Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# CDA-YOLO: An Improved YOLOv11 Model for PCB Defect Detection

**Anlong Yuan**
Sichuan University
**Jingmin Li**
jingminli@scu.edu.cn

Sichuan University

**Article**

**Keywords:** PCB defect detection, YOLOv11, Small object detection, Dual-scale convolution, Multi-scale feature fusion, Parameter-free average attention

**Posted Date:** August 28th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10685349/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

**Version of Record:** A version of this preprint was published at Scientific Reports on October 19th, 2026.
See the published version at https://doi.org/10.1038/s41598-026-74471-x.

# CDA-YOLO: An Improved YOLOv11 Model for PCB Defect Detection

Anlong Yuan & Jingmin Li*✉

Printed circuit boards (PCBs) are fundamental to modern electronic systems, and their manufacturing quality directly determines the performance and reliability of the final product. However, existing YOLO-based defect detection models still face critical challenges, including high miss rates for tiny defects and frequent false positives caused by complex background textures. To address these issues, we propose CDA-YOLO, an enhanced YOLOv11 framework incorporating three key innovations. First, a Cross-level Bidirectional Feature Pyramid Network (CBiFPN) is designed, which effectively alleviates the loss of fine-grained details of small targets as features propagate through deep layers. Second, a lightweight Dual-Scale Convolution module (DualScaleConv) is proposed, which employs a parallel branch structure to simultaneously capture local details and global contextual information. Finally, a parameter-free average attention mechanism (PfAAM) is integrated into the C3k2 module to enhance the feature response in defect-related regions and suppresses background noise. Experimental results demonstrate that the proposed model achieves competitive performance on the PCB defect dataset, with mAP@0.5 reaching 96.1% and mAP@0.5:0.95 reaching 56.0%, corresponding to improvements of 2.8% and 3.4% over the baseline YOLOv11, respectively. These results validate that CDA-YOLO offers an efficient and robust solution for industrial-grade PCB defect detection.

**Keywords** PCB defect detection, YOLOv11, Small object detection, Dual-scale convolution, Multi-scale feature fusion, Parameter-free average attention

School of Mechanical Engineering, Sichuan University, Chengdu 610065, China.
Corresponding author: Jingmin Li. ✉email: jingminli@scu.edu.cn

## Introduction

As a core component of electronic products, the printed circuit board (PCB) directly determines circuit performance and product reliability[1-3]. With the trend of electronic products toward higher density, miniaturization, and lighter weight, the complexity of PCB circuits continues to increase, leading to a higher probability of various tiny defects during manufacturing. Common PCB defects include open circuits, short circuits, missing holes, mouse bites, spurs, and spurious copper. If these defects are not accurately detected at an early stage, they may cause system failures, resulting in significant economic losses and safety hazards. Therefore, developing high-precision and high-efficiency automatic PCB defect detection methods is of great engineering value and practical significance[4].

Traditional PCB defect detection mainly relies on manual visual inspection or traditional machine vision methods. Manual inspection suffers from low efficiency, high susceptibility to fatigue, and strong subjectivity, making it difficult to meet the real-time and consistency demands of modern large-scale production[5]. Traditional machine vision methods are limited by the representational power of hand-crafted features and do not generalize well when facing complex backgrounds and diverse defect morphologies[6]. Recently, breakthroughs in deep learning have provided new solutions to these challenges. Object detection algorithms based on convolutional neural networks (CNNs), with their powerful automatic feature learning capabilities, have demonstrated great potential in the field of industrial defect detection[7].

Among the numerous CNN-based object detection algorithms, the YOLO (You Only Look Once) series has become one of the most popular foundational frameworks in industrial vision inspection due to its excellent balance between speed and accuracy. Since YOLOv1 pioneered the unification of object detection into a regression problem, the series of algorithms has continued to iterate[8]. YOLOv3 incorporates the idea of a Feature Pyramid Network (FPN) to make detections at three different scales[9].

YOLOv5 offered an easy-to-use lightweight version[10], YOLOv8 adopted C2f, anchor-free, and decoupled-head designs[11]. YOLOv11 further improved multi-scale fusion with C3k2 and C2PSA modules, incorporated depth-wise separable detection heads to reduce redundancy, achieving higher accuracy than YOLOv8 with fewer parameters and stronger generalization, thus providing a stronger foundation for tiny PCB defect detection[12]. The network architecture of YOLOv11 is shown in Fig. 1.

In recent times, YOLO algorithms aimed at improving PCB defect detection have yielded some results. Zhang et al.[13] proposed YOLO-VCL, which incorporates the C3K2-VSSD module in the backbone to capture local features and global context, achieving an mAP@0.5 of 98.4% and mAP@0.5:0.95 of 70.3%. However, these results were obtained on a relatively high-performance computing platform. He and Xie[14] proposed ABF-YOLO, which incorporates an axial attention module to capture horizontal and vertical long-range dependencies and a bidirectional feature fusion (BiFusion) module for adaptive multi-scale feature aggregation, attaining an mAP@0.5 of 97.7%. Elsharkawy[15] introduced Efficient Multi-Scale Attention (EMA), Content-Aware ReAssembly of Features (CARAFE), and Efficient IoU (EIoU) loss into YOLOv11, substantially improving detection accuracy and real-time performance for minute defects.

However, directly applying generic YOLO models to PCB defect detection still faces several challenges. First, PCB defects are extremely small in size; some defects occupy only a few pixels, making this a typical small-object detection problem that demands high spatial detail preservation capability from the model. Second, the background of PCB images is highly complex. Dense traces, pads, and textural information such as silk-screen characters severely interfere with defect feature extraction, leading to false positives and missed detections. Although existing YOLO models perform excellently on generic object detection tasks, their feature extraction and multi-scale fusion capabilities are still insufficient when targeting high-density, small-object scenarios with complex backgrounds, such as PCB inspection.

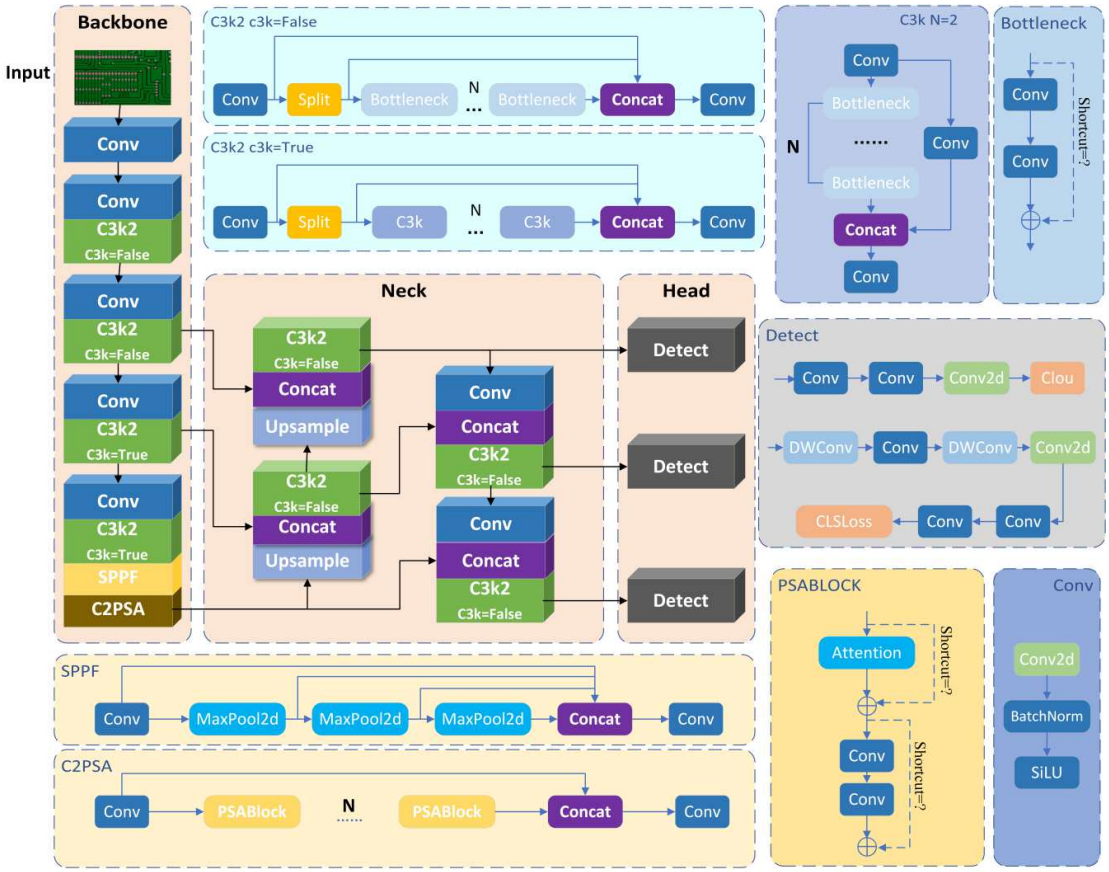



**Fig. 1**. YOLOv11 network structure.

To address the issues above, we propose an improved YOLOv11 model for PCB defect detection. The main contributions are as follows:

(1) A Cross-level Bidirectional Feature Pyramid Network (CBiFPN) is designed. By introducing skip connections between non-adjacent layers and extending the feature pyramid to the high-resolution P2 level, CBiFPN effectively enhances multi-scale feature fusion and alleviates the degradation of fine-grained details of small objects during deep propagation.

(2) A lightweight Dual-Scale Convolution module (DualScaleConv) is proposed. Using a parallel branch structure, it simultaneously captures local details and global contextual information, significantly improving the discriminability of tiny defects while substantially reducing the parameter count.

(3) A parameter-free average attention mechanism (PfAAM) is integrated into the C3k2 block. Through progressive feature focusing and channel-spatial collaborative

attention, it enhances feature responses in defect-related regions and suppresses background noise.

## Methodology

### Network architecture of CDA-YOLO

YOLOv11 is one of the most advanced real-time object detection models in the YOLO series. Its network architecture mainly consists of three parts: Backbone, Neck, and Head.

Fig. 2 illustrates the CDA-YOLO network architecture, which is based on YOLOv11. The Backbone is responsible for extracting features from the input image. By integrating the Parameter-free Average Attention Module (PfAAM) into the original C3k2 blocks of YOLOv11, it forms the C3k2_PfAAM module, thereby enhancing feature representation for tiny defects.

Meanwhile, a lightweight Dual-Scale Convolution module (DualScaleConv) is embedded between the Backbone and the Neck. Using a parallel branch structure, it simultaneously captures local details and global contextual information, laying the foundation for subsequent object detection. Building on this, the Neck adopts a Cross-level Bidirectional Feature Pyramid Network (CBiFPN) to perform further feature fusion and multi-scale information transfer on feature maps, effectively enhancing the semantic consistency and fine-grained spatial information of multi-scale features, and helping the network adapt to variations in defect sizes and the challenges posed by complex backgrounds.

The Head is responsible for the final defect classification and bounding box regression tasks, outputting the category and location of objects. These three improved modules work synergistically, enabling CDA-YOLO to significantly improve detection accuracy and robustness for tiny PCB defects while maintaining real-time inference efficiency.

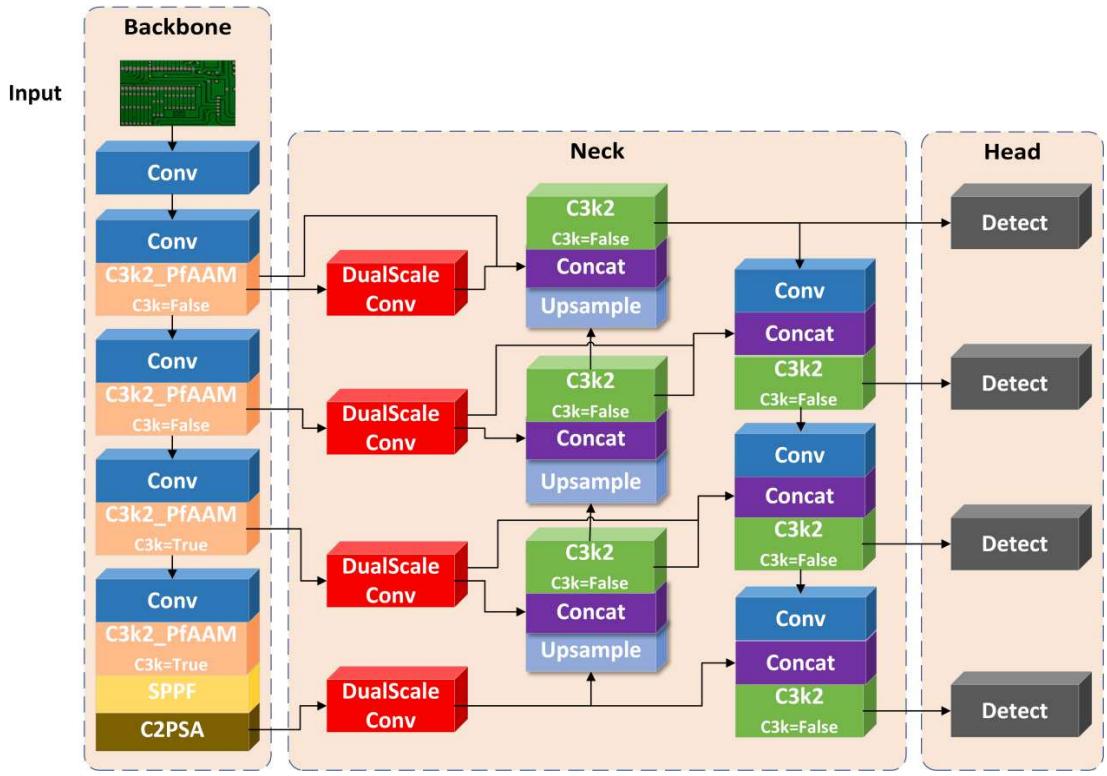



**Fig. 2**. CDA-YOLO overall network structure.

## DualScaleConv module

PCB defects are mostly tiny in size, with cluttered contours and severe interference from dense circuit traces. They are typical small targets that are prone to missed detections and false positives, necessitating the enhancement of the network's multi-scale feature extraction capability. While the standard 3×3 convolution effectively extracts local features, its fixed receptive field limits the modeling of multi-scale contextual information. To address this issue in a simple yet effective way, we propose a lightweight DualScaleConv module, which aims to capture both local details and global context with lower computational cost, thereby improving the representation of small defects and their surrounding environment.

As illustrated in Fig. 3, the DualScaleConv module first splits the input feature map equally along the channel dimension. Let the input feature be $\mathbf{X}\in\mathrm{R}^{C\times H\times W}$; it is divided into $\mathbf{X}_1\in\mathrm{R}^{C/2\times H\times W}$ and $\mathbf{X}_2\in\mathrm{R}^{C/2\times H\times W}$. The Local Detail Branch employs a standard 3×3 convolution (dilation rate $d$=1) to preserve fine-grained local detail extraction. The Contextual Perception Branch adopts a 3×3 dilated convolution with

the same kernel size but a dilation rate of $d$=2, which significantly enlarges the receptive field without increasing parameter count, capturing a broader range of contextual information[16,17]. The outputs of the two branches are concatenated along the channel dimension to produce the final feature map $\mathbf{Y}\in\mathrm{R}^{C\times H\times W}$. Equation (1) shows this process.

$$\mathbf{Y}=\mathrm{Concat}\left(\mathrm{Conv}_{3\times3}^{d=1}(\mathbf{X}_1),\ \mathrm{Conv}_{3\times3}^{d=2}(\mathbf{X}_2)\right) \tag{1}$$

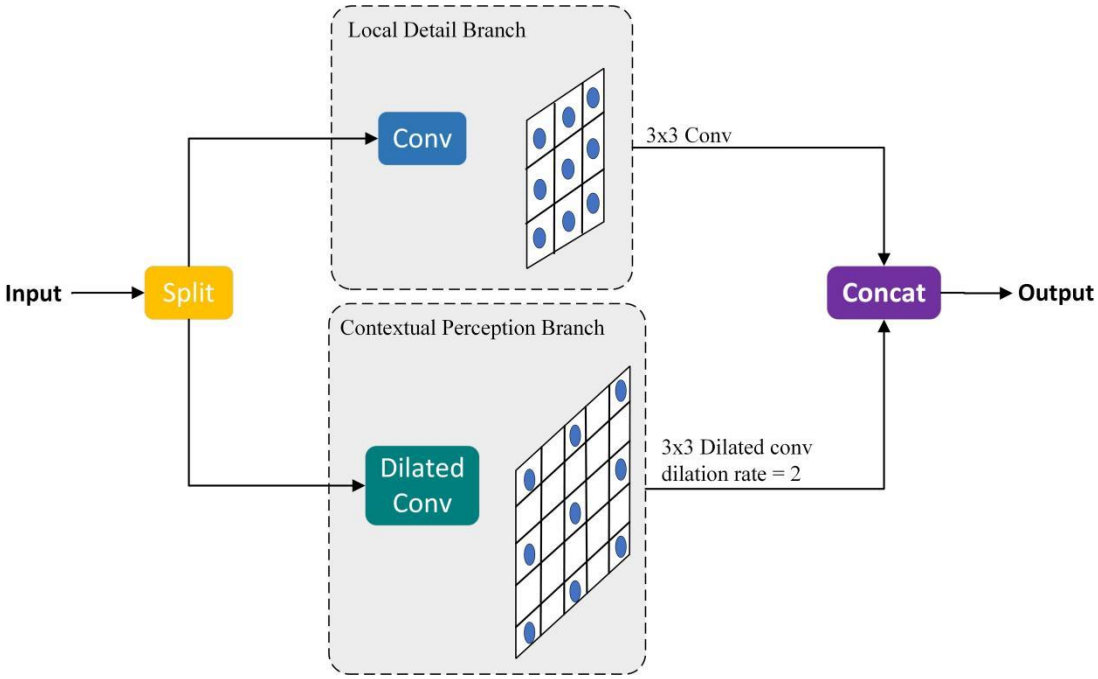



**Fig. 3**. DualScaleConv module structure.

A standard 3×3 convolution requires $C_{\mathrm{in}}\times C_{\mathrm{out}}\times3\times3$ parameters. In DualScaleConv, the input channels are grouped; each group has half the channels, and each group independently performs a 3×3 convolution (with output channels also halved). Therefore, the total number of parameters is approximately $2\times((C/2)\times(C/2)\times9)=\frac{C^2}{2}\times9$, compared to $C^2\times9$ for a standard convolution. Consequently, DualScaleConv reduces the parameter count by about 50%, significantly lowering the model's computational burden.

The Local Detail Branch (d=1) maintains the original receptive field and excels at capturing local high-frequency information such as defect edges and textures. The

Contextual Perception Branch (d=2) expands the receptive field to an equivalent 5×5 range via dilated convolution, enabling the perception of contextual information around the defect. This dual-branch design is particularly beneficial for tiny defects on PCBs (e.g., broken lines, short circuits): the local branch ensures that fine defect details are not missed, while the contextual branch provides discriminative background information to help distinguish genuine defects from noise. Tiny targets such as micro-cracks, cold solder joints, and scratches typically occupy only a few pixels, lacking sufficient features for learning. Therefore, incorporating contextual information can provide additional features.

In summary, DualScaleConv provide complementary multi-scale receptive fields with fewer parameters, especially in scenarios involving small targets and complex backgrounds.

## CBiFPN

In the baseline YOLOv11, the neck network adopts a Path Aggregation Network (PANet)[2,12,18] structure, which enables multi-scale feature fusion through a top-down and bottom-up pathway. Although PANet improves the flow of semantic and localization information, it primarily relies on linear sequential connections, which can lead to the gradual dilution of fine-grained spatial information (crucial for small object detection) from the original input image as the propagation path lengthens within deep networks. To further enhance multi-scale feature fusion in the neck network of YOLOv11, we propose a Cross-level Bidirectional Feature Pyramid Network (CBiFPN). This module is inspired by the efficient bidirectional cross-scale connections introduced in EfficientDet[19], while introducing several key enhancements to the original PANet topology.

To further enhance cross-level information flow, we introduce skip connections between non-adjacent layers. This design allows gradients to backpropagate more directly and enables deep feature maps to retain more original spatial details from shallower networks. Specifically, features from the backbone's shallow layer (e.g., P2) are directly fused with deeper neck outputs. This design preserves fine-grained

spatial details and alleviates the degradation of small object features caused by deep stacking.

We insert the DualScaleConv module between the backbone and the neck network, which refines the input feature maps before they enter the bidirectional fusion paths. This convolutional layer reduces channel-wise redundancy and enhances the alignment of features across different scales, leading to more stable and effective multi-scale fusion.

Finally, we extend the original feature pyramid down to level 2 (P2) to incorporate higher-resolution feature maps. The inclusion of the P2 detection head significantly improves the model's sensitivity to small and fine-grained PCB defects, such as spur, missing hole, and mouse bite. The overall structure of CBiFPN is illustrated in Fig. 4.

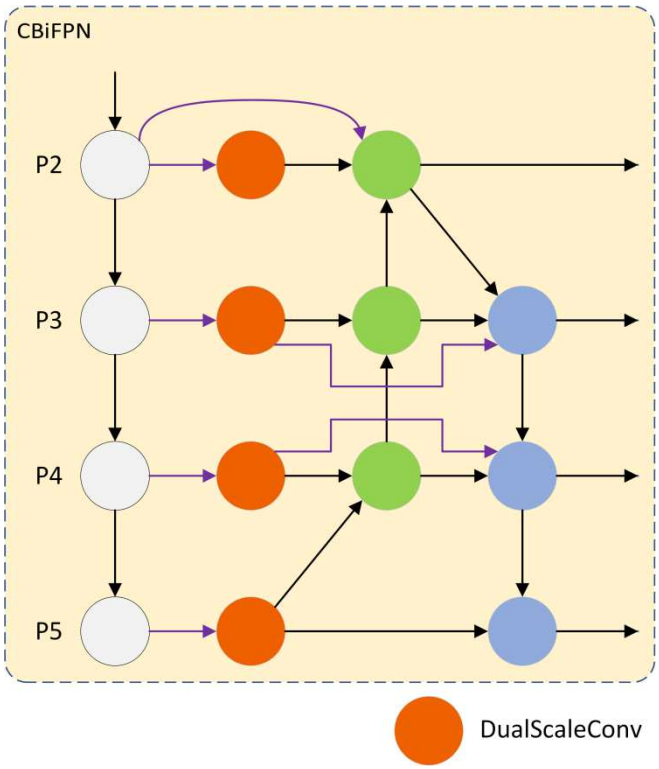



**Fig. 4**. Neck network structure.

## C3k2_PfAAM module

The C3k2 block improves feature extraction by utilizing smaller convolution kernels, enabling the model to capture intricate details essential for identifying small or partially visible PCB defects. However, in high-density PCB images, background

noise and irrelevant regions often dilute the representation of defect features. To address this, we integrate the PfAAM[20] mechanism to refine the feature maps.

This paper introduces a novel C3k2_PfAAM[21] module. This module synergistically combines the fine-grained feature extraction capacity of the C3k2 block with the targeted focus provided by the Parameter-free Average Attention Module (PfAAM), establishing a robust architecture for complex industrial inspection scenarios.

PfAAM generates 3D attention weights using only simple average pooling and element-wise multiplication, introducing no trainable parameters; thus it almost no increases model size or inference latency. As illustrated in Fig. 5, C3k2_PfAAM retains the cross-stage partial structure and consists of two convolutional layers (*cv1*, *cv2*), a stack of Bottleneck (or C3k) blocks, and a PfAAM attention module.

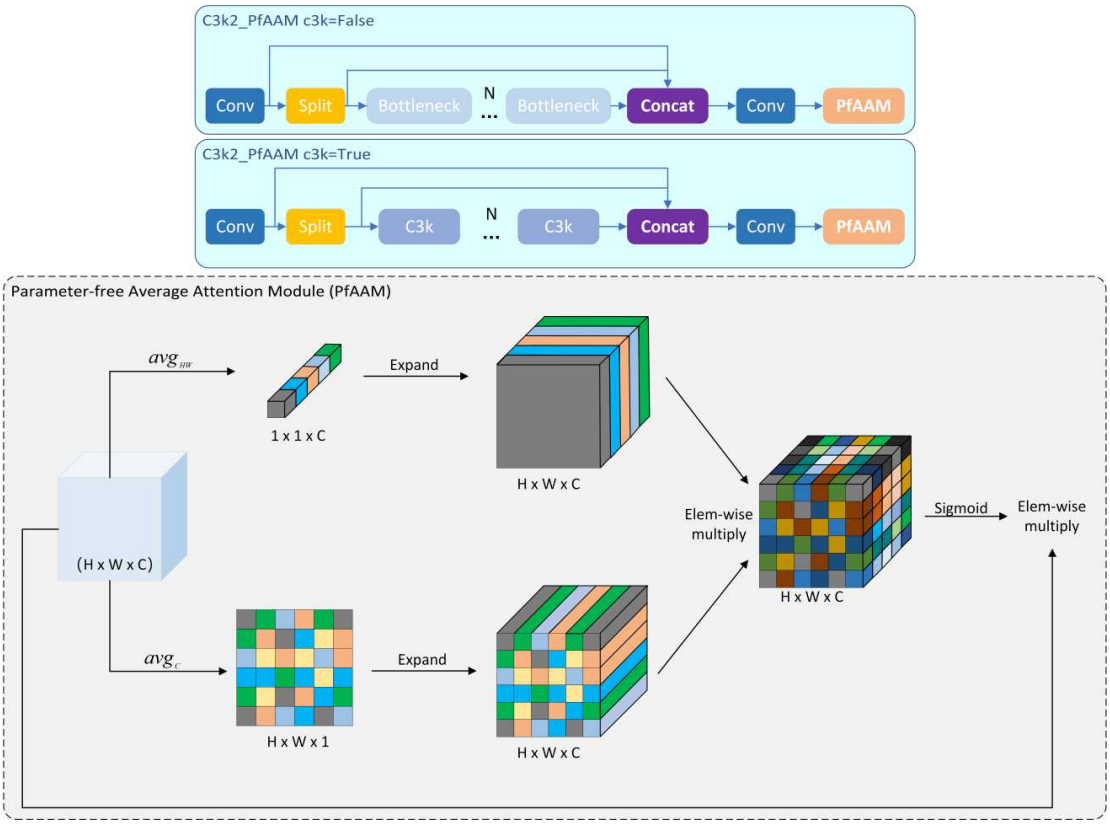



**Fig. 5**. C3k2_PfAAM block overall architecture.

Given an input feature map $\mathbf{X}\in\mathrm{R}^{C\times H\times W}$, the module first expands the channel dimension to $2C^{'}$ (where $C^{'}=e\cdot C_{\text{out}}$ and $e$ is the expansion ratio) via *cv1*, then splits it into two branches along the channel dimension. One branch is fed into a series of

Bottleneck blocks for deep feature extraction, while the other serves as a residual shortcut connection. After passing through $n$ Bottleneck blocks, the output $\mathbf{F}_{\text{bottleneck}}$ is concatenated with the shortcut features, yielding $\mathbf{F}_{\text{cat}} \in \mathrm{R}^{(2+n)C' \times H \times W}$, which is then compressed to the target output channel number $C_{\text{out}}$ by *cv2*.

The PfAAM module then applies parameter-free average attention to this intermediate feature, as shown in Eq. (2), Eq. (3) and Eqs. (4).

$$\mathbf{A}_{\text{ch}} = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{F}_{\text{tmp}}(i,j,:) \quad \text{(channel average)} \tag{2}$$

$$\mathbf{A}_{\text{sp}} = \frac{1}{C_{\text{out}}} \sum_{k=1}^{C_{\text{out}}} \mathbf{F}_{\text{tmp}}(:,:,k) \quad \text{(spatial average)} \tag{3}$$

$$\mathbf{A} = \sigma(\mathbf{A}_{\text{ch}} \otimes \mathbf{A}_{\text{sp}}), \mathbf{X}_{\text{out}} = \mathbf{A} \odot \mathbf{F}_{\text{tmp}}, \tag{4}$$

$\mathbf{F}_{\text{tmp}} = \text{Conv}_{1\times1}(\mathbf{F}_{\text{cat}})$, $\sigma$ is the Sigmoid function, $\otimes$ denotes broadcasted element-wise multiplication, and $\odot$ represents element-wise weighting. The final output $\mathbf{X}_{\text{out}}$ not only preserves the original feature information but also enhances the response of defect-related regions while suppressing background noise.

## Experimentation

### Dataset

As shown in Fig. 6, this experiment uses the printed circuit board (PCB) defect dataset released by the Open Lab on Human Robot Interaction, Peking University (https://robotics.pkusz.edu.cn/resources/dataset). The dataset comprises a total of 693 PCB images, encompassing six types of representative defects in the PCB manufacturing process: Missing hole, Mouse bite, Open circuit, Short, Spur, and Spurious copper. However, the total number of images and the quantity within each category are insufficient to meet the training requirements, and there is a disparity with real-world industrial inspection scenarios. To enrich features and enhance model robustness, data augmentation techniques are employed to expand the sample set. Specifically, horizontal and vertical flipping are applied to the original images, resulting in 1386 images. The flipping operation generates abundant new positional information, effectively enhancing the generalization ability of the model.

Subsequently, Gaussian blur is applied to 15% of these images to simulate common sensor noise encountered during actual industrial acquisition. The dataset is partitioned into a training set, a validation set, and a test set at a ratio of 7:2:1.

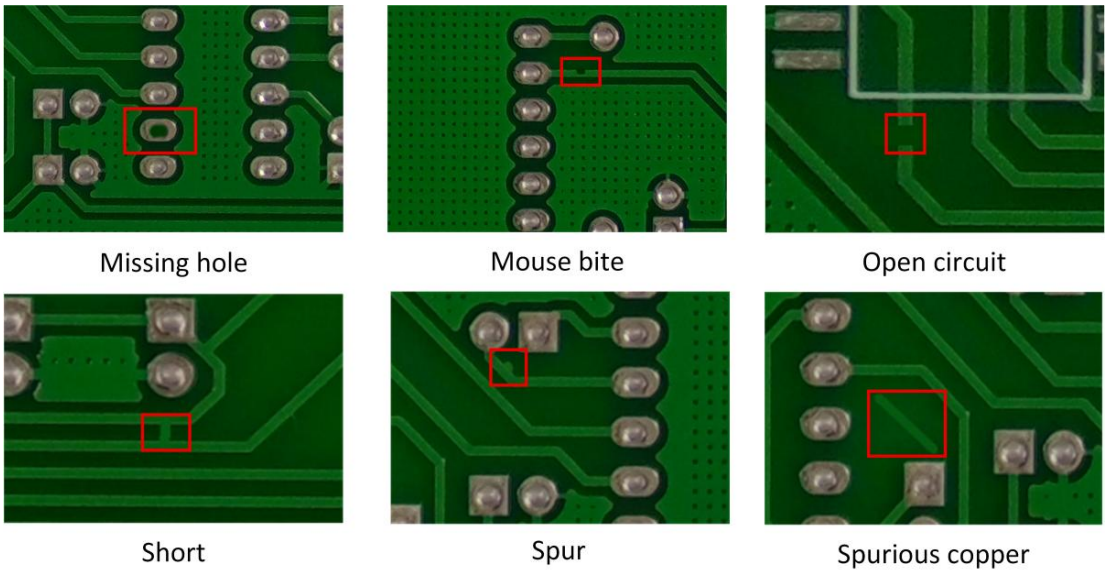



**Fig. 6.** Examples of PCB defects.

## Experimental environment and training parameters

The experiments were conducted on the Windows 11 operating system. The detailed experimental environment configuration is shown in Table 1.

| Category | Configuration |
|---|---|
| OS | Windows 11 |
| CPU | Intel 11th Gen i5-11400H @ 2.70GHz |
| GPU | NVIDIA GeForce RTX 3050 Ti Laptop (4GB VRAM) |
| RAM | 16 GB |
| Language | Python 3.11 |
| Framework | PyTorch 2.5.0 |
| CUDA version | 11.8 |

**Table 1.** Experimental environment configuration.

To ensure a fair comparison, the model is trained for 300 epochs without using pre-trained weights. The batch size was set to 8 due to hardware memory limitations. An early stopping mechanism (patience) was set to 50 epochs, terminating training early if the validation loss did not improve for 50 consecutive epochs to prevent overfitting. The main training parameters are listed in Table 2.

| Parameter | optimizer | lr0 | lrf | momentum | weight_decay | epochs | batch size | imgsz | warmup_epochs |
|---|---|---|---|---|---|---|---|---|---|
| Value | AdamW | 0.001 | 0.01 | 0.937 | 0.0005 | 300 | 8 | 640 | 3 |

**Table 2.** Training parameter configuration.

To improve model generalization and mitigate overfitting, a series of data augmentation strategies were adopted during training. These strategies include geometric transformations, color space enhancements, and image mixing methods.

## Evaluation indicators

To systematically evaluate the performance of the proposed model in PCB defect detection, we adopt Precision, Recall, Average Precision (AP), mean Average Precision (mAP), and mAP50-95 as the primary evaluation metrics. Additionally, we also adopts the number of model parameters (Parameters) and computational complexity (GFLOPs) as evaluation metrics to assess the model's performance in terms of resource consumption and inference efficiency.

Precision measures the proportion of correctly detected positive samples among all samples predicted as positive, reflecting the model's false detection rate. Recall measures the proportion of correctly detected positive samples among all actual positive samples, reflecting the model's missed detection rate. They are defined as in Eq. (5) and Eqs. (6).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

where TP (True Positive) is the number of correctly detected positive samples, FP (False Positive) is the number of negative samples incorrectly detected as positive, and FN (False Negative) is the number of positive samples missed by the model.

Average Precision (AP) is defined as the area under the precision-recall (P-R) curve, where recall is on the x-axis and precision on the y-axis. The closer the AP value is to 1, the better the model performs on that class. It is calculated in Eq. (7).

$$AP = \int_0^1 P\,(R)\,dR \tag{7}$$

The mean Average Precision (mAP) is the arithmetic mean of AP values over all object classes, reflecting the overall detection performance of the model across multiple categories. It is defined as in Eq.(8).

$$mAP = \frac{1}{C}\sum_{i=1}^{C} A\,P_i \tag{8}$$

where C is the total number of classes and $AP_i$ is the Average Precision of the i-th class.

mAP50 refers to the mAP calculated with an IoU threshold of 0.5, which is a common benchmark for evaluating object detection models. mAP50-95 refers to the average mAP calculated over multiple IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05, providing a stricter evaluation of the model's ability to precisely localize objects.

$$mAP_{50} = mAP|_{IoU=0.5} mAP_{50-95} = \frac{1}{10}\sum_{i=0}^{9} m\,AP|_{IoU=0.5+i\times 0.05} \tag{9}$$

# Results and analysis

## Comparison experiments

1）DualScaleConv

To evaluate the design rationale of various convolutional operations in the DualScaleConv module, this study compares three branch configurations: A

(Proposed): the first branch employs a standard 3×3 convolution, while the second uses a 3×3 dilated convolution with a dilation rate d=2; B: both branches employ a standard 3×3 convolution; C: both branches use standard 3×3 convolutions. The results are shown in Table 3.

| Plan | Branch Convolution | Params (M) | GFLOPs | mAP@0.5 (%) | mAP@0.5:0.95 (%) |
|---|---|---|---|---|---|
| A (Ours) | Standard + Dilated | 3.15 | 12.8 | **95.4** | 55.5 |
| B | Standard + Depthwise Separable | 2.69 | 10.5 | 94.9 | 55.3 |
| C | Standard + Standard | 3.15 | 12.8 | 95.1 | 56.0 |

**Table 3.** The comparison of different plans of branch convolution.

The experimental results demonstrate that, compared to Scheme C which does not employ dilated convolution, Scheme A utilizing dilated convolution achieves a 0.3% improvement in mAP@0.5, confirming the effectiveness of dilated convolution in expanding the receptive field and enhancing multi-scale feature extraction. Although Scheme B adopts depthwise separable convolution for lightweight design, its detection accuracy is inferior to that of Scheme A.

2） C3k2_PfAAM

To validate the effectiveness of the proposed C3k2_PfAAM module and its optimal insertion position within the YOLOv11 framework, a series of comparative experiments are designed in this section. Since using a single module alone is insufficient to explore the adaptability of different attention mechanisms to the already improved network, we employ four lightweight attention modules

respectively on top of the previously constructed strong baseline model (i.e., the YOLOv11 integrated with CBiFPN and DualScaleConv modules), to fairly evaluate their contributions to multi-scale feature enhancement.

First, we integrate attention mechanisms into the C3k2 module in the backbone, with results presented in Table 4.

| Module | mAP@0.5 (%) | mAP@0.5:0.95 (%) | P (%) | R (%) | Params (M) | GFLOPs |
|---|---|---|---|---|---|---|
| Strong baseline | 95.4 | 55.5 | 95.3 | 92.5 | 3.15 | 12.8 |
| C3k2_SimAM[22] | 95.8 | 55.8 | 96.3 | 91.0 | 3.11 | 12.8 |
| C3k2_MCAttn[23] | 95.5 | 55.3 | 94.3 | 91.7 | 3.29 | 13.1 |
| C3k2_ELA[24] | 95.6 | 55.9 | 96.0 | 90.4 | 3.12 | 12.8 |
| **C3k2_PfAAM (Ours)** | **96.1** | **56.0** | 96.2 | 90.6 | 3.11 | 12.8 |

**Table 4.** Comparison of different attention modules on the strong CBiFPN + DualScaleConv baseline.

Then, the different C3k2 modules were applied separately to the backbone and neck for comparison. The results are shown in the Fig. 7.

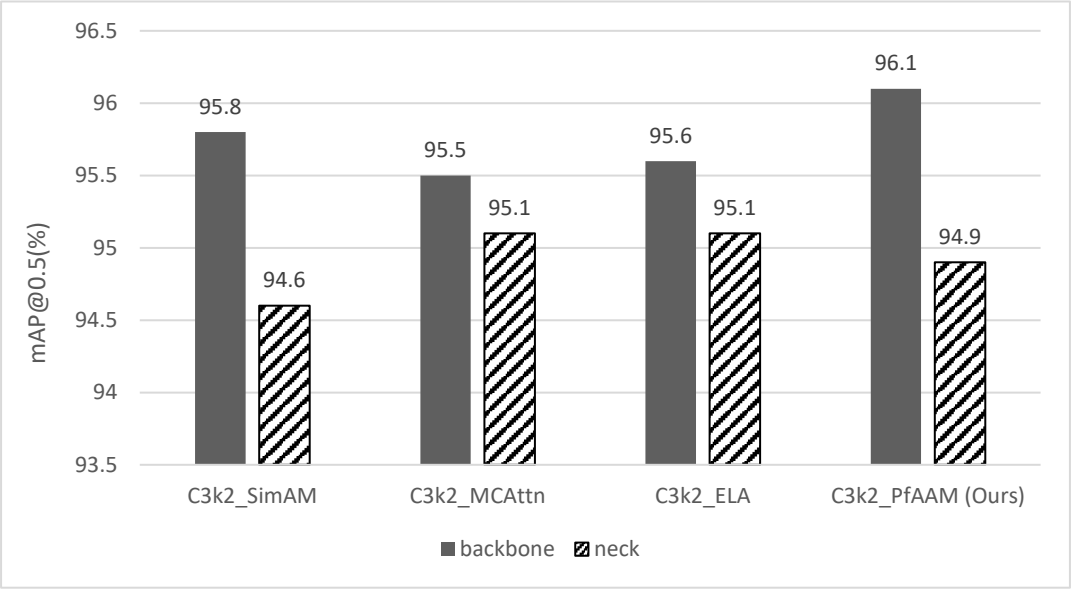



**Fig. 7**. mAP@0.5(%) comparison of improved module at different insertion positions.

As shown in Table 4, the C3k2_PfAAM module outperforms all other attention variants in every evaluation metrics. Specifically, in terms of mAP@0.5, C3k2_PfAAM achieves 96.1%, which is 0.3%, 0.6%, 0.5% and 0.7% higher than C3k2_SimAM, C3k2_MCAttn, C3k2_ELA and *Strong baseline* respectively. Since the introduced attention mechanisms are parameter-free and lightweight, they introduce negligible additional computational cost or parameters. The experimental results indicate that through its progressive feature focusing and channel-spatial collaborative attention mechanism, PfAAM can more effectively suppress complex background interference and enhance feature responses of tiny defects, thereby achieving a better balance between detection accuracy and model efficiency.

As shown in Fig. 7, using an attention-integrated C3k2 module in the backbone achieves better detection performance. Further analysis indicates that parameter-free global attention (SimAM, PfAAM) has very low computational cost and is suitable for computation-sensitive mid-to-deep backbone layers to enhance global semantics efficiently. In contrast, local multi-scale attention (ELA, MCAttn), despite introducing a few parameters, effectively enhances spatial details and multi-scale feature interactions, making these modules more suitable for low-level neck layers to improve small object and fine edge representation. Consequently, when applied in the neck, ELA and MCAttn outperform SimAM and PfAAM. However, overall, using an attention-integrated C3k2 module in the backbone achieves better detection performance.

3） Comparison of different methods

Table 5 demonstrates our model's superior performance over current object detection methods on the datasets. The proposed model achieves significant advantages across multiple evaluation metrics, including mAP50, mAP50-95, Precision and Recall. Compared with the baseline model YOLOv11n, our model improves mAP50 and mAP50-95 by 2.8% and 3.4% respectively. These results fully demonstrate the improvement of detection accuracy achieved by our strategies. In contrast, although

YOLOv6n has a larger number of parameters, its detection accuracy is lower than that of the model presented in this paper. Similarly, YOLOv5n and YOLOv10n exhibit certain advantages in computational cost, but their detection accuracy is significantly lower than that of our model, making them inadequate to meet the high demands for both precision and real-time performance in industrial scenarios.

| Model | mAP50 (%) | mAP50-95 (%) | Params (M) | GFLOPs | P% | R% |
|---|---|---|---|---|---|---|
| YOLOv3-tiny | 89.3 | 49.1 | 12.1 | 18.9 | 92.2 | 82.6 |
| YOLOv5n | 93.3 | 51.6 | 2.50 | 7.1 | 93.9 | 89.7 |
| YOLOv6n[25] | 93.0 | 51.0 | 4.23 | 11.8 | 94.5 | 89.6 |
| YOLOv8n | 94.4 | 52.8 | 3.01 | 8.1 | 95.1 | 90.6 |
| YOLOv10n[26] | 91.4 | 50.8 | 2.27 | 6.5 | 91.7 | 84.8 |
| YOLOv11n | 93.3 | 52.6 | 2.59 | 6.4 | 94.5 | 89.3 |
| YOLOv12n[27] | 93.4 | 51.0 | 2.56 | 6.3 | 95.1 | 90.1 |
| Ours | 96.1 | 56.0 | 3.11 | 12.8 | 96.2 | 90.6 |

**Table 5.** Comparison of different methods

## Ablation experiment

Table 6 presents ablation experiments that analyze YOLOv11 with CBiFPN, DualScaleConv, and C3k2_PfAAM modules, focusing on their impact on mAP50 and mAP50-95. The baseline YOLOv11 model starts with mAP50 of 93.3% and mAP50-95 of 52.6%. Introducing only the CBiFPN module enhances multi-scale feature fusion, raising mAP@0.5 to 94.8% and mAP@0.5:0.95 to 54.1%, while slightly increasing the number of parameters, demonstrating the effect of CBiFPN in cross-scale information interaction. Adding only the DualScaleConv module improves the collaborative representation of local and global features through dual-scale convolutions, achieving a mAP@0.5 of 94.0% and mAP@0.5:0.95 of 52.7%. Integration of the C3k2_PfAAM module maintains lightweight properties while enhancing sensitivity to small targets, significantly increasing mAP@0.5 and

mAP@0.5:0.95 to 94.1% and 53.1%, reducing parameters to 2.54M, demonstrating a favorable balance between accuracy and efficiency. Finally, integrating all three modules achieves the best performance: mAP@0.5 reaches 96.1%, mAP@0.5:0.95 reaches 56.0%, improving by 2.8% and 3.4% over the baseline, respectively. Meanwhile, the number of parameters is compressed to 3.1M and computational cost is 12.8 GFLOPs. These results fully demonstrate the complementarity and effectiveness of the three proposed modules in improving detection accuracy, enhancing feature representation, and maintaining real-time performance.

| Model | CBiFPN | DualScale Conv | C3k2_PfAAM | mAP@0.5 (%) | mAP@0.5:0.95 (%) | Params (M) | GFLOPs |
|---|---|---|---|---|---|---|---|
| YOLOv11 (baseline) | | | | 93.3 | 52.6 | 2.59 | 6.4 |
| A | ✓ | | | 94.8 | 54.1 | 2.79 | 11.0 |
| B | | ✓ | | 94.0 | 52.7 | 3.03 | 7.7 |
| C | | | ✓ | 94.1 | 53.1 | 2.54 | 6.3 |
| D | ✓ | ✓ | | 95.4 | 55.5 | 3.15 | 12.8 |
| E | ✓ | | ✓ | 94.1 | 53.1 | 2.75 | 11.0 |
| F | | ✓ | ✓ | 94.4 | 52.2 | 2.99 | 7.7 |
| Ours | ✓ | ✓ | ✓ | **96.1** | 56.0 | 3.11 | 12.8 |

**Table 6.** Ablation study results for different modules.

## Visualization

In the field of object detection, visualization aids in interpreting model behavior. Fig. 8 presents a qualitative comparison between the proposed method and the baseline YOLOv11 model in terms of object detection results, while Fig. 9 employs Grad-CAM++ to generate heatmaps that intuitively illustrate the model's focus on target regions[28]. Grad-CAM++ computes higher-order derivative weights for each

pixel within every feature map channel, enabling more precise localization of positively contributing regions and offering stronger discriminative ability in multi-object scenarios compared to standard Grad-CAM[29,30]. It can be observed that, compared to YOLOv11, the proposed method produces more complete and accurately localized detection bounding boxes. Moreover, the heatmap results indicate that the proposed method concentrates effectively on genuine target regions while suppressing background interference, which is benefited from Grad-CAM++'s emphasis on positive gradient information and its pixel-wise weighting strategy. Overall, the proposed method achieves more complete and reliable object detection in complex scenes, along with notably improved detection confidence scores.

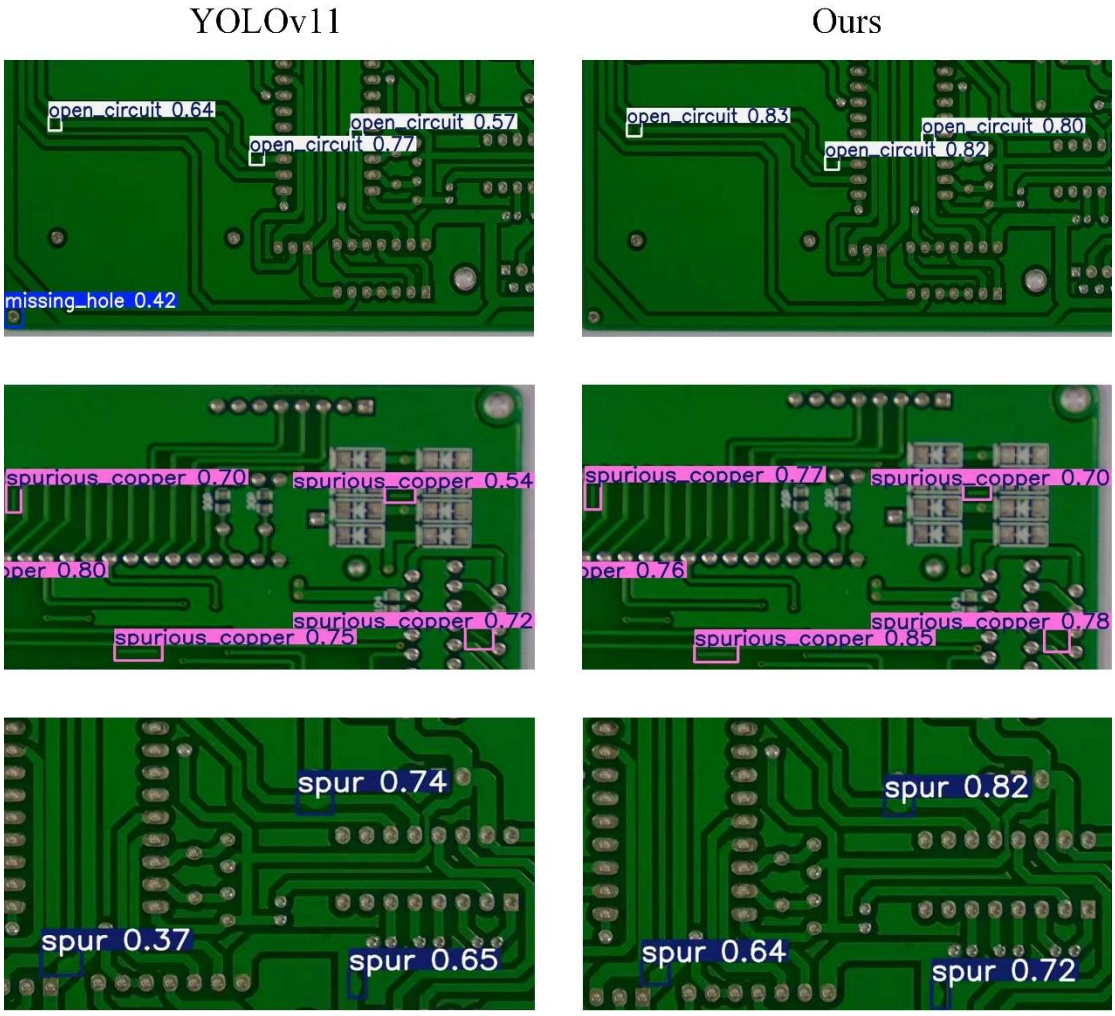



**Fig. 8**. The comparison results of our method with YOLOv11 are shown.

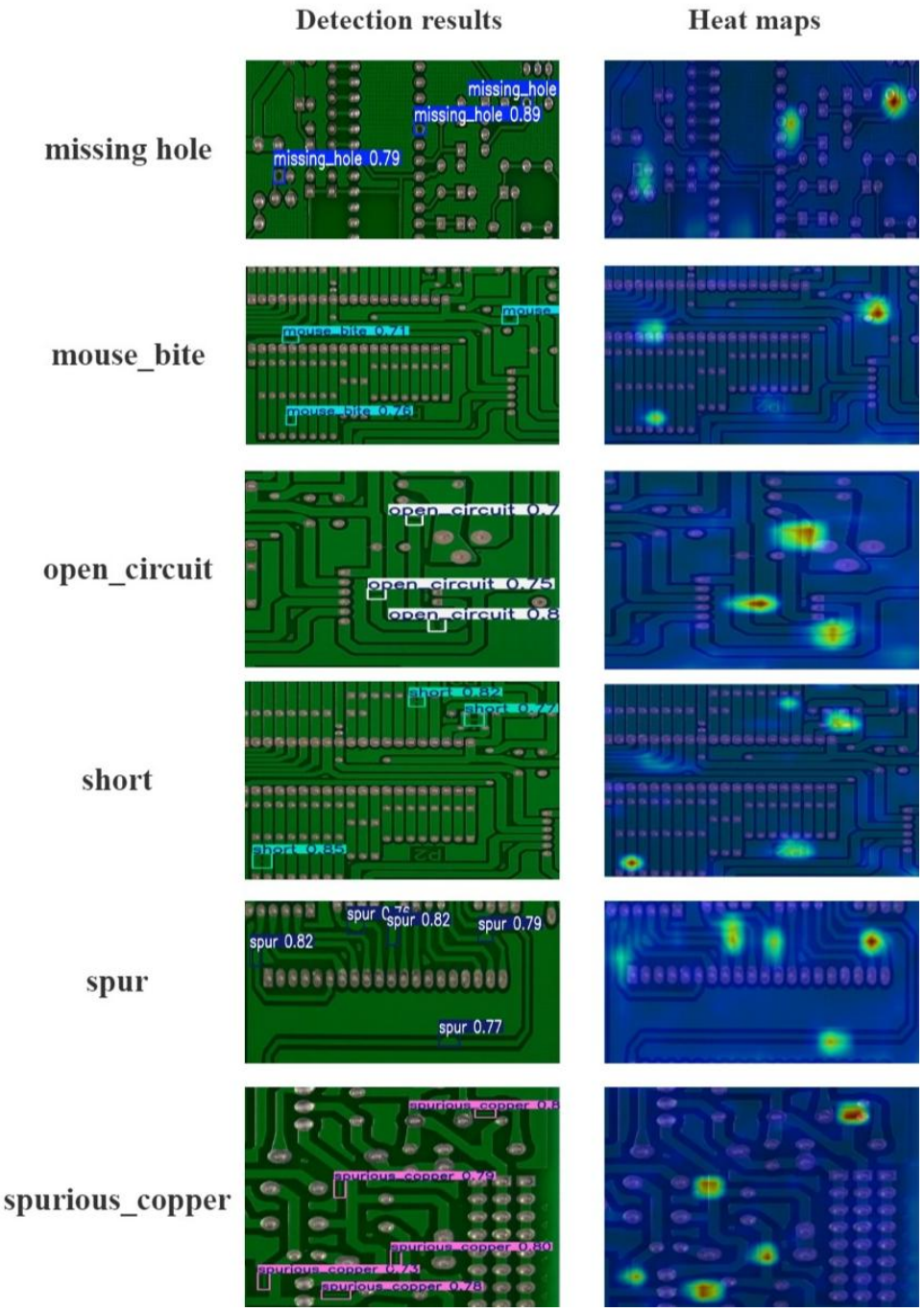



**Fig. 9**. Heatmap visualizations of CDA-YOLO detection results using Grad-CAM++.

## Conclusion

This paper proposes CDA-YOLO, an improved YOLOv11 model for PCB defect detection, to address the challenges of high missed-detection rate of tiny defects and high false-positive rate caused by complex background textures. First, we design a Cross-level Bidirectional Feature Pyramid Network (CBiFPN). Second, we propose

DualScaleConv, which uses parallel branches to capture local details and global context. Third, a parameter-free average attention mechanism (PfAAM) is integrated into the C3k2 module. Together, these enhancements improve multi-scale fusion and small-target detection.

Experiments show that CDA-YOLO achieves mAP@0.5 of 96.1% and mAP@0.5:0.95 of 56.0%, improving by 2.8% and 3.4% over YOLOv11. Ablation and comparison experiments validate the effectiveness and complementarity of each module. Visualization analysis shows that the proposed method can better focus on defects and generate more accurate detection bounding boxes, significantly reducing missed detections and false positives. Future work includes further compressing the model parameters and computational complexity for edge deployment, exploring more efficient attention mechanisms and lightweight architectures, and extending our method to other industrial vision inspection scenarios, such as semiconductor wafer defect detection and surface mount technology (SMT) solder joint inspection.

## Data availability

The datasets analysed during the current study are available at https://robotics.pkusz.edu.cn/resources/dataset

## References


1. Kiernich, A., Kalenik, J., Stęplewski, W., Kościelski, M., Chołaj. A. Impact of Particular Stages of the Manufacturing Process on the Reliability of Flexible Printed Circuits. *Sensors* **25**, 140. https://doi.org/10.3390/s25010140 (2025).
2. Lang, D., Lv, Z. SEPDNet: simple and effective PCB surface defect detection method. *Sci Rep* **15**, 10919. https://doi.org/10.1038/s41598-024-84859-2 (2025).
3. Reinhardt Seidel, Ben Rachinger, Nils Thielen, Konstantin Schmidt, Sven Meier, Jörg Franke. Development and validation of a digital twin framework for SMT

manufacturing. *Computers in Industry* **145**, 103831. https://doi.org/10.1016/j.compind.2022.103831 (2023).

4. Ling, Q. & Isa, N. A. M. Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey. *IEEE Access* **11**, 15921–15944. doi:10.1109/ACCESS.2023.3245093 (2023).
5. Nguyen, H. D., Cheng, D., Wang, X., Shi, Y. & Wen, B. Data-Efficient Deep Learning for Printed Circuit Board Defect Detection Using X-Ray Images. *IEEE Trans. Instrum. Meas* **74**, 1–14. doi:10.1109/TIM.2025.3529089 (2025).
6. Zihan, H., Lian, Y., Wang, Y. & Lu, Z. A comprehensive review of research on surface defect detection of PCBs based on machine vision. *Results in Engineering* **27**, 106437. https://doi.org/10.1016/j.rineng.2025.106437 (2025).
7. Zheng, J., Sun, X., Zhou, H., Tian, C. & Qiang, H. Printed circuit boards defect detection method based on improved fully convolutional networks. *IEEE Access* **10**, 109908–109918. doi: 10.1109/ACCESS.2022.3214306 (2022).
8. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Redmon, J. et al.) 779–788 (IEEE, 2016).
9. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767 (2018).
10. Khanam, R. & Hussain, M. What is YOLOv5: a deep look into the internal features of the popular object detector. arXiv preprint arXiv:2407.20892. https://doi.org/10.48550/arXiv.1804.02767 (2024).
11. Varghese, R. & Sambath, M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness. In 2*024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)* 1–6 (IEEE, 2024)
12. Khanam, R. & Hussain, M. Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725. https://doi.org/10.48550/arXiv.2410.17725 (2024).

13. Zhang, W., Chen, W., Ni, Y., Wang, S. & Dang, B. YOLO-VCL: a lightweight and high-precision PCB defect detection network model. *J Real-Time Image Proc* **23**, 61. https://doi.org/10.1007/s11554-026-01858-4 (2026).
14. He, X. & Xie, M. ABF-YOLO: A Fine-Grained PCB Defect Detection Framework Integrating Axial Attention and Bidirectional Feature Fusion. *IEEE Trans. Consum. Electr.* **71** (3), 8562-8570. doi: 10.1109/TCE.2025.3597022 (2025).
15. Elsharkawy, Z.F. Enhanced YOLOv11 framework for high precision defect detection in printed circuit boards. *Sci Rep* **15**, 42550. https://doi.org/10.1038/s41598-025-27415-w (2025).
16. Yu, F. & Koltun, V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122. https://doi.org/10.48550/arXiv.1511.07122 (2015).
17. Sun, Y. et al. MDSF-YOLO: Advancing object detection with a multiscale dilated sequence fusion network. *IEEE Trans. Neural Netw. Learn. Syst.* **37** (3), 1236–1247. doi:10.1109/TNNLS.2025.3617122 (2026).
18. Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. Path Aggregation Network for Instance Segmentation. In *2018 IEEE /CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Liu, S. et al.) 8759–8768 (IEEE, 2018). doi: 10.1109/CVPR.2018.00913.
19. Tan, M., Pang, R. & Le, Q. V. EfficientDet: Scalable and Efficient Object Detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Tan, M. et al.) 10778-10787 (IEEE, 2020) doi: 10.1109/CVPR42600.2020.01079.
20. Körber, N. Parameter-Free Average Attention Improves Convolutional Neural Network Performance (Almost) Free of Charge. arXiv preprint arXiv: 2210.07828. https://doi.org/10.48550/arXiv.2210.07828 (2022).
21. Zhang, W., Lin, B., Wei, S. & Wu, J. Small Target Detection of Surface Defects on PCB Boards: An Improved YOLO Method Integrating Attention Mechanism

and Multi-Scale Feature Focusing. *Artif. Intell. Eng. Des. Anal. Manuf.* **40,** e2. https://doi.org/10.1017/S0890060426100213 (2026).

22. Yang, L., Zhang, R., Li, L. & Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning (ICML)* (eds. Meila, M. & Zhang, T.) 11863-11874 (PMLR, 2021).
23. Dai, W. et al. Exploiting Scale-Variant Attention for Segmenting Small Medical Objects. *IEEE Trans. Neural Netw. Learn. Syst.* **37** (7), 3161-3175. doi: 10.1109/TNNLS.2025.3645355 (2026).
24. Xu, W., Wan, Y. & Zhao, W. ELA: efficient location attention for deep convolution neural networks. *J Real-Time Image Proc* **22**, 140. https://doi.org/10.1007/s11554-025-01719-6 (2025).
25. Li, C. et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv preprint arXiv:2209.02976. https://doi.org/10.48550/arXiv.2209.02976 (2022).
26. Wang, A. et al. YOLOv10: real-time end-to-end object detection. arXiv preprint arXiv:2405.14458. https://doi.org/10.48550/arXiv.2405.14458 (2024).
27. Tian, Y., Ye, Q. & Doermann, D. YOLOv12: attention-centric real time object detectors. In *Advances in Neural Information Processing Systems 38 (NeurIPS)*. https://doi.org/10.48550/arXiv.2502.12524 (2025).
28. Chattopadhay, A., Sarkar, A., Howlader, P. & Balasubramanian, V. N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* 839-847 (IEEE, 2018) doi: 10.1109/WACV.2018.00097.
29. Selvaraju, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)* 618–626 (IEEE, 2017) doi: 10.1109/ICCV.2017.74.

30. Persada, A. G., Ardiyanto, I., Sasongko, M. B. & Nugroho, H. A. A Review of CAM-Based Visual Explanation on Diabetic Retinopathy. *IEEE Access* **14**, 13737-13767. doi: 10.1109/ACCESS.2026.3656888 (2026).

## Figure Legends

**Fig. 1**. YOLOv11 network structure. The backbone extracts multi-scale features via convolutional and C3k2 modules; the neck fuses features using PANet; the head produces final detections.

**Fig. 2**. CDA-YOLO overall network structure. It consists of a backbone with C3k2_PfAAM modules, a DualScaleConv module inserted between backbone and neck, a CBiFPN neck for enhanced multi-scale fusion, and the original YOLOv11 detection head.

**Fig. 3**. DualScaleConv module structure. The input feature map is split along the channel dimension into two branches: one uses a standard 3×3 convolution (d=1) to capture local details, the other uses a dilated 3×3 convolution (d=2) to capture global context. The outputs are concatenated.

**Fig. 4**. Neck network structure. It introduces skip connections between non-adjacent layers, incorporates DualScaleConv for feature refinement, and extends the feature pyramid down to the high-resolution P2 level to improve small-defect detection.

**Fig. 5**. C3k2_PfAAM block overall architecture. It retains the cross-stage partial design with two convolutional layers and a stack of bottleneck blocks, followed by a parameter-free average attention module (PfAAM) that generates 3D attention weights via average pooling and element-wise multiplication.

**Fig. 6.** Examples of PCB defects. Six typical defect types are shown: missing hole, mouse bite, open circuit, short, spur, and spurious copper.

**Fig. 7**. mAP@0.5(%) comparison of improved module at different insertion positions. Comparison of mAP@0.5 (%) when applying improved C3k2 modules at different insertion positions (backbone and neck). The results show that placing

attention-enhanced C3k2 in the backbone generally yields better detection performance.

**Fig. 8**. The comparison results of our method with YOLOv11 are shown. Qualitative detection comparison between the proposed method and YOLOv11. Our method produces more complete and accurately localized bounding boxes with higher confidence scores, reducing missed detections and false positives.

**Fig. 9**. Heatmap visualizations of CDA-YOLO detection results using Grad-CAM++. The model focuses effectively on genuine defect regions while suppressing background interference, demonstrating improved discriminability.

## Table Legends

**Table 1.** Experimental environment configuration. Lists the operating system, CPU, GPU, RAM, Python, PyTorch, and CUDA versions used in all experiments.

**Table 2.** Training parameter configuration. Details the optimizer, initial learning rate (lr0), final learning rate factor (lrf), momentum, weight decay, epochs, batch size, image size, and warmup epochs.

**Table 3.** The comparison of different plans of branch convolution. Schemes A (standard + dilated), B (standard + depthwise separable), and C (standard + standard) are compared in terms of parameters, GFLOPs, mAP@0.5, and mAP@0.5:0.95.

**Table 4.** Comparison of different attention modules on the strong CBiFPN + DualScaleConv baseline. SimAM, MCAttn, ELA, and the proposed PfAAM are evaluated for mAP@0.5, mAP@0.5:0.95, Precision, Recall, parameters, and GFLOPs.

**Table 5.** Comparison of different methods. Includes YOLOv3-tiny, YOLOv5n, YOLOv6n, YOLOv8n, YOLOv10n, YOLOv11n, YOLOv12n, and our CDA-YOLO, reporting mAP@0.5, mAP@0.5:0.95, parameters, GFLOPs, Precision, and Recall.

**Table 6.** Ablation study results for different modules. Evaluates the individual and combined contributions of CBiFPN, DualScaleConv, and C3k2_PfAAM on the baseline YOLOv11 in terms of mAP@0.5, mAP@0.5:0.95, parameters, and GFLOPs.

## Author contributions

Anlong Yuan wrote the main manuscript text and prepared Figs. 1, 2, 3, 4, 5, 6, 7, 8 and 9. Jingmin Li was responsible for the overall direction and supervision of the paper. All authors reviewed the manuscript.

## Funding

No funding was received for this manuscript.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information