Hindawi
International Journal of Photoenergy
Volume 2024, Article ID 6113260, 13 pages
https://doi.org/10.1155/2024/6113260

Research Article

PA-YOLO-Based Multifault Defect Detection Algorithm for PV Panels

Wang Yin, Zhao Jingyong, Xie Gang, Zhao Zhicheng, and Hu Xiao

School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan, Shanxi 030024, China

Correspondence should be addressed to Zhao Jingyong; s202215110525@stu.tyust.edu.cn

Received 11 October 2023; Revised 8 January 2024; Accepted 24 January 2024; Published 8 February 2024

Academic Editor: Qiliang Wang

In recent years, solar photovoltaic (PV) energy, as a clean energy source, has received widespread attention and experienced rapid growth worldwide. However, the rapid growth of PV power deployment also brings important challenges to the maintenance of PV panels, and in order to solve this problem, this paper proposes an innovative algorithm based on PA-YOLO. First, we propose to use PA-YOLO's asymptotic feature pyramid network (AFPN) instead of YOLOv7's backbone network to support direct interactions of nonadjacent layers and avoid large semantic gaps between nonadjacent layers. For the occlusion problem of dense targets in the dataset, we introduce a repulsive loss function, which successfully reduces the occurrence of false detection situations. Finally, we propose a customized convolutional block equipped with an EMA mechanism to enhance the perceptual and expressive capabilities of the model. Experimental results on the dataset show that our proposed model achieves excellent performance with an average accuracy (mAP) of 94.5%, which is 6.8% higher than YOLOv7. In addition, our algorithm also succeeds in drastically reducing the model size from 71.3 MB to 48.4 MB, which well demonstrates the effectiveness of the model.

1. Introduction

With the evolution of the global energy situation, the urgent need for renewable energy highlights the limitations of fossil fuels and their adverse impact on the environment [1]. Therefore, it has become imperative to seek alternative renewable energy solutions [2]. Solar photovoltaic (PV) technology is being widely emphasized and applied as a clean and renewable energy solution. However, the issue of routine maintenance of photovoltaic panels has become more prominent in the context of the dramatic expansion of PV deployment, where photovoltaic (PV) panels are exposed to a wide range of potential failure types and defects during actual operation. These include, but are not limited to, phenomena such as hot spots, fragmentation, and shading. These problems may trigger energy loss and system efficiency degradation or even, in extreme cases, lead to system failure. Therefore, it is important to use accurate and efficient methods to detect defects in PV panels to ensure the reliability and stability of the PV system. This proactive approach enables early detection, timely intervention, and subsequent remediation, thus ensuring the reliability and stability of the PV system.

However, the distribution environment of large-scale photovoltaic (PV) power stations is complex, covering a large area, and is more cluttered by the influence of terrain, and manual inspection requires a lot of time and energy. The traditional methods for detecting defects in PV panels, such as visual inspection, infrared (IR) thermography [3], Canny and Sobel edge detection operator, and electrical testing, have been widely used in practical applications. However, these methods have some limitations, such as the relatively single type of faults detected and insufficient sensitivity to tiny or hidden defects. With the continuous development of artificial intelligence and machine learning technologies, automated PV panel defect detection methods have become a hot area in research and industry. These methods utilize computer vision, image processing, and data analysis techniques to enable the detection and classification of PV panel defects in an efficient and accurate manner at the same time.

With the development of convolutional neural networks (CNN) and deep learning in the field of image processing,

various deep learning methods have achieved good results in PV panel defect detection. Zhang et al. [4] proposed an improved deep convolutional neural network- (DCNN-) based model. This approach first utilizes the dense cross-stage partial Darknet (DCP-Darknet) network for efficient feature extraction, enhanced feature reuse, and reduced overfitting. Then, efficient fusion of features is achieved by designing a new module that combines a cross-stage feature fusion strategy and deep separable convolution applied to each node of the path aggregation network (PAN). In addition, efficient channel attention (ECA) mechanism is introduced to the PAN. Based on the above methods, Zhang et al. design an efficient algorithm for surface defect detection. Lu et al. [5] proposed a dual-channel convolutional neural network (DCCNN) for automatic diagnosis of PV module faults that automatically extracts key features and weighs these features for fault classification. Menghao and Hongwei [6] integrated image preprocessing, migration learning, and an enhanced feature extraction network into the original faster R-CNN framework for infrared image detection, resulting in a hot spot defect detection model. Winston et al. [7] used feed-forward backpropagation neural networks and support vector machines (SVMs) to identify defects in a variety of photovoltaic (PV) modules, including microcracks and hot spots. In contrast to these image-based approaches, some studies have adopted data-driven methods for PV fault detection. For instance, Madeti and Singh [8] proposed a k-nearest neighbors (kNN) rule-based photovoltaic (PV) system string-level fault detection and diagnosis technique. This technique is capable of detecting and classifying multiple fault types in PV systems in real time, including disconnection faults, line-to-line (L-L) faults, partially shaded with and without bypass diodes, and partially shaded with reverse bypass diode faults. Chen et al. [9] introduced an innovative modeling technique utilizing the extreme learning machine (ELM) using current-voltage (IV) curves collected under different operating scenarios. This novel modeling approach characterizes the electrical properties of PV modules, providing fast training and powerful generalization capabilities. Abbas and Zhang [10] proposed an intelligent system using adaptive neuro-fuzzy inference (ANFIS) for efficient PV fault detection and classification by deploying the trained ANFIS model into the grid partition. The deep learning-based detection technique significantly improves the accuracy compared to traditional image processing methods. However, the detection stability of the method is poor under different weather conditions, and there are challenges in detecting faults in some dense and small targets. In addition, fault detection involves multiple types of faults, and current research focuses on optimizing one of the methods, which leads to industrial applications that are still in the early stages of development.

In order to improve detection accuracy, one-stage target detection models, especially the YOLO series, were developed. Among them, YOLOv7 excels in the field of target detection. It adopts multiscale feature fusion and convolutional neural network structure to achieve excellent target localization and recognition. The end-to-end design and optimized loss function further enhance target detection efficiency and accuracy. Therefore, YOLOv7 is suitable to be used as a research benchmark as a single-stage target detection model with fewer parameters and higher performance [11]. However, for the current dataset, the native YOLOv7 suffers from the problems of large model size and low accuracy for target detection such as dense hot spots. Therefore, this paper proposes a more efficient single-level target detection model: the PA-YOLO (the name PA-YOLO is derived from the letters P and A in the asymptotic feature pyramid network (AFPN) and EMA mechanism).

The main contributions of this study are as follows:

(1) Introducing asymptotic feature pyramid network (AFPN) to enhance the direct interaction of nonadjacent layers, we adopt an improved asymptotic feature pyramid network (AFPN) to replace the original feature fusion network in YOLOv7. This improvement helps to avoid the problem of large semantic differences between different levels and the loss or degradation of feature information

(2) Customized convolution equipped with EMA (efficient multiscale attention) mechanism is proposed, which automatically learns and selects key features in the image to enhance the model's ability to perceive and recognize defects in PV panels. Compared to the traditional attention mechanism, this approach avoids introducing more network complexity, thus ensuring the efficiency and practicality of the model

(3) Replacing the traditional CIOU loss function with the repulsion loss function significantly improves the detection accuracy in the case of occlusion. By using the repulsion loss function, the leakage rate of dense targets in the dataset is reduced and the performance of target detection is further improved

(4) Balanced training is introduced, and the trained PA-YOLO model achieves 94.5% mAP on the PV panel dataset; meanwhile, the network structure is compressed from 104 layers to 70 layers, which greatly simplifies the complexity of the model structure, and the comprehensive performance is better than that of other target detection networks

2. PA-YOLO Algorithm

2.1. YOLOv7 Structure. YOLOv7, as an outstanding representative of the YOLO series of target detection models, introduces a new label assignment method called "coarse-to-fine guided label assignment" [12]. The emergence of this method solves the key problem faced in dynamic label assignment, i.e., how to assign dynamic target labels to the outputs of different branches. Second, in YOLOv7, the model reparameterization technique effectively merges multiple computational modules into a single entity during the inference phase. This approach not only helps to build a more streamlined and efficient inference model but also reduces the computational burden while still maintaining

the high accuracy of the model. By cleverly applying reparameterized modules, YOLOv7 is able to leverage its optimized performance across different architectures and adapt to diverse application scenarios. The strategy of combining module-level and model-level aggregation further enhances YOLOv7's robust performance and accuracy in target detection tasks.

ELAN plays a key role in YOLOv7 by optimizing the gradient paths of deeper networks, thereby significantly enhancing the learning and convergence capabilities of the model. Through an innovative approach, ELAN extends, reorganizes, and merges the channels and computational blocks of the network, which not only enhances the learning ability of the network but also maintains the integrity of the original gradient paths. In addition, ELAN ensures that the model is able to learn a more diverse set of features by effectively steering individual sets of computational blocks, which directly improves the overall accuracy and robustness of the model. ELAN skillfully finds a balance between expanding the number of computational blocks and maintaining the stability of the network, allowing the network to maintain good scalability without sacrificing performance. With this innovation, YOLOv7 significantly improves the detection accuracy of the model, enabling it to outperform many other object detectors on the COCO dataset, including YOLOR, YOLOX, Scaled-YOLOv4, YOLOv5, DETR, and Deformable DETR. Meanwhile, this new label assignment method allows YOLOv7 to better understand and handle targets of different scales, shapes, and locations, which improves the accuracy and stability of the detection, as shown in Figure 1 for the model structure of YOLOv7.

YOLOv7 receives input images of solar panels taken by infrared cameras that contain various types of defects, followed by a backbone network that uses a deep convolutional neural network to extract key features from the input images. It operates in a layered fashion, starting with basic textures and edges in the early layers and moving to more complex patterns that represent potential defects in the solar panel. The network is trained to recognize subtle differences between normal panel features and anomalies. The neck acts as an intermediary to enhance and consolidate the features extracted by the backbone network. It uses techniques such as feature pyramid networks to combine high-resolution details with high-level semantic information to ensure that subsequent detection heads have rich feature representations at different scales. This step is crucial for detecting defects of different sizes and severities. The detector head part is responsible for the final defect detection and classification. It consists of multiple detection heads that operate at different scales to accommodate the various sizes of defects that may exist on the solar panel. Each detection head predicts the bounding box and associated confidence score, indicates the presence and location of the defect, and classifies the defect type.

Although YOLOv7 has made significant strides in the technological advancement of object detection, its complex 104-layer network structure still includes numerous convolutional operations. This design may cause the network to lose some valuable data while processing the information, especially in the case of photovoltaic panel fault detection where the capture of details and overall features appears to be insufficient. This loss of information particularly affects fault types that are similar in shape, color, and size, such as the "hot spot" fault shown in Figure 2(b), which consists of multiple hot spots in close proximity and is visually very similar to the "battery string" fault. The key to distinguishing between these two faults lies in their location: the "battery string" fault is generally located at the edge of the panel, as shown in Figure 2(a), covering about one-third of the panel area, while the "hot spot" is usually distributed more randomly. Therefore, in order to effectively distinguish the "hot spot" faults in Figure 2(b), we need to take into account the information around the faults and the characteristics of the whole panel for a comprehensive evaluation, which is important for reducing the misdetection between similar faults.

2.2. AFPN. In the PA-YOLO framework (e.g., Figure 3), the asymptotic feature pyramid network (AFPN) adeptly integrates features across various levels [13], specifically targeting low-level, high-level, and top-level features. The structure of AFPN, depicted in Figure 4, utilizes black arrows to signify convolution processes and blue arrows to indicate adaptive spatial fusion. The initial phase of feature fusion involves extracting the last layer of features from each level of the backbone network, culminating in a diverse set of features at different scales, identified as C2, C3, C4, and C5.

The design of the AFPN in the PA-YOLO framework is strategically tailored to bridge the semantic gap between nonadjacent layers, a prevalent challenge in object detection methods that rely on feature pyramid networks. The fusion process commences with the integration of the lower layer features, C2 and C3, into the feature pyramid network. This foundational step is pivotal, setting the stage for the subsequent integration of features. The methodology then progressively incorporates higher-level features, with C4 being added next, followed by the inclusion of the topmost layer, C5. This methodical layer-by-layer integration is key to diminishing the semantic gap, thereby enhancing the efficacy of the feature fusion.

In essence, the AFPN in PA-YOLO begins by merging the foundational bottom layer features (C2 and C3), then progressively integrates the more complex layer feature (C4), and culminates with the fusion of the highest layer feature (C5), representing the most abstract level of features. This incremental fusion process is instrumental in harmonizing the semantic content of features across different layers in the AFPN, effectively addressing the challenges posed by direct fusion of semantically disparate layers.

In the process of multilevel feature fusion, ASFF is utilized to assign different spatial weights to features of different levels, which effectively enhances the importance of key levels.

Presented in Figure 5 is an illustration of feature fusion at three different levels, fusing 3 levels of features with $x_{ij}^{n\longrightarrow l}$ denoting the feature vector at the position of level n to level l. The resultant vector denoted as y_{ij}^{l} is obtained by

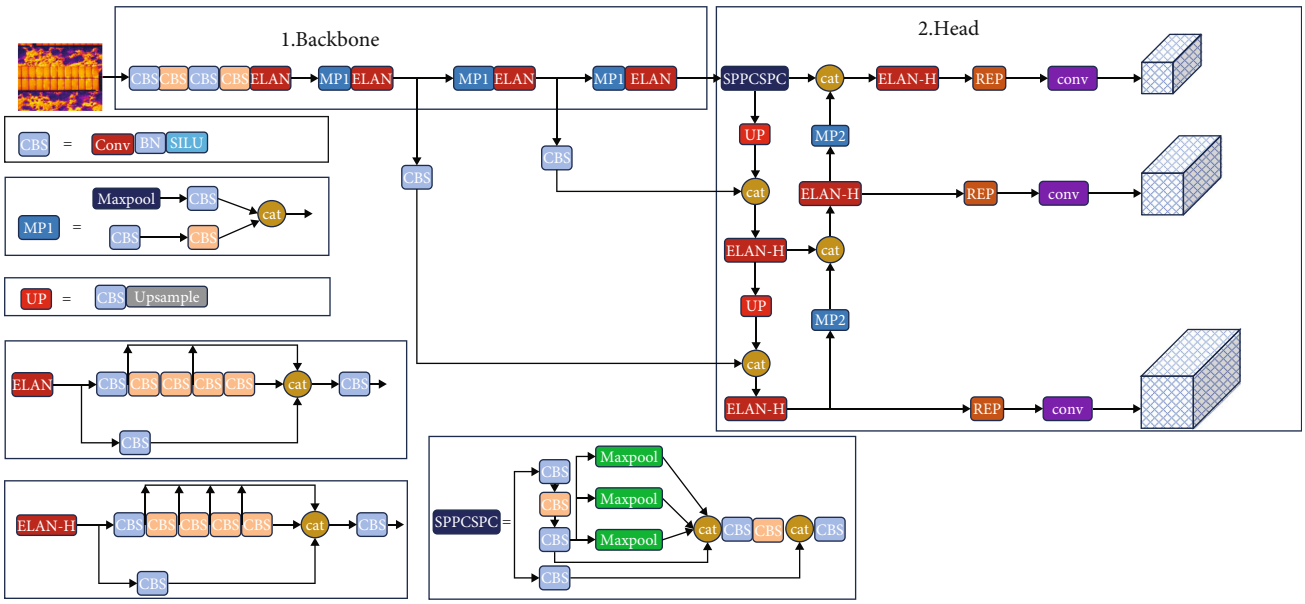
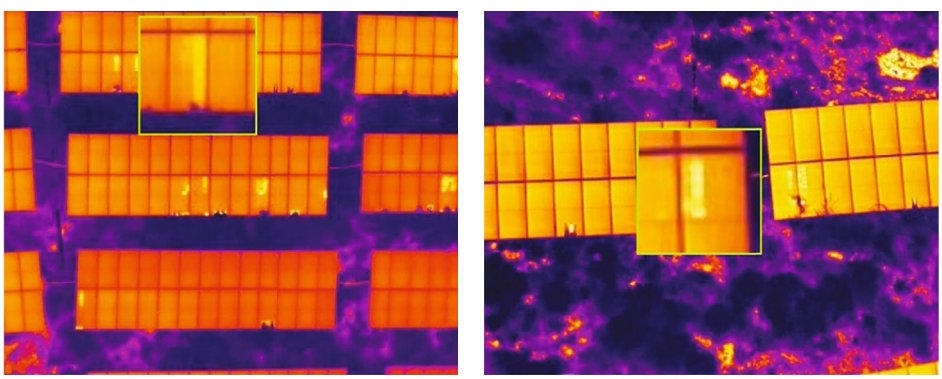


Figure 1: YOLOv7 structure diagram.



(a) (b)

Figure 2: Similar faults.

adaptive spatial fusion of features at multiple levels with a linear combination of the feature vectors $x_{ij}^{1\longrightarrow l}$, $x_{ij}^{2\longrightarrow l}$, and $x_{ij}^{3\longrightarrow l}$ as in

$$y_{ij}^{l} = \alpha_{ij}^{l} \cdot x_{ij}^{1\longrightarrow l} + \beta_{ij}^{l} \cdot x_{ij}^{2\longrightarrow l} + \gamma_{ij}^{l} \cdot x_{ij}^{3\longrightarrow l}, \tag{1}$$

where α_{ij}^{l}, β_{ij}^{l}, and γ_{ij}^{l} denote the feature space weights of the three classes, subject to the constraints of $\alpha_{ij}^{l} + \beta_{ij}^{l} + \gamma_{ij}^{l} = 1$.

The adaptiveness of the AFPN is achieved through the assignment of spatial weights (α_{ij}^{l}, β_{ij}^{l}, and γ_{ij}^{l}) to features at different levels, ensuring that shallow and deep features contribute optimally to the final fused feature vector y_{ij}^{l}. These weights are learned during the training process and are constrained such that the sum of weights at any given position (i, j) for a level l equals to 1, i.e., $\alpha_{ij}^{l} + \beta_{ij}^{l} + \gamma_{ij}^{l} = 1$.

For instance, consider a scenario where shallow features need to be emphasized due to their fine-grained spatial information. In this case, the network might learn to assign higher weights to α_{ij}^{l}, compared to β_{ij}^{l} and γ_{ij}^{l}, which correspond to deeper features. Conversely, if the context provided by deeper features is more critical, β_{ij}^{l} and γ_{ij}^{l} might receive higher weights. The specific values of these parameters are dynamically adjusted during the training process based on the loss function and the backpropagation signals, reflecting the importance of each feature level for the detection task at hand.

Regarding parameter learning attenuation, the network relies on the adaptive nature of the learning process itself to fine-tune these parameters. The optimization process inherently adjusts the contribution of each feature level over time as it minimizes the loss, without a predefined decay schedule. The stage-specific implementation of adaptive spatial fusion modules allows for a flexible and

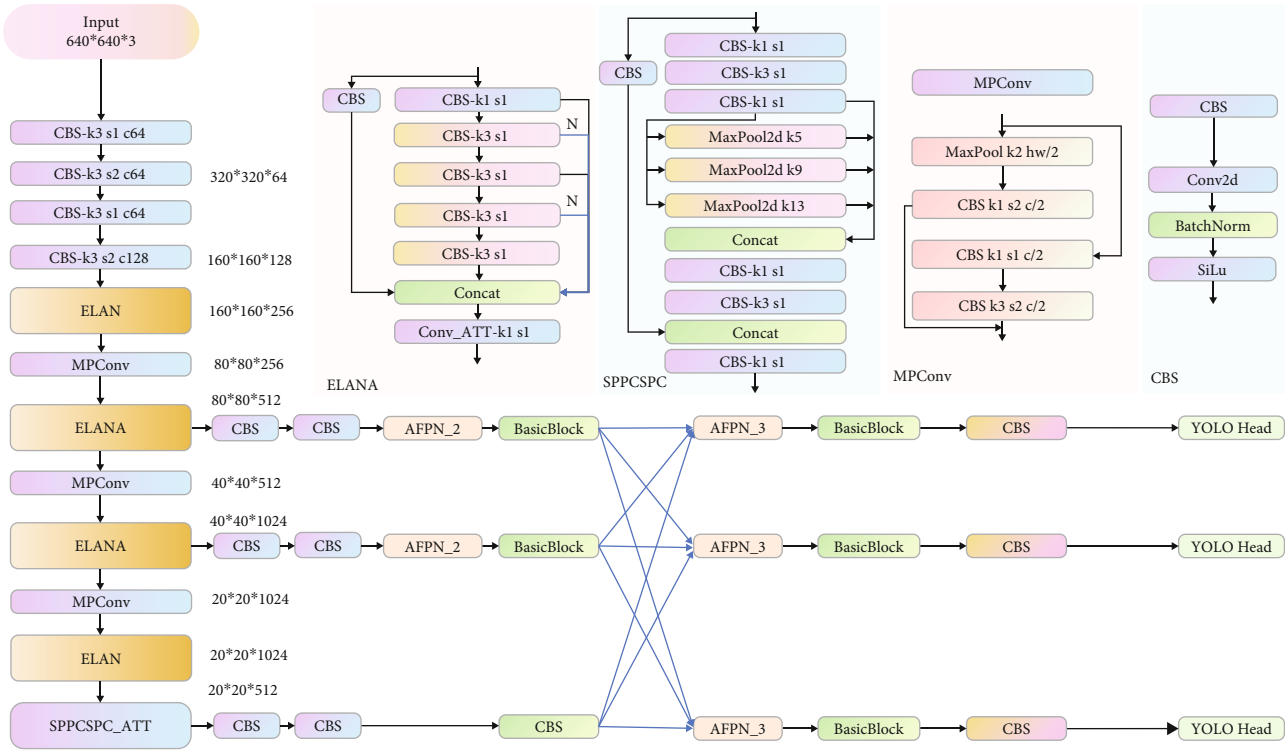


Figure 3: PA-YOLO structure.

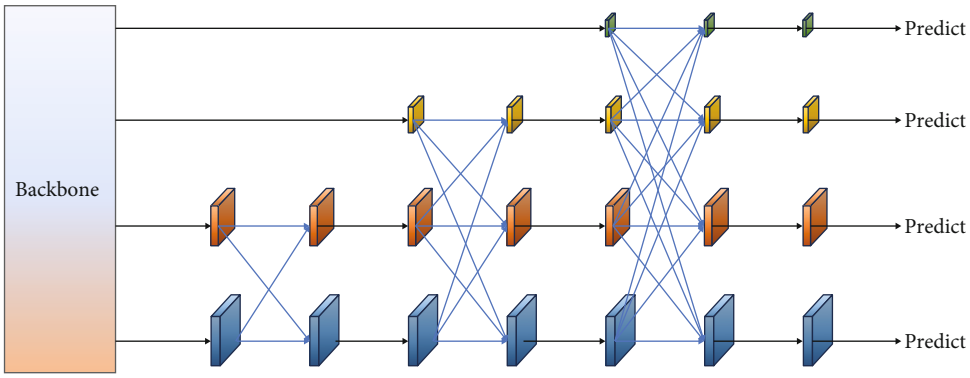


Figure 4: Asymptotic feature pyramid network (AFPN) architecture.

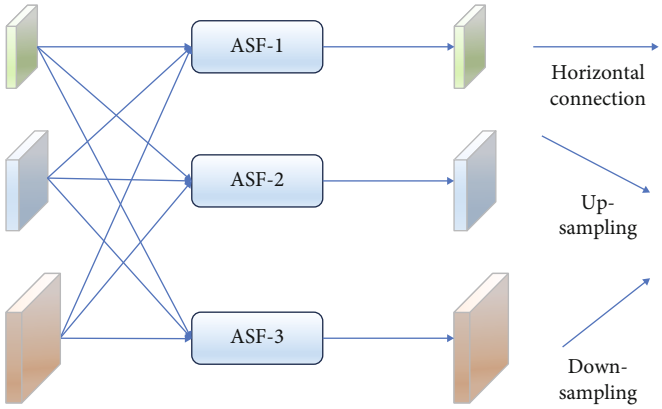


Figure 5: Adaptive spatial fusion operation.

tailored approach to feature fusion, catering to the unique requirements of different stages in the network. This design choice is made to ensure that the network can effectively learn and adapt the contribution of each feature level without the need for an explicit attenuation mechanism.

2.3. Loss Function. The choice of loss function in an object detection network plays a key role in determining its detection accuracy, as it directly affects the training and optimization process of the model. An effective loss function can enhance the model's ability to fit training data, thereby improving detection accuracy. In the YOLO model family, commonly used bounding box regression losses include IoU, GIoU, CIoU, and DIoU losses [14]. These loss functions quantify the difference between the predicted and target boxes by taking into account factors such as overlap, centroid distance, and aspect ratio. Take the CIoU loss function used by default in YOLOv7 as an example:

$$\begin{aligned} \mathrm{CIOU} &= \mathrm{IOU} - \left(\frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v\right), \\ v &= \frac{4}{\pi^2}\left(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h}\right)^2, \\ \alpha &= \frac{v}{(1-\mathrm{IOU}) + v}. \end{aligned} \tag{2}$$

The CIOU loss function has a high sensitivity to the bounding box, when there are outliers in the sample may lead to the model in the training process does not converge or oscillation phenomenon, and secondly, the CIOU loss of the parameter accuracy requirements are high; otherwise, it will not only lead to the network training speed being slow but also increase the computational cost of the network.

Traditional loss functions (e.g., CIoU), while effective in many cases, exhibit limitations when dealing with datasets containing poor-quality instances or dense objects. These limitations manifest themselves in higher sensitivity to outliers, which can lead to training challenges such as nonconvergence or oscillations. In addition, the accuracy requirements of CIoU can lead to slower network training and increased computational costs. We introduce a repulsion loss function in the PA-YOLO model specifically to address these limitations. The repulsion loss function contains both attraction and repulsion terms and is designed to handle dense target scenes more efficiently. This function not only minimizes the distance between the predicted frame and the actual target to improve the accuracy of target detection but also maintains the distance between the predicted frame and other objects or predicted frames. This dual approach is crucial in datasets with dense target occlusion, where traditional loss functions (e.g., CIoU) are difficult to function. Therefore, the repulsion loss function provides a more robust solution for object detection in complex scenes, overcoming the challenges posed by traditional loss functions. By incorporating this function, our PA-YOLO model improves the detection accuracy and training efficiency, especially in object-dense scenes, which are common in real-world applications.

Consequently, in this paper, the repulsion loss function [15] is adopted to focus on coping with the dense target occlusion problem. Specifically, the paper proposes two types of repulsion loss, namely, RepGT loss and RepBox loss.RepGT loss penalizes the prediction frames directly and prevents them from transferring to other ground-truth objects, while RepBox loss requires each prediction frame to keep a certain distance from other prediction frames of different targets, so as to reduce the dependence of detection results on nonmaximum suppression (NMS) and effectively deal with the dense occlusion problem. Repulsion loss is defined as in

$$L = L_{\mathrm{Attr}} + \alpha * L_{\mathrm{RepGT}} + \beta * L_{\mathrm{RepBox}}, \tag{3}$$

$$L_{\mathrm{Attr}} = \frac{\sum_{P\in P^+}\mathrm{Smooth}_{L1}\left(B^P, G^P_{\mathrm{Attr}}\right)}{|P^+|}, \tag{4}$$

$$L_{\mathrm{RepGT}} = \frac{\sum_{P\in P^+}\mathrm{Smooth}_{\ln}\left(\mathrm{IoG}\left(B^P, G^p_{\mathrm{Rep}}\right)\right)}{|P^+|}, \tag{5}$$

$$L_{\mathrm{RepBox}} = \frac{\sum_{i\neq j}\mathrm{Smooth}_{\ln}\left(\mathrm{IoU}\left(B^{P_i}, B^{P_j}\right)\right)}{\sum_{i\neq j}1\left[\mathrm{IoU}\left(B^{P_i}, B^{P_j}\right) > 0\right] + \varepsilon}, \tag{6}$$

$$\mathrm{IoG}\left(B^P, G^P_{\mathrm{Rep}}\right) = \frac{\mathrm{area}\left(B^P \cap G^P_{\mathrm{Rep}}\right)}{\mathrm{area}\left(G^P_{\mathrm{Rep}}\right)}, \tag{7}$$

where L_{Attr} is an attraction term that requires the prediction box to be close to its designated target and L_{RepGT} and L_{RepBox} are the repulsion terms that require the prediction box to be far away from other real objects around it and other prediction boxes with different designated targets, respectively. The coefficients α and β are weights to balance the auxiliary losses. For L_{Attr}, the Smooth_{L1} distance is chosen as a measure of the attraction term as in Equation (4). In calculating the attraction term loss, first for each candidate frame $P \in P+$ ($P+$ denotes the set of candidate frames), it is assigned to the actual target frame with which it has the maximum intersection-to-union (IoU) ratio, which is considered as its assigned target frame, i.e., $G^P_{\mathrm{Attr}} = \arg\max_{G\in G}\mathrm{IoU}(G, P)$. Subsequently, the attractor loss can be computed by comparing the prediction frame BPs of the candidate frames with their assigned target frames. The goal of this loss is to improve the performance and accuracy of target detection by minimizing the attractor loss to induce the predictor frame to locate the target object more precisely. This approach helps to minimize the distance between the prediction frame and the actual target, leading to better target matching. The meanings represented by the other parameters in the formula are as follows: G^P_{Attr}: ground-truth box assigned as the designated target for

proposal P; $\text{Smooth}_{\ln}$: a smooth $L1$ distance metric used for measuring the difference between the predicted box B^P and its designated ground-truth box G^P_{Attr}; IoG: intersection over ground-truth, measuring the overlap between the predicted box B^P and the repulsion ground-truth object G^P_{Rep}; G^P_{Rep}: ground-truth object that has the largest IoU region with proposal P, excluding its designated target; B^{P_i}, B^{P_j}: predicted boxes for different proposals; $1[\text{IoU}(B^{P_i}, B^{P_j}) > 0]$: an indicator function that is active only when there is an overlap between the predicted boxes; ε: a small constant to prevent division by zero; $\text{area}(B^P \cap G^P_{\text{Rep}})$: the intersection area between B^P and G^P_{Rep}; and $\text{area}(G^P_{\text{Rep}})$: the area of the ground-truth box.

2.4. Custom Convolution with EMA Mechanism. The datasets used in this study are from different power stations, and the datasets involve the influence of different weather, geographic environment, shooting height, and other factors in the process of collection, so there are various differences in the manifestation of the same fault type on different power stations, for example, the infrared camera, when shooting; because of the influence of the ground temperature, the picture will show a red or purple color difference, and these differences in color will give the model to bring a certain degree of misdetection, or the different flight altitude of the UAV when collecting data will lead to the same fault type; there is a difference in scale. The "reweight" operation in the EMA mechanism can enhance those features that are more important for the task at hand (e.g., defect detection). For example, when detecting small defects or irregular color differences, the model can pay more attention to small changes in those features. The "reweight" operation effectively fuses features at different scales to enhance or suppress certain features by applying the attention map derived from the input feature map to the original feature map. This reweighting step is crucial in the attention mechanism as it allows the network to focus its attention on the more relevant parts of the input data, which in the case of the EMA module involves channel information and spatial information. Essentially, the "reweighting" here is the application of the learned attentional map to the original input features, enhancing the sensitivity of the model to modulate them for subsequent processing.

The EMA mechanism [16] is a novel and efficient multiscale attention module. Its main goal is to maintain the integrity of information from different channels without significantly increasing the computational overhead. This is achieved by reshaping certain channels into batch dimensions and organizing the channel dimensions into multiple subfeatures. This reconstruction ensures that spatial semantic features are evenly distributed within each feature group. The overall structure of the EMA model is shown in Figure 6. To ensure the efficiency of the model without adding unnecessary complexity, we seamlessly integrate the EMA mechanism into the 1×1 convolution layer, creating a custom convolution block with attention function. This custom convolution block (which we call Conv-ATT) is shown in Figure 7.

EMA divides any given feature input $X \in R^{C \times H \times W}$ into G subfeatures across the channel dimension directions in order to learn different semantics. EMA employs three parallel pathways to extract attention weight descriptors from the grouped feature graph. Two of these pathways are part of the 1×1 branching, while the third belongs to the 3×3 branching. This design is aimed at capturing channel dependencies and reducing computational demands effectively. To achieve this, two 1D global average pooling operations are utilized within the 1×1 branch. These pooling operations encode channel information along both spatial directions independently. Conversely, in the 3×3 branch, only a single 3×3 kernel is employed to capture multiscale feature representations. This strategic configuration optimizes the balance between capturing essential channel dependencies and managing computational resources. After two 1D global average pooling operations, a similar processing method to the CA mechanism is applied, and finally, the original intermediate feature maps are aggregated using the learned attention map weights of the two parallel routes as the final output [17]. Then, the global spatial information is encoded in the output of the 1×1 branch using 2D global average pooling, and the output of the smallest branch is directly converted to the shape $R_1^{1 \times C//G} \times R_3^{C//G \times HW}$, and 2D global pooling operation of the corresponding dimension before the joint activation mechanism of the channel features as follows [18]:

$$z_c = \frac{1}{H \times W} \sum_j^H \sum_i^W x_c(i, j). \tag{8}$$

To improve computational efficiency, we applied a Softmax function for linear transformation adaptation on the output of 2D global average pooling [19]. By combining the output of parallel processing with matrix dot product operations, we generated the first spatial attention map. At the same time, we also encoded the global spatial information of the 3×3 branches using 2D global average pooling to obtain the second spatial attention map. Finally, we fused the set of spatial attention weight values computed from the output feature maps in each group from the two generated spatial attention maps using the sigmoid function. This approach helps to capture pixel-level pairwise relationships and highlights the global context. Through cross-spatial information aggregation, we are able to capture long-range dependencies and embed precise location information into the EMA. In this way, the fusion of contextual information at different scales allows the CNN to better focus on the pixel-level details of high-level feature maps.

3. Experimental Results and Analysis

3.1. Experimental Environment. In our experiments, we used the PyTorch framework for the study under the following conditions: Ubuntu 20.04 operating system, Python 3.8, PyTorch version 1.10.0, and RTX A5000 GPU with 24 GB of graphics memory and 30 G of RAM. Trained using SGD

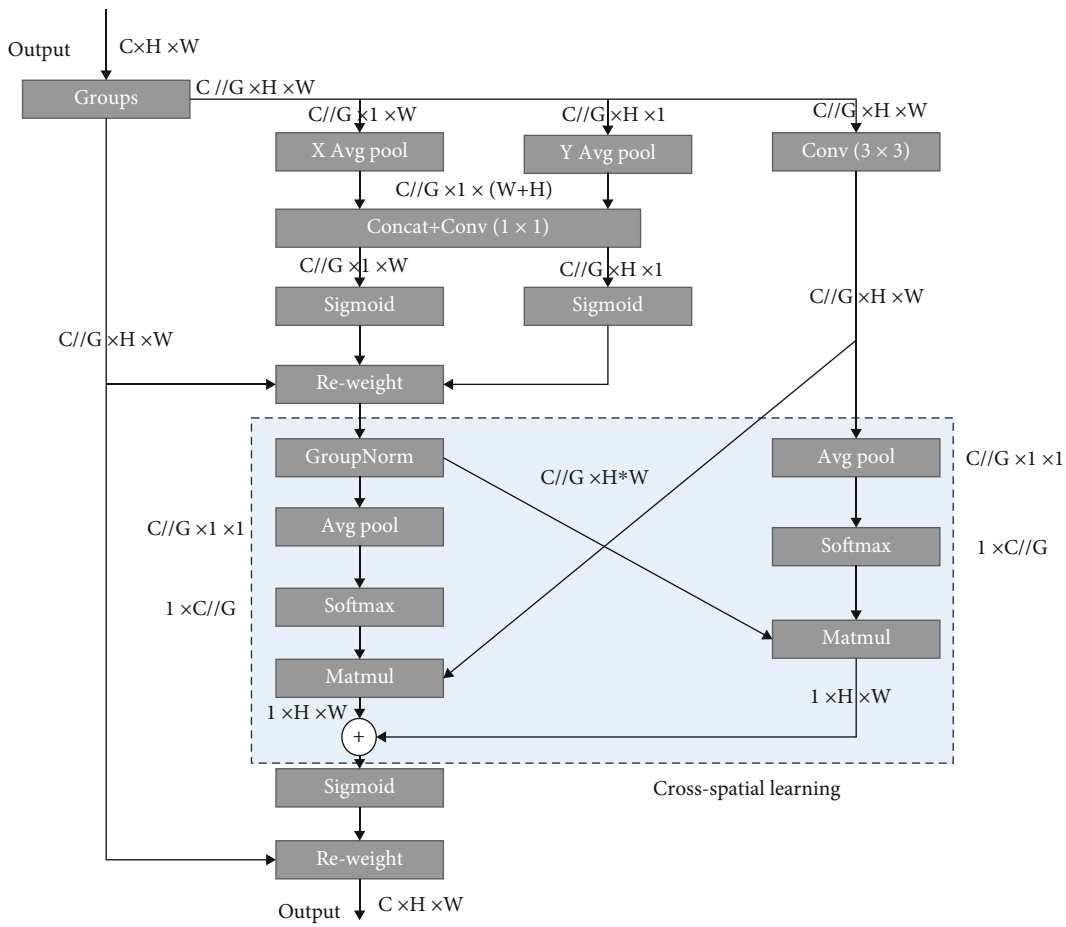


Figure 6: EMA model structure.

Figure 7: Conv-ATT convolution.

optimizer, input image is 640×640, learning rate is 0.01, batch size is 32, and epochs are 100.

3.2. Datasets. In this study, we utilized infrared photovoltaic panel images acquired by our group using a drone. These images were obtained from 13 large power stations, forming a dataset of about 8,900 infrared images (image resolution information is 640×512). Given that temperature variations lead to changes in PV panel color [20], we carefully selected 2,549 images from the original dataset as the basis of our study with the aim of improving detection accuracy. During the annotation process, we manually labeled these images using the LabelImg tool. Specifically, we categorized and labeled four types of faults: breaks, hot spots, plant shadows, and battery strings. The shatter and string categories were defined based on the size of the PV module, while plant shadows and hot spots were labeled based on the size and location of the fault point. It is worth noting that we purposely excluded positive samples with unclear pixels [21] to prevent overfitting of the neural network. The final dataset is in YOLO format and contains a total of 25,017 labels. Figure 8 gives an example of the distribution of labels in the dataset.

Figure 9 illustrates the characteristics of the four fault types mentioned above. Figure 9(a) is the fragmentation, which is characterized by a large number of densely distributed, irregular hot spots. Figure 9(b) is the hot spot, which manifests itself as small, bright hot spots with sharp, usually square edges. Figure 9(c) is the vegetation shading, which sometimes shades only one PV panel and sometimes shades multiple PV panels due to irregular vegetation growth. Figure 9(d) is the cell strings, usually rectangular distribution, accounting for 1/3 or 2/3 of the whole module. Considering the for case of images and labels, we divide the dataset into three parts according to 7 : 2 : 1, with 70% for training, 20% for validation, and 10% for testing.

3.3. Evaluation Indicators. In order to assess the performance difference of the network model for various types of faulty images under the same experimental conditions as well as the false and missed detections, we used three evaluation metrics, AP, recall, and mAP, to measure the comprehensive

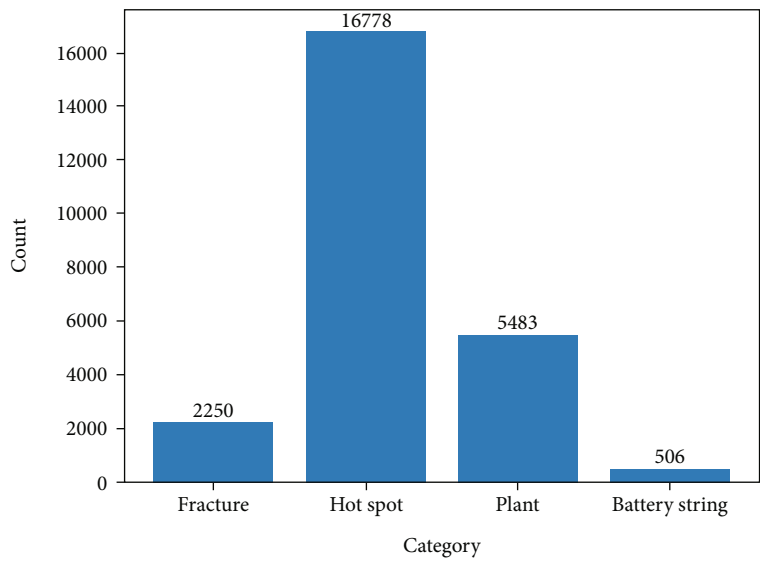


Figure 8: Label distribution map.

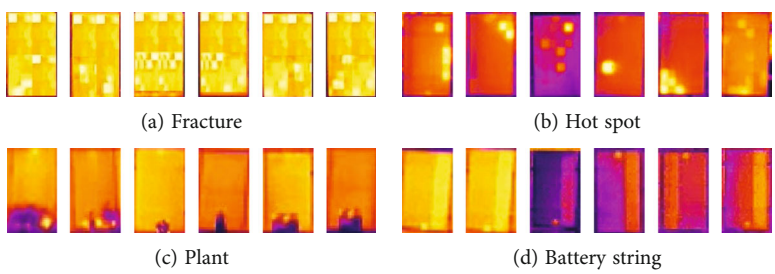


Figure 9: Example of a fault.

performance of the network. The formulas for these metrics are as follows:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}}, \\ \text{AP} &= \int_0^1 \text{precision}(\text{recall})\,\text{d}(\text{recall}), \\ \text{mAP} &= \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i. \end{aligned} \tag{9}$$

In the formula, precision indicates the total number of positive samples identified correctly over the total number of samples identified as positive samples; recall is the total number of positive samples identified correctly over the total number of samples identified as positive samples; TP is the number of positive samples predicted to be positive samples; FN indicates the number of positive samples predicted to be negative samples; FP is the number of negative samples predicted to be positive samples; TN is the number of negative samples predicted to be negative samples; AP is the average precision, which is the area circumscribed by PR curves and the axes of each category; and mAP is the averaging of the AP values for all categories.

3.4. Test Results and Analysis

3.4.1. Comparison with Other Algorithms. In order to validate the state-of-the-art of the algorithms proposed in this paper, we selected Lite$^{our}_{g}$-YOLOv5, YOLOv5-s, YOLOX-m [22], and YOLOv7 for comparison experiments. These experiments used the same equipment, dataset, and data augmentation methods, while keeping the training and test sets in equal proportions. The experiments were performed for 100 iterations, and the best results were selected for testing. Table 1 lists the mAP, recall (recall), and AP values for each fault type for the different algorithms.

As can be seen from Table 1, under the same experimental conditions, the algorithm PA-YOLO proposed in this paper has a significantly higher mean accuracy (mAP) value of 94.5% compared with other models such as Lite$^{our}_{g}$-YOLOv5 (83.1%), YOLOv5-s (83.4%), YOLOX-m (85%), and the standard YOLOv7 (87.7%). This significant difference in mAP proves the excellent performance of PA-YOLO. This significant difference in mAP demonstrates the excellent performance of PA-YOLO. In addition, PA-YOLO's recall increased by 11.9% compared to YOLOv7. The higher recall indicates that PA-YOLO is better able to

Table 1: Comparison of different network models.

Algorithm	mAP	R	Fracture	Hot spot	Plant	PV panel string	FPS
Lite^{our}_{g}-YOLOv5	83.1	77.1	86.8	68.4	80.7	96.6	65
YOLOv5-s	83.4	77.3	96.1	60.7	83.7	93	62
YOLOX-m	85	79.5	75	80	93	92	78
YOLOv7	87.7	79.9	98.7	70.3	89.9	91.9	80
PA-YOLO	94.5	91.8	99.4	88.6	92.7	97.3	83

Note: data in the table are expressed as percentages except FPS. The unit of FPS is frames per second. "Battery string" and "PV panel string" in Table 1 represent the same meaning.

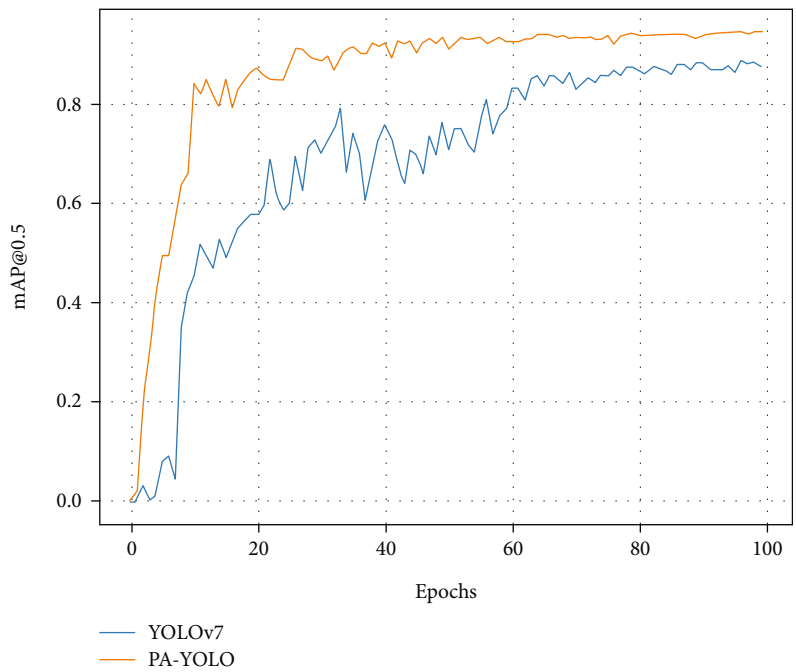


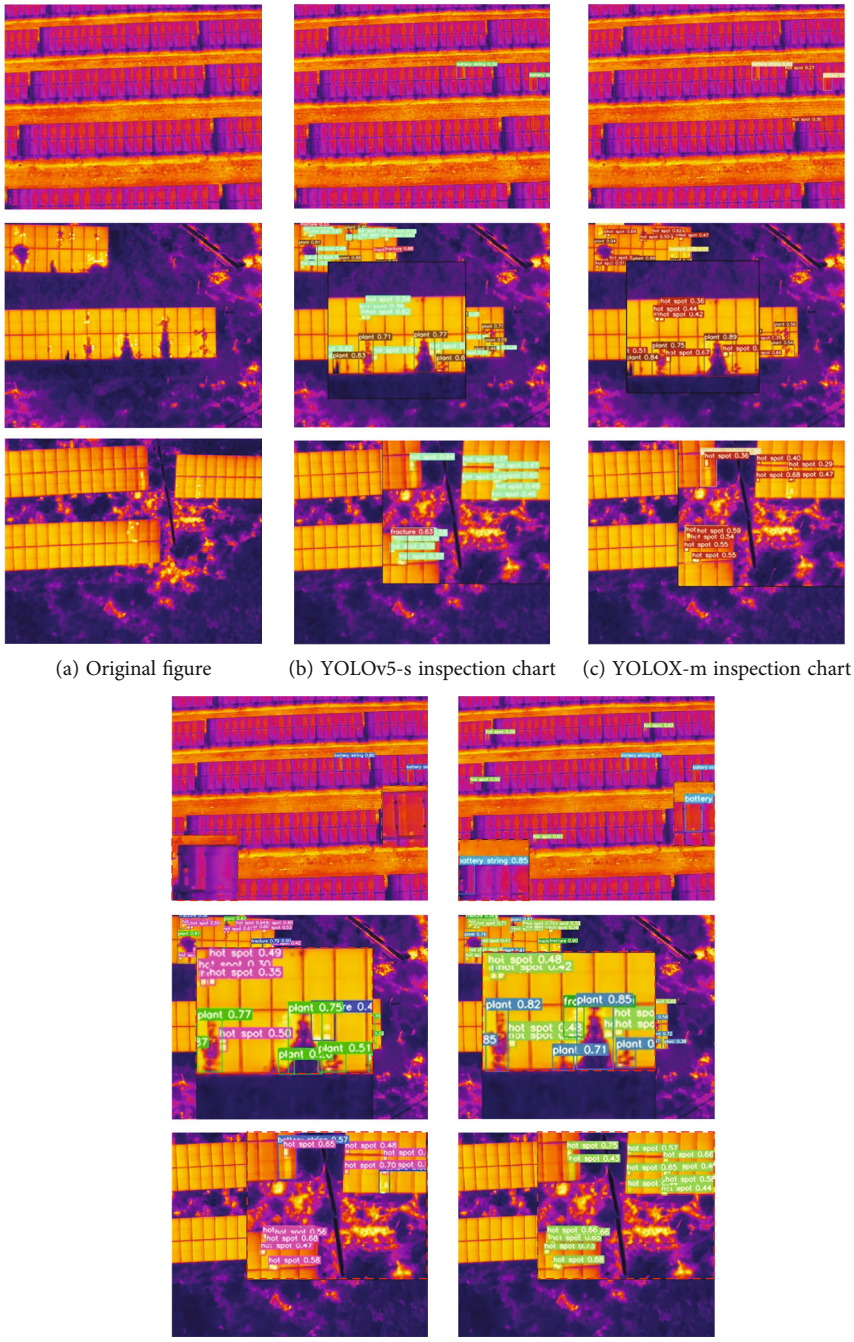
Figure 10: Comparison of mAP curves.

capture the target objects in the image, which reduces missed detections and lowers the risk of missing important information [23]. To visualize the performance improvement, Figure 10 shows a comparison of the network mAP curves before and after the improvement. It is clear that the PA-YOLO network outperforms the original YOLOv7 network on this dataset, showing higher stability and detection capabilities. In conclusion, PA-YOLO is an efficient algorithm with significant mAP values and higher recall that outperforms other models including YOLOv7 in a given experimental setting.

In order to visualize the superiority of the proposed algorithms in this paper, Figure 11 shows some comparative results of YOLOv5, YOLOX, and YOLOv7 algorithms with the algorithms in this paper on the test set. As can be seen from the bottom of Figure 11(a), there is only one kind of defect "hot spot" on the photovoltaic panel, while there are different degrees of misdetection and omission in Figures 11(b) and 11(c), and there are also two misdetections in Figure 11(d), which misdetect the hot spot as "battery string," as shown in the enlarged image in the red curve box in Figure 11(d). Second, there are a large number of missed detections in the detection diagrams in the middle of Figures 11(b) and 11(c). In the middle picture of Figure 11(d), YOLOv7 detects one plant shade as multiple ones and misses one plant shade and misdetects the "hot spot" as "broken," as shown in the blue box, and misdetected "hot spot" as "cracked," as shown in the blue box. The top images in Figures 11(b)–11(d) have missed detections of "battery strings." The above YOLOv7 network in the detection of various errors in this paper's algorithm PA-YOLO has been effectively resolved, greatly reflecting the superiority of this paper's algorithm.

3.4.2. Ablation Experiment. In order to fully verify the performance of each module in the PA-YOLO algorithm, ablation experiments are conducted on repulsion loss, asymptotic feature pyramid network (AFPN), and convolutional block with EMA mechanism under the same hyperparameters and experimental environments, and the results of the experiments are shown in Table 2.

From the experimental results, it can be seen that the mAP value and recall (R) of each module are improved when the ablation experiments of individual modules are



(a) Original figure (b) YOLOv5-s inspection chart (c) YOLOX-m inspection chart

(d) YOLOv7 inspection chart (e) PA-YOLO inspection chart

Figure 11: Comparison of test results.

Table 2: Effectiveness of different modules on model detection.

Repulsion loss	AFPN	Conv-ATT	R (%)	Model size (M)	mAP (%)
			79.8	71.3	87.7
√			87.9	71.3	91.1
	√		88.2	48.4	92.4
		√	85.0	71.8	92.2
√	√	√	91.8	48.4	94.5

performed, and the mAP value of the fused PA-YOLO is improved by 6.8% compared to YOLOv7, which is a significant improvement over the original model. The experimental results show that replacing the original feature fusion network of YOLOv7 with the asymptotic feature pyramid (AFPN) structure greatly compresses the size of the model while improving the network recall, indicating that the introduction of the new feature fusion network is imperative for improving the network detection accuracy. The addition of the repulsion loss greatly improves the detection accuracy of the occlusion faults, e.g., Figure 11(d). For example, in the

middle YOLOv7 detection map, there is a "fracture" defect that is partially obscured by the "plant," which leads to the omission of the YOLOv7 network, while the "fracture" fault type is successfully detected in the PA-YOLO network. "Meanwhile, the inclusion of repulsion loss is essential to improve the detection accuracy of dense small targets in the dataset. The introduction of convolutional blocks with EMA mechanism can make the network pay better attention to the global features and prevent the network from misdetecting the features outside the PV panels as a certain fault type, which can effectively reduce the misdetection rate of the network [24].

4. Conclusion

To address the challenge of PV panel fault detection, we reconfigure the YOLOv7 network to include an asymptotic feature pyramid network (AFPN) as the backbone for feature fusion. In addition, we propose a novel convolutional block with an attention mechanism, which can be seen from the data derived from the ablation experiments to greatly enhance the network's attention to global features. The introduction of repulsion loss (RL) plays a key role in improving the detection accuracy of occluded targets and accelerating network convergence. Experimental results show that our model achieves the highest mean average precision (mAP) under consistent experimental conditions, while the model size is significantly reduced. This reduction not only helps to improve detection performance but also facilitates deployment in embedded systems. In the future, we will aim to further compress the model size and continuously optimize it to ensure superior detection accuracy even in distributed PV panel application scenarios.

Data Availability

The data of this study is taken from the field photography of the cooperating enterprises and labeled manually by the researchers of the group. Due to the reason of data nondisclosure, the data of this study cannot be open-sourced, but it can be obtained from the corresponding author upon reasonable request.

Disclosure

This manuscript is an extension of a preprint that has been previously published (Yin Wang, Jingyong Zhao, Yan Yihua, Zhao Zhicheng, and Hu Xiao, 2023/6/27) on Research Square (https://assets.researc-hsquare.com/files/rs-3118568/v1_covered_13f2798f-cf97-449b-a2c5-3544932c1783.pdf?c=1693999504). The current submission includes substantial enhancements, novel findings, and additional analysis.

Conflicts of Interest

All authors of the study have no conflict of interest.

Acknowledgments

This study was supported by the Key R&D Project of Shanxi Province (Project No. 202102020101005), Scientific and Technological Achievement Transformation Guidance Project of Shanxi Province (grant no. 202204021301059), and Key Research and Development Plan of Shanxi Province (grant no. 202202150401007).

References

[1] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," *Energy Strategy Reviews*, vol. 24, pp. 38–50, 2019.

[2] K. Hansen, C. Breyer, and H. Lund, "Status and perspectives on 100% renewable energy systems," *Energy*, vol. 175, pp. 471–480, 2019.

[3] S. Wei, *Intelligent Detection of Photovoltaic Module Hot Spot Based on Infrared Image*, Zhejiang University, Hangzhou, 2020.

[4] D. Zhang, X. Hao, L. Liang, W. Liu, and C. Qin, "A novel deep convolutional neural network algorithm for surface defect detection," *Journal of Computational Design and Engineering*, vol. 9, no. 5, pp. 1616–1632, 2022.

[5] X. Lu, P. Lin, S. Cheng et al., "Fault diagnosis model for photovoltaic array using a dual-channels convolutional neural network with a feature selection structure," *Energy Conversion and Management*, vol. 248, article 114777, 2021.

[6] G. U. O. Menghao and X. U. Hongwei, *Research on Hot Spot Defect Detection of Infrared Thermal Images Based on Faster RCNN①*, Computer System and Application, 2019.

[7] D. P. Winston, M. S. Murugan, R. M. Elavarasan et al., "Solar PV's micro crack and hotspots detection technique using NN and SVM," *IEEE Access*, vol. 9, pp. 127259–127269, 2021.

[8] S. R. Madeti and S. N. Singh, "Modeling of PV system based on experimental data for fault detection using kNN method," *Solar Energy*, vol. 173, pp. 139–151, 2018.

[9] Z. Chen, H. Yu, L. Luo et al., "Rapid and accurate modeling of PV modules based on extreme learning machine and large datasets of I-V curves," *Applied Energy*, vol. 292, article 116929, 2021.

[10] M. Abbas and D. Zhang, "A smart fault detection approach for PV modules using adaptive neuro-fuzzy inference framework," *Energy Reports*, vol. 7, pp. 2962–2975, 2021.

[11] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7464–7475, Vancouver, BC, Canada, 2023.

[12] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944, Honolulu, HI, USA, 2017.

[13] G. Yang, J. Lei, and Z. Zhu, "AFPN: asymptotic feature pyramid network for object detection," 2023, http://arxiv.org/abs/2306.15988.

[14] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: faster and better learning for bounding box regression," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[15] X. Wang, T. Xiao, Y. Jiang, S. Shao, J. Sun, and C. Shen, "Repulsion Loss: Detecting Pedestrians in a Crowd," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7774–7783, Salt Lake City, UT, USA, 2018.

[16] D. Ouyang, S. He, G. Zhang et al., "Efficient Multi-Scale Attention Module with Cross-Spatial Learning," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, Rhodes Island, Greece, 2023.

[17] Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13708–13717, Nashville, TN, USA, 2021.

[18] W. Gong, H. Chen, Z. Zhang, M. Zhang, and H. Gao, "A data-driven-based fault diagnosis approach for electrical power DC-DC inverter by using modified convolutional neural network with global average pooling and 2-D feature image," *IEEE Access*, vol. 8, pp. 73677–73697, 2020.

[19] B. Gao and L. Pavel, "On the properties of the softmax function with application in game theory and reinforcement learning," 2017, http://arxiv.org/abs/1704.00805.

[20] Y. A. Mindzie, J. Kenfack, V. Joseph, U. Nzotcha, D. M. Djanssou, and R. Mbounguen, "Dynamic performance improvement using model reference adaptive control of photovoltaic systems under fast-changing atmospheric conditions," *International Journal of Photoenergy*, vol. 2023, Article ID 5703727, 20 pages, 2023.

[21] L. Zhao, L. Zhi, C. Zhao, and W. Zheng, "Fire-YOLO: a small target object detection method for fire inspection," *Sustainability*, vol. 14, no. 9, p. 4930, 2022.

[22] Z. Ge, S. Liu, and F. Wang, "Yolox: exceeding yolo series in 2021," 2021, http://arxiv.org/abs/2107.08430.

[23] J. Jerome Vasanth, S. Naveen Venkatesh, V. Sugumaran, and V. S. Mahamuni, "Enhancing photovoltaic module fault diagnosis with unmanned aerial vehicles and deep learning-based image analysis," *International Journal of Photoenergy*, vol. 2023, Article ID 8665729, 17 pages, 2023.

[24] F. M. A. Mazen, R. A. A. Seoud, and Y. Shaker, *Deep Learning for Automatic Defect Detection in PV modules using Electroluminescence Images*, IEEE Access, 2023.