing. The 2023 report from the International Energy Agency (IEA) indicates that the spot prices for solar photovoltaics have decreased by nearly 50% year-over-year, with the production capacity being three times that of 2021. Solar photovoltaics account for three-quarters of the global new renewable energy additions, with an expected global supply reaching 1100 GW by the end of 2024. By 2028, nearly 60% of the new installation capacity is projected to be in China[1]. Therefore, photovoltaic power generation technology, benefiting from significant reductions in component costs and extensive deployment, has become the most favored form of renewable energy generation in the future.

The development of photovoltaic power generation and its initial investment returns are primarily dependent on the performance and lifespan of photovoltaic cells. However, due to their crystalline structure, these cells are highly susceptible to human-induced disturbances during manufacturing, transportation, and installation processes[2]. Moreover, photovoltaic power plants are often located in harsh environments such as Gobi and semi-arid deserts, which expose them to natural elements like storms and hail[3]. These elements can lead to defects

such as debris accumulation, grid rupture, and concealed cracks. Literature[4] suggests that defects or failures in photovoltaic cells can result in a loss of up to 18.9% in total output power. Notably, while microcracks initially have a minimal impact on power generation, they may expand and degrade over time, resulting in power loss rates ranging from 0.9% to 42.8%[5]. Additionally, this degradation can trigger hotspot effects[6], leading to panel damage, increased fire risk, and a significant reduction in power generation efficiency[7,8]. Therefore, the quality and health status of photovoltaic cells are closely related to power generation efficiency and economic benefits[9]. To ensure the normal operation of photovoltaiccells, the adoption of effective defect detection methods is crucial. Currently, defect detection in photovoltaic cells is primarily classified into three categories: manual inspection, optical image-based detection, and other techniques. The advantages and disadvantages of each method are summarized in Table 1.

Manual inspection methods are operationally complex and offer low localization accuracy, making them unsuitable for large-scale applications. Electroluminescence (EL) imaging, favored for its non-destructive nature and low cost, is easily integrated into production lines.Recently, computer vision has been widely used in the detection of surface defects in photovoltaic cells. The effective integration of EL imaging and computer vision not only emphasizes the characteristics of the defect area but also extracts more comprehensive feature information, thereby effectively overcoming the limitations of traditional methods and enhancing the quality and efficiency of photovoltaic defect detection.

Computer vision methods are primarily categorized into three groups: artificial design rules, traditional machine learning, and deep learning techniques. The first two methods, though providing preliminary results in defect detection, have limitations due to their manually designed feature extraction processes. These processes do not adequately capture the effective features of defects, particularly under conditions of multiple defect uncertainties, leading to suboptimal outcomes. Deep learning, with its superior learning capabilities, has achieved significant success in classifying, detecting, and segmenting photovoltaic cell surface defects. Literature [18] proposed a lightweight defect detector based on the EfficientNet-B0 backbone, utilizing a graph channel attention module to enhance global information modelling in CNNs. Literature[19] introduced the YOLO-PV algorithm, a YOLO-based object detection method that reduces reliance on deep information extraction and focuses on refining low-level defect details. Literature[20] developed a semantic segmentation model for the analysis of electroluminescence images of photovoltaic components.

Compared to existing methods for detecting defects in photovoltaic cells from EL images, this paper attempts to achieve precise detection and localization by incorporating attention mechanisms, merging high and low-level semantic features, and improving the loss function. Our contributions are as follows:

•The SENet attention reduces background noise in the backbone network, while bicubic interpolation in PAFPN preserves key features.

•The ASFF strategy adjusts feature weights across scales, improving multi-scale processing and enhancing performance.

•The EIoU replaces IoU to improve bounding box aspect ratio computation, and VariFocal loss addresses sample imbalance, boosting accuracy.

The proposed model was evaluated on the PVEL-AD dataset through quantitative and qualitative analyses, ablation studies, small object detection performance, and loss variation.

Table 1 Photovoltaic cell defect detection method comparison table

Category	Methods	Strengths	Weaknesses
Visual detection	Manual Visual Inspection[10]	Fast and easy to operate.	The efficiency is low, the experience and subjective awareness of the inspectors will affect the results, and the internal defects of the battery are difficult to detect with the naked eye.
Optical detections	Infrared heat detection[11]	Non-destructive testing, quantitative measurement, high resolution.	Equipment is expensive, and large-scale detection leads to low efficiency.
	Electroluminescent detection[12]	Non-destructive testing, low cost, good repeatability, defects are easy to detect.	The image appears stray light interference and random noise.
	Photoluminescence detection[13]	Non-contact, non-destructive testing, fast detection speed, high resolution.	The specifications of the pulse light source are strict, the cost is high, and the laser cannot be irradiated uniformly.
	UV fluorescence imaging[14]	Non-destructive testing, snail tracks and cracks are easy to detect.	Long time exposure and sufficient light source irradiation, some defect types cannot be detected.
Other detections	Acoustic testing[15]	High detection accuracy for specific defect types.	Only detecting a single defect has great limitations.
	Electronic interferometry[16]	Good detection of external defects such as scratches and chips.	The detection effect on internal defects such as hidden cracks and broken grids is poor.
	Electrical characteristic detection[17]	High real-time performance, can be used for quantitative calculation.	High cost and inability to accurately detect the size and location of defects.

The organization of this paper is as follows: Section 2 provides a review of the literature on surface defects in photovoltaic cells. Section 3 describes the dataset, framework, and enhancements made to the algorithm of the model. Section 4 discusses the experimental design, results, and evaluation of performance. Section 5 concludes the paper and suggests directions for future research.

2 Related works

In deep learning, neural networks enhance the accuracy of detection by extracting refined textural features of defects along with more abstract semantic features. In the past, researchers have proposed many excellent detection methods based on deep learning for detecting surface defects in photovoltaic cells, which can primarily be categorized into two types: two-stage and single-stage object detection.

In the field of two-stage object detection networks, literature[21] proposes a novel Complementary Attention Network (CAN) that is embedded within the Region Proposal Network of Faster R-CNN to extract more refined information on defect areas. Literature[22] employs the ResNet-101-FPN backbone to train Mask R-CNN, meeting the requirements for automatic detection of multiple types of defects and their application in actual production lines. Furthermore, literature[23] introduces a CNN-based decoupled two-stage object detection framework that enhances detection performance through Multi-Head Aggregation (MHA) blocks and Local Non-Local (LNL) blocks. However, the processing speed of two-stage detection networks is slow, which may not meet the demands of actual production environments.

In comparison, single-stage object detection networks, such as the YOLO series, have the advantages of speed and real-time performance. By detecting objects of various sizes on feature maps of different dimensions, they address the issue of diverse defect scales in EL images, effectively enhancing both detection speed and accuracy. Literature[24] fine-tunes the model through transfer learning, demonstrating that the YOLO single-stage detection method is superior to the two-stage methods. Literature[25] incorporates a Global Attention Mechanism (GAM) into the backbone network of YOLOv5, enhancing the image feature extraction capability. Literature[26] proposes a defect detection method under background interference, adapting channel characteristics for reweighting and adjusting monitoring signals at different scales. Literature[27] improves feature extraction capability and suppresses background noise by introducing DCNv2 and attention mechanisms, achieving a detection accuracy of 89.64%. Literature[28] introduces a Cross-Space Multi-Scale Attention (EMA) mechanism after the C2Dense module, focusing on pixel-level features to enhance useful information for defect detection. Literature[29] designs a multi-scale attention mechanism that refines multi-scale features through attention, thereby enhancing classification and detection performance. These methods primarily integrate various attention mechanisms into the backbone to suppress background information and improve model detection accuracy. Some researchers focus on improving feature fusion in the model fusion layer. Literature[30] designs a novel YOLO model to capture local connections between images and feature maps, extracting more detailed edge features of similar faults and enhancing small object detection accuracy. Literature[31] introduces a backbone neck structure called BPFPN to enhance multi-scale feature fusion capability, improving the detection of small defects with an accuracy of 88.8%. Literature[32] adds adaptive feature space fusion to the existing feature fusion structure in YOLOv5. Literature[33] designs a self-fusion network (SeFNet) that replaces the classification layer in HRNet, better integrating multi-resolution feature information in image models. In object detection networks, the loss functions used for defect localization and classification also merit attention. Literature[34] replaces the original cross-entropy loss function with focal zoom loss to enhance network feature extraction without increasing computational load. Literature[35] uses a focal loss function to train a lightweight defect detector based on EfficientNet-B0, addressing multi-defect category imbalance. Literature[36] considers DIoU loss function for bounding box overlap areas and the distance between centers, minimizing the normalized distance between prediction and target boxes. Literature[37] replaces the default NMS with Distance Intersection over Union (DIoU), considering overlap areas and center distances between two boxes, thereby generating more accurate bounding boxes. Literature[38] introduces the Wise-IOU loss function, further improving model detection accuracy and achieving excellent results.

The aforementioned enhanced model can suppress complex background information in photovoltaic cell electroluminescence (EL) images and improve the fusion capability of multi-scale features. However, while ensuring that the model's detection speed and deployment performance meet practical requirements, enhancing the accuracy of defect detection still faces several challenges: (1) Crystal structures resembling defects may mislead the model's judgement. (2) The types and shapes of defects vary, and overlapping occurrences of multiple different defects exist. (3) Small-sized defects have weaker feature expression capabilities, making localization and identification more difficult.

3 Dataset and methods

3.1 Introduction of dataset

The PVEL-AD dataset is collected from actual industrial manufacturing, with a higher original resolution of 1024×1024, and defect types are more diverse and comprehensive, such as cracks, broken grids, black cores, misalignment, thick lines and other types of defects[39],

experts annotated the location and category of defects. From the perspective of the dataset, several challenges still remain in training an effective defect detection model:

(1) Defect categories are imbalanced, with 3403 instances of "finger" and 157 instances of "star_crack". This significant difference in numbers may result in overfitting problems. The distribution of the number of different defects is illustrated in Fig.1 below.

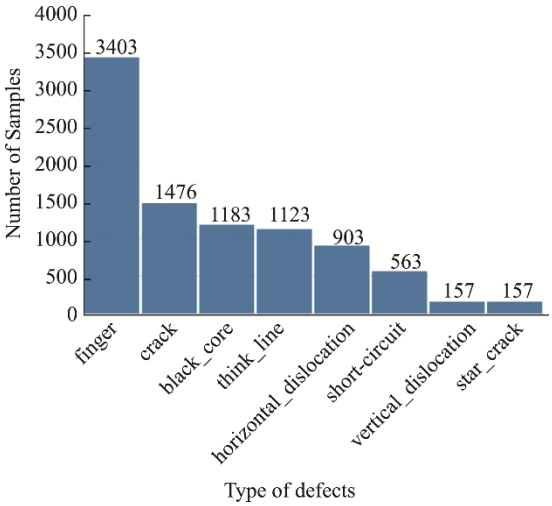

Fig.1 Distribution map of the number of defects in the data set.

(2) The image background is complex and non-uniform, featuring structures like grid lines and grains. Notably, the grain structures vary in size, are randomly distributed, and closely resemble defects, increasing the likelihood of false detections. Existing methods struggle to suppress the interference from complex backgrounds in defect detection.

(3) Defects on photovoltaic cells vary widely in type and shape, often appearing in multiples with uncertain sizes, shapes, and distributions on the same cell. The "crack" category, in particular, exhibits high randomness in its appearance.

(4) Locating and identifying small target defects is extremely challenging due to their small scale and limited visual information. These defects often have weak feature expression and can be overshadowed by larger targets, leading to their neglect during the learning process.

Fig.2 visually illustrates the characteristic shapes of defects: the brightly illuminated areas are defect-free, while the darker areas or lines indicate the presence of defects, gate lines, and grains. It is important to note that grains, although visible, are not defects. In Fig.2, grid lines are marked with green boxes, grains with red boxes, grain false defects (which closely resemble cracks) with blue boxes, and linear cracks with yellow boxes.

3.2 Methodology

We propose an enhanced photovoltaic cell defect detection network based on the modified YOLOX, as shown in Fig.3: The network utilizes the CSPDarknet architecture to extract feature maps from EL images, while the defect features are reinforced and irrelevant information is suppressed through the SENet attention mechanism. Subsequently, the upsampling performance in the PAFPN network is improved using bicubic interpolation, and the ASFF module is employed for adaptive learning of multi-scale features to augment fusion. Ultimately, the EIoU loss function is employed to optimize the IoU, and the VariFocal loss is replaced with binary cross-entropy to enhance model precision and convergence speed.

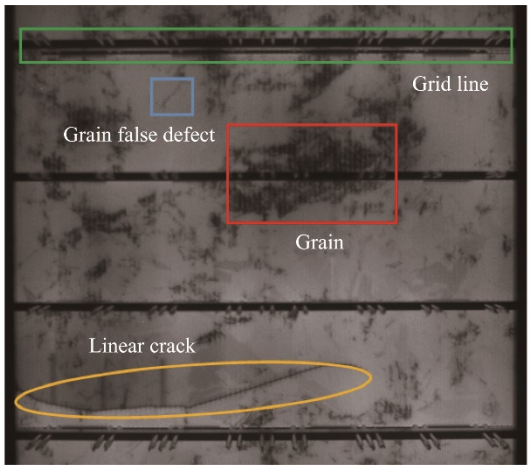

Fig.2 EL image with non-uniform texture background.

(1) YOLOX

YOLOX is a single-stage target algorithm, because of its flexibility, high efficiency and generalization ability, and it is widely used in the field of object detection. The YOLOX network structure is divided into three components: Backbone, Neck and Head.

The backbone network is based on CSPDarknet and employs multiple modules such as Focus, CBS, CSP, and SPP to extract features, enhancing the model's ability to recognize targets of various sizes. The neck network integrates the FPN and PAN structures to facilitate efficient information transfer and fusion across different levels. The head network improves detection accuracy and convergence speed by optimizing the classification and regression tasks through a decoupled head structure.

(2) Transfer Learning

To mitigate model underfitting resulting from insufficientdefect samples, this study introduces a model-based transfer learning approach[40]. Pre-trained model parameters are transferred to the training of the new dataset, addressing issues related to random parameters initialization. Moreover, to expedite model adjustment efficiency, a freezing training strategy is adopted during the training process. The backbone network is frozen, keeping the feature extraction network unchanged and conserving video memory.

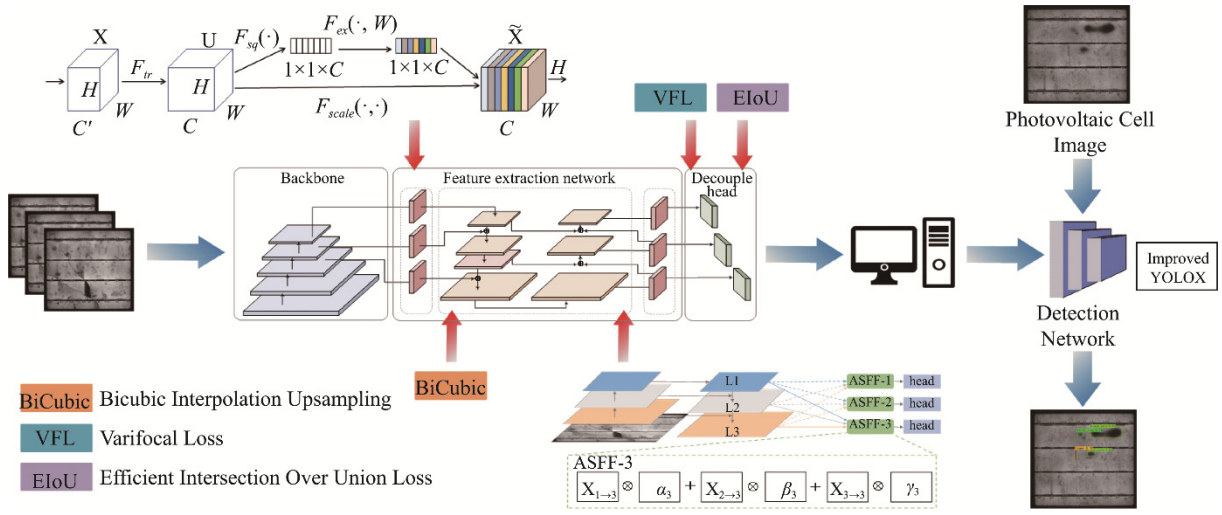

Fig.3 Defect detection network of photovoltaic cells based on improved YOLOX model.

(3) SENet Attention

Structural defects such as grid lines and particulates in photovoltaic cells increase the complexity of detection. Consequently, we introduced the SENet (Squeeze-and-Excitation Network)[41] attention mechanism to enhance the contrast between the defect core and its background, thereby reducing background noise. The architecture is depicted in Fig.4.

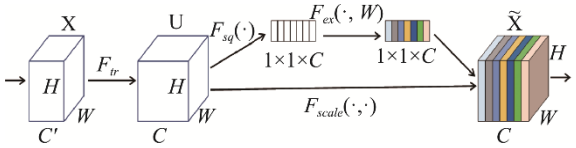

Fig.4 SENet network structure diagram.

Firstly, the Squeeze operation compresses the feature channel into a series of real numbers with a global receptive field, the original feature map is compressed from $H\times W\times C$ to $1\times 1\times C$. The specific details of the Squeeze operation can be formulated as

$$u_c = v_c \times X = \sum_{s=1}^{c} v_c^s \times X^s \tag{1}$$

$$z_c = F_{sq}(u_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{2}$$

Among them, v_c^s represents the c-th convolution kernel; X^s represents the sth input; H and W represent the height and width of the feature map, c represents the number of channels.

Secondly, Excitation uses two fully connected layers to perform a non-linear transformation on the results after Squeeze to build the interconnection between channels. The details can be formulated as

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma(W_2 g(W_1 z)),\ W_1, W_2 \in R^{\frac{C}{r}\times r} \tag{3}$$

$$\widetilde{X} = F_{scale}(u_c, v_c) = u_c \cdot v_c \tag{4}$$

Among them, S_c represents the weight coefficient; u_c represents a two-dimensional matrix; r represents the scaling coefficient.

Finally, weights are calculated using the sigmoid activation function and features are weighted through channel-wise multiplication, effectively extracting defect characteristics on the surface of photovoltaic cells and suppressing background interference.

(4) Bicubic Interpolation Upsampling

In the PAFPN architecture, nearest neighbor interpolation is employed for upsampling, which, although fast, tends to result in the loss of pixel values. To minimize the loss of critical features during the sampling process, a bicubic interpolation algorithm is utilized. This algorithm performs cubic interpolation calculations on the grayscale values of the 16 surrounding points of the sample points, carefully considering the rate of change in grayscale values between adjacent points. This approach optimizes image details, reduces distortion, and enhances the precision and effectiveness of the upsampling.

(5) Adaptive Spatial Feature Fusion

In the process of defect feature extraction, the position information of small target defects is easy to disappear. Therefore, in our study, we introduce the adaptive space feature fusion[42] (ASFF) strategy, adding at the end of PAFPN. The ASFF structure adaptively learns the weight of each position on each feature layer, suppresses the inconsistency of different scales in the feature extraction network, and highlights the important position of small target defect features in feature layer fusion. The ASFF algorithm structure is shown in Fig.5.

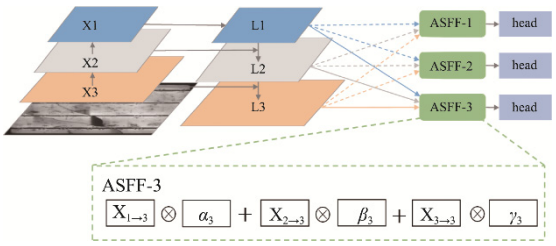

Fig.5 ASFF fusion network structure diagram.

In the feature scaling part, the three enhanced feature layers L1, L2, and L3 obtained by the PAFPN extraction network are arranged in order from small to large. The larger the scale, the smaller the number of channels. In Fig.6, ASFF-3 is taken as an example.

In the adaptive fusion part, the weight parameter α、β、γ is obtained through a 1×1 convolution operation, used for the fusion of features at different scales. This weight parameter is multiplied by the feature maps L1, L2, and L3 and then summed to form the fused feature ASFF−3. The description of the above process can be formulated as

$$y_{ij}^{l} = a_{ij}^{l} * x_{ij}^{1\to l} + \beta_{ij}^{l} * x_{ij}^{2\to l} + \gamma_{ij}^{l} * x_{ij}^{3\to l} \tag{5}$$

Among them, y_{ij}^{l} represents the new feature map obtained through ASFF feature fusion. $x_{ij}^{1\to l}$, $x_{ij}^{2\to l}$ and $x_{ij}^{3\to l}$ are the map vector of the first, second and third feature layer to the l feature layer, respectively. α_{ij}^{l}, β_{ij}^{l} and γ_{ij}^{l} are the feature mapping weight of three different feature layers at position (i, j) to the l feature layer, and the calculation process of the weight α_{ij}^{l} can be formulated as

$$\alpha_{ij}^{l} = \frac{e^{\lambda_{\alpha ij}^{l}}}{e^{\lambda_{\alpha ij}^{l}} + e^{\lambda_{\beta ij}^{l}} + e^{\lambda_{\gamma ij}^{l}}} \tag{6}$$

Which is the control parameter obtained by using a 1×1 two-dimensional convolutional layer to process input features of three different scales.

Compared to the original structure, the ASFF structure enables more effective fusion of features across different scales through backpropagation, resulting in improved model performance with minimal additional overhead.

(6) Loss Function Improvement

In the YOLOX, the loss function is mainly composed of three modules: bounding box positioning loss, classification loss and confidence loss.

1) Bounding box localization loss improvement

The YOLOX uses the intersection over union (IoU) [43] to calculate the bounding box positioning loss. However, the IoU function primarily exhibits two drawbacks: 1) It only considers the overlapping area between the predicted bounding box A and the actual bounding box B; 2) When there is no overlap between A and B, the loss is 0. In our research, IoU is replaced by an efficient intersection over union loss (EIOU) function[44] to solve these problems. The expression of the EIoU loss function can be formulated as

$$\begin{aligned} L_{EIoU} &= L_{IoU} + L_{Dis} + L_{Asp} \\ &= 1 - IoU + \frac{\rho^2(o, o^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \end{aligned} \tag{7}$$

In the equation, $\rho^2(o, o^{gt})$ represents the distance between the centers of two frames, o denotes the center of the predicted frame, o^{gt} is the center of the true frame, (w^{gt}, h^{gt}) and (w, h) correspond to the dimensions of the true frame and the predicted frame respectively, c is the minimum corner distance encompassing both boundaries, and (c_w, c_h) refers to the dimensions of this minimum bounding box.

Traditional IoU loss functions fall short in regression precision when dealing with various randomly distributed defects on the surfaces of photovoltaic cells. The EIoU loss function, however, enhances the model's convergence speed and regression accuracy, making it more suitable for practical applications.

2) Classification Loss and Confidence Loss Improvement

The YOLOX network uses the binary cross-entropy loss function[45] to calculate the classification loss and confidence loss. Due to the imbalance in sample categories and the high similarity between defect targets and the photovoltaic cell background, the loss function struggles to differentiate defect targets from the background and is ineffective in handling target overlap and occlusion.

The VFL (Varifocal loss, VFL)[46] loss function can effectively alleviate the problems of uneven distribution of positive and negative samples and few defective samples by asymmetrically treating the loss caused by positive and negative samples. The function can be formulated as

$$VFL(p,q) = \begin{cases} -q(q\log(p) + (1-q)\log(1-p)), q > 0 \\ -\alpha p^{\gamma}\log(1-p), q = 0 \end{cases} \tag{8}$$

In the formula (8), p is the predicted output value of the YOLOX network, and q is the real value. By reducing the loss of negative samples with q=0 through the γ factor in the VFL, while keeping the loss of positive samples unchanged, the loss information is balanced to enhance model accuracy.

4 Results

4.1 Experimental settings

The dataset is randomly divided into training, validation, and test sets according to a ratio of 7:1.5:1.5. Therefore, in all experiments, the number of instances in the training set is 3150, the validation set is 675, and the test set is 675.

The proposed model is written in Python under the framework of PyTorch, and the experiment is run on Intel(R) Xeon(R) Gold 5317 CPU @ 3.00GHz processor, 504GB memory, and A100 80GB GPU platform. Some specific details of the experimental procedure are described below. During the experiment, mixed-precision training was used to reduce video memory by nearly half. We set a probability of 25% to use the mosaic and mix up data enhancement method[47] to enrich the detection background information and increase the number of overlapping defects.

4.2 Performance metrics

The performance of the network model is mainly

judged by three aspects: detection accuracy, detection speed, and deployment difficulty.

In our study, mAP is the most commonly used evaluation standard for target detection tasks. It represents the average value of n defect categories AP, and AP is the integral of precision to recall on the Precision-Recall curve. They can be formulated as

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

$$AP = \int_0^1 P(R)dR \tag{10}$$

TP (true positive) refers to predicting positive samples as positive, FN (false negative) predicts positive samples as negative, and FP (false positive) predicts negative samples as Positive class. Therefore, Recall and Precision can be formulated as

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

The mAP@0.5:0.95 shown in the experimental evaluation section is the average mAP on different IoU thresholds (from 0.5 to 0.95, step size 0.05). Among them, S , M and L represent the average accuracy for small, medium and large targets, where the small target: the area of the defective target is less than 32×32 pixels (1024 pixels), the medium target: the area of the defective target is between 32×32 and 96×96 pixels (1024 to 9216 pixels), and large target: the area of the defective target is greater than 96×96 pixels (more than 9216 pixels). In this way, we can evaluate the detection performance of the model from multiple perspectives for defects of different scales.

In our study, the frame rate (frame per second, FPS) indicator is used to evaluate the detection speed of the model. FPS is the number of detected images per second. Furthermore, we use model parameters (Parameters) and floating point operations per second (FLOPs) to evaluate the complexity of the network model.

4.3 Evaluation

This chapter will introduce the experimental results of our proposed model on the PVEL-AD dataset, from five different perspectives: quantitative experiments, qualitative experiments, ablation experiments, the detection performance of the model in small target defect types, and the loss trend graph of the model. First of all, in the quantitative experiment, our model was compared with the commonly used methods and other improved methods in recent years. Secondly, in the qualitative experiment, the detection result map of our model is compared with the real map. Thirdly, in the ablation experiment, the effectiveness of each improved module is jointly verified by combining the experimental result data and the heat map. Then, to highlight the power of the improved model, it is compared with other methods on the three defect types that are more difficult to identify. Finally, in order to highlight the stability, convergence and efficiency of the improved model, the loss trend graph of the model is analyzed.

(1) Quantitative Analysis

We selected the more commonly used models from the single-stage and two-stage for comparison. Among them, the single-stage target detection algorithm are: YOLOV4, YOLOV5, YOLOX, Fcos, RetinaNet. Two-stage target algorithms are: SPP-Net, Fast RCNN, Faster-RCNN. The test results of different models are shown in Table 2.

From the above experimental results, it can be clearly seen that the overall performance of the YOLOX model is superior to other basic models. In terms of detection accuracy of YOLOX, the mAP and mAP@0.5:0.95 indicators are 93.8% and 61.2%, respectively. The amount of parameters and calculations of YOLOX are 8.940M and 26.772G respectively, farlower.

Table 2 Results of performance comparisons with commonly used models

Model	Train	Detection accuracy/%					Speed	Difficulty	
		mAP	mAP@ 0.5:0.95	mAP@ 0.5:0.95_S	mAP@ 0.5:0.95_M	mAP@ 0.5:0.95_L	FPS	Total params	GFLOPS
YOLOV4	S	84.0	46.9	11.3	28.7	45.8	55.3386FPS	63.975M	142.021G
YOLOV5	S	90.4	55.1	23.3	41.2	54.9	91.5746FPS	46.669M	114.679G
Fcos	S	80.2	48.4	22.8	49.3	48.0	58.9931FPS	32.127M	161.410G
RetinaNet	S	60.6	39.8	13.1	38.3	42.2	53.1913FPS	36.475M	165.967G
SPP-Net	T	82.3	54.0	23.3	45.3	56.0	58.5612FPS	45.285M	82.747G
Fast RCNN	T	87.7	57.6	24.8	48.3	59.7	42.7465FPS	57.53M	192.514G
Faster-RCNN	T	93.0	61.6	25.7	50.6	63.1	28.7104FPS	136.832M	401.862G
YOLOX	S	93.8	61.2	30.6	45.0	61.8	77.3538FPS	8.940M	26.772G
YOLOX (No-TL)	S	91.4	56.2	24.3	41.7	56.2	76.5121FPS	8.940M	26.772G
Our model	S	96.7	66.1	39.2	56.2	68.4	71.4741FPS	14.424M	35.223G

than others. The overall performance of the YOLOX model is undoubtedly the best. It is worth noting that "No-TL" indicates that the YOLOX model did not apply transfer learning. Although the pre-trained weights originated from the COCO dataset, which has a vastly different sample distribution from photovoltaic cell defects, the performance with transfer learning was still 2.4% higher than without it, demonstrating that transfer learning significantly enhances model performance. Therefore, our research makes improvements based on YOLOX with transfer learning. Compared with YOLOX, our model has increased by 2.9% in mAP and achieved a 4.9% increase in mAP@0.5:0.95. The detection accuracy has been significantly improved, especially for the recognition task of small targets, which can be proved by the mAP@0.5:0.95_S indicator. It is worth noting that the detection speed of our model is 71.4741FPS, which is only a small increase compared to YOLOX. Due to the addition of the attention mechanism and the adaptive feature fusion network, the parameters and calculations of our model have increased to 14.424M and 35.223G respectively, but these indicators are still far smaller than other models. So, in terms of deployment difficulty, our model is still highly competitive. The experimental results prove that our schemes such as introducing adaptive spatial feature fusion structure, adding SENet attention mechanism, and improving loss function are correct, can effectively alleviate the problems of lack of defect samples and complex background.

(2) Qualitative Analysis

In order to more intuitively show that our improved method is effective compared with other models, we start from the perspective of qualitative analysis. We selected three representative EL images from the test set. Fig.6 shows the real label and the actual detection results of several better models on the surface defects of photovoltaic cells. It needs to be explained that the four columns represent the real labels of defects, the prediction results of Faster-RCNN, YOLOX, and our models.

The first row in Fig.6 shows that all three models correctly detected all defects, but our improved model effectively enhanced the confidence of all defects, proving that our improvement of the confidence loss function is scientific and correct. In the second row, the two prediction boxes of "finger" and "crack" coincide, so Faster-RCNN and YOLOX have a bad situation of missing the "finger" defect. However, our model improves the bounding box localization loss function and reasonably solves the overlap problem of two boxes, so it can effectively identify the "finger" defect in the "crack" detection box. In the third row, because of the complex background and the interference of non-defect impurities, the Faster-RCNN and YOLOX models reappeared in the case of missed detection, and all small target "finger" defects cannot be identified. Nevertheless, our model is effective because of its SENet attention mechanism and ASFF feature fusion layer which can highlight the features of defect areas. In conclusion, the improved method suppresses the interfering impurities and detects almost all defects effectively.

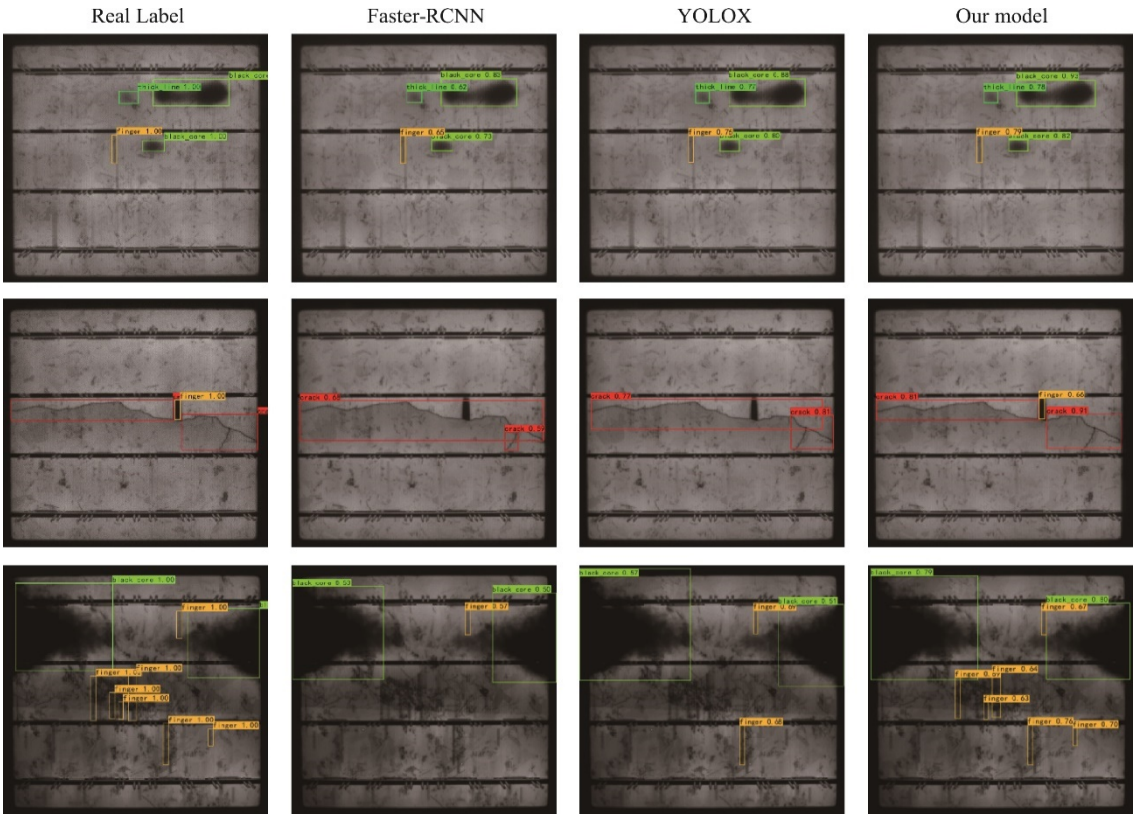

Fig.6 The actual forecast graph of each model.

(3) Ablation Studies

In order to confirm the effectiveness of each component in our improved model, we use ablation experiments for analysis. The experimental results are shown in Table 3. On the basis of the original YOLOX, the improvement scheme is gradually added. From the experimental results, we can see that all the indicators of detection accuracy are generally improved. Especially, the detection performance of small and medium targets is the most prominent: mAP@0.5:0.95_S increased by 8.6%, mAP@0.5:0.95_M increased 11.2%. Although the amount of parameters and calculations under the improvement are slightly insufficient compared with those before the improvement, the detection performance of the model for various defects is greatly enhanced. It fully demonstrates that this improvement method is scientific and rea-sonable.

At the same time, Fig.7 shows the prediction results and corresponding heat map of each degree of the improved model. It is clearly seen that the YOLOX basic model can only recognize one defect point in the image. Although there are thermal distributions in other areas in the image, the thermal value is too low to reach the threshold set by the experiment, so a large number of defect areas are lost. During the sampling stage on PAFPN, the bicubic interpolation algorithm is used tominimize the loss of key feature points. The model's focus on the area of defects widens, which can be confirmed by the diffusion of regions in the heatmap. After improving the classification and localization loss functions in the network, the model enhances its confidence in its detection results,while emphasizing the importance of defect detection results, while emphasizing the importance of defect locations more prominently. However, more focused areas do not mean better ability. In order to reduce the influence of the interference background, highlight defect feature areas, and avoid blind recognition of the model, our model adds an attention mechanism and a feature fusion network. The improving effects can be seen in columns 5 and 6.

(4) Model Performance on Small Target Defect Types

In order to highlight the recognition performance of our model for small target defects, we choose Faster-RCNN, YOLOV5, and the basic YOLOX to compare with our improved model. The performance of each model is compared on the four most representative types. The four types of defects are: "crack", "thick_line", "finger" and "star_crack". The evaluation indicators include AP, Precision, Recall and F1-score. It should be noted that the score_threhold of Precision, Recall and F1-score are all set to 0.5.

Table 3 Ablation experiment results

Model	Detection accuracy /%					Speed	Difficulty	
	mAP	mAP@ 0.5:0.95	mAP@ 0.5:0.95_S	mAP@ 0.5:0.95_M	mAP@ 0.5:0.95_L	FPS	Total params	GFLOPS
YOLOX	93.8	61.2	30.6	45.0	61.8	77.3538FPS	8.940M	26.772G
YOLOX+BIU	94.0	62.1	32.4	46.5	61.2	77.5652FPS	8.940M	27.406G
YOLOX+BIU +VFL	94.3	62.4	32.7	47.8	60.9	79.1329FPS	8.940M	27.406G
YOLOX+BIU +VFL+EIoU	95.1	63.1	33.3	52.9	62.8	79.9891FPS	8.940M	27.406G
YOLOX+BIU +VFL+EIoU+SENet	95.8	63.9	37.7	56.6	64.0	75.5763FPS	8.984M	27.409G
YOLOX+BIU +VFL+EIoU+SENet+ASFF	96.7	66.1	39.2	56.2	68.4	71.4741FPS	14.424M	35.223G

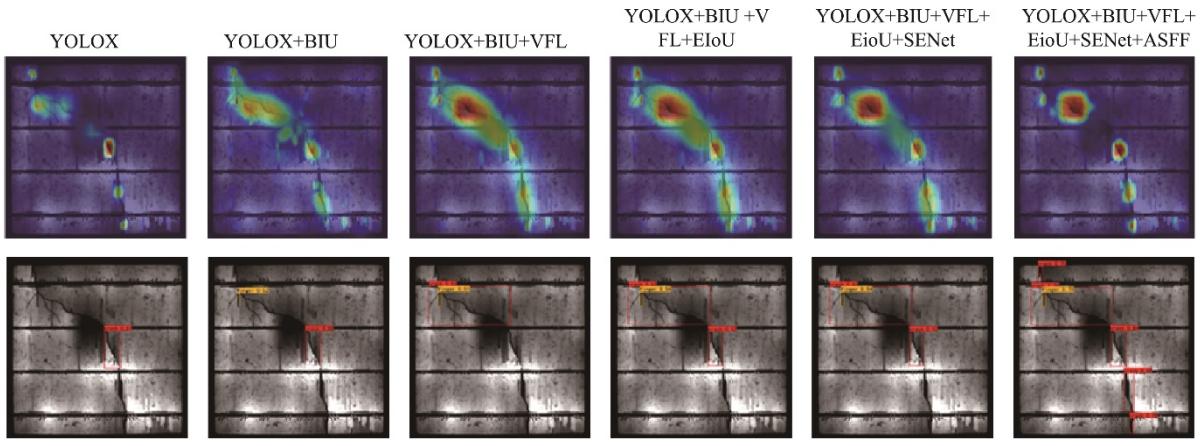

Fig.7 The heat map and actual prediction map of each model in the ablation experiment.

It can be seen from the AP value diagram in Fig.8 that the our model has improved the detection performance of four representative defects, especially the "crack" defect, which has greatly increased on the basis of YOLOX. Observing the whole figure, the overall performance of our model is enhanced. Although the detection results of the "star_crack" type do not rise but fall, other types of defects are optimistic. The detection results show that our solution to the small target detection problem is very effective.

(5) The Loss Trend Graph of Our Model

In order to verify the effectiveness of the loss function module improvement and the performance of the model during training, Fig.9 shows the loss function change diagrams of the YOLOX model and our model, respectively. In the first 50 epoch freezing phase, by comparing the loss change graphs of the two models, it is obvious that the loss of the original YOLOX model on the training set drops to about 5.5, and the loss on the test setdrops to about 4.8, while our model dropped to about

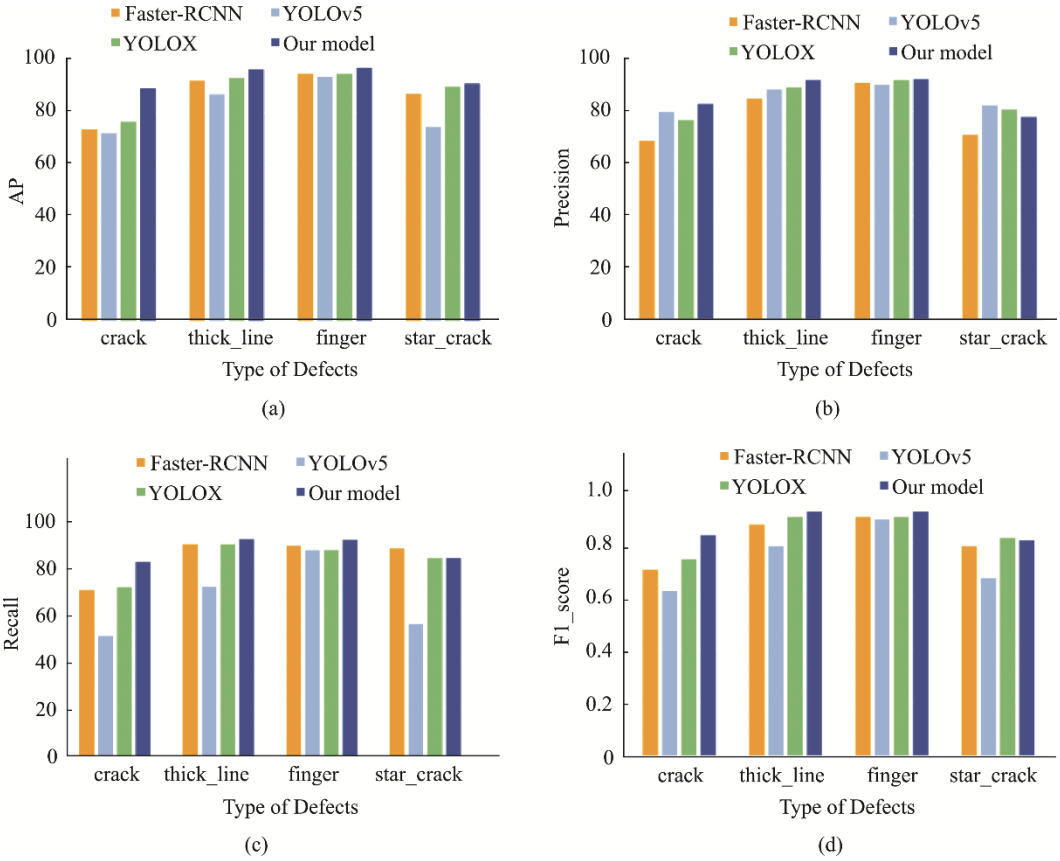

Fig.8 Different indicators on small target defect recognition performance chart of each model: (a)AP; (b)Precision; (c)Recall; (d)F1-score.

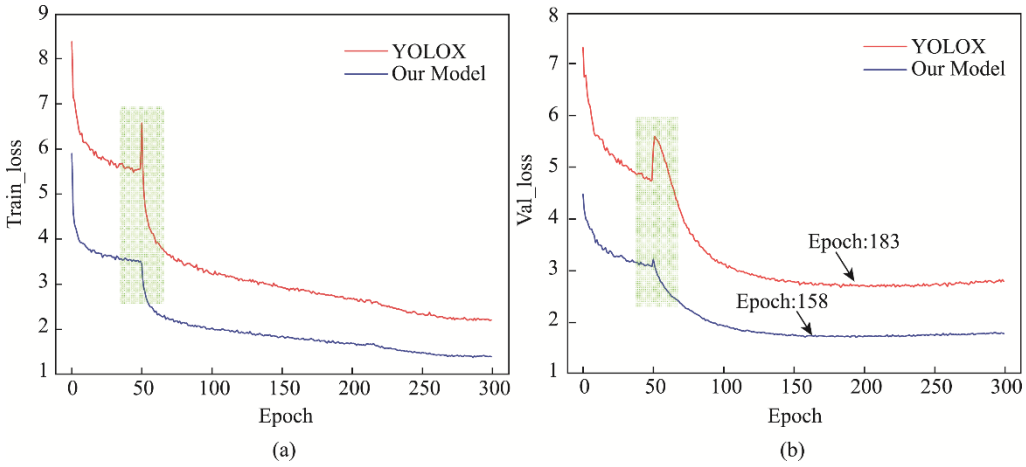

Fig.9 Comparison of training convergence graphs before and after YOLOX improvement on train and validation set: (a) The loss of train set; (b) The loss of validation set.

3.4 and 2.9 respectively, and the accuracy improved significantly.

Comparing the two figures, it can be found that our model converges faster during training, tends to be flat at 158 epoch, the final training loss is 1.393, and the verification loss is 1.783. In contrast, the original YOLOX only flattens out at 183 epochs, the final training loss is 2.214, and the verification loss is 2.789. Therefore, it is fully proved that the stability and convergence of the improved model are significantly improved.

5 Conclusions

This paper introduces a framework based on an enhanced YOLOX model that identifies and localizes defects effectively, even with limited sample sizes. The framework focuses on highlighting defective area features while minimizing complex background distractions. Experimental results demonstrate that the proposed improved YOLOX outperforms other detection methods based on deep learning in terms of detection accuracy, speed, and ease of model deployment. It exhibits certain advantages in defect detection tasks.

In future work, further research should focus on incorporating more defect types and EL image data to enhance the model's generalization capabilities and more effectively assess the operational health of photovoltaic power plants. Additionally, advances in EL sensors and high-performance computing hardware enable the development of portable field detection applications. These advancements offer potential improvements in the efficiency and accuracy of defect detection in real-world settings.

Author Contributions:

Junjie WANG: Writing – original draft, Software, Conceptualization, Validation. Li BI: Supervision, Funding acquisition, Writing – review & editing. Xunde Ma: Writing – original draft, Visualization. Pengxiang Sun: Visualization, Formal analysis.

Funding Information:

This work was supported by the National Natural Science Foundation of China under Grant 62266034 and the Ningxia Natural Science Foundation Key Program under Grant 2023AAC02011.

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Conflict of Interest:

The authors declare no competing interests.

Dates:

Received 14 March 2024; Accepted 21 May 2024; Published online 30 June 2024

References

[1] International Energy Agency (2024), Renewables 2023, IEA, Paris https://www.iea.org/reports/renewables-2023, Licence: CC BY 4.0.

[2] S. Chen, H. Wu, Y. Meng, Y. Wang, X. Li, and C. Zhang, "Reliable Detection Method of Variable Series Arc Fault in Building Integrated Photovoltaic Systems Based on Nonstationary Time Series Analysis," IEEE Sensors Journal, vol. 23, no. 8, pp. 8654-8664, 2023.

[3] D. Korkmaz and H. Acikgoz, "An efficient fault classification method in solar photovoltaic modules using transfer learning and multi-scale convolutional neural network," Engineering Applications of Artificial Intelligence, vol. 113, p. 104959, 2022.

[4] Steven K. Firth, Kevin J. Lomas, Simon J. Rees, A simple model of PV system performance and its use in fault detection, Solar Energy 84 (4) (2010) 624-635.

[5] Köntges M., Kunze I., Kajari-Schröder S., Breitenmoser X., Bjørneklett B. The risk of power loss in crystalline silicon based photovoltaic modules due to micro-cracks. Sol. Energy Mater. Sol. Cells, 95 (4) (2011), pp. 1131-1137

[6] Dhimish M. Micro cracks distribution and power degradation of polycrystalline solar cells wafer: Observations constructed from the analysis of 4000 samples. Renew Energy 2020;145:466-77.

[7] Zhehan Yi, Amir H. Etemadi, Fault detection for photovoltaic systems based on multi-resolution signal decomposition and fuzzy inference systems, IEEE Trans. Smart Grid 8 (3) (2016) 1274-1283.

[8] Mohammed Khorshed Alam, Faisal Khan, Jay Johnson, Jack Flicker, A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques, IEEE J. Photovolt. 5 (3) (2015) 982-997.

[9] A.Ahmad, Y. Jin, C. Zhu, I. Javed, A. Maqsood, and M. W. Akram, "Photovoltaic cell defect classification using convolutional neural network and support vector machine," IET Renewable Power Generation, vol. 14, no. 14, pp. 2693-2702, 2020.

[10] M. Quintana, D. King, T. McMahon, and C. Osterwald, "Commonly observed degradation in field-aged photovoltaic modules," in Conference Record of the Twenty-Ninth IEEE Photovoltaic Specialists Conference, 2002., 2002: IEEE, pp. 1436-1439.

[11] X. Guo and J. Cai, "Optical stepped thermography of defects in photovoltaic panels," IEEE Sensors Journal, vol. 21, no. 1, pp. 490-497, 2020.

[12] T. Fuyuki, H. Kondo, T. Yamazaki, Y. Takahashi, and Y. Uraoka, "Photographic surveying of minority carrier diffusion length in polycrystalline silicon solar cells by electroluminescence," Applied Physics Letters, vol. 86, no. 26, p. 262108, 2005.

[13] T. Trupke, R. Bardos, M. Schubert, and W. Warta, "Photoluminescence imaging of silicon wafers," Applied Physics Letters, vol. 89, no. 4, p. 044107, 2006.

[14] G. C. Eder, P. Grillberger, Y. Voronko, and K. Knöbl, "UV-fluorescence measurements as tool for the detection of degradation effects in PV-modules," 2018.

[15] L. Meng, D. Nagalingam, C. Bhatia, A. Street, and J. Phang, "Distinguishing morphological and electrical defects in polycrystalline silicon solar cells using scanning electron

acoustic microscopy and electron beam induced current," Solar energy materials and solar cells, vol. 95, no. 9, pp. 2632-2637, 2011.

[16] E. A. Zarate, C. G. Treviño-Palacios, R. Rodríguez-Vera, and H. J. Puga-Soberanes, "Defect detection in metals using electronic speckle pattern interferometry," Solar energy materials and solar cells, vol. 88, no. 2, pp. 217-225, 2005.

[17] D. Stellbogen, "Use of PV circuit simulation for fault detection in PV array fields," in Conference Record of the Twenty Third IEEE Photovoltaic Specialists Conference-1993 (Cat. No. 93CH3283-9), 1993: IEEE, pp. 1302-1307.

[18] Liu, Q., et al., An efficient CNN-based detector for photovoltaic module cells defect detection in electroluminescence images. SOLAR ENERGY, 2024. 267.

[19] Meng, S. Xu, L. Wang, Y. Gong, X. Zhang, and Y. Zhao, "Defect object detection algorithm for electroluminescence image defects of photovoltaic modules based on deep learning," Energy Science & Engineering, vol. 10, no. 3, pp. 800-813, 2022.

[20] Pratt, L., D. Govender, and R. Klein, Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation. RENEWABLE ENERGY, 2021. 178: p. 1211-1222.

[21] B. Su, H. Chen, P. Chen, G. Bian, K. Liu, and W. Liu, "Deep learning-based solar-cell manufacturing defect detection with complementary attention network," IEEE Transactions on Industrial Informatics, vol. 17, no. 6, pp. 4084-4095, 2020.

[22] Y. Zhao, K. Zhan, Z. Wang, and W. Shen, "Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line," Progress in Photovoltaics: Research and Applications, vol. 29, no. 4, pp. 471-484, 2021.

[23] Luo, J., et al., FPCB Surface Defect Detection: A Decoupled Two-Stage Object Detection Framework. IEEE Transactions on Instrumentation and Measurement, 2021. 70: p. 1-11.

[24] P. Li, S. Shan, P. Zeng and H. Wei, "Improved Yolov5 Algorithm for Surface Defect Detection of Solar Cell," 2023 35th Chinese Control and Decision Conference (CCDC), Yichang, China, 2023, pp. 3601-3605.

[25] S. Zhao, H. Chen, C. Wang, and S. Shi, "SNCF-Net: Scale-aware neighborhood correlation feature network for hotspot defect detection of photovoltaic farms," Measurement, vol. 206, p. 112342, 2023.

[26] Zhang M, Yin L. Solar cell surface defect detection based on improved yolo v5. IEEE Access. 2022;10:80804-80815.

[27] Zhu, J.H., et al., C2DEM-YOLO: improved YOLOv8 for defect detection of photovoltaic cell modules in electroluminescence image. Non-destructive Testing And Evaluation, 2024.

[28] S. Zhao, H. Chen, C. Wang,S. Shi，"SNCF-Net：Scale-aware neighborhood correlation feature network for hotspot defect detection of photovoltaic farms", Measurement, vol. 206, 2023.

[29] Acikgoz H. An automatic detection model for cracks in photovoltaic cells based on electroluminescence imaging using improved YOLOv7[J]. SIGNAL IMAGE AND VIDEO PROCESSING, 2024, 18(1): 625-635.

[30] Chen H, Song M, Zhang Z, et al. Detection of surface defects in solar cells by bidirectional-path feature pyramid group-wise attention detector. IEEE Trans Instrum Meas. 2022; 71:1-9.

[31] Mazen FMA, Seoud RAA, Shaker YO. Deep learning for automatic defect detection in pv modules using electroluminescence images. IEEE Access. 2023.

[32] Zhao X L, Song C H, Zhang H F, et al. HRNet-based automatic identification of photovoltaic module defects using electroluminescence images[J]. Energy, 2023, 267.

[33] X. Hu, B. Zhao, T. Sun, S. Cao, S. Fan, and H. Xing, "A detecting algorithm for occlusion on the surface of photovoltaic modules based on improved YOLOv5," in 2021 China Automation Congress (CAC), 2021: IEEE, pp. 3131-3135.

[34] Zhang, J.X., et al., A lightweight network for photovoltaic cell defect detection in electroluminescence images based on neural architecture search and knowledge distillation. Applied Energy, 2024. 355.

[35] Chen A, Li X, Jing H, et al. Anomaly Detection Algorithm for Photovoltaic Cells Based on Lightweight Multi-Channel Spatial Attention Mechanism. Energies, 2023.

[36] Mazen F M A, Seoud R a A, Shaker Y O. Deep Learning for Automatic Defect Detection in PV Modules Using Electroluminescence Images[J]. IEEE ACCESS, 2023, 11: 57783-57795.

[37] Zhao Y, Zhang Y, Han L. PV Cell Defect Detection Based on Improved YOLOv7 Modeling[C]. 2nd International Conference on Mechatronics and Automation Technology, ICMAT 2023, October 27, 2023, 2024: 552-558.

[38] X. Chen, T. Karin, and A. Jain, "Automated defect identification in electroluminescence images of solar modules," Solar Energy, vol. 242, pp. 20-29, 2022.

[39] B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection," IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 404-413, 2022.

[40] P. Tománek, P. Skarvada, R. Macku, and L. Grmela, "Detection and localization of defects in monocrystalline silicon solar cell," Advances in Optical Technologies, vol. 2010, 2010.

[41] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132-7141.

[42] Liu, S., D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," arXiv, 2019.

[43] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," in Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 516-520.

[44] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," Neurocomputing, vol. 506, pp. 146-157, 2022.

[45] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980-2988.

[46] U. A. Shah, S. Chen, G. M. G. Khalaf, Z. Jin, and H. Song, "Wide bandgap Sb2S3 solar cells," Advanced Functional Materials, vol. 31, no. 27, p. 2100265, 2021.

[47] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," arXiv preprint arXiv:1710.09412, 2017.