

PCB Defect Detection Algorithm Combining Cascaded-Wise Residual and Dynamic Sparse Attention

Weipeng Fan^{1,2}, Jinmin Peng^{1,2,*} and Song Lan^{3,*}

¹ Key Laboratory of Intelligent Processing Technology and Equipment in Fujian Province, Fujian University of Technology, Fuzhou, 350118, China

² College of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou, 350118, China

³ Fujian Polytechnic of Water Conservancy and Electric Power, Yong'an, 366000, China

INFORMATION

Keywords:

PCB
defect detection
CREC
minor defects
YOLOv7

DOI: 10.23967/j.rimni.2024.10.56504

Revista Internacional
Métodos numéricos
para cálculo y diseño en ingeniería

RIMNI



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

In cooperation with
CIMNE[®]

PCB Defect Detection Algorithm Combining Cascaded-Wise Residual and Dynamic Sparse Attention

Weipeng Fan^{1,2}, Jinmin Peng^{1,2,*} and Song Lan^{3,*}

¹Key Laboratory of Intelligent Processing Technology and Equipment in Fujian Province, Fujian University of Technology, Fuzhou, 350118, China

²College of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou, 350118, China

³Fujian Polytechnic of Water Conservancy and Electric Power, Yong'an, 366000, China

ABSTRACT

Detecting defects is crucial to ensuring the quality of printed circuit board (PCB) products. Due to the diminutive nature of surface defects on PCBs, current detection algorithms struggle to extract small defect features accurately, leading to a propensity for missed detections. To tackle these challenges, we propose a PCB defect detection algorithm that builds upon the YOLOv7 algorithm with enhancements. Firstly, we integrate the proposed CREC module into the backbone network to enhance the capture of local features about minor defects. Secondly, we propose the integration of a multi-scale feature fusion module, SPPB, within the head network to selectively activate channels or positions related to minor defects in the feature map, thereby enhancing the accuracy of local feature extraction for minor defects. Subsequently, the algorithm is endowed with higher efficiency in learning small defect features with the help of a new loss function, MPNWD. Finally, a small target detection layer P2 is added to enrich the contextual information in order to facilitate the algorithm to understand the relationship between small defects and their surrounding regions. Experimental results demonstrate the effectiveness of the enhanced YOLOv7 algorithm in testing the PCB_DATASET defect dataset, achieving a detection accuracy (mAP) of 99.3%, surpassing YOLOv7 by 5.4%, and outperforming other algorithms in terms of detection accuracy.

OPEN ACCESS

Received: 24/07/2024

Accepted: 19/11/2024

Published: 20/04/2025

DOI

10.23967/j.rimni.2024.10.56504

Keywords:

PCB
defect detection
CREC
minor defects
YOLOv7

1 Introduction

Printed circuit boards (PCBs) are essential components in electronic product manufacturing, and their quality and reliability greatly influence product performance and longevity [1]. Traditional methods for detecting surface defects in PCBs include manual visual inspections, electrical testing, and infrared scanning. However, these approaches often encounter issues such as missed detections, false alarms, and inefficiencies. The advancement of computer vision technology has led to the exploration

*Correspondence: Jinmin Peng, Song Lan (pjmpyt@163.com, lansong@126.com). This is an article distributed under the terms of the Creative Commons BY-NC-SA license

of various algorithms for PCB defect detection that enhance accuracy and efficiency while lowering labor costs. Currently, deep learning algorithms for PCB surface defect detection are primarily divided into one-stage and two-stage object detection frameworks based on their target localization methods. The two-stage algorithms include prominent models such as Fast-RCNN [2], Faster-RCNN [3], and DetectoRS [4]. For instance, to achieve near-real-time simultaneous detection of various damage types, Cha et al. proposed a visual inspection algorithm based on Faster-RCNN for PCB defect detection, which outperformed traditional CNN approaches [5]. The one-stage object detection algorithms encompass the YOLO (You Only Look Once) series, SSD [6], EfficientDet [7], anchor-based RetinaNet [8], CenterNet [9], and RepPoints [10] (which does not use anchor points). The YOLO series, being a one-stage detection algorithm, has been widely adopted across various applications and offers significant advantages over other methods in terms of accuracy, speed, and ease of deployment [11].

YOLOv7 [12], the seventh generation of the YOLO algorithm detector for all kinds of defect detection, has shown good performance. In PCB defect detection, most defects can be detected, but for the mouse bite, spur, and other minor defects in the PCB, only a few pixel points are detected, and the difference with the background is negligible. At present, based on the network detection of YOLOv7, it is still challenging to find out the cause, as follows:

(1) Deep networks struggle to capture localization and spatial relationships accurately. Minor defects such as mouse bites and spur in PCBs typically manifest in specific localized regions of images (e.g., in the interstices of two parallel arterial roads) and there are detection difficulties with small differences from the background. (2) Key feature information extraction for minor defects in PCBs is inadequate. Due to the small proportion of small defect images in PCBs (mouse bites and burrs only account for 3×3 pixels in 640×640 pixel PCBs), deep neural networks fail to effectively remove redundant background information during feature extraction, thus resulting in the network's low sensitivity to minor defect information. (3) Contextual information is difficult to retain effectively, causing the relationship between the tiny defect and the surrounding contextual information to become complicated and difficult for the algorithm to understand.

This article proposes a novel YOLOv7 network algorithm to tackle the challenges above. Initially, a backbone network structure was devised to resolve the challenge of deep networks inadequately capturing localization and spatial relationships, introducing a Cascaded Residual with an Efficient CSP (CREC) residual module featuring cascaded attention. Subsequently, a multi-scale feature fusion module SPPB (Spatial Pyramid Pooling BiFormer) is proposed to embed into the head network. Secondly, to tackle the challenge of incomplete extraction of crucial feature information regarding PCB minor defects, we propose the MPNWD loss function to weigh the significance of defects. Lastly, to combat the challenges of poor scale adaptability and complexity in context algorithms, we introduced the small object detection layer, denoted as P2. Fig. 1 illustrates the enhanced YOLOv7 network structure.

The remainder of the paper is structured as follows: Section 2 provides an overview of related research in PCB defect detection. Section 3 details the enhanced algorithm, highlighting four key improvements. Next, Section 4 outlines the experimental setup and offers a thorough analysis of the results obtained. Lastly, Section 5 summarizes the contributions, limitations, and future research directions of this study.

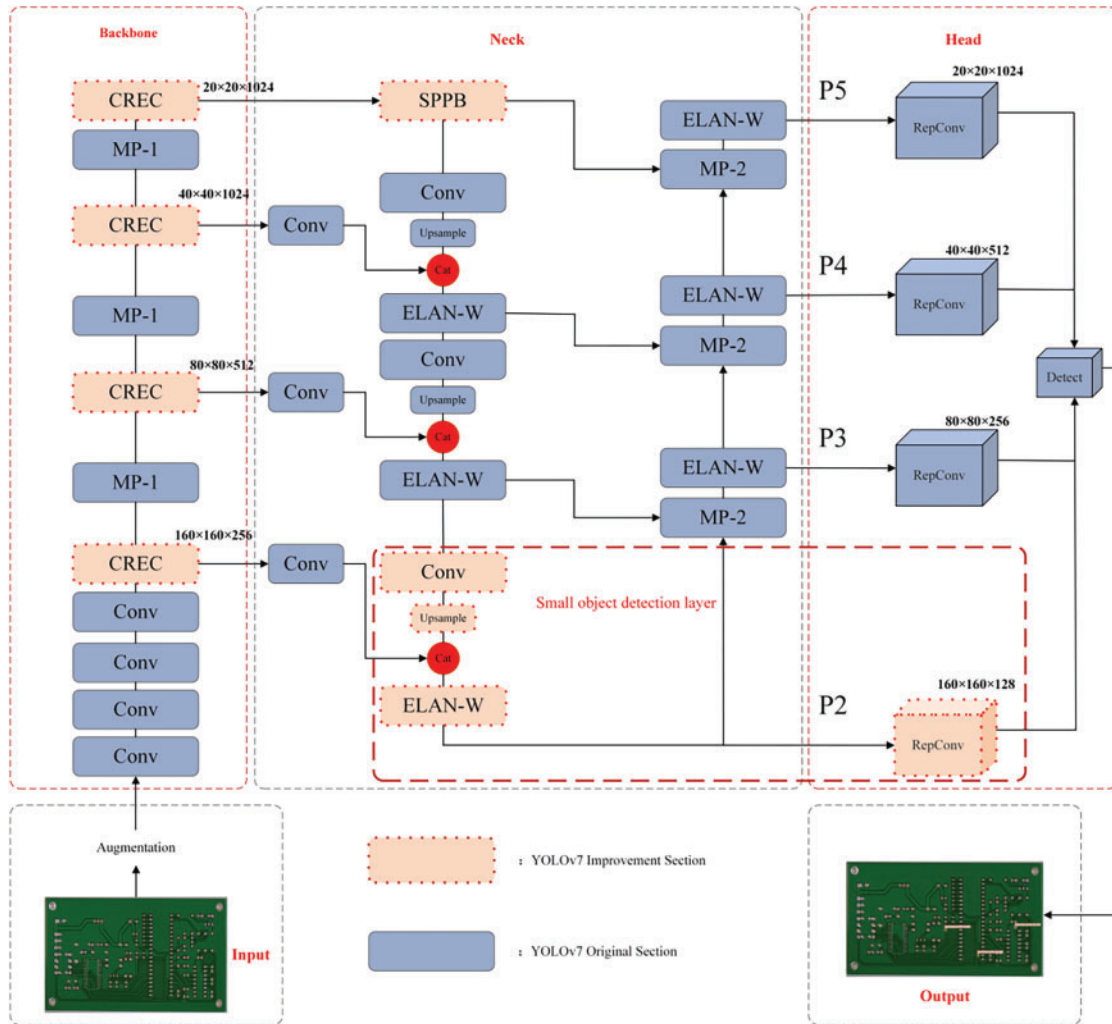


Figure 1: Improved YOLOv7 algorithm structure

2 Related Work

With the advancement of deep learning techniques for detecting PCB surface defects, numerous detection algorithms have been developed and implemented. Cha introduced a real-time crack segmentation algorithm known as SDDNet, which is widely used in detecting PCB losses [13]. Li et al. proposed innovative priority cloning technology and cost-sensitive Bayesian network classifiers [14] to address the imbalanced category distribution of training data and varying misclassification costs.

The YOLO algorithm series has become widely adopted in applications for detecting PCB defects. Liu et al. [15] selected YOLOv4 as their foundational algorithm and proposed a new loss function based on GIoU regression. While this adjustment improved target detection recall, the original YOLOv4 algorithm exhibited a redundant structure and high computational demands. Tang et al. [16] enhanced YOLOv5 by adding a small target detection layer, allowing the algorithm to concentrate on more information regarding small targets. They improved small target localization capabilities by introducing a depth-separable convolutional compression algorithm. This targeted detection layer focuses on small targets and optimizes the regression process of the predicted and detected frames

using the EIou loss function. In response to the challenges of weak spatial feature extraction and inadequate recognition of small targets in PCB defect detection with the original YOLOv7 algorithm, Yang et al. [17] incorporated the SwinV2_TDD module and the MFSA mechanism into YOLOv7. They replaced part of the ELAN (Efficient Layer Aggregation Network) module to enhance the extraction of local PCB features and avoid over-compression of the feature map. The mAP of the improved network reached 98.74%. Meanwhile, Chen et al. [18] employed FasterNet as the backbone network for YOLOv7, facilitating more efficient spatial feature extraction through a novel partial convolution. Meanwhile, the CBAM attention mechanism is added into YOLOv7, which strengthens the feature expression ability of the algorithm, and at the same time, the recognition accuracy of PCB tiny defects is greatly improved, and the mAP of this enhanced algorithm reaches 97.5%.

The current PCB defect detection methods utilizing the YOLOv7 algorithm have shown promising results. Nonetheless, there is a need to enhance its accuracy for detecting minor defects within complex backgrounds and subtle variations from the background. To tackle these issues, this paper introduces a novel PCB defect detection algorithm that combines cascaded residuals with dynamic sparse attention, enabling the detection of small defects against the intricate backgrounds typically found in PCBs.

3 Improving the YOLOv7 Network Algorithm

3.1 Improving the Backbone Network

The ELAN module within the YOLOv7 backbone network involves numerous convolutions that require extensive parameter calculations. However, it struggles to accurately capture localization and spatial relationships when extracting minor defects, making it challenging to differentiate between these defects and the background. The extraction of small defect features could be improved. Drawing inspiration from YOLOv8's c2f module [19], we developed a CSP module named CREC to replace the ELAN module in the backbone network. The CREC module utilizes cascaded residuals and cascade attention mechanisms to enhance local feature extraction, establish spatial correlations, and improve the detection of minor defects in PCBs. Fig. 2 illustrates the CREC module.

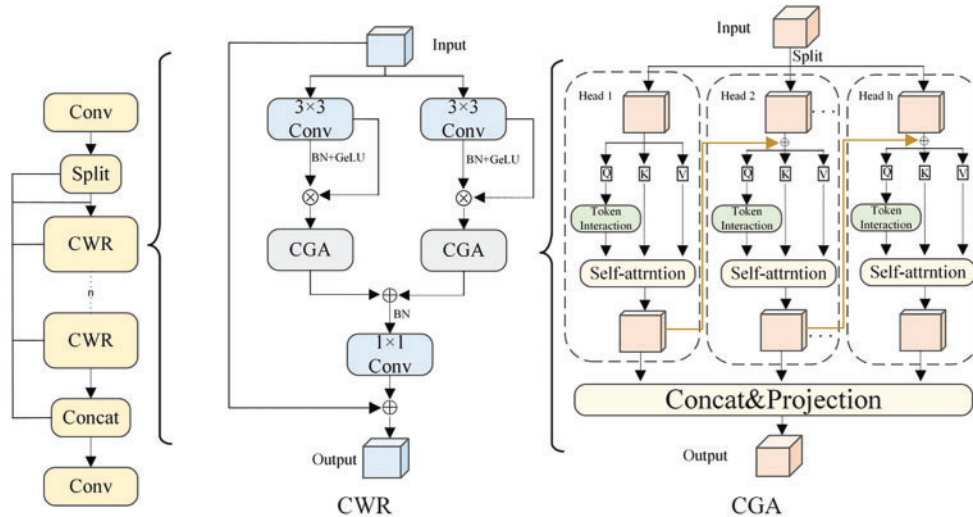


Figure 2: CREC module

The CREC module employs the widely recognized CSP method of slicing the input tensor $X \in \mathbb{R}^{n \times c \times h \times w}$ to splice it with the original tensor. This operation preserves the rich gradient information.

The procedure is described in Eq. (1).

$$\begin{aligned} Y_1 &= CWR_n(X_1), \\ Y_2 &= CWR_n(X_2) + \dots + CWR_2(X_2) + CWR_1(X_2) + X_2, \\ Y &= [Concat(Y_1, Y_2)] + X_2 \end{aligned} \quad (1)$$

$CWR_n(\cdot)$ represents the n th CWR module, $Concat(\cdot)$ features that connect two branches, $[\cdot]$ represents 1×1 convolution operation, $+$: residual connection.

The residual structure not only preserves more detailed information of PCB defect features but also significantly contributes to reducing the algorithm's computational complexity.

The CWR residual module obtains more complete defect characterization information in the following two steps, followed by the fusion of feature maps generated from the multi-scale receptive domain.

Step 1: This step involves generating relevant residual features from the input features, referred to as regional residualization. By utilizing the region residuals, the image is precisely segmented and PCB defect feature extraction is performed in each segmented region. This method allows tiny defects to be recognized more accurately.

Step 2: Applying the cascading attention mechanism to feature regions with different scale information and applying the feeling field to a single channel reduces the inflow of unnecessary tiny defect feature information and captures more spatial relationships.

Finally, the feature information from these two steps is fused with the original feature information by point-by-point convolution.

CGA(Cascaded Group Attention) [20] level group attention is expressed as:

$$\begin{aligned} \tilde{X}_{ij} &= Attn(X_{ij}W_{ij}^Q, X_{ij}W_{ij}^K, X_{ij}W_{ij}^V), \\ \tilde{X}_{i+1} &= Concat\left[\tilde{X}_{ij}\right]_{j=1:h} W_i^P \end{aligned} \quad (2)$$

The j th head calculates self attention on X_{ij} , which is the j th segmentation of input feature X_i , i.e., $X_i = [X_{i1}, X_{i2}, \dots, X_{ih}]$, and $1 \leq j \leq h$. h is the total number of Heads $W_{ij}^Q, W_{ij}^K, W_{ij}^V$ and k are projection layers that map the input features into different subspaces, while W_i^P is a linear layer that projects the connected output features back to the same dimension as the input.

Integrating the CREC module into the YOLOv7 backbone network improves the accuracy of localization and spatial relationships, thereby highlighting the differences between minor PCB defects and their backgrounds.

3.2 Spatial Pyramid Pooling Structure SPPB

Multiple convolutional layer networks in SPPCSPC within YOLOv7 exhibit slow convergence, employ small pooling kernels, and possess large target receptive fields, thereby hindering the selective activation of channels or positions corresponding to minor PCB defects within the feature map. Algorithms, the spatial relationship of minor defects poses a challenge, leading to inaccurate extraction of features about minor PCB defects. To address these challenges, we introduce a novel spatial pyramid pooling structure, SPPB, inspired by SimCSPSPF in YOLOv6 v3.0 [21]. The SPPB structure efficiently integrates multi-scale information from feature maps of different sizes by applying pooling operations on feature maps of various scales. It facilitates cross-channel feature fusion among feature maps of

different scales to comprehend minor defects in images from diverse levels and perspectives, thus enhancing the algorithm's capacity for detecting minor defects, as illustrated in Fig. 3. This module is compared with the SPPCSPC module.

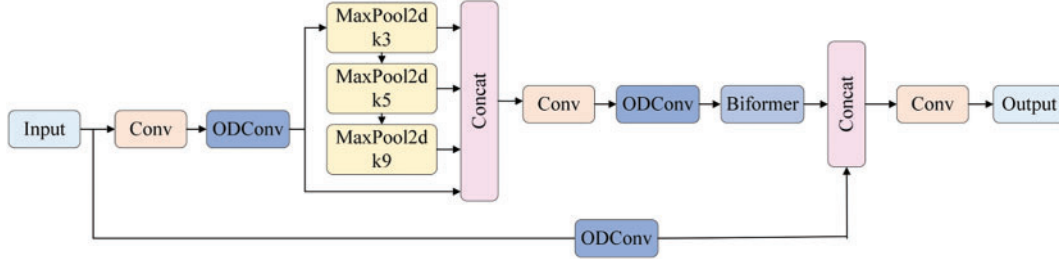


Figure 3: Spatial pyramid pooling structure SPPB

(1) By replacing 3×3 convolutions with multi-dimensional dynamic convolution in ODConv and eliminating the convolution layer prior to the pooling layer, we simplify the architecture to utilize only a single convolution layer along with the ODConv layer. This approach reduces computational complexity and parameter size, thereby accelerating the network's convergence rate. ODConv complementary attention mechanisms in four dimensions in molding space [22]. By progressively multiplying convolutions across positional, channel, filter, and kernel dimensions—each with its own attention levels—the convolution operation captures variations across different input dimensions, improving the ability to extract rich contextual information. The output of ODConv can be represented by the following formula:

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x \quad (3)$$

In the formula, $\alpha_{wi}, \alpha_{fi}, \alpha_{ci}, \alpha_{si}$ ($i = 1, 2, \dots, n$) represents four types of attention, and the convolutional kernel W_i ($i = 1, 2, \dots, n$) is variable.

(2) Pooling feature maps of various scales individually and then cascading the results produces a fixed-length feature vector. Information sharing during pooling reduces the required number of pooling operations, resulting in faster computation. Simultaneously, the pooling kernel (3, 5, 9) is replaced with the original (5, 9, 13) to align the receptive field of the expanded pooling kernel with the micro-scale. Enhancing the feature extraction capability of minor defects in PCBs improves the accuracy of small defect detection.

(3) The study introduces the BiFormer [23] attention mechanism. BiFormer, a Transformer [24] algorithm variant, incorporates a dynamic sparse attention mechanism into the original Transformer architecture. This mechanism selectively activates channels or positions in the feature map, enabling spatial relationships among minor defects in PCBs to be analyzed. Consequently, this enhances the algorithm's comprehension of the contextual relationship between minor defects and their surroundings, thus improving detection accuracy. BiFormer achieves dynamic and query-adaptive sparsity via a Bi-Level Routing Attention (BRA) mechanism, depicted in Fig. 4.

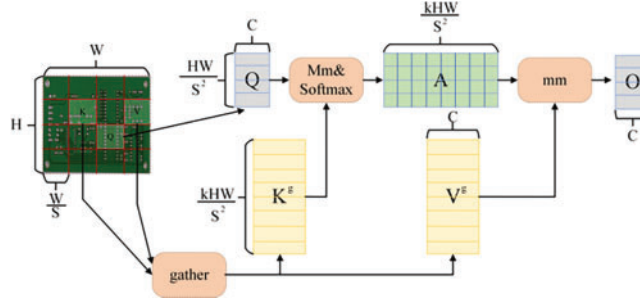


Figure 4: BRA structure diagram

Its mathematical expression is shown in Eq. (4).

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \quad (4)$$

Among them, O is a tensor, and Q, K, V are the eigenvectors $X \in R^{H \times W \times C}$ corresponding to the input feature map $X^r \in R^{S^2 \times \frac{HW}{S^2} \times C}$ obtained through linear transformation; K^g, V^g is generated by collecting all K and V in the critical area k . $\text{LCE}(V)$ is an additional item for enhancing local context.

A new hierarchical visual converter, BiFormer, was constructed using BRA as the core. BiFormer employs a pyramid structure of the fourth order, illustrated in Fig. 5.

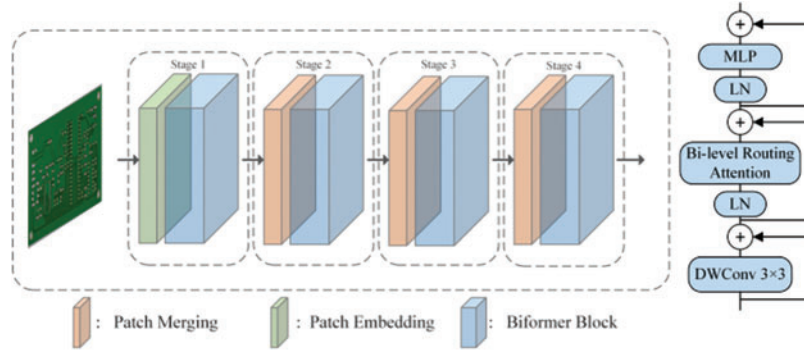


Figure 5: Fourth order pyramid structure

Initially, BiFormer utilizes overlapping block embedding, which is followed by a block merging module in the subsequent stages (from the second to the fourth). This approach aims to reduce the input spatial resolution while increasing channel capacity. Afterward, a series of continuous BiFormer blocks are used for feature transformation.

3.3 Optimization of Loss Function

To give greater emphasis to the small defect areas of PCBs and help the algorithm better focus on learning the features of minor defects, it is proposed to use a proportional combination of the NWD (Normalized Wasserstein Distance) [25] loss function and MPDIoU [26]. This combination aims to derive the CIoU loss function for the MPNWD-optimized YOLOv7, ultimately enhancing the accuracy of PCB defect detection. The specific design process for MPNWD is outlined as follows:

For the loss function, there are real bounding boxes B_{gt} and predicted bounding boxes B_{prd} , $B_{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt})$, $B_{prd} = (x_1^{prd}, y_1^{prd}, x_2^{prd}, y_2^{prd})$, and the inputted image width w and height h have

IoU calculation expressions as shown in Eq. (5).

$$IoU = \frac{B_{gt} \cap B_{prd}}{B_{gt} \cup B_{prd}} \quad (5)$$

The CIoU loss function in YOLOv7 considers three essential factors: the overlap area, the distance between center points, and the aspect ratio. However, it neglects the center position, which restricts the algorithm's ability to accurately capture the bounding boxes of small targets and, as a result, lowers detection accuracy. To tackle this limitation, MPDIoU has been introduced. Its expression is as follows:

$$L_{MPDIoU} = 1 - IoU + \frac{d_1^2}{h^2 + w^2} + \frac{d_2^2}{h^2 + w^2} \quad (6)$$

$$\begin{aligned} d_1^2 &= (x_1^{prd} - x_1^{gt})^2 + (y_1^{prd} - y_1^{gt})^2, \\ d_2^2 &= (x_2^{prd} - x_2^{gt})^2 + (y_2^{prd} - y_2^{gt})^2 \end{aligned} \quad (7)$$

The MPDIoU loss function incorporates the width, height, and center position of the bounding box during regression. For small targets, it mitigates the difficulties in distinguishing defects from the background by employing multi-scale regression, which reduces the impact of small bounding box sizes on detection accuracy.

However, when using the MPDIoU loss function during backpropagation, a lack of overlap between the predicted and actual bounding boxes significantly hampers the algorithm's ability to update its parameters effectively. Determining the similarity between predicted and actual bounding boxes that encompass one another poses certain challenges. To address this issue, the NWD loss function is introduced.

NWD evaluates the bounding box using a two-dimensional Gaussian distribution, measuring the similarity between the predicted and actual boxes through their respective Gaussian distributions. The calculation is defined by Eqs. (8) and (9).

$$W_2^2(N_{B_{prd}}, N_{B_{gt}}) = \left\| \left(\left[cx^{prd}, cy^{prd}, \frac{w^{prd}}{2}, \frac{h^{prd}}{2} \right]^T, \left[cx^{gt}, cy^{gt}, \frac{w^{gt}}{2}, \frac{h^{gt}}{2} \right]^T \right) \right\|_2^2 \quad (8)$$

$$NWD(N_{B_{prd}}, N_{B_{gt}}) = \exp \left(-\frac{\sqrt{W_2^2(N_{B_{prd}}, N_{B_{gt}})}}{C} \right) \quad (9)$$

Among them, C is a constant, and $W_2^2(N_{B_{prd}}, N_{B_{gt}})$ is the second-order Wasserstein distance of the two-dimensional Gaussian distribution $N_{B_{prd}}$ and $N_{B_{gt}}$ between the predicted box and the true box. (cx^{prd}, cy^{prd}) and (cx^{gt}, cy^{gt}) are the center coordinates of the predicted box and the true box, respectively. (w^{prd}, h^{prd}) is the width and height of the predicted box, and (w^{gt}, h^{gt}) is the width and height of the actual box.

The NWD loss function can be matched according to the probability to solve the problem of low detection accuracy caused by the small difference between PCB tiny defects and the background.

The new loss function, MPNWD, is created by proportionally combining MPDIoU and NWD, as demonstrated in Eq. (10). These enhancements to the loss function enable the algorithm to concentrate more on learning the characteristics of minor defects, improving its ability to differentiate between defects and the background, and consequently increasing the detection accuracy of minor PCB defects.

$$L_{MPNWD} = \alpha L_{MPDIoU} + (1 - \alpha) NWD(N_{B_{prd}}, N_{B_{gt}}) \quad (10)$$

where α is the proportional coefficient, $0 \leq \alpha \leq 1$.

3.4 Add Small Object Detection Layer P2

The scale span of defects in the PCB defect dataset is large, and the proportion of minor defects is exceptionally high. In the YOLOv7 network, when inputting images with a size of 640×640 , the P3, P4, and P5 detection layers can only generate feature maps of 80×80 , 40×40 , and 20×20 . However, minor PCB defects with only a few pixels of “mouse bites and burrs” cannot be detected in the P3 detection layer of 80×80 . Consequently, a specialized P2 detection layer for small objects is introduced to generate a 160×160 feature map. As shown in Fig. 6, the 160×160 feature layer detects minor defects such as mouse bites and burrs.

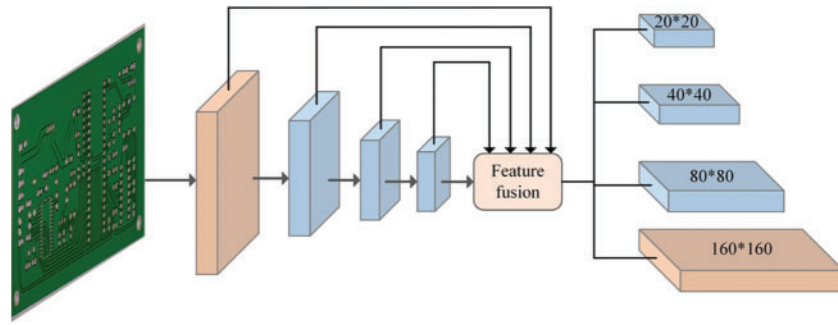


Figure 6: Increase small object detection layer

In contrast, the 80×80 and 40×40 feature layers detect medium-sized defects such as short circuits, copper impurities, and open circuits. The 20×20 feature layer is dedicated to detecting more significant defects such as missing holes. Including an additional detection layer fulfills the need to identify defects across different scales precisely. It enhances contextual understanding, aiding in discerning the relationship between minor defects and their surroundings, thus elevating the accuracy of small defect detection in PCBs.

4 Experiment and Analysis

4.1 Experimental Detail Settings

Tables 1 and 2 show the experimental details of this study. Our experiments did not use any pre-trained weights.

Table 1: Experimental environment

Configure	Setting
CPU	i7-10700F 2.90 GHz
GPU	NVIDIA GeForce RTX 3050
Accelerated computing framework	CUDA 11.7
Deployment environment	Python 3.10.11
Optimizer	SGD
Operating systems	Windows 10
Deep learning framework	PyTorch 2.0.0

Table 2: Configuration of experimental parameter

Parameters	Setting
Input image size	640 × 640
Epoch	300
Parameter learning rate	0.001(First 200epoch), 0.0001(Post 100epoch)
Batch size	8

4.2 PCB Defect Image Dataset

The defective dataset utilized in this study comes from the Open Laboratory of Intelligent Robotics at Peking University [27]. This dataset consists of 693 defect images, categorized into six classes: missing holes, spurious copper, short circuits, mouse bites, open circuits, and spurs. Due to the limited sample size of the original dataset, various data augmentation techniques were employed to enhance it. These techniques include blurring, brightness adjustment, random cropping, rotation, translation, and mirroring. Blurring reduces the detail in the images, creating variations that differ from the original yet remain highly correlated, thereby increasing the diversity of the training set and aiding the algorithm in learning richer features. Brightness adjustment alters the image's brightness, simulating variations caused by different environmental lighting conditions, which improves the algorithm's adaptability to varying brightness levels. Random cropping extracts regions of random size and position from the images, allowing the algorithm to better adapt to transformations. Image rotation adjusts the image by a specified angle around its center, enabling the algorithm to identify defects from various perspectives. Translation shifts the image, while mirroring flips it either horizontally or vertically. After applying these augmentation techniques, the dataset was expanded to 9009 images, which were randomly divided into training and testing sets in an 8:2 ratio. The distribution of the expanded defect categories is presented in Table 3.

Table 3: Dataset details

Defect type	Example of defects	Number of original images	Number of images after expansion
Missing_hole		115	1495
Spurious_copper		116	1508
Short		116	1508
Mouse_bite		115	1495
Open_circuit		116	1508
Spur		115	1495
Total number	—	693	9009

4.3 Performance Evaluation and Indicators

We use precision, recall, and mean precision(mAP) as evaluation indicators for PCB defect detection effectiveness, as shown in Eqs. (11) and (12).

$$P = \frac{T_p}{T_p + F_p} \times 100\%,$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (11)$$

$$AP = \int_0^1 P(R) dR,$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (12)$$

where T_p denotes the count of correctly detected defects, F_p indicates the count of incorrectly detected defects, and F_N signifies the count of undetected defects. n represents the total number of predicted bounding box categories.

4.4 Optimization Strategy Ablation Experiment

In order to evaluate the impact of each improved module on the PCB defect detection algorithm, ablation experiments were conducted based on the original YOLOv7. The algorithm obtained by replacing SPPCSPC with SPPB in the YOLOv7 network is referred to as YOLOv7-1. The inclusion of the MPNWD loss function in the YOLOv7-1 configuration results in the YOLOv7-2 algorithm. Replacing the ELAN module with the CREC module in YOLOv7-2 leads to the YOLOv7-3 algorithm. Lastly, the algorithm incorporating a P2 small object detection layer into YOLOv7-3 represents the proposed method in this paper.

4.4.1 PCB Defect Detection Algorithm Ablation Experiment

We conducted ablation experiments on the algorithm in this article, and the results are shown in Table 4.

Table 4: Ablation experiment

Algorithm	Augmentation	SPPB	MPNWD	CREC	P2	$P/\%$	$R/\%$	$mAP/\%$
YOLOv7	✓					96.2	85.4	93.9
YOLOv7-1	✓	✓				97.6	86.8	95.5
YOLOv7-2	✓	✓	✓			98.8	92.2	97.3
YOLOv7-3	✓	✓	✓	✓		99.6	91.9	98.7
OURS	✓	✓	✓	✓	✓	99.9	93.9	99.3

Note: ✓ represents adding the module.

The comparison between YOLOv7 and YOLOv7-1 clearly shows that integrating the SPPB module results in a 1.6% increase in accuracy and a 1.4% rise in recall rate, alongside a notable improvement in mean Average Precision (mAP). This suggests that the algorithm can effectively learn specific local features associated with minor defects. When evaluating YOLOv7-1 against YOLOv7-2, it becomes apparent that the MPNWD boundary box regression loss function significantly boosts the network's recognition accuracy, achieving an enhancement of 1.8%. Although the increase in mAP is modest, the issues related to subtle differences between backgrounds and defects have been somewhat alleviated. A further comparison between YOLOv7-2 and YOLOv7-3 indicates that substituting the ELAN module with the CREC module for feature extraction yields a 0.8% improvement in accuracy. Despite a slight decrease in recall rate by 0.3%, likely due to a reduction in parameter count, the mAP saw an increase of 1.5%, demonstrating the CREC module's superior ability to capture localization and spatial relationships. Lastly, the comparison of YOLOv7-3 with the proposed algorithm highlights that the addition of the P2 detection layer boosts the mAP to 99.3% and significantly enhances the recall rate by 2%. This reflects a marked improvement in detecting minor defects, such as mouse bites and spurs.

4.4.2 Algorithm P-R Curve Analysis

We analyze the detection performance of the algorithm through the P-R curve. The extent of fluctuation in the P-R curve illustrates the algorithm's stability when handling various data samples; smaller fluctuations indicate greater stability. Fig. 7 presents the P-R curves for six types of defects, derived from testing the dataset with different algorithms. Notably, the P-R curve for each type of PCB defect in this algorithm fully encompasses the curves of the other algorithms, demonstrating that its detection accuracy is superior. When comparing the P-R curves from the YOLOv7 algorithm, the YOLOv7+SPPB network shows larger fluctuations in the P-R curve for "short circuit" and "spurs" defect detection tasks. Conversely, the YOLOv7 algorithm exhibits more substantial fluctuations in the P-R curve for "mouse bite" defect detection. By replacing the ELAN module with the CREC module and adding the small object detection layer P2, YOLOv7 addresses the issues associated with significant P-R curve fluctuations and limited curve envelope area that arise when employing a single optimization strategy.

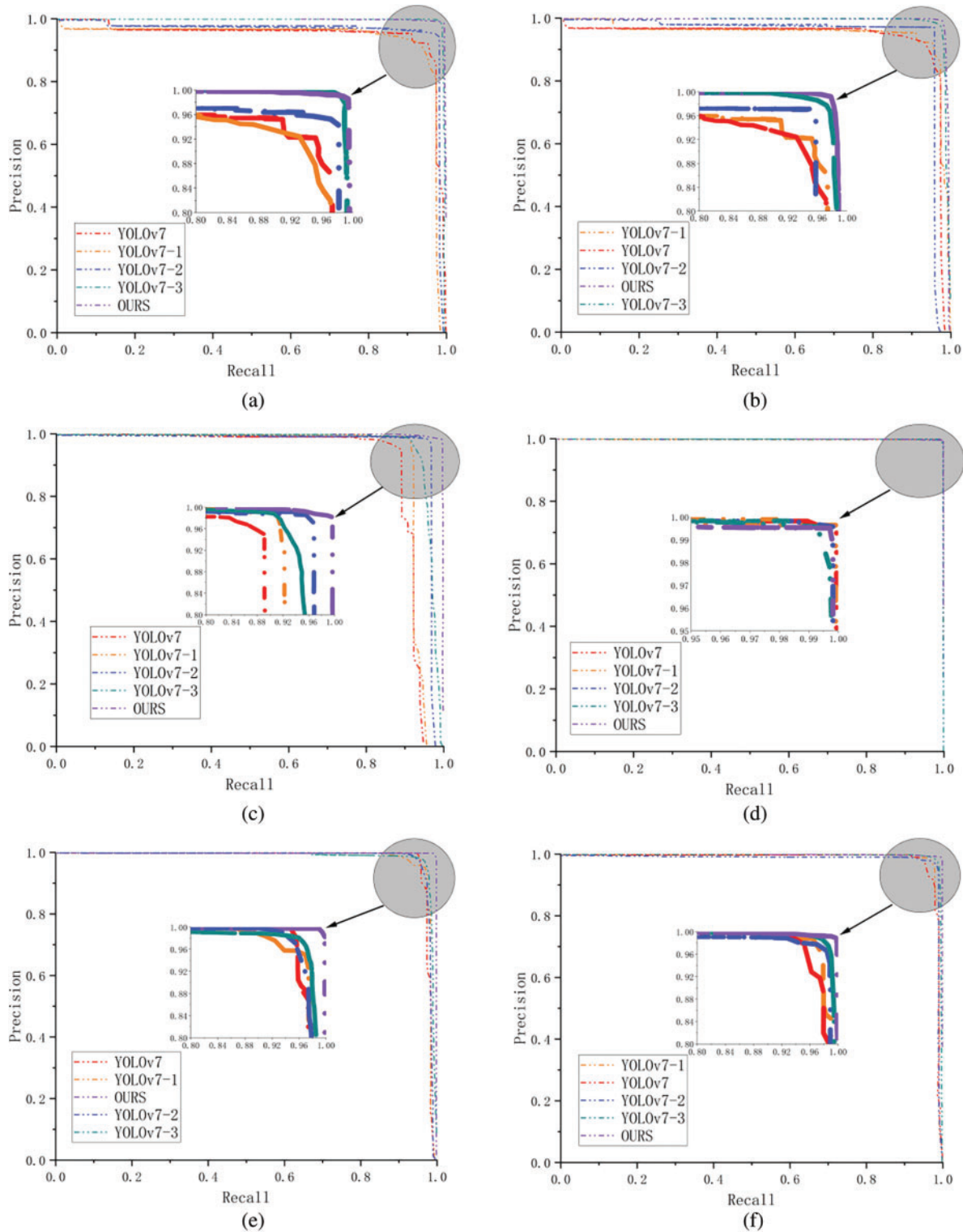


Figure 7: P-R curves of different defect types. (a) P-R curve of short defect; (b) P-R curve of open circuit defect; (c) P-R curve of spur defects; (d) P-R curve of defect with missing hole; (e) P-R curve of mouse bite defect; (f) P-R curve of spurious copper defects

4.4.3 Performance Analysis of CREC Structure

We carried out comparative experiments involving the original YOLOv7 network, the YOLOv7 network featuring two configurations of the CREC module, and the YOLOv7 network with four different types of CREC structures. The results of these experiments are presented in [Table 5](#).

Table 5: CREC performance analysis

Experiment	$R/\%$	$P/\%$	$mAP/\%$
YOLOv7	85.45	96.22	93.96
YOLOv7 + 2CREC	89.36	98.10	97.36
YOLOv7 + 4CREC	90.9	99.21	97.82

[Table 5](#) demonstrates the effectiveness of the proposed CREC algorithm. (1) The results indicate that in the original YOLOv7 network, replacing two ELAN modules with CREC led to an increase of 1.88% in the P-value, 3.91% in the R-value, and 3.40% in the mAP value. This suggests that incorporating the CREC structure into YOLOv7 enhances the detection accuracy of PCB defects. (2) Furthermore, when comparing the network with four ELAN modules to the one with just two ELAN modules, the YOLOv7 network showed an increase of 1.11% in the p -value, 1.54% in the R-value, and 0.46% in the mAP value. These findings confirm the high detection accuracy of the CREC module within the YOLOv7 framework.

4.4.4 MPNWD Proportional Coefficient Experiment

The experiment uses YOLOv7 as the baseline model. The proportion of NWD in MPNWD is determined to achieve the best detection of the algorithm. The experimental results are shown in [Table 6](#).

Table 6: MPNWD proportional coefficient experiment

Algorithm	NWD proportion/%	$P/\%$	$R/\%$	$mAP/\%$
YOLOv7 + MPNWD	0	96.2	85.4	93.9
	50	97.2	88.2	94.8
	60	97.2	88.5	95.1
	70	97.5	89.1	95.6
	80	97.8	90.6	95.8
	90	98.1	91.2	96.1
	95	97.6	90.4	95.4
	100	97.5	89.4	94.9

The results in [Table 6](#) show that the accuracy, recall, and precision of the algorithm all improve with the increase of NWD proportion. Specifically, the YOLOv7 + MPNWD algorithm has the highest accuracy, accuracy, and recall at a proportion of 90%, with a maximum accuracy of 96.1%, a maximum recall of 91.2%, and a maximum accuracy of 98.1%. However, when the proportion of NWD exceeds 90%, the algorithm's accuracy declines, and the recall and accuracy decrease accordingly.

Therefore, experiments have shown that the optimal proportion of NWD is 90%, as the target of “missing holes” in PCBs is more significant than other defects. The algorithm performance will decrease if the proportion of NWD exceeds 90%.

4.4.5 Comparative Experiment on Attention Mechanisms

We compared the attention mechanisms of BiFormer with those of SE, CA, and CBAM, emphasizing the advantages of our approach. The results of these experiments are presented in [Table 7](#).

Table 7: Experimental results of different attention mechanisms

Algorithm	$R/\%$	$P/\%$	$mAP/\%$
YOLOv7-SPPB-SE	86.2	94.2	91.5
YOLOv7-SPPB-CA	86.3	96.8	94.2
YOLOv7-SPPB-CBAM	86.4	97.1	94.3
YOLOv7-SPPB	86.8	97.6	95.5

The BiFormer mechanism outperforms other attention mechanisms in terms of recall, accuracy, and precision. The BiFormer attention mechanism utilized in this study enhances the algorithm’s ability to comprehend the contextual relationships between minor defects and their surrounding areas. This effectively emphasizes the critical features of the target area, allowing the algorithm to concentrate more on the PCB defect regions while reducing training losses. In contrast, other attention mechanisms tend to focus on some background targets even when highlighting PCB defects.

4.4.6 Different Types of Defect Detection Experiments

To verify the detection effect of this paper’s algorithm on different PCB defects, we carried out comparison experiments between YOLOv7 and this paper’s algorithm. The experimental results are shown in [Table 8](#).

Table 8: Results of different types of defects

Class	YOLOv7			OURS		
	$P/\%$	$R/\%$	$mAP/\%$	$P/\%$	$R/\%$	$mAP/\%$
Missing hole	99.8	99.6	99.5	100	99.9	99.5
Spurious copper	99.3	90.6	96.2	100	92.9	99.5
Short	97.9	85.9	94.8	99.5	93.0	99.5
Mouse bite	91.7	76.2	89.6	99.2	91.1	99.1
Open circuit	97.5	83.8	94.5	99.9	95.1	99.5
Spur	90.7	75.6	88.6	99.9	91.3	98.7
Average	96.2	85.4	93.9	99.9	93.9	99.3

As can be seen from [Table 8](#), compared with the YOLOv7 algorithm, this paper’s algorithm has improved the detection accuracy on each kind of defect, especially on Mouse bite and Spur defects,

which have the most obvious improvement, with 9.5% and 10.1% improvement in detection accuracy, respectively.

4.5 Comparison of the Effectiveness of Commonly Used Algorithms

To verify whether the algorithm proposed in this article has good performance, it was compared with YOLOv3 [28], YOLOv5, Faster-RCNN, YOLOv8, DETR [29], RT-DETR [30], Gold-YOLO [31], YOLOv9 [32], and YOLOv10 [33] in the same experimental environment. mAP , P , and R were used as evaluation indicators for algorithm performance. The experimental results are shown in Table 9.

Table 9: Comparison table of results for different algorithms

Algorithm	$R/\%$	$P/\%$	$mAP/\%$
YOLOv3	86.2	94.2	91.5
YOLOv5	86.3	96.8	93.2
Faster-RCNN	86.8	97.6	95.5
YOLOv8	87.5	97.5	94.8
DETR	89.5	96.2	93.8
RT-DETR	91.2	98.6	96.2
Gold-YOLO	90.8	97.8	95.2
YOLOv9	87.2	97.6	94.3
YOLOv10	90.6	98.1	95.3
OURS	93.9	99.9	99.3

An analysis of the data presented in Table 9 reveals that our algorithm outperforms others in terms of mAP , accuracy, and recall. The mAP of our method reaches 99.3%, surpassing the other algorithms by at least 3.1%. As illustrated in the comparison shown in Fig. 8, when detecting mouse bite defects with minimal manifestations, our algorithm successfully identifies all three defects. It achieves high confidence levels and accurately pinpoints the locations of these defects. In contrast, other algorithms demonstrate subpar performance, with lower detection rates, accuracy, and confidence in identifying the defects.

In addition, the algorithm proposed in this paper is compared with TDD-net [34], YOLOv8-F [35], YOLOv7-Tiny [36], YOLO-HMC [37], MSD-YOLO [38], YOLOv7-S [17] for the PCB DATA SET dataset. Variations in detection accuracy across these algorithms can be attributed to differences in factors such as hardware environment, training dataset, training parameters, and training methodologies. As presented in Table 10, our algorithm outperforms TDD-net by 0.40% in mAP , YOLOv8-F by 2.70%, YOLOv7-Tiny by 0.70%, and YOLOv7-S by 0.56%, while achieving performance comparable to MSD-YOLOv5. In terms of parameter count and computational complexity, the proposed method reduces these aspects by 20% and 32%, respectively, when compared to MSD-YOLOv5. This indicates that the proposed algorithm not only matches the detection performance of MSD-YOLOv5 but also achieves a substantial reduction in network complexity. Overall, the analysis demonstrates that, in the context of small defect detection in PCBs, the enhanced YOLOv7 algorithm outperforms other detection algorithms in terms of accuracy and is better suited to the requirements of this specific task.

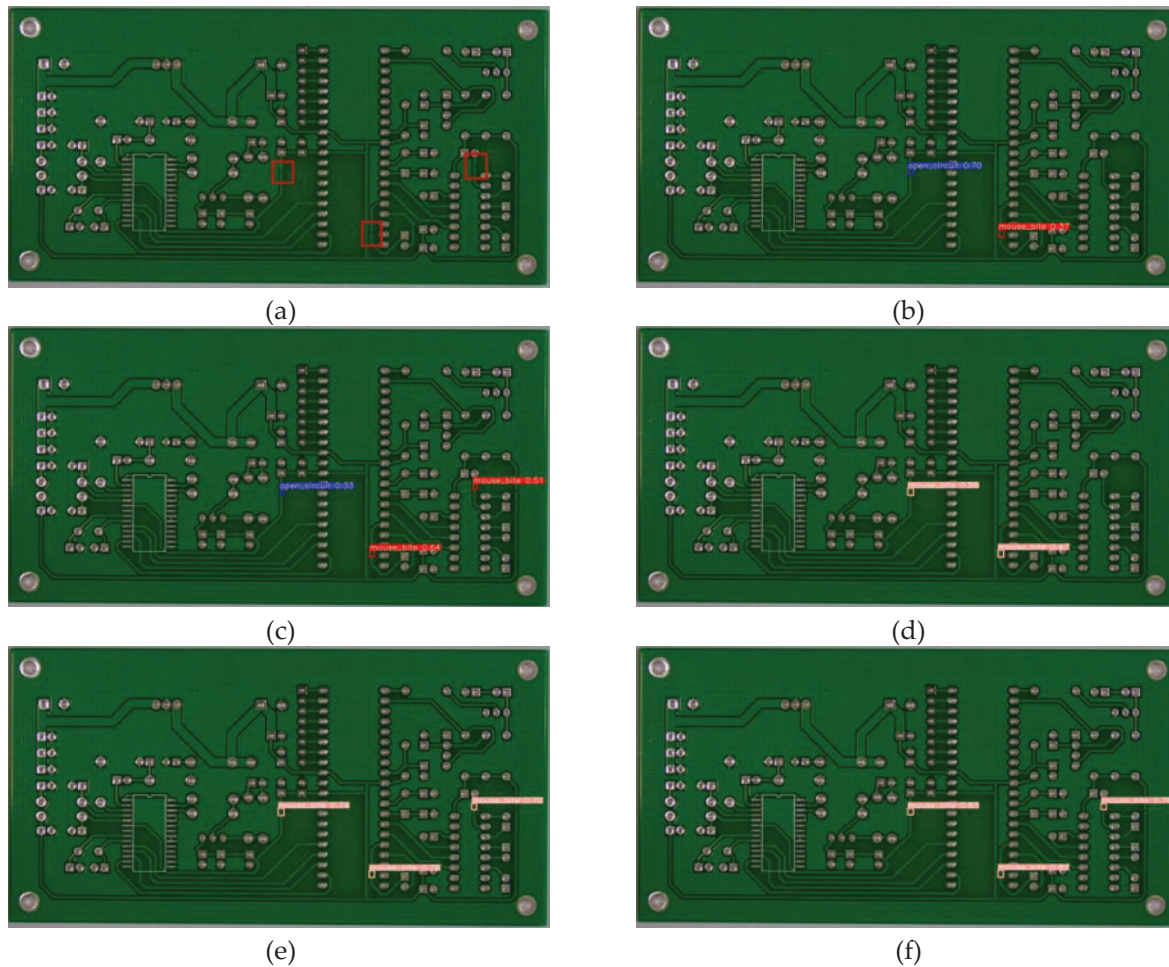


Figure 8: Comparison of different algorithms for detecting mouse bite defects. (a) Original; (b) YOLOv3; (c) YOLOv5; (d) Faster-RCNN; (e) YOLOv8; (f) OURS

Table 10: The detection accuracy results of different PCB defect detection algorithms

Algorithm	Network Structure	<i>mAP</i> /%
TDD-net (2019)	Faster-RCNN + feature pyramid network	98.90
YOLOv8-F (2023)	YOLOv8 + CFM feature fusion network	96.60
YOLOv7-Tiny (2023)	YOLOv7-Tiny + SE + PRelu activation function + REPVGG	98.60
YOLO-HMC (2024)	YOLOv5 + horNet baykbone + MCBAM + CARAFE content reorganization	98.60
MSD-YOLO (2022)	YOLOv5 + mobileNet-v3 + SE	99.30
YOLOv7-S (2023)	YOLOv7 + swinV2_TDD + MFSA	98.74
OURS	YOLOv7 + CREC + SPPB + MPNWD + P2	99.30

5 Conclusion

In this paper, an algorithm for detection of tiny defects in circuit boards is proposed. First new loss function MPNWD can be used in datasets with multiple small targets. The multi-scale feature fusion module, SPPB, replaces the original SPPCSPC module, and this tuning selectively activates specific channels or locations in the feature map, in addition, the CREC module replaces the ELAN module for more effective advanced feature fusion. Finally, a small target detection layer P2 is added at the end of the model to improve the detection accuracy. The improved algorithm achieves a detection accuracy of 99.9%, with a mean Average Precision (mAP) of 99.3%, marking an improvement of 3.7 and 5.4 percentage points, respectively, over the original model. Although this enhancement significantly improves PCB defect detection accuracy, further studies are required to optimize parameter efficiency and detection speed. Future developments will aim to preserve high detection accuracy while reducing model weight, and plans are in place to deploy the algorithm on factory edge equipment to accelerate detection speed.

Acknowledgement: None.

Funding Statement: This work was supported by the Fuxiaquan National Independent Innovation Demonstration Zone High End Flexible Intelligent Packaging Equipment Collaborative Innovation Platform Project (2023-P-006); 2022 Fujian Provincial Key Project for Science and Technology Innovation (2022G02007). The funders had no role in the study design; in the collection, analysis or interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Author Contributions: Weipeng Fan drafted the manuscript and developed the methodology, conducted the investigation, designed and trained the models, and edited the manuscript. Jinmin Peng contributed to the formal analysis and conceptualization, project administration. Song Lan contributed to data collection and processing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All data that support the findings of this study are included within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Tian X, Zhao L, Dong H. Application of image processing in the detection of printed circuit board. In: 2014 IEEE Workshop on Electronics, 2014; Ottawa, ON, Canada; p. 157–9.
2. Girshick R. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision, 2015; Santiago, Chile; p. 1440–8.
3. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2016;39(6):1137–49. doi:10.1109/TPAMI.2016.2577031.
4. Qiao S, Chen L, Yuille A. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution. In: 2021 IEEE/CVF Conference Computer Vision and Pattern Recognition, 2021; Nashville, TN, USA; p. 10213–24.

5. Cha Y, Choi W, Suh G, Mahmoudkhani S, Büyükoztürk O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Comput Aided Civ Infrastruct Eng.* 2018;33(9):731–47. doi:10.1111/mice.12334.
6. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: single shot multibox detector. In: *Computer Vision-ECCV 2016: 14th European Conference, 2016 Oct 11–14; Amsterdam, The Netherlands*; p. 21–37.
7. Tan M, Pang R, Le Q. EfficientDet: scalable and efficient object detection. In: *2020 IEEE/CVF Conference Computer Vision Pattern Recognition, 2020; Seattle, WA, USA*; p. 10781–90.
8. Zhang S, Chi C, Yao Y, Lei Z, Li S. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: *2020 IEEE/CVF Conference Computer Vision Pattern Recognition, 2020; Seattle, WA, USA*; p. 9759–68.
9. Duan K, Bai S, Xie L, Qi H, Huang Q, Tian Q. CenterNet: keypoint triplets for object detection. In: *2019 IEEE/CVF International Conference Computer Vision, 2019; Seoul, Republic of Korea*; p. 6569–78.
10. Yang Z, Liu S, Hu H, Wang L, Lin S. RepPoints: point set representation for object detection. In: *2019 IEEE/CVF Conference Computer Vision, 2019; Seoul, Republic of Korea*; p. 9657–66.
11. Ren C, Zhang G, Jeong D, Hao L. CNN-based multi-object tracking networks with position correction and IMM in intelligent transportation system. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería.* 2023;39(4):1–12. doi:10.23967/j.rimni.2023.11.001.
12. Wang C, Bochkovskiy A, Liao H. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *2023 IEEE/CVF Conference Computer Vision Pattern Recognition, 2023; Vancouver, BC, Canada*; p. 7464–75.
13. Choi W, Cha Y. SDDNet: real-time crack segmentation. *IEEE Trans Ind Electron.* 2019;67(9):8016–25. doi:10.1109/TIE.2019.2945265.
14. Zhang H, Jiang L, Li C. CS-ResNet: cost-sensitive residual convolutional neural network for PCB cosmetic defect detection. *Expert Syst Appl.* 2021;185:115673. doi:10.1016/j.eswa.2021.115673.
15. Liu X, Hu J, Wang H, Zhang Z, Lu X, Sheng C, et al. Gaussian-IoU loss: better learning for bounding box regression on PCB component detection. *Expert Syst Appl.* 2022;190:116178. doi:10.1016/j.eswa.2021.116178.
16. Tang J, Liu S, Zhao D, Tang L, Zou W, Zheng B. PCB-YOLO: an improved detection algorithm of PCB surface defects based on YOLOv5. *Sustainability.* 2023;15(7):5963. doi:10.3390/su15075963.
17. Yang Y, Kang H. An enhanced detection method of PCB defect based on improved YOLOv7. *Electronics.* 2023;12(9):2120. doi:10.3390/electronics12092120.
18. Chen B, Dang Z. Fast PCB defect detection method based on fasternet backbone network and CBAM attention mechanism integrated with feature fusion module in improved YOLOv7. *IEEE Access.* 2023;11:95092–103. doi:10.1109/ACCESS.2023.3311260.
19. Varghese R, Sambath M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness. In: *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024; Chennai, India*; p. 1–6.
20. Liu X, Peng H, Zheng N, Yang Y, Hu H, Yuan Y. EfficientViT: memory efficient vision transformer with cascaded group attention. In: *2023 IEEE/CVF Conference Computer Vision Pattern Recognition, 2023; Vancouver, BC, Canada*; p. 14420–30.
21. Li C, Li L, Geng Y, Jiang H, Cheng M, Zhang B, et al. YOLOv6 v3.0: a full-scale reloading. *arXiv:2301.05586.* 2023.
22. Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution. *arXiv:2209.07947.* 2022.
23. Zhu L, Wang X, Ke Z, Zhang W, Lau R. BiFormer: vision transformer with bi-level routing attention. In: *2023 IEEE/CVF Conference Computer Vision Pattern Recognition, 2023; Vancouver, BC, Canada*; p. 10323–33.

24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017; Long Beach, CA, USA; p. 5998–6008.
25. Wang J, Xu C, Yang W, Yu L. A normalized Gaussian Wasserstein distance for tiny object detection. arXiv:2110.13389. 2021.
26. Ma S, Xu Y. MPDIoU: a loss for efficient and accurate bounding box regression. arXiv:2307.07662. 2023.
27. Huang W, Wei P, Zhang M, Liu H. HRIPCB: a challenging dataset for PCB defects detection and classification. J Eng. 2020;2020(13):303–9. doi:10.1049/joe.2019.1183.
28. Redmon J. YOLOv3: an incremental improvement. arXiv:1804.02767. 2018.
29. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: European Conference on Computer Vision, 2020; Glasgow, UK; p. 213–29.
30. Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, et al. DETRs beat YOLOs on real-time object detection. In: 2024 IEEE/CVF Conference Computer Vision Pattern Recognition, 2024; Seattle, WA, USA; p. 16965–74.
31. Wang C, He W, Nie Y, Guo J, Liu C, Wang Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. In: Advances in Neural Information Processing Systems 36 (NeurIPS 2024), 2024; Vancouver, BC, Canada; p. 1234–45.
32. Wang C, Yeh IH, Liao H. YOLOv9: learning what you want to learn using programmable gradient information; arXiv:2402.13616. 2024.
33. Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. YOLOv10: real-time end-to-end object detection. arXiv:2405.14458. 2024.
34. Ding R, Dai L, Li G, Liu H. TDD-net: a tiny defect detection network for printed circuit boards. CAAI Trans Intell Technol. 2019;4(2):110–6. doi:10.1049/trit.2019.0019.
35. Lan H, Zhu H, Luo R, Ren Q, Chen C. PCB defect detection algorithm of improved YOLOv8. In: 2023 8th International Conference on Image, Vision and Computing (ICIVC), 2023; Nanjing, China; p. 178–83.
36. Yang B, Qu Z. A PCB defect detection algorithm based on improved YOLOv7-tiny. In: 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2023; Xi'an, China; p. 1450–4.
37. Yuan M, Zhou Y, Ren X, Zhi H, Zhang J, Chen H. YOLO-HMC: an improved method for PCB surface defect detection. IEEE Trans Instrum Meas. 2024;73:1–11. doi:10.1109/tim.2024.3351241.
38. Zhou G, Yu L, Su Y, Xu B, Zhou G. Lightweight PCB defect detection algorithm based on MSD-YOLO. Cluster Comput. 2024;27(3):3559–73. doi:10.1007/s10586-023-04156-x.