IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 November 2023, accepted 12 December 2023, date of publication 15 December 2023,
date of current version 20 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343566

RESEARCH ARTICLE

PDDD-Net: Defect Detection Network Based on Parallel Attention Mechanism and Dual-Channel Spatial Pyramid Pooling

TINGTING SUI AND JUNWEN WANG
School of Electronic Information Engineering, Shanghai Dianji University, Shanghai 201306, China
Corresponding author: Tingting Sui (suitt@sdju.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 62103256.

ABSTRACT Owing to the small size of the defect pixel area and poor defect-background contrast issues in industrial images, noise and missed detection can easily occur. Therefore, automated defect detection is both necessary and challenging. To address these issues, with parallel attention mechanism (PAM) and dual-channel spatial pyramid pooling-fast block (DC_SPPF), a novel defect detection network, namely, PDDD-Net, is proposed in this paper. First, to make the detection network emphasize small defect areas better, the PAM block is proposed to be embedded into YOLOv5 to obtain more low-level visual features and improve the detection accuracy of microdefects. Meanwhile, by fusing multi-channel features, the DC_SPPF block is proposed to replace the raw spatial pyramid pooling-fast block to acquire richer discriminative features of the defect areas. Finally, The soft non-maximum suppression (Soft-NMS) module is used to undertake the feature candidate box filtering task in YOLOv5 to reduce missed detection. Two public datasets are adopted for the model evaluation: the Tianchi aluminum profile defects dataset (APDDD) and the power line insulator dataset (CPLID). The experimental results indicate that the proposed PDDD-Net network exhibits remarkable defect detection performance compared with other related detection methods.

INDEX TERMS Defect detection, parallel attention mechanism, spatial pyramid pooling-fast, deep learning, attention fusion.

I. INTRODUCTION

In intelligent industrial manufacturing scenarios [1], [2], [3], the product surface is inevitably affected by factors such as the processing technology, environmental temperature and manual operation errors, which lead to porosity and scratches. Therefore, defect detection is an important step in evaluating the quality of industrial products [3]. Through automatic defect detection, problematic products can be effectively prevented from entering the market and causing major accidents [4]. In the real world, defect detection tasks are mostly completed through manual visual inspection. However, such detection methods are time-consuming and labor-intensive [5]. In addition, manual visual inspection relies heavily on the expert's experience in defect detection, which can easily lead to noise and missed detection. Therefore, developing an automated defect detection network to provide objective and effective quality control and performance evaluation will contribute to the intelligent production of industrial products.

Currently, computer vision (CV)-based methods are frequently used to automatically detect defects in industrial images. Existing CV-based defect detection methods are mainly divided into two categories: 1) machine learning-based methods, and 2) deep learning-based methods. Machine learning-based methods often use artificial features such as edges, HOG [6], [7], SIFT [8], [9], etc. to achieve defect detection. However, artificial features mostly rely on subjective experience. Meanwhile, there is great similarity among the various defect features, which makes it difficult to find the best artificial features to describe different defects. Deep learning-based methods do not require human

The associate editor coordinating the review of this manuscript and approving it for publication was Marco Martalo.

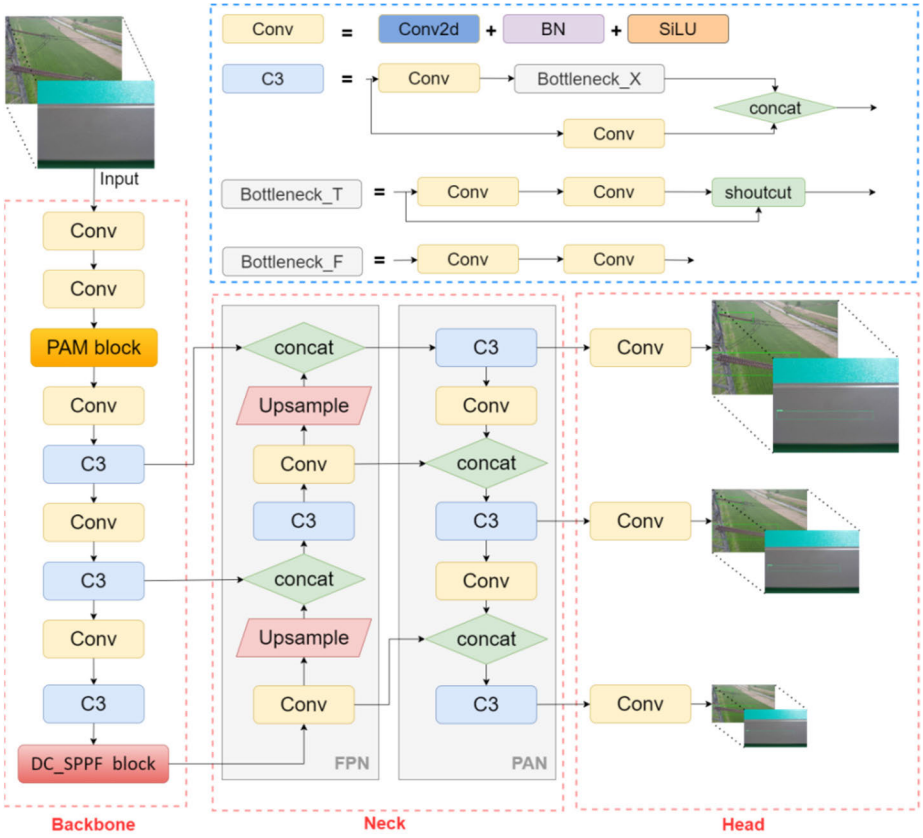

FIGURE 1. The framework of the proposed defect detection network.

intervention in the feature extraction process [10], [11] and can automatically extract discriminative features from defect areas through a convolutional neural network (CNN) to detect defects. Therefore, deep learning-based methods have achieved outstanding results in industrial defect detection tasks.

There are two main types of defect detection methods based on deep learning: one-stage framework-based method and two-stage framework-based method. The one-stage framework-based detection method adopts a shared network strategy to perform defect localization and identification tasks. The typical framework is the YOLO-family [12], [13], [14], such as YOLO, YOLOv5, YOLOv8, etc. The YOLO-family has a fast detection speed and can satisfy the basic detection accuracy requirements of industrial defect detection tasks. To achieve effective defect detection, the two-stage framework-based detection method consists of two sub-networks: candidate box extraction block, validation and classification block. The two-stage framework-based detection method can achieve good defect detection accuracy [15], [16], [17]. However, its detection efficiency is far from meeting the industrial requirements. In addition, a detection method with a tightly coupled two-stage framework can only achieve good detection performance when the features of the object area and background are significantly different. However, this restriction is difficult to achieve in actual industrial defect detection scenarios. Therefore, one-stage framework-based methods are more suitable for the detection of defects in industrial images.

One-stage framework-based methods have been applied for industrial defect detection. Among them, YOLOv5 [18], [19] is a widely used network with high detection efficiency and detection accuracy on datasets such as COCO and PASCAL Visual Object Class (VOC). Therefore, YOLOv5 is adopted as the baseline network for the defect detection in this study. However, there are two challenges in defect detection on the industrial product surface: 1) The defect-background contrast is poor. 2) The pixel area of the defect is small. Considering these two challenges, the features obtained via YOLOv5 are difficult to fully represent the location and semantic information of defects, which can easily lead to noise and missed detection. Therefore, it is important to design a multi-attention fusion mechanism to obtain more discriminative features and achieve sound detection performance by overcoming the aforementioned limitations of the existing detection methods.

To address these issues, a defect detection method based on a PAM block and a DC_SPPF block, namely, the PDDD-Net, is proposed, as illustrated in Fig. 1. On the public datasets APDDD and CPLID, the PDDD-Net network performs better for defect identification and defect location with small defect areas and poor defect-background contrast issues compared

with other advanced detection methods. The main contributions of this paper are as follows:

1) With a YOLOv5 structure, an end-to-end defect detection network is designed for accurate defect identification and location.
2) A pluggable PAM block is proposed and inserted into the feature extraction process of the proposed PDDD-Net to obtain more low-level features and enhance the detection performance against small defect areas issue.
3) To fully utilize the local feature maps, the DC_SPPF block is proposed to replace the original SPPF block to maximize the preservation of the discrimination of the extracted features to improve the detection accuracy.
4) Soft_NMS is set as the screening module for candidate detection boxes in the proposed PDDD-Net to output the object locations. In addition, the proposed PDDD-Net is applied to two public datasets APDDD and CPLID with small defect areas and poor defect-background contrast issues to demonstrate competitive performance.

The remainder of this paper is organized as follows. Section II briefly describes related work. Section III describes the proposed network in detail. Section IV discusses the experimental results and analysis. The conclusions are presented in Section V.

II. RELATED WORK

The related works are reviewed in two aspects in this section. Firstly, the development of machine learning-based defect detection methods is overviewed briefly, then the researches of deep learning-based defect detection methods are summarized.

A. MACHINE LEARNING-BASED DEFECT DETECTION METHODS

Machine learning-based defect detection methods are often used in the field of industrial automation quality control. Sha et al. [20] combined HOG and optimized support vector machines (SVM) to design an intelligent diagnosis method. Faced with curvilinear surface defects, Ma et al. [21] proposed a surface defect detection method based on improved Gabor filters and achieved good detection accuracy. Daghigh et al. [22] employed the k-nearest neighbor (k-NN) algorithm to provide a model for predicting the size, thickness, and location of penny-shaped defects in composite laminates. Aiming at the surface defects of industrial materials, Liu et al. [23] proposed a Haar-Weibull-variance model for steel surface defect detection in an unsupervised manner. Zhang et al. [24] proposed an on-line defect detection method for aluminum alloys in robotic arc welding based on the random forest and arc spectrum. To quickly satisfy defect recognition, Liu et al. [25] proposed a classification model based on OTSU and the random forest algorithm. Machine learning-based defect detection methods exhibit high computational efficiency and good detection accuracy to reduce the workload of inspectors. However, these methods can only achieve good defect detection performance when the background, shooting angle, and global threshold of the industrial images are consistent. This indicates that the environmental conditions significantly affect the effectiveness of defect detection. In addition, machine learning-based methods rely on special prior knowledge, which also affects the detection results. Therefore, machine learning-based defect detection methods cannot effectively satisfy the current defect detection requirements in the field of intelligent manufacturing.

B. DEEP LEARNING-BASED DEFECT DETECTION METHODS

Deep learning-based methods have achieved significant performance in the field of object detection [26], [27]. This can render defect feature extraction of industrial images more objective and effective. Typically, convolutional neural network (CNN) based architectures are widely adapted for feature extraction and defect localization. Therefore, a deep learning-based method is used in this paper to achieve feature extraction and defect detection in industrial defect images. Deep learning-based defect detection methods are used to achieve object localization and identification, which are mainly divided into two categories: two-stage framework-based methods, and one-stage framework-based methods.

The most famous two-stage framework is the R-CNN family [28], [29], [30]. The methods in the R-CNN family recommend thousands of candidate boxes for detection in the region proposal network, which is very time-consuming. Chen et al. improved the faster region-based convolutional neural network (Faster R-CNN) by embedding Gabor kernels, which can effectively alleviate the problem of unclear textures [31]. Guo et al. proposed a method based on an improved Mask R-CNN to detect defects on the surface of photovoltaic panels [32]. To monitor the occurrence of maize spodoptera frugipeda in a timely manner, an end-to-end detection model, namely, the Pest Region CNN (Pest R-CNN) was proposed, which yielded good insect detection performance [33]. The methods in the R-CNN family have high detection accuracy, but the detection speed cannot meet industrial requirements. Meanwhile, the images should have significant difference in the feature representations between the background and defect areas to guarantee the detection accuracy. If the feature representations of the previous two areas are relatively similar, then the detection effect is poor.

The typical detection method based on the one-stage framework is the YOLO family [34]. The methods in the YOLO family divide images into small grids for the regression and prediction of bounding boxes to obtain the object frames. Shi et al. used the Bayesian model to optimize YOLOv3 for facial recognition [35]. Mekhalfi et al. used YOLOv5 to conduct research on crack circle detection and counting [36]. These studies indicate that YOLOv5 exhibits

excellent detection potential for both within-domain and cross-domain scenarios. Therefore, we adopt YOLOv5 as the basic method for defect detection in industrial images However, the methods in the YOLO family may overlook some detailed features. To address this issue, researchers have attempted to add attention mechanisms to these methods.

The attention mechanism can extract different features and utilize contextual information, which helps to analyze complex scene information quickly and efficiently. Qi et al. [37] added the Squeeze and Excitation module (SE) [38] to YOLOv5 and proposed an improved SE-YOLOv5 network for the recognition of tomato virus diseases, effectively improving the recognition performance. Zhou et al. proposed an object detection method based on YOLOv5 and the convolutional block attention module (CBAM), which effectively improved the detection accuracy for construction waste [39]. The CBAM module [40], [41] can serialize attention feature map information from both the channel attention module (CAM) and the spatial attention module (SAM), and can be embedded into any backbone network to improve performance. Dong et al. proposed a parallel hybrid attention mechanism based on the YOLO network (PHAM-YOLO) for automatic defect detection [42]. PHAM-YOLO includes two parallel attention mechanism,which is CBAM module and coordinate attention module. However, the CBAM block fixes the processing order of the attention mechanism. In the serial attention mechanism, the former one filters out some features, which makes the feature maps that the latter focuses on already missing some defect information, causing CBAM to easily lose key features. Therefore, this study makes a major contribution to research on the defect detection network combined with the parallel attention mechanism of CAM and SAM.

In addition, YOLOv5 often uses datasets with low-resolution images (such as CIFAR-10) and generates good detection results. However, in industrial scenarios, defect detection is usually performed on high-resolution images, which leads to a poor detection performance of YOLOv5. One of the reasons is that the spatial pyramid pooling-fast module (SPPF) in YOLOv5 only use max-pooling branches to achieve an adaptive size output. Rather than improving the detection accuracy, astrous SPPF (ASPPF) [43] and simplified SPPF (SimSPPF) [44] have been proposed to improve the detection speed performance of SPPF. To improve the detection accuracy, we adopt a dual-channel approach to improve the SPPF model to preserve more image information while scaling the image size.

III. PROPOSED METHOD

A. NETWORK ARCHITECTURE

Inspired by the high-precision detection effect of YOLOv5, a novel defect detection method, PDDD-Net, has been proposed for automatic industrial defect detection. Fig. 1 shows the entire network architecture of the proposed defect detection method.

As shown in Fig. 1, PDDD-Net consists of three parts: a baseline network, a PAM block, and a DC_SPPF block. The baseline network consists of a combination of the backbone, neck, and head. Backbone is adopted to extract multi-level features from the input images through convolutional neural networks. Neck is a module that processes and fuses features extracted by the backbone. The main body of the head module consists of three detectors that perform defect detection based on feature maps of different scales output by the neck module. The PAM block is proposed to obtain more discriminative features and make PDDD-Net networks pay more attention to defect areas. The DC_SPPF block is proposed to fully integrate the features of the spatial pyramids.

B. BASELINE NETWORK

YOLOV5 is a typical deep defect detection network that achieves excellent detection results in various scenarios. Here, it is used as the baseline network of the proposed PDDD-Net defect detection network, consisting of the backbone, neck, and head, as shown in Fig. 1.

The backbone module consists of *C3* module and *Conv* module. A combination of slicing and gradient calculation is applied to feature maps to achieve deep-level feature extraction and reduce computational costs.

The neck module is a structure that combines the FPN block and the PAN block. First, the FPN block is used to transmit high-level semantic features from top to bottom through upsampling. Then, the PAN block is used to transfer low-level localization features from bottom to top for cross-stage hierarchical feature fusion to enhance the descriptive ability of the features.

In the head module, generalized intersection over union (GIoU) loss and non-maximum suppression (NMS) are combined to detect prediction boxes with different scales.

C. PAM BLOCK

Although YOLOv5 has achieved excellent detection results, it exhibits various limitations in scenarios with small defect areas and poor defect-background contrast issues. To overcome these issues and enable PDDD-Net to focus better on defect areas, the PAM block is proposed to extract more discriminative features from the defect areas.

In a deep detection network, the quality of the features extracted from the defect areas affects the detection accuracy for small defects. The convolutional block attention module (CBAM) has shown high values for classification and recognition tasks. However, CBAM sequentially generates attention feature maps from the perspectives of channel and space, which causes the later attention mechanism to focus on feature maps that have already lost some information. This serialized attention mechanism causes CBAM to miss the key features. Therefore, a parallel attention mechanism (PAM) that includes a spatial attention module (SAM) and a channel attention module (CAM) is proposed to capture discriminative features. Fig 2 shows the network architecture of the PAM block.

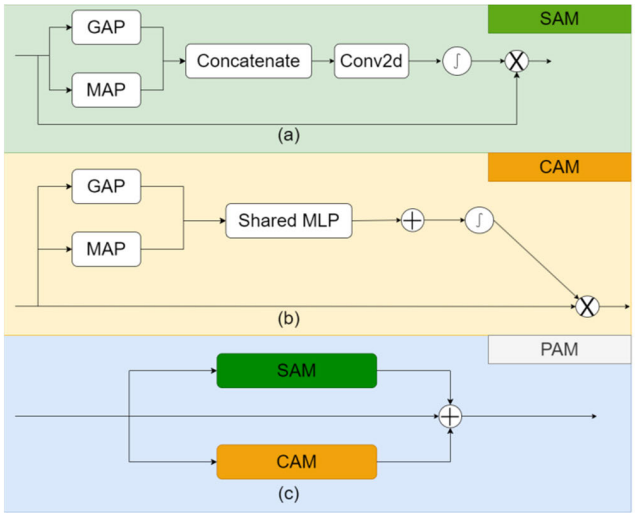

FIGURE 2. Network structure of the PAM block.

1) SPATIAL ATTENTION MODULE

To better utilize and fuse the cross spatial information of different local feature maps, as shown in Fig. 2(a), the feature maps are applied with the global average pooling operation (GAP) and maximum pooling operation (MAP) along the channel direction of each feature point. These feature maps are then stacked to generate an effective feature descriptor, which is sequentially convolved and connected using the standard convolutional layers and sigmoid activation function to yield a two-dimensional spatial attention weight. Finally, the weight matrix is fused with the raw feature maps to obtain spatial attention feature maps through matrix multiplication.

2) CHANNEL ATTENTION MODULE

As shown in Fig. 2(b), both GAP and MAP are performed on the input feature maps simultaneously. Then, the shared fully connected layer (Shared MLP) and sigmoid activation function are used to fuse these two pooling results to generate the channel attention weight. Finally, the raw feature maps are fused with the weight matrix through matrix multiplication to form channel attention feature maps.

3) RESIDUAL SHORTCUT LINK

Inspired by ResNet [45], [46], two attention channels are added to the raw feature maps through a residual shortcut link to maximize the feature parameter information.

4) PARALLEL MECHANISM

To eliminate the impact of the order in which the attention mechanisms are executed, the serialization operation order of CBAM is changed to the parallel mode, as shown in Fig. 2(c). To overcome the problem of inconsistent dimensions, the spatial attention feature maps and the channel attention feature maps are matrix multiplied with the raw feature maps through a residual shortcut link. Finally, the feature maps of the two parallel branches are fused to form the PAM feature maps.

D. DC_SPPF BLOCK

Because the object of interest may generate the maximum pixel value, MAP may effectively preserve the area of interest in the image. However, in spatial pyramid pooling-fast block, using only MAP channels results in singularity and loss of features. GAP can preserve the global features of an image to highlight background information. When encountering a poor defect-background contrast issue, the DC_SPPF block is proposed based on MAP and GAP to achieve the advantages of two pooling operations, as shown in Fig. 3.

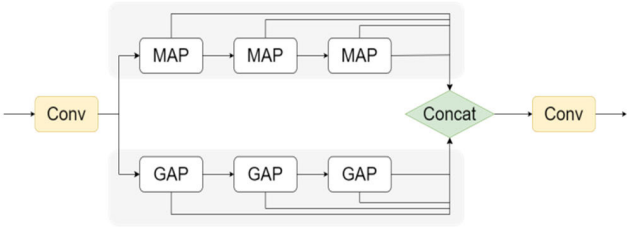

FIGURE 3. The network structure of the DC_SPPF block.

As shown in Fig. 3, the GAP branch is added to the traditional SPPF block. By parallelizing the branches of MAP and GAP in spatial pyramid pooling-fast block, discriminative features can be obtained to improve the detection accuracy of defects with poor defect-background contrast issue.

E. SOFT-NMS BLOCK

The redundancy removal operation of the detection box is an important part of defect detection networks, which directly affects the training effectiveness. NMS filters the prediction boxes based on a fixed threshold, as shown in (1). Using threshold judgment, object detection boxes with high confidence are reserved, whereas false detection boxes with low confidence are suppressed. When the spatial distance between the predicted detection box and the detection box in the optimal detection box set M is small, NMS filters out the prediction box with higher mutual coverage. However, in industrial scenarios, a situation in which multiple defect areas are crowded together is prone to occur, resulting in missed detection.

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_t \\ 0, & iou(M, b_i) \geq N_t \end{cases} \tag{1}$$

where $iou(\cdot)$ is used to calculate the intersection and union ratios of two detection boxes. S_i is the score calculated by using the classifier for each box. M denotes the set that stores the optimal detection boxes. b_i is the current detection box and N_t is the filtering threshold.

By contrast, Soft-NMS abandons the hard threshold mechanism of NMS. Soft-NMS multiplies the scores of different prediction boxes using a weight function, as shown in (2). This function attenuates the score of the detection box that overlaps the optimal detection box in M. The higher the overlap with the optimal detection box in M, the more severe is the attenuation of the current detection box score.

Therefore, when two defect areas are adjacent, the detection box for the defect is missed owing to the excessive overlapping area.

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_t \\ s_i G(iou(M, b_i)), & iou(M, b_i) \geq N_t \end{cases} \quad (2)$$

where the Gaussian function is used as the weight function $G(\cdot)$.

IV. EXPERIMENT RESULTS AND ANALYSIS

The proposed PDDD-Net detection network is designed with PyTorch 1.13 and CUDA10.1. Meanwhile, to accelerate the training and learning process of the deep learning models, relevant experiments were conducted under an NVIDIA GeForce 2080Ti card with 11GB memory. The detailed parameter settings of the proposed network are listed in Table 1. Through a detailed experimental analysis and comparison, the detection performance of the proposed PDDD-Net defect detection network has been evaluated.

First, details of the experimental dataset and evaluation indicators are provided. Second, the proposed PDDD-Net detection network is evaluated through an ablation study. Third, the effectiveness of the PAM block, DC_SPPF block and Soft_NMS is discussed. Meanwhile, the time consumption analysis of the method is also conducted. Finally, the effectiveness of the proposed defect detection method is compared with that of the current advanced methods.

TABLE 1. Parameter settings for the proposed PDDD-Net.

Index	Parameter	Value
1	Training rounds	200
2	Initial learning rate	0.01
3	Batch size	16
4	Optimizer	AdamW
5	Activation Function	Sigmoid

A. DATASETS

To evaluate the defect detection effectiveness and scalability of the proposed PDDD-Net defect detection method, two public datasets are used for various detection experiments: the aluminum profile surface detection database (APDDD) and the power line insulator dataset (CPLID). These are two typical defect datasets with small defect areas and poor defect-background contrast issues.

APDDD: APDDD is a public dataset collected in a real industrial environment, with 3005 defect images. This dataset has 10 types of defects: non-conductive, jet, scratch, mottled, orange peel, paint bubble, bottom leakage, dirty spots, corner leakage, and pit. The number of images in each category ranges from 240 to 370.

CPLID: CPLID is a public dataset that includes two categories of insulators: normal insulators and insulators with defects. The number of normal insulator images is 600. The number of images of defective insulator images is 248.

The sample images for the APDDD and CPLID are shown in Fig. 4 and Fig. 5, respectively. For model training and evaluation, they are divided into a training set and a testing set at a ratio of 60:40.

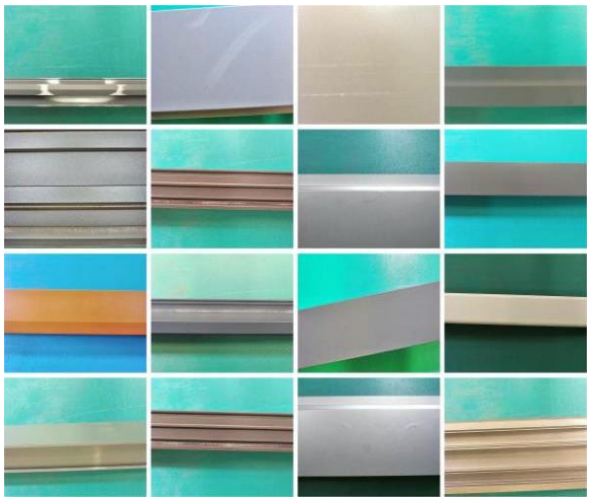

FIGURE 4. Sample images of APDDD.

FIGURE 5. Sample images of CPLID.

B. IMAGE PREPROCESSING

The training process for deep detection networks requires a large amount of data. However, the raw data of these two datasets could not satisfy the training requirements of the model. Therefore, image preprocessing is performed on the images within the dataset, including general data augmentation technology and Mosaic technology.

General data augmentation technology: This technology includes image flipping, image rotation, and adjustment of the contrast and brightness.

Mosaic: Four images are randomly selected and cropped into four sub-images. Subsequently, only one sub-image from each image is selected. Finally, these four sub-images are concatenated into a new defect image.

Image preprocessing technology is applied to APDDD and CPLID. Fig. 6 and Fig. 7 shows the sample images after image preprocessing.

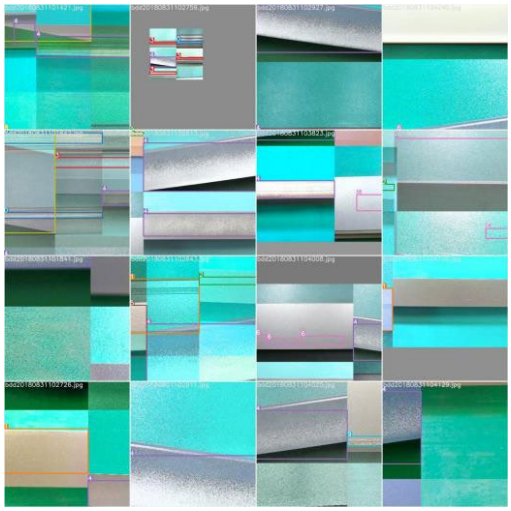

FIGURE 6. Sample images of APDDD after image preprocessing.

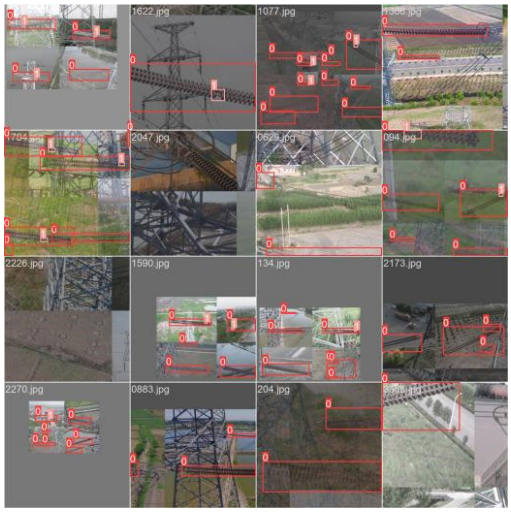

FIGURE 7. Sample images of CPLID after image preprocessing.

C. EVALUATION INDICATORS

To better validate the detection performance of the proposed PDDD-Net model, some evaluation indicators are introduced for model evaluation, namely precision (P), recall (R) and mean average precision (mAP). These evaluation indicators are defined as follows:

$$P = \frac{T_p}{T_p + F_p} \tag{3}$$

$$R = \frac{T_p}{T_p + F_n} \tag{4}$$

$$AP = \int_0^1 PRdR \tag{5}$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \times 100\% \tag{6}$$

where (T_p, F_p) denotes the numbers of true positives and false positives, respectively. (T_n, F_n) denotes the numbers of true negatives and false negatives, respectively. Additionally, n denotes the class number. It should be noted that mAP@0.5:0.95 is used in this study, which represents the average area of all P-R curves when their $iou(\cdot)$ results are between 0.5 and 0.95.

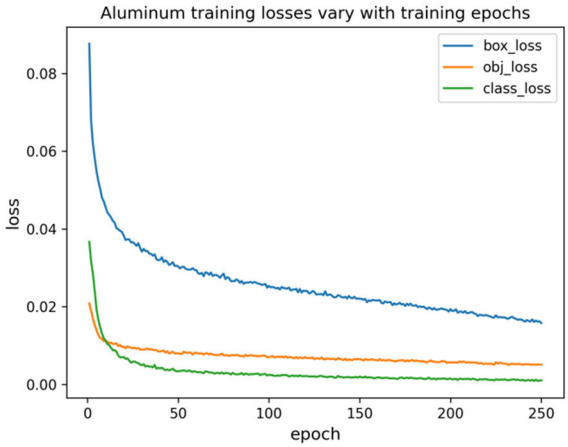

FIGURE 8. The training loss curve of PDDD-Net detection network on APDDD.

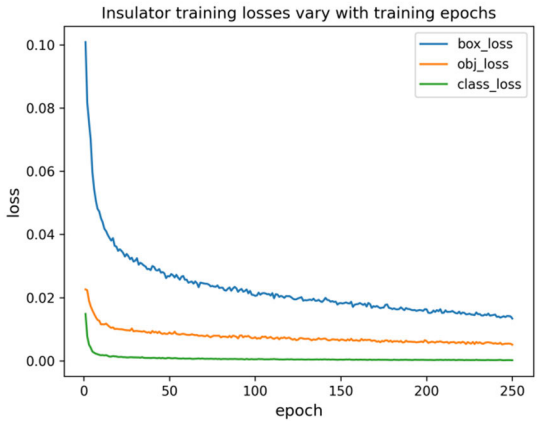

FIGURE 9. The training loss curve of PDDD-Net detection network on CPLID.

D. GENERALIZATION EXPERIMENT

To some extent the generalization ability of the defect detection networks is related to the epoch. We simply study effects of different training epochs of the PDDD-Net detection network on the training loss. Fig. 8 and Fig. 9 show that the loss value of the defect detection model tends to stabilize for APDDD and CPLID datasets when the epoch is 200. Moreover, the difference between the loss value when the epoch is 250 and the loss value when the epoch is 200 is very small, indicating that their training effects are almost the same. However, the cost of training duration and parameter capacity has increased significantly with the increase of epoch. Therefore, setting the epoch to 200 is appropriate.

E. EFFECTS OF THE PAM BLOCK

To better demonstrate the feature extraction performance of the proposed PAM block, some advanced attention mechanisms, including Squeeze and Extraction (SE) and CBAM, are inserted into YOLOv5 for comparison.

From Table 2 and Table 3, the PAM block exhibits the highest *mAP*, which indicates that the PAM block has a better detection ability for defects with only a small proportion of pixels. For defect detection in the industrial field, indicator *R* is more important than indicator *P*. In addition, from Fig. 10 and Fig. 11, we can observe that the *R* value of the PAM block has certain advantages, indicating that the PAM block has better recall and can more comprehensively detect defect areas.

TABLE 2. The evaluation indicators of different attention blocks on APDDD.

Indicator	Baseline	Baseline+SE	Baseline+CBAM	Baseline+PAM (Proposed)
mAP	55.7 %	55.7 %	55.8 %	57.1 %
FPS	103	100	94	91

TABLE 3. The evaluation indicators of different attention blocks on CPLID.

Indicator	Baseline	Baseline+SE	Baseline+CBAM	Baseline+PAM (Proposed)
mAP	80.3 %	80.7 %	80.0 %	81.2 %
FPS	95	89	82	88

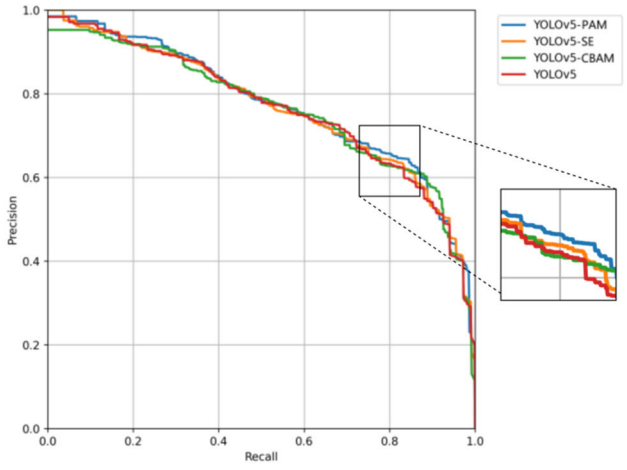

FIGURE 10. The P-R curve of different attention mechanisms on APDDD.

F. ABLATION STUDY

To better evaluate the detection performance of the proposed PDDD-Net detection network, an ablation study is conducted for the effectiveness analysis in the paper. The PAM block is adopted to fully utilize the advantages of the two parallel attention mechanisms to better address the issue of small defect pixel area. DC_SPPF is adopted to fully utilize the advantages of the two pooling operations to address the poor defect-background contrast issue. Soft-NMS is used to address the issue of missed detection. The effects of Soft-NMS, PAM block, and DC_SPPF block on APDDD and CPLID is evaluated through an ablation study. Table 4 and Table 5 list the defect detection results for fair experimental comparison.

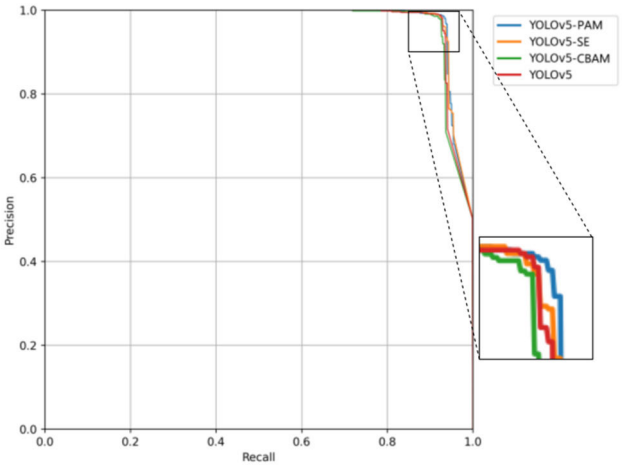

FIGURE 11. The P-R curve of PDDD-Net detection network on CPLID.

TABLE 4. Ablation study on APDDD.

Methods	*mAP*	*FPS*
Baseline	55.7 %	103
Baseline+Soft-NMS	60.3 % (+4.6 %)	92
Baseline+DC_SPPF	57.6 % (+1.9 %)	98
Baseline+PAM	57.1 % (+1.4 %)	91
Baseline+Soft-NM+PAM	60.7 % (+5.0 %)	90
Baseline+Soft-NM+PAM+DC_SPPF (Proposed PDDD-Net)	61.5 % (+5.8 %)	90

TABLE 5. Ablation study on CPLID.

Methods	*mAP*	*FPS*
Baseline	80.3 %	95
Baseline+Soft-NMS	82.4 % (+2.1 %)	77
Baseline+DC_SPPF	81.7 % (+1.4 %)	90
Baseline+PAM	81.2 % (+0.9 %)	88
Baseline+Soft-NM+PAM	82.6 % (+2.3 %)	77
Baseline+Soft-NM+PAM+DC_SPPF (Proposed PDDD-Net)	83.3 % (+3.0 %)	76

The experimental results show that the proposed PDDD-Net with different settings sacrifices only a small amount of time efficiency and achieves the highest *mAP*. Combined with Soft-NMS, PAM block or DC_SPPF block, the proposed PDDD-Net detection network acquire higher *mAP* values, indicating that integrating these blocks into the PDDD-Net can improve detection capabilities.

Meanwhile, combined with the Soft-NMS, PAM block and DC_SPPF block, the proposed PDDD-Net detection network generates the highest *mAP* on APDDD and CPLID, indicating the effectiveness of the proposed detection network in small defect areas and poor defect-background contrast issues.

Moreover, defect detection experiments are conducted using different improved models, as shown in Fig. 12. It can be seen that YOLOv5 and some improved models (line 4 - line 7 in Fig. 12) yield noise and missed detection. In contrast, the proposed PDDD-Net detection network combined with Soft-NMS, PAM block and DC_SPPF block achieves the best detection effect.

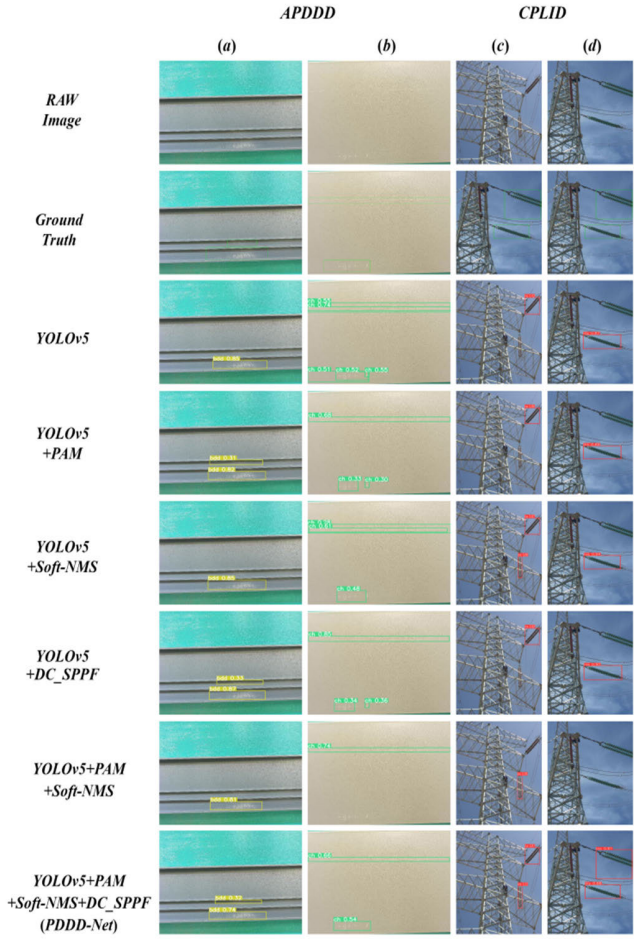

FIGURE 12. The defect detection results generated by different models.

G. PERFORMANCE VISUALIZATION AND COMPARISON

To better demonstrate the defect detection performance of the proposed PDDD-Net detection network, various advanced defect detection models have been used as comparative methods, such as Faster R-CNN, YOLOv3, YOLOv5, YOLOv7, and YOLOv8. To fairly evaluate the detection performance, the same runtime environment and experimental data are applied to the model evaluation. Meanwhile, samples from the test sets in APDDD and CPLID are used as the experimental data for the performance analysis.

TABLE 6. Comparison experiments on different detection models on APDDD and CPLID.

Set	Methods	*mAP*	*FPS*
APDDD	Faster R-CNN	49.2 %	7
APDDD	YOLOv3	51.5 %	90
APDDD	YOLOv5	55.7 %	103
APDDD	YOLOv7	47.9 %	161
APDDD	YOLOv8	54.1 %	87
APDDD	OURS (PDDD-Net)	61.5 %	90
CPLID	Faster R-CNN	76.4 %	13
CPLID	YOLOv3	79.2 %	74
CPLID	YOLOv5	80.3 %	95
CPLID	YOLOv7	79.6 %	131
CPLID	YOLOv8	81.9 %	128
CPLID	OURS (PDDD-Net)	83.3 %	76

Table 6 and Fig. 13 show the defect detection results of different detection methods on APDDD and CPLID. From Fig. 13, we can observe that the images in APDDD exhibit small defect areas and poor defect-background contrast issues. Fig. 13(a) - (c) shows that other advanced models are prone to exhibit noise detection and missed detection. Compared to other methods, the proposed PDDD-Net can more accurately locate defects in aluminum profile surface images, even if the defect area is small. The images in CPLID have a complex-background issue, whereas PDDD-Net can accurately locate the position of insulators or defects, as shown in Fig. 13(d) - (f).

In addition, Table 6 shows that compared to other advanced detection models, the proposed PDDD-Net detection network has higher *mAP* values on APDDD and CPLID, indicating that the proposed method is feasible for object detection tasks with small defect areas and poor defect-background contrast issues.

H. DISCUSSION OF THE PAM BLOCK

To better combine the local and spatial information of the image, parallel operations on the SAM and CAM branches are adopted in the PAM block. Table 2 and Table 3 indicates that the PAM block is helpful for defect detection, mainly because the attention mechanism can fully utilize the advantages of SAM and CAM to focus on more key features related to the defect areas.

To illustrate the feature extraction effect of the PAM block, heatmaps are used to exhibit the regions of interest of different attention mechanisms. Fig. 14 reveals that if the image has a complex background or poor defect-background contrast issue, the attention area of the SE block for the image is

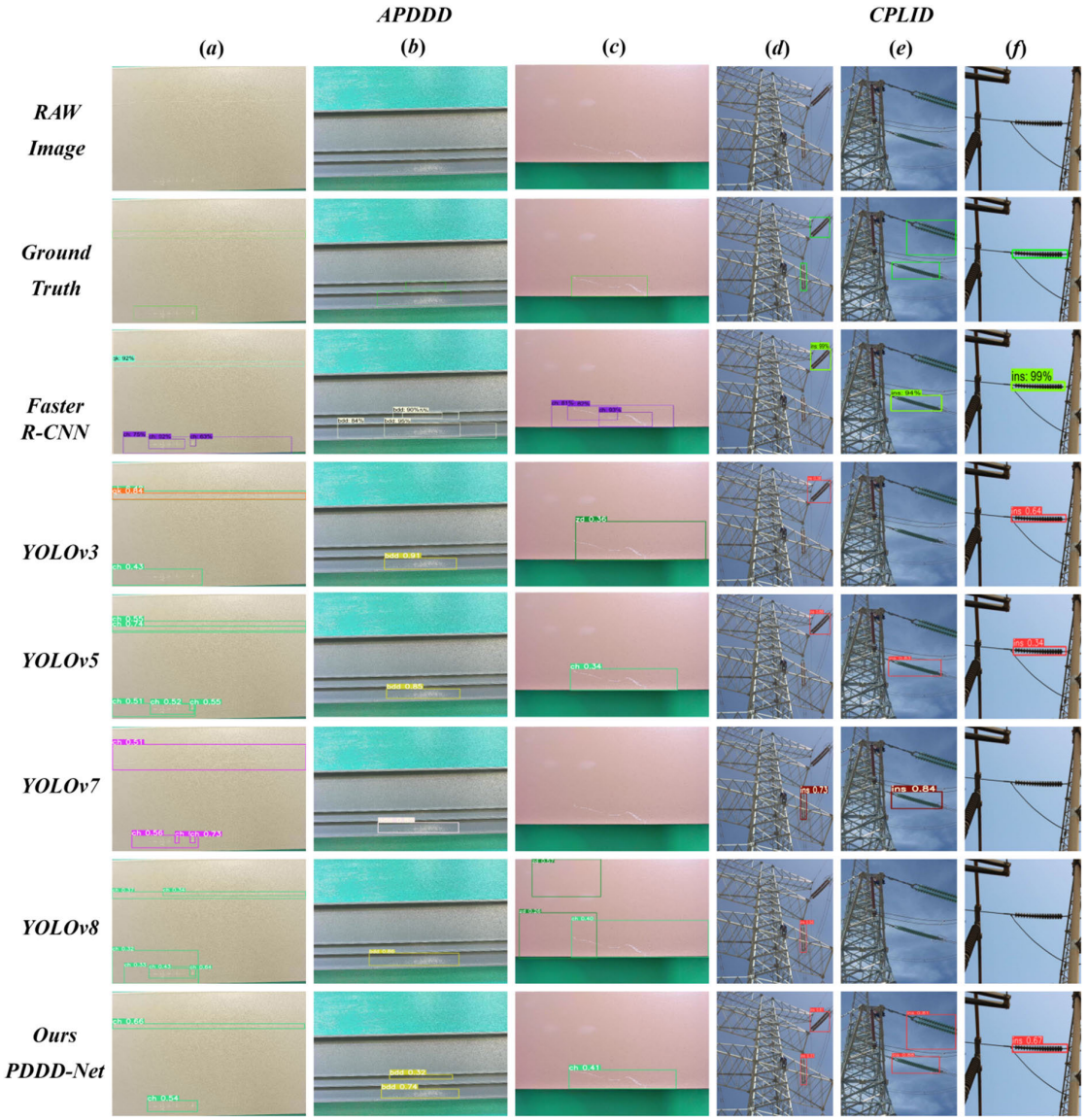

FIGURE 13. The defect detection performance on APDDD and CPLID.

easily affected and becomes more divergent. CBAM block is susceptible to complex background or small defect area issue, leading to missed detection. In contrast, the PAM block proposed in this paper can effectively focus on the defect areas. For images with small defect areas and poor defect-background contrast issues in the APDDD set, the PAM block can focus on the defect areas of the aluminum profile surface. For images with complex backgrounds in the CPLID set, the PAM block can filter out interference areas and noise, and focus more on insulators and defect areas.

I. TIME CONSUMPTION ANALYSIS

Computational efficiency is also a key indicator for defect detection. The processing time of the proposed PDDD-Net detection network is analyzed in this paper. An ablation study is conducted based on APDDD and CPLID datasets, and the processing time of each module are listed in Table 4 and Table 5, respectively. According to Table 4 and Table 5, we can observe that the processing time of PDDD-Net on the aluminum profile surface defect image is 90 *FPS* and the processing time of PDDD-Net on the power line insulator image is 76 *FPS*, which are slightly longer than those of the baseline network.

Meanwhile, with the proposed PAM block and DC_SPPF block, the defect detection accuracy is significantly improved without sacrificing excessive computational efficiency. Table 6 indicates that although some advanced methods have significant advantages in terms of processing efficiency, they sacrifice significant detection accuracy in exchange for detection efficiency, resulting in poor detection accuracy. However, the detection efficiency of the proposed PDDD-Net network not only meets the basic processing requirements of

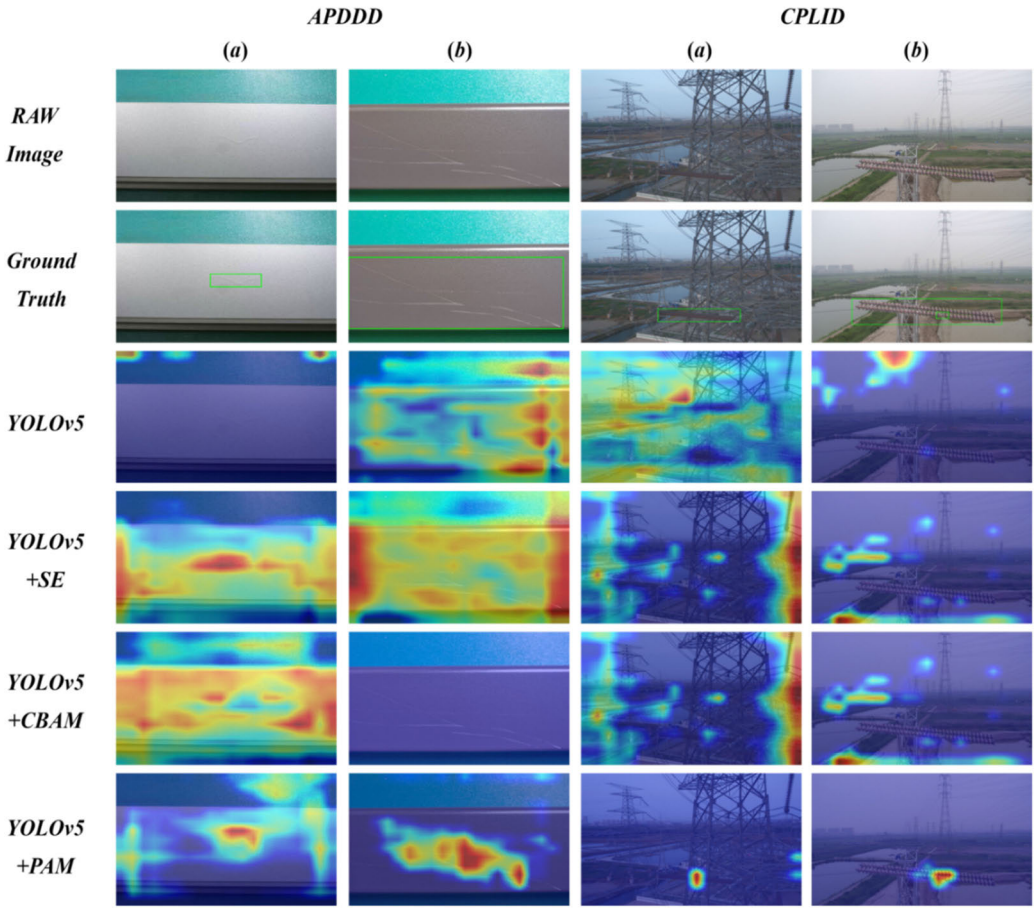

FIGURE 14. The heatmaps of the regions of interest obtained via different attention mechanisms.

industrial detection tasks, but also achieves the best detection accuracy compared to other advanced methods.

V. CONCLUSION

To facilitate the quality control of industrial products and the development of defect repair plans, a new detection method called PDDD-Net is proposed in this paper to automatically perform defect detection task. On two public datasets, APDDD and CPLID, the proposed PDDD-Net shows excellent defect detection performance against small defect areas and poor defect-background contrast issues. The main contributions of this paper are as follows.

1) With YOLOv5 network structure and Soft_NMS, an end-to-end defect detection network is designed to automatically and accurately detect defects.

2) To enable the detection network to focus better on the pixel-level features of small defects, a pluggable PAM block is proposed to obtain more discriminative features from local feature maps to improve the detection accuracy of small defects.

3) To fully utilize the local feature maps, the DC_SPPF block containing two pooling operations is proposed for incorporation into PDDD-Net to obtain richer features and improve the detection accuracy.

In the future, we will proceed this research to build a high-precision and efficient defect detection network.

REFERENCES

[1] Z.-W. Li, X.-Z. Liu, F. Yang, and L.-W. Zhang, "Mud pumping defect detection of high-speed rail slab track based on track geometry data," *J. Transp. Eng. A, Syst.*, vol. 148, no. 6, Jun. 2022, Art. no. 04022023.

[2] Y. Wu, L. Li, B. Wang, Z. Zhu, T. Gao, Q. Xie, Q. Chen, and Q. Xiao, "Rapid detection of defect structures in graphene by the machine learning," *Modern Phys. Lett. B*, vol. 36, no. 15, May 2022, Art. no. 2250081.

[3] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *Int. J. Adv. Manuf. Technol.*, vol. 94, nos. 9–12, pp. 3465–3471, Feb. 2018.

[4] I. Kim, Y. Jeon, J. W. Kang, and J. Gwak, "RAG-PaDiM: Residual attention guided PaDiM for defects segmentation in railway tracks," *J. Electr. Eng. Technol.*, vol. 18, no. 2, pp. 1429–1438, Dec. 2023.

[5] Y.-F. Hsieh, J.-H. Ye, N.-J. Wu, and Q.-C. Hsu, "Intelligent automatic deburring system by integrating palletizing robot with image and vibration sensors," *Sensors Mater.*, vol. 33, no. 3, pp. 933–993, Mar. 2021.

[6] X. Hong, L. Huang, S. Gong, and G. Xiao, "Shedding damage detection of metal underwater pipeline external anticorrosive coating by ultrasonic imaging based on HOG + SVM," *J. Mar. Sci. Eng.*, vol. 9, no. 4, p. 364, Mar. 2021.

[7] X. Wang and Z. Zhang, "On-line defect detection and classification of latex gloves," *J. Phys., Conf.*, vol. 1575, no. 1, Jun. 2020, Art. no. 012103.

[8] C. Dunderdale, W. Brettenny, C. Clohessy, and E. E. van Dyk, "Photovoltaic defect classification through thermal infrared imaging using a machine learning approach," *Prog. Photovoltaics, Res. Appl.*, vol. 28, no. 3, pp. 177–188, Mar. 2020.

[9] A. Hikmat, K. Afdel, and I. Bakkouri, "Automatic detection of stellate lesions in digital mammograms using multi-scale SIFT," *J. Pharmacy Pharmacol.*, vol. 8, no. 1, pp. 24–34, Jan. 2020.

[10] H. Zhang, L. Jiang, and C. Li, "CS-ResNet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection," *Exp. Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115673.

[11] W. Liqun, W. Jiansheng, and W. Dingjin, "Research on vehicle parts defect detection based on deep learning," *J. Phys., Conf.*, vol. 1437, no. 1, Jan. 2020, Art. no. 012004.

[12] J. Liu, X. Zhu, X. Zhou, S. Qian, and J. Yu, "Research on vehicle parts defect detection based on deep learning," *Electronics*, vol. 11, no. 10, p. 1561, 2022

[13] Y. Xie, W. Hu, S. Xie, and L. He, "Surface defect detection algorithm based on feature-enhanced YOLO," *Cognit. Comput.*, vol. 15, no. 2, pp. 565–579, Mar. 2023.

[14] K. Patel, V. Patel, V. Prajapati, D. Chauhan, A. Haji, and S. Degadwala, "Safety helmet detection using YOLO v8," in *Proc. 3rd Int. Conf. Pervasive Comput. Social Netw. (ICPCSN)*, Jun. 2023, pp. 22–26.

[15] B. Wei, K. Hao, L. Gao, and X.-S. Tang, "Detecting textile micro-defects: A novel and efficient method based on visual gain mechanism," *Inf. Sci.*, vol. 541, pp. 60–74, Dec. 2020.

[16] Y. Liu, Y. Liu, P. Zhang, Q. Zhang, L. Wang, R. Yan, W. Li, and Z. Gui, "Spark plug defects detection based on improved faster-RCNN algorithm," *J. X-Ray Sci. Technol.*, vol. 30, no. 4, pp. 709–724, Aug. 2022.

[17] M. Chen, F. Bai, and Z. Gerile, "Special object detection based on mask RCNN," in *Proc. 17th Int. Conf. Comput. Intell. Secur. (CIS)*, Nov. 2021, pp. 128–132.

[18] Y. Wang, Q. Li, Y. Liu, and C. Wang, "Insulator defect detection based on improved YOLOv5 algorithm," in *Proc. IEEE 12th Data Driven Control Learn. Syst. Conf. (DDCLS)*, May 2023, pp. 770–775.

[19] X. Wang and G. Cheng, "Research on surface defect detection of disc harrow based on YOLOv5," in *Proc. IEEE 3rd Int. Conf. Inf. Technol., Big Data Artif. Intell. (ICIBA)*, vol. 3, May 2023, pp. 802–808.

[20] Y. Sha, Z. He, J. Du, Z. Zhu, and X. Lu, "Intelligent detection technology of flip chip based on H-SVM algorithm," *Eng. Failure Anal.*, vol. 134, Apr. 2022, Art. no. 106032.

[21] J. Ma, Y. Wang, C. Shi, and C. Lu, "Fast surface defect detection using improved Gabor filters," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1508–1512.

[22] V. Daghigh and M. Naraghi, "Machine learning-based defect characterization in anisotropic materials with IR-thermography synthetic data," *Compos. Sci. Technol.*, vol. 233, Mar. 2023, Art. no. 109882.

[23] K. Liu, H. Wang, H. Chen, E. Qu, Y. Tian, and H. Sun, "Steel surface defect detection using a new Haar–Weibull-variance model in unsupervised manner," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 10, pp. 2585–2596, Oct. 2017.

[24] Z. Zhang, Z. Yang, W. Ren, and G. Wen, "Random forest-based real-time defect detection of Al alloy in robotic arc welding using optical spectrum," *J. Manuf. Processes*, vol. 42, pp. 51–59, Jun. 2019.

[25] C. Z. Liu, X. Wang, L. X. Chen, H. Guo, R. Luo, and Y. C. Zhou, "Surface defect recognition of fibreboard based on random forest," *Scientia Silvae Sinicae*, vol. 56, no. 11, pp. 121–126, 2018.

[26] G. Vashishtha and R. Kumar, "Unsupervised learning model of sparse filtering enhanced using Wasserstein distance for intelligent fault diagnosis," *J. Vibrat. Eng. Technol.*, vol. 11, no. 7, pp. 2985–3002, Oct. 2023.

[27] A. Kumar, G. Vashishtha, C. P. Gandhi, H. Tang, and J. Xiang, "Tacho-less sparse CNN to detect defects in rotor-bearing systems at varying speed," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104401.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[29] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[30] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[31] M. Chen, L. Yu, C. Zhi, R. Sun, S. Zhu, Z. Gao, Z. Ke, M. Zhu, and Y. Zhang, "Improved faster R-CNN for fabric defect detection based on Gabor filter with genetic algorithm optimization," *Comput. Ind.*, vol. 134, Jan. 2022, Art. no. 103551.

[32] S. Guo, Z. Wang, Y. Lou, and H. Lin, "Detection method of photovoltaic panel defect based on improved mask R-CNN," *J. Internet Technol.*, vol. 23, no. 2, pp. 397–406, Mar. 2022.

[33] L. Du, Y. Sun, S. Chen, J. Feng, Y. Zhao, Z. Yan, X. Zhang, and Y. Bian, "A novel object detection model based on faster R-CNN for spodoptera frugiperda according to feeding trace of corn leaves," *Agriculture*, vol. 12, no. 2, p. 248, Feb. 2022.

[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[35] D. Shi and H. Tang, "A new multiface target detection algorithm for students in class based on Bayesian optimized YOLOv3 model," *J. Electr. Comput. Eng.*, vol. 2022, pp. 1–12, Jan. 2022.

[36] M. L. Mekhalfi, C. Nicolo, Y. Bazi, M. M. A. Rahhal, N. A. Alsharif, and E. A. Maghayreh, "Contrasting YOLOv5, transformer, and EfficientDet detectors for crop circle detection in desert," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[37] J. Qi, X. Liu, K. Liu, F. Xu, H. Guo, X. Tian, M. Li, Z. Bao, and Y. Li, "An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease," *Comput. Electron. Agricult.*, vol. 194, Mar. 2022, Art. no. 106780.

[38] S. Mekruksavanich, P. Jantawong, and A. Jitpattanakul, "A lightweight deep convolutional neural network with squeeze-and-excitation modules for efficient human activity recognition using smartphone sensors," in *Proc. 2nd Int. Conf. Big Data Analytics Practices (IBDAP)*, Aug. 2021, pp. 23–27.

[39] Q. Zhou, H. Liu, Y. Qiu, and W. Zheng, "Object detection for construction waste based on an improved YOLOv5 model," *Sustainability*, vol. 15, no. 1, p. 681, Dec. 2022.

[40] R. Fan and Z. Qiu, "Improved YOLOv5 algorithm based on CBAM attention mechanism," in *Proc. Int. Conf. Frontiers Artif. Intell. Mach. Learn. (FAIML)*, Jun. 2022, pp. 229–233.

[41] C. Yang, C. Zhang, X. Yang, and Y. Li, "Performance study of CBAM attention mechanism in convolutional neural networks at different depths," in *Proc. IEEE 18th Conf. Ind. Electron. Appl. (ICIEA)*, Aug. 2023, pp. 1373–1377.

[42] H. Dong, M. Yuan, S. Wang, L. Zhang, W. Bao, Y. Liu, and Q. Hu, "PHAM-YOLO: A parallel hybrid attention mechanism network for defect detection of meter in substation," *Sensors*, vol. 23, no. 13, p. 6052, Jun. 2023.

[43] C. Zheng, "Stack-YOLO: A friendly-hardware real-time object detection algorithm," *IEEE Access*, vol. 11, pp. 62522–62534, 2023.

[44] J. Wu and L. Wang, "Face mask detection based on improved YOLOv3 algorithm," in *Proc. 3rd Int. Conf. Neural Netw., Inf. Commun. Eng. (NNICE)*, Feb. 2023, pp. 201–207.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[46] L. Zhan, W. Li, and W. Min, "FA-ResNet: Feature affine residual network for large-scale point cloud segmentation," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, Apr. 2023, Art. no. 103259.

TINGTING SUI received the B.S. degree in computer science and technology, the M.S. degree in computer application technology, and the Ph.D. degree in information management and information technology from Shanghai Maritime University, Shanghai, China, in 2011, 2013, and 2016, respectively. She is currently a Lecturer with Shanghai Dianji University. Her research interests include pattern recognition, computer vision, and large-scale data analysis.

JUNWEN WANG received the bachelor's degree in vehicle engineering from the Nanhang Jincheng College, in 2019. He is currently pursuing the master's degree in electrical engineering with the School of Electronic Information Engineering, Shanghai Dianji University.

His research interests include object detection, machine vision, and computer-based research on deep learning.