Academic Journal of Emerging Technologies
ISSN Print: 3104-4417
ISSN Online: 3104-4425
CD ISSN: 3104-4433
https://doi.org/10.63313/AJET.9035

A PCB Micro-Defect Detection Method Based on Mult-Scale Feature Enhancement and Background Suppression

Yu Cao
School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
Email: 2843172786@qq.com

How to cite this paper: Cao, Y. (2026). A PCB Micro-Defect Detection Method Based on Mult-Scale Feature Enhancement and Background Suppression. *Academic Journal of Emerging Technologies*, 2(2), 29-50. ISSN Print: 3104-4417; ISSN Online: 3104-4425. https://doi.org/10.63313/AJET.9035
Published: 2026-03-04

Abstract

To address the issues of strong background texture interference, small defect scales, and the tendency for feature loss in the detection of defects on Printed Circuit Boards (PCBs) within complex industrial environments—which lead to high miss rates and false positive rates in existing detection algorithms—this paper proposes a PCB defect detection method based on multi-scale feature enhancement and background suppression. An end-to-end detection network is introduced. First, a multi-scale contextual feature enhancement module is designed. By constructing a parallel atrous convolutional pyramid with different dilation rates, it captures global context rich in semantic information. This is combined with shallow, high-resolution feature maps to fully preserve the edge and texture details of defects, addressing the information loss caused by down-sampling in deep networks. Second, a background decoupling and suppression attention mechanism is introduced. This mechanism separates foreground responses from background noise in the feature maps, generating a discriminative attention weight map. It adaptively enhances the response values in defect regions while suppressing interfering responses from complex circuit traces and copper foil backgrounds. Finally, an improved localization loss function is incorporated to optimize the regression accuracy of the predicted bounding boxes. Experimental results on the public PCB defect dataset PKU-Market-PCB demonstrate that the proposed method achieves a mean Average Precision (mAP50) of 96.8%, which is a 4.3 percentage point improvement over the mainstream algorithm YOLOv8n (92.5%). Particularly for defects such as nicks and pinholes, the recall rate is improved by more than 5.8%. Furthermore, the model achieves a detection speed of 85 frames per second, and its parameter count is reduced by 22% compared to the baseline, meeting the real-time requirements of industrial production lines. The method proposed in this paper effectively suppresses complex background interference and enhances the feature representation for multi-scale defects, significantly improving detection accuracy while maintaining detection speed, demonstrating good robustness and industrial application value.

Keywords

PCB defect detection; multi-scale feature enhancement; background suppression; attention mechanism; deep learning

1. Introduction

As the core carrier of electronic products, the quality of Printed Circuit Boards (PCBs) directly determines the performance and reliability of the final electronic devices. With the rapid development of fields such as consumer electronics, automotive electronics, and communication equipment, electronic products are evolving towards lighter, thinner, and more precise forms. Consequently, the design and manufacturing of PCBs are becoming increasingly complex—characterized by ever-increasing circuit density, more compact component arrangements, and significantly enhanced structural precision. In this context, any minor defects during the PCB production process, such as burrs, mouse bites, nicks, and pinholes, can lead to circuit functional failure or even serious product quality incidents. According to industry statistics, over 60% of poor soldering issues in connectors originate from undetected micro-level height defects. These "invisible defects" represent blind spots for traditional inspection methods. Therefore, achieving efficient and accurate PCB defect detection has become a critical step in the electronics manufacturing sector to ensure product quality and reduce production costs. [1]

However, the detection of defects on PCBs in complex industrial environments faces multiple challenges. Firstly, the defect scale is extremely small—the pixel occupancy of some defects (e.g., pinholes, mouse bites) is less than 3% of the image, making their feature information during multiple down-sampling operations in deep networks[2]. Secondly, background interference is severe—the dense circuitry, pads, and characters on the PCB surface form a complex textural background. The contrast between defects and this background is often low, leading to insufficient feature extraction. Thirdly, the industrial environment is complex and variable—factors such as uneven illumination, camera imaging noise, and differences in materials and processes among different PCB manufacturers further exacerbate the difficulty of detection. Additionally, industrial production imposes stringent requirements on detection speed (typically needing to exceed 30 FPS), making it a persistent research challenge to balance accuracy and efficiency[3].

To address these issues, academia and industry have conducted continuous and in-depth research on PCB defect detection technology. Based on the technological development trajectory, existing methods can be categorized into three main classes: traditional image processing-based methods, traditional machine learning-based methods, and deep learning-based methods.

Early research primarily relied on traditional image processing techniques. These

methods identify defects using handcrafted features such as image segmentation, edge detection, and template matching[4]. For example, methods based on threshold segmentation and mathematical morphology were used for solder joint inspection; geometric feature-based methods for hole detection achieved fast localization; image subtraction and normalized cross-correlation were also applied to PCB defect detection. The advantage of these methods lies in their high computational efficiency and lack of requirement for extensive training data, making them still valuable in resource-constrained scenarios. However, their limitations are equally evident: handcrafted features have weak generalization ability, are sensitive to illumination changes, struggle with complex background interference, and cannot effectively detect defects[5].

With the development of machine learning, researchers began introducing traditional machine learning classifiers into PCB defect detection. Algorithms like Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN) were used to classify extracted handcrafted features. For instance, a PCB detection system based on SURF features and Random Forest achieved defect recognition; ensemble learning methods were also applied to solder joint defect classification. While these methods represented an improvement over pure image processing, they still fundamentally relied on the quality of feature engineering and had limited representational power for defects, failing to meet the demands of high-precision detection[6].

In recent years, the rapid advancement of deep learning technology has brought revolutionary breakthroughs to PCB defect detection. Convolutional Neural Networks (CNNs) can automatically learn multi-level feature representations, significantly improving detection accuracy and robustness. Current mainstream deep learning detection methods can be divided into two-stage algorithms (e.g., Faster R-CNN) and one-stage algorithms (e.g., YOLO series, SSD). Two-stage algorithms first generate candidate regions and then classify and regress them, offering high accuracy but relatively slower speed. One-stage algorithms perform end-to-end prediction directly, achieving a better balance between speed and accuracy, making them more suitable for industrial real-time detection scenarios[7].

Addressing the specific challenge of PCB defect detection, researchers have made improvements from multiple dimensions. In terms of feature fusion, the YOLOv5-TDD algorithm proposed by Liao Xinting et al. added a shallow feature fusion branch in the neck network, directly introducing high-resolution feature maps from the early output of the backbone into the detection branch[8]. This preserves the edge and texture information of defects, achieving 99.12% mAP on the PCB_DATASET and improving the detection accuracy of mouse bite defects from 94.88% to 99.59%. In terms of attention mechanisms, the SE-SiLU module dynamically assigns channel weights, giving higher weight to channels containing defect information while suppressing background noise responses. Exploring

Transformer architectures, the MFSF-DETR model proposed by Zhang Yonghong et al. employed multi-scale feature shift fusion and entangled Transformer modules, achieving 98.1% detection accuracy on the PKU-Market-PCB dataset with a speed of 120.2 FPS. The YOLO-DTS model proposed by Ou et al. designed a dual-branch Transformer down-sampling module to capture both global context and local details, combined with a deformable attention mechanism to handle irregular defect shapes, improving mAP by 7.5% over the baseline. Furthermore, to address the sensitivity of small object localization, novel loss functions such as NWD (Normalized Wasserstein Distance) have been introduced to replace traditional IoU, effectively mitigating the impact of positional shifts of defects on detection accuracy.

In summary, existing research has made significant progress in the field of PCB defect detection, but shortcomings remain: First, most methods focus on improving a single technique and fail to systematically address the two core issues of multi-scale feature enhancement and background suppression simultaneously. Second, in complex industrial environments, false detections and missed detections caused by factors like illumination variation and background interference are not yet completely resolved. Third, how to further reduce model computational complexity while maintaining high accuracy to facilitate edge deployment remains a bottleneck[9].

To address the above problems, this paper proposes a PCB defect detection method based on multi-scale feature enhancement and background suppression. The method innovates in two dimensions: first, by designing a multi-scale contextual feature enhancement module that captures multi-scale semantic information through an atrous convolutional pyramid and combines it with shallow high-resolution feature maps to preserve the detailed texture of defects; second, by introducing a background decoupling and suppression attention mechanism that adaptively enhances responses in defect regions and suppresses complex background interference by generating an attention weight map. This paper aims to achieve high-precision, real-time detection of PCB defects in complex industrial environments through the aforementioned technical approaches, providing an effective solution for quality control in the electronics manufacturing sector.

2. Overview of PCB Defect Detection

Common PCB defects include open circuits, short circuits, burrs, nicks, pinholes, mouse bites, copper slivers, etc.

2.1. PCB Manufacturing Process and Defect Causes

The manufacturing of Printed Circuit Boards (PCBs) is a complex process involving multiple disciplines and numerous steps. Its quality is influenced by various factors such as material properties, process parameters, and equipment status. According to statistics from the Fifth Electronics Research Institute of the Ministry of Industry

and Information Technology on PCB failure cases between 2019 and 2022, production process defects are the primary cause of PCB failures, accounting for as high as 69%. An in-depth understanding of each stage of the manufacturing process and the types of defects that may be introduced is of great significance for the subsequent targeted design of detection algorithms.

Overview of the PCB Manufacturing Process PCB manufacturing typically involves over 200 processes. The core process flow can be summarized in the following main stages: material cutting, inner layer circuit patterning, lamination, drilling and through-hole plating, outer layer circuit patterning, solder mask and legend printing, surface finishing, and finally, shaping and inspection.

Correlation Analysis between Defect Cause sand Detection Algorithms.

Firstly, defects (such as nicks, pinholes, and burrs) are primarily generated during the circuit patterning stage. Their characteristics manifest as minor local edge variations. The pixel occupancy of such defects in an image is typically less than 3%, imposing a high demand on the multi-scale perception capability of detection algorithms.

Secondly, some defects (e.g., hole breakout, lifted pads) exhibit distinct three-dimensional structural features. They may be difficult to distinguish from the background in2Dimagingandrequirecontextualinformationforaccuratejudgment.

Thirdly, defects arising from different process stages possess different morphological characteristics. Etching defects often manifest as irregular edge variations, drilling defects appear as anomalies in circular regions, and surface finishing defects are reflected as grayscale or color differences in pad areas. This necessitates that detection algorithms can adapt to diverse defect morphologies.

Finally, statistical data showing that nearly 70% of failures originate from production process defects highlights the significant economic value of enhancing online inspection capabilities and detecting process anomalies promptly. The defect detection method studied in this paper is precisely designed to target the most challenging defects encountered in the aforementioned process stages.

2.2. Detection Challenges in Industrial Environments

Although deep learning technology has achieved significant progress in the field of PCB defect detection, its deployment and application in real-world industrial scenarios still face a series of challenges. These challenges primarily stem from the complexity of the industrial environment, the unique characteristics of the defects themselves, and the stringent requirements of production takt times. A thorough understanding of these challenges is a prerequisite for designing robust detection algorithms.

Complexity of the Imaging Environment Lighting conditions on industrial shop floors are complex and variable, directly impacting image quality. On one hand, the copper foil, pads, and solder mask on a PCB surface have different reflective

properties. Illumination from a single light source can easily create localized highlights or shadow areas, leading to significant contrast variations for defects in different regions of the same board. On the other hand, interference from ambient light is difficult to completely isolate—factors such as fluctuating production line lighting, imaging blur caused by equipment vibration, and camera sensor noise all contribute to variability in the quality of captured images. This inconsistency in imaging conditions requires detection algorithms to possess good illumination robustness and anti-interference capabilities.

Strong Interference from Background Texture The densely arranged circuits, pads, and characters on a PCB surface constitute an extremely complex textural background. Taking High-Density Interconnect (HDI) boards as an example, line widths/spacings can be as small as 30-50μm, presenting highly repetitive geometric patterns under a microscope. The difference between defects (like nicks and burrs) and normal circuit edges often lies only in local pixel grayscale mutations, making them. Furthermore, significant differences exist in substrate colors (e.g., green, black, blue solder masks) and circuit layouts among PCBs from different manufacturers and models, making it difficult for models trained on a single dataset to generalize across different scenarios.

Minuteness of Defect Scale One of the most challenging issues in PCB defect detection is the extremely small scale of defects. Taking a typical industrial camera resolution (e.g., 5 megapixels) imaging a 50mm × 50mm PCB area as an example, a 0.1mm × 0.1mm pinhole defect occupies only about 4×4 pixels in the image. According to research statistics, when the pixel occupancy of a target in an image is below 3%, the recall rate of mainstream detection algorithms drops sharply. How to preserve shallow detailed features while maintaining deep semantic information is a core problem that multi-scale feature fusion needs to solve.

Scarcity of Defect Samples On industrial production lines, defective products are low-probability events, with yield rates typically exceeding 95% or even 99%. This makes constructing large-scale, high-quality defect sample datasets extremely difficult. On one hand, the distribution of defect types is highly imbalanced—some rare defects (e.g., hole breakout, interlayer shorts) may not be encountered for months. On the other hand, obtaining annotated real defect images requires significant manual effort and cost. Sample scarcity leads to insufficient model training, making it prone to overfitting or missing unseen defects.

Real-time Requirements for Detection Speed PCB production lines usually operate in a continuous flow manner, and the inspection cycle needs to match the line speed. Taking a typical SMT line as an example, the conveyor belt speed can reach 0.5 m/s, requiring the complete inspection of a single PCB to be completed within hundreds of milliseconds. This imposes extremely high demands on the algorithm's inference speed—typically needing to exceed 30 FPS. However, high-accuracy models often come with significant computational complexity, making it difficult to meet

real-time requirements on edge computing devices. Achieving a balance between accuracy and speed is a key challenge for industrial deployment.

Stringency of Detection Metrics Unlike general object detection tasks, industrial defect detection has an extremely low tolerance for miss rates—a single undetected defect could lead to a batch recall or even customer complaints. Therefore, recall is often prioritized over precision. Additionally, requirements for localization accuracy are more stringent: minor between the predicted bounding box and the actual defect could lead to misclassification of the defect type or inability to pinpoint the location accurately during repair.

In summary, PCB defect detection in complex industrial environments is essentially a multi-objective optimization problem involving high-precision, low-miss-rate, real-time response for extremely small targets against a background of strong interference. The challenges mentioned above are coupled, and no single technical approach can comprehensively solve them. Breakthroughs are needed synergistically across multiple levels, including image acquisition, feature representation, and model design. This is precisely the starting point for the research in the subsequent chapters of this paper.

2.3. Machine Vision and Image Acquisition System

Reliable image acquisition is the primary step in PCB defect detection. Image quality directly dictates the upper limit of subsequent detection algorithms—if defect information is lost or severely distorted in the raw image, no algorithm, no matter how advanced, can effectively detect it. This section elaborates on the key design points of an image acquisition system for PCB defect detection from three aspects: industrial camera and lens selection, optical illumination design, and image preprocessing.

2.3.1. Industrial Camera and Lens Selection

The industrial camera is the core component of an image acquisition system, and its performance parameters directly determine the resolution, speed, and dynamic range of imaging.

- Sensor Type Selection: Current mainstream industrial cameras utilize CMOS or CCD sensors. CMOS sensors dominate the PCB inspection field due to their higher frame rates, lower power consumption, and cost. For high-precision inspection needs, global shutter CMOS cameras can be chosen to avoid image distortion under high-speed motion.
- Resolution Determination: Camera resolution must be determined based on inspection accuracy requirements. According to the sampling theorem, the minimum defect size should cover at least 3×3 pixels. Taking the detection of a 0.1mm pinhole as an example, if the field of view is 50mm×50mm, the required resolution is no less than (50/0.1×3) = 1500 pixels, meaning a camera with at

least 2 megapixels is needed. In practical applications, to balance accuracy and cost, cameras with 5 to 12 megapixels are most common.

- Frame Rate Requirements: The camera frame rate must match the production line conveyor speed. If the line speed is 0.5 m/s and each PCB is 200mm long, the inspection cycle is 0.4 seconds per board, requiring a camera frame rate of no less than 2.5 FPS. Considering the need for multiple image stitching or repeated captures, the actual selected frame rate is typically above 30 FPS.
- Lens Selection: The lens focal length determines the working distance and field of view. Telecentric lenses are favored in precision measurement due to their low distortion and absence of parallax—the magnification remains constant even if the object has slight height variations, which is crucial for measuring critical dimensions like line width and pad diameter. For general defect detection, high-performance fixed focal length lenses can suffice.

2.3.2. Optical Illumination System Design

The illumination system is the most critical factor affecting image quality. For different PCB defect types and surface characteristics, appropriate lighting methods and light source colors must be selected.

- Lighting Method Selection: For PCB defect detection, multi-angle combination lighting strategies are often employed. For instance, when detecting line nicks and burrs, dark-field illumination can be used to enhance edge contrast; for detecting pad oxidation, coaxial illumination is needed to eliminate reflection interference from the copper foil; for through-hole defects, backlight illumination can be combined to highlight the hole contours.
- Light Source Color Selection: Light of different wavelengths has varying penetration and reflection characteristics on PCB materials. Red light (630nm) penetrates the green solder mask well, enhancing the contrast between circuits and the substrate; blue light (470nm) has a shorter wavelength, suitable for detecting scratches; white light provides the most comprehensive spectral information and is suitable for most scenarios. In practice, a complementary color light source can be selected based on the PCB substrate color to maximize defect contrast.

2.3.3. Image Preprocessing Techniques

Raw acquired images often suffer from issues like noise and uneven illumination. Preprocessing is needed to improve image quality, creating favorable conditions for subsequent detection algorithms.

- Image Denoising: Industrial camera sensor noise and ambient light interference can introduce salt-and-pepper noise or Gaussian noise into images. Common denoising methods include median filtering (effective for removing salt-and-pepper noise), Gaussian filtering (smoothing Gaussian noise), and

bilateral filtering (edge-preserving denoising). For PCB images, bilateral filtering can remove noise while preserving circuit edge information, making it more practical.

- Illumination Unevenness Correction: Due to uneven light source distribution or differences in PCB surface reflectivity, images often exhibit locally overexposed or underexposed areas. Common correction methods include: homomorphic filtering (separating illumination and reflectance components), top-hat transform (extracting and subtracting the background illumination distribution), and Contrast Limited Adaptive Histogram Equalization (CLAHE) (enhancing local contrast). Among these, CLAHE is particularly effective in enhancing the contrast between defects and the background.
- Geometric Correction: Camera mounting errors or PCB conveyor misalignment can cause image rotation or perspective distortion. By pre-setting fiducial marks (Mark points) and calculating affine or perspective transformation matrices, the image can be corrected to a standard position, facilitating subsequent template comparison or defect localization.
- Image Enhancement: To address the low contrast of defects, methods like gradient enhancement and Laplacian sharpening can be used to highlight edge information. In recent years, enhancement algorithms based on Retinex theory have also been introduced, which simulate the human visual system to enhance details while maintaining color consistency.

Machine vision and image acquisition systems serve as the "eyes" of PCB defect detection. Through reasonable selection of cameras and lenses, careful design of illumination schemes, and effective application of preprocessing techniques, defect information can be maximally separated from the complex industrial environment, providing high-quality input images for subsequent deep learning detection algorithms. However, hardware-level optimization alone cannot fully overcome the challenges of background interference and feature extraction, which is the focus of the algorithmic research in the following sections of this paper.

2.4. Deep Learning-Based Object Detection Algorithms

The introduction of deep learning technology has brought revolutionary breakthroughs to the field of object detection. Convolutional Neural Networks can automatically learn multi-level feature representations, avoiding the tedious manual feature design process of traditional methods. Based on differences in detection pipelines, existing deep learning object detection algorithms can be categorized into three main classes: two-stage detection algorithms, one-stage detection algorithms, and Transformer-based detection algorithms. This section systematically analyzes the principles, evolution, and applicability of each type for the task of PCB defect detection.

2.4.1. Two-Stage Detection Algorithms

The core idea of two-stage detection algorithms is "coarse-to-fine"—the first stage generates region proposals, and the second stage classifies and refines the positions of these proposals. The pioneering work of this type was R-CNN (Region-based Convolutional Neural Network). Subsequently, Faster R-CNN introduced the Region Proposal Network (RPN), enabling end-to-end training by unifying region proposal generation and feature extraction within the same network, significantly improving detection efficiency.

Faster R-CNN first extracts image features through a backbone network (e.g., ResNet, VGG) to generate shared feature maps. The RPN slides a window over the feature maps, anchor boxes of different scales and aspect ratios, and outputs region proposals. Then, an ROI pooling layer maps region proposals of varying sizes to fixed-dimensional feature vectors. Finally, fully connected layers perform classification and bounding box regression.

From an accuracy perspective, two-stage algorithms have a natural advantage for defects (like nicks and pinholes)—the region proposal mechanism allows limited computational resources to focus on potential defect areas, avoiding a search for targets across the entire image. However, PCB production lines typically require detection speeds above 30 FPS, which two-stage algorithms struggle to meet. Furthermore, defects are easily filtered out as background during the RPN stage, leading to increased miss rates. Therefore, while two-stage algorithms are still used in academic research, the industry tends to prefer one-stage algorithms for their superior speed.

2.4.2. One-Stage Detection Algorithms

One-stage detection algorithms discard the region proposal step and perform dense sampling and prediction directly on feature maps, achieving end-to-end one-shot detection. The YOLO (You Only Look Once) series and SSD (Single Shot MultiBox Detector) are representative works of this type. One-stage algorithms have a significant speed advantage, making them more suitable for industrial real-time detection scenarios.

- YOLO Series Evolution: YOLO treats object detection as a regression problem, dividing the image into a grid, with each grid cell directly predicting bounding boxes and class probabilities. YOLOv2 introduced anchor boxes and batch normalization, improving accuracy. YOLOv3 adopted the DarkNet-53 backbone and multi-scale prediction, further enhancing small object detection capabilities. YOLOv4 and YOLOv5 integrated advanced structures like CSPNet and PANet, achieving a better balance between speed and accuracy. YOLOv8 and subsequent versions continue to optimize network architecture, loss functions, and training strategies, becoming the most widely used object detection framework in the industry.
- SSD Algorithm Characteristics: SSD sets default boxes on feature maps of

different scales, using shallow high-resolution feature maps to detect small objects and deep low-resolution feature maps to detect large objects, achieving multi-scale detection.

- Advantages and Disadvantages Analysis: The main advantage of one-stage algorithms is their detection speed—no region proposal generation, simple network structure, and high inference efficiency. However, their accuracy is generally slightly lower than two-stage algorithms, especially for detecting small objects. Reasons include: first, one-stage algorithms require dense sampling across the entire image, leading to extreme positive-negative sample imbalance, where a large number of easy negative samples dominate the training process. Second, targets suffer severe feature loss in deep feature maps, while shallow feature maps, though retaining detail information, lack sufficient semantic information Applicability to PCB Defect Detection: One-stage algorithms are the mainstream choice for PCB industrial inspection, as their speed advantage meets production line real-time requirements. However, they face the following key issues for defect detection:

Firstly, Feature Loss Due to Down-sampling. Standard one-stage detectors rely on strided convolutions for rapid down-sampling. For defects occupying only 5-10 pixels, this aggressive down-sampling strategy often erodes their defining features before effective network encoding. Although Feature Pyramid Networks (FPN) attempt to mitigate this through multi-scale fusion, the fusion occurs after features have already been significantly lost in the backbone network.

Secondly, Mismatch between Anchor Boxes and Defect Scales. The sizes of predefined anchor boxes are typically designed for general objects and deviate from the actual scales of defects, leading to sparse positive samples and regression difficulty.

Thirdly, Severe Background Interference. The complex background formed by circuits on the PCB surface results in low distinguishability between defects and the background. The dense sampling strategy of one-stage algorithms can easily misclassify numerous background areas as targets.

In response to these issues, researchers have proposed various improvements. For example, YOLO-PICO introduces an Expansion Attention module to enhance spatial detail retention in lightweight models. YOLOv5-TDD adds a shallow feature fusion branch in the neck network, directly introducing shallow high-resolution feature maps into the detection branch, significantly improving the recall rate of defect

2.4.3. Transformer-Based Detection Algorithms

After the tremendous success of the Transformer model in natural language processing due to its self-attention mechanism, it was introduced into computer vision tasks. DETR (Detection Transformer) is the first end-to-end framework applying Transformers to object detection, treating detection as a set prediction

problem and discarding components like anchor box design and non-maximum suppression.

DETR extracts features using a CNN backbone, flattens the feature maps, and inputs them into a Transformer encoder-decoder structure. The decoder interacts with the encoded features through a set of learnable object queries, directly outputting the prediction set. Through bipartite matching loss, the network learns the correspondence between predictions and ground truth end-to-end.

To address the challenge of small object detection, researchers have proposed various improvements. Deformable DETR introduces a deformable attention mechanism, focusing attention on key sampling points around reference points, reducing computational complexity and accelerating convergence. HyDeTr generates spatial priors through a density map prediction module, guiding object queries to focus on high-density areas, ensuring effective querying of densely distributed small targets. WDFS-DETR employs wavelet transforms and two-stage feature enhancement, improving inference speed while maintaining detection accuracy, achieving 50.1 FPS on a Jetson Orin Nano. DAWDet designs a dynamic multi-branch framework based on adaptive wavelet enhancement, using wavelet transform down-sampling layers to specifically capture low-frequency global and high-frequency local detail information of small targets.

- Applicability to PCB Defect Detection: The application of Transformer detectors in PCB defect detection is still in the exploratory stage. Their advantage lies in: global context modeling helps distinguish real defects from background noise—for example, an isolated small dot might be a pinhole, but a small dot on a circuit edge could be a normal process feature. However, their limitations are equally evident: high computational cost makes it difficult to meet industrial real-time requirements; the feature representation capability for small targets still needs improvement. Recent research suggests that through the introduction of multi-scale feature alignment and lightweight design, Transformer detectors have the potential to achieve a better balance between accuracy and speed.

2.4.4. Evaluation Metrics for Detection Algorithms

The performance evaluation of object detection algorithms involves multiple dimensions, mainly including detection accuracy, detection speed, and model complexity.

- Intersection over Union (IoU): Measures the overlap between the predicted bounding box and the ground truth box, defined as the ratio of the area of their intersection to the area of their union. An IoU threshold (e.g., 0.5) is typically set to determine if a detection is correct.
- Precision and Recall: Precision measures the proportion of samples predicted as positive that are actually positive. Recall measures the proportion of all

actual positive samples that are correctly detected. For defect detection tasks, recall is often prioritized over precision—missing a single defect could lead to a batch recall, whereas false positives can be mitigated by manual re-inspection.

- Mean Average Precision (mAP): The most commonly used comprehensive performance metric. It calculates the Average Precision (AP) across different recall levels and then averages over all classes. Common variants include mAP50 (mAP at an IoU threshold of 0.5) and mAP50:95 (average mAP across different IoU thresholds from 0.5 to 0.95).
- Specific Metrics for Defects: Small object detection faces unique evaluation challenges—positional significantly impacts IoU, making traditional metrics inadequate for accurately reflecting model performance. Researchers have proposed the Normalized Wasserstein Distance (NWD) as an alternative loss function. It is scale-invariant, responds smoothly to positional shifts, and can measure similarity between non-overlapping boxes. Additionally, metrics like Size-Normalized Average Precision (SNAP) have been proposed to quantify detection accuracy per unit parameter, assessing the parameter efficiency of a model.
- Detection Speed and Model Complexity: Common metrics include Frames Per Second (FPS), the number of parameters, and Floating Point Operations (FLOPs). Industrial deployment requires a balance between accuracy and speed—typically demanding detection speeds above 30 FPS, while the model size must be compatible with the storage and computational constraints of edge computing devices.

3. A PCB Micro-Defect Detection Method Based on Multi-Scale Feature Enhancement and Background Suppression

3.1. Overall Network Architecture

- Backbone: A lightweight backbone network (e.g., improved YOLOv8n, MobileNetV4, or GhostNet) is selected as the fundamental feature extractor to ensure detection speed.
- Neck: An improved feature pyramid structure (e.g., BiFPN or a custom fusion structure) is introduced to achieve sufficient fusion of multi-scale features.
- Detection Head: A dedicated detection branch is designed for defects, or small object perception capability is added to existing detection heads.

3.2. Multi-Scale Feature Enhancement Module

In the task of PCB defect detection, the diversity of defect scales is the primary challenge. Taking the PKU-Market-PCB dataset as an example, the pixel dimensions of defects such as notches and pinholes are typically only 5×5 to 15×15 pixels, while defects like burrs and mouse bites may occupy larger areas. Standard convolutional

neural networks, while expanding the receptive field by stacking convolutional and pooling layers, continuously reduce the resolution of feature maps, leading to severe loss of spatial details of defects in deep networks. To address this issue, this section designs a Multi-scale Feature Enhancement Module (MFEM) aimed at fusing semantic information and detail information from different levels to provide rich feature representations for subsequent detection.

3.2.1. Motivation for Module Design

Feature Pyramid Networks (FPN) and their variants, which fuse deep semantics with shallow details through top-down pathways, have become common solutions for multi-scale object detection. However, standard FPN has the following limitations: First, although shallow feature maps contain rich edge and texture information, they need to pass through multiple layers before fusing with deep features, during which detail information further attenuates. Second, the fusion method of FPN is relatively simple (usually element-wise addition or channel concatenation), failing to fully consider the importance differences of features at different levels. Third, for defects, contextual information is crucial for distinguishing real defects from background noise, yet the standard FPN has limited ability to expand the receptive field.

To address these issues, MFEM is improved from two dimensions: first, establishing shortcut connections from shallow features to deep layers, directly introducing high-resolution feature maps into the detection branch; second, introducing multi-branch atrous convolution to capture multi-scale contextual information without reducing resolution.

3.2.2. Module Architecture Design

The overall structure of MFEM mainly consists of three sub-modules: the shallow feature guidance branch, the atrous convolution context branch, and the adaptive feature fusion unit.

- Shallow Feature Guidance Branch: The two feature maps with the highest resolution in the backbone network are selected, and their channel numbers are unified to 256 through 1×1 convolution, preserving their rich edge texture and spatial location information. Unlike traditional FPN, these shallow features not only participate in top-down fusion but are also directly sent to the detection head via shortcut connections, ensuring that the detail information of defects is retained in the final prediction layer.
- Atrous Convolution Context Branch: Four parallel atrous convolutions with different dilation rates (set to 1, 2, 3, and 5, respectively) are applied on the deep feature maps. This design enables the convolutional kernels to obtain exponentially expanded receptive fields without increasing the number of parameters, capturing rich information from local details to global context. For

defects, contextual information is particularly critical—for example, an isolated small dot might be a pinhole, but a small dot on the edge of a circuit could be a normal process feature, requiring a larger context for discrimination.

- The multi-branch atrous convolution can be formally expressed as:

$$F_{fusion} = \sum_{i=1}^{N} \alpha_i \cdot \mathrm{Conv}_{1\times1}(F_i) \quad (1)$$

Here,αi=Softmax(Wi·GAP(Fi)) represents the adaptively learned fusion weights, GAP denotes global average pooling, and W_i are learnable parameters. In this way, the network can dynamically adjust the contributions of features from different levels based on the content of the input image, making the fused features more discriminative.

3.2.3. Module Architecture Design

The structure of BSAM is shown in Figure 3.Y, which mainly consists of three parts: a background feature estimation branch, a feature decoupling unit, and attention recalibration.

Background Feature Estimation Branch: First, global average pooling is performed on the input feature map $\boldsymbol{F_{in}} \in \mathbb{R}^{C\times H\times W}$ to aggregate global information across spatial dimensions.

3.2.4. Analysis of Module Advantages

MFEM is specifically optimized for the characteristics of the PCB micro-defect detection task:

First, by directly introducing shallow features into the detection branch, it preserves the edge and texture information of micro-defects to the greatest extent, addressing the issue of detail loss in deep networks.

Second, the multi-branch atrous convolution expands the receptive field while maintaining feature map resolution, enabling the model to distinguish genuine defects from false positives by incorporating contextual information, thereby effectively reducing the false detection rate.

Third, the adaptive feature fusion mechanism endows the network with the ability to dynamically adjust feature weights based on the input content, enhancing the model's adaptability and generalization capability for defects of different scales.

In summary, MFEM provides richer and more discriminative feature representations for subsequent defect classification and localization through multi-dimensional feature enhancement.

3.3. Background Suppression Attention Mechanism

In the task of PCB micro-defect detection, complex background interference is another major factor affecting detection accuracy. The densely distributed circuits, pads, and characters on the PCB surface constitute a highly structured textural

background, while the difference between micro-defects (such as nicks and pinholes) and normal circuit edges often manifests only as local pixel grayscale mutations. This low-contrast characteristic makes it easy for models to misclassify background textures as defects (causing false positives) or to weak defects within the background (leading to missed detections). To address this issue, this section designs a Background Suppression Attention Mechanism (BSAM), which explicitly models background features and suppresses their responses, guiding the network to focus on genuine defect regions.

3.3.1. Design Motivation

Standard convolutional neural networks treat all regions in an image indiscriminately during feature extraction. When the background texture is complex, the network invests substantial computational resources in learning the statistical features of the background, which diminishes its focus on defect regions. Although existing attention mechanisms (such as SENet and CBAM) can enhance important features, they lack targeted suppression of background noise.

We observe that the background in PCB images is highly repetitive and structured—the circuit textures on the same board exhibit similarity within local areas. This characteristic makes background features amenable to modeling. Based on this, the core idea of BSAM is: first, model the common characteristics of the background, and then subtract the background component from the original features, forcing the network to focus on anomalous regions that are inconsistent with the background.

3.3.2. Module Architecture Design

The structure of BSAM is shown in Figure 3.Y, which mainly consists of three parts: a background feature estimation branch, a feature decoupling unit, and attention recalibration.

Background Feature Estimation Branch: First, global average pooling is performed on the input feature map $\boldsymbol{F_{in}} \in \mathbb{R}^{\boldsymbol{C\times H\times W}}$ to aggregate global information across spatial dimensions:

$$F_{gap} = \mathrm{GAP}(F_{in}) \in \mathbb{R}^{C\times 1\times 1} \tag{2}$$

Subsequently, two fully connected layers (incorporating a reduction ratio r) are used to learn the common feature representation of the background:

$$F_{bg} = W_2 \cdot \mathrm{ReLU}\left(W_1 \cdot F_{gap}\right) \tag{3}$$

where $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$, $W_2 \in \mathbb{R}^{C\times\frac{C}{r}}$. The resulting $F_{bg} \in \mathbb{R}^{C\times 1\times 1}$ can be regarded as the channel-wise description vector of the "typical background" in the current feature map.

Feature Decoupling Unit: The original feature map is decoupled from the background features to extract the foreground defect component:

$$F_{fg} = F_{in} - \alpha \cdot \left(F_{bg} \otimes \mathbf{1}_{H \times W}\right) \tag{4}$$

where $F_{bg} \otimes \mathbf{1}_{H \times W}$ denotes broadcasting the background vector to a 2D feature map of the same size as F_{in}, and α is a learnable suppression coefficient (initialized to 0.1 and updated during network training). This operation subtracts the common background component from the channel features at each spatial location, with the remainder representing potential defect responses that significantly differ from the background.

Attention Recalibration: Spatial attention weighting is applied to the decoupled foreground features to further highlight defect regions:

$$M_{spatial} = \sigma\left(\mathrm{Conv}_{7\times7}\left(\left[\mathrm{AvgPool}\left(F_{fg}\right); \mathrm{MaxPool}\left(F_{fg}\right)\right]\right)\right) \tag{5}$$

$$F_{out} = F_{in} \otimes M_{spatial} \tag{6}$$

where $\left[\mathrm{AvgPool}(F_{fg}); \mathrm{MaxPool}(F_{fg})\right]$ represents channel concatenation, σ is the Sigmoid activation function, and $\otimes$ denotes broadcast multiplication. The final output F_{out} enhances the response in defect regions and suppresses background interference while retaining the integrity of the original features.

3.3.3. Integration with the Baseline Network

BSAM is designed as a plug-and-play module that can be embedded at multiple key positions within the detection network. In this paper, it is placed after the outputs of each layer of the Feature Pyramid Network, applying background suppression to the fused multi-scale feature maps. This enables the subsequent detection heads to perform defect classification and localization based on "purified" features.

3.3.4. Analysis of Module Advantages

BSAM is tailored to address the characteristics of complex PCB backgrounds:

Firstly, by explicitly modeling common background features and decoupling them from the original features, it achieves direct suppression of background interference, making it more targeted than traditional attention mechanisms.

Secondly, the learnable suppression coefficient α allows the network to adaptively adjust the suppression strength—increasing suppression when the background is highly consistent and decreasing it when the background itself varies, thereby avoiding information loss due to excessive suppression.

Thirdly, the integration of spatial attention provides secondary enhancement to the decoupled foreground features, further improving the model's sensitivity to micro-defects.

Experimental results demonstrate that the introduction of BSAM significantly enhances the model's robustness to complex backgrounds, reduces the false detection rate, and performs particularly well in areas with high-density circuitry.

4. Image Enhancement and Defect Detection Method for Complex Lighting Environments

In practical industrial scenarios, complex and variable lighting conditions are a significant factor affecting the accuracy of PCB defect detection. The PCB surface is composed of various materials such as copper foil, solder mask, and pads, each with distinct reflective properties. Under a single light source, this often leads to issues like local overexposure, shadow occlusion, and uneven brightness in the captured images. Experiments show that under non-uniform illumination, even advanced detection algorithms can experience a 5-8 percentage point drop in recall rate, with a particularly pronounced impact on micro-defects that already have low contrast. To address this problem, this chapter proposes an image enhancement method tailored for complex lighting environments. This method serves as a preprocessing module before detection to improve the quality of input images.

4.1. Analysis of Illumination Degradation and Enhancement Framework

4.1.1. Types of Illumination Degradation

Illumination degradation in PCB images primarily manifests in three forms: local overexposure, uneven illumination, and low-light noise.

4.1.2. Overview of the Enhancement Framework

The image enhancement method proposed in this chapter is based on the Retinex theory, which decomposes an image into an illumination component and a reflectance component. These components are processed separately and then fused.

4.2. Adaptive Image Enhancement Based on Retinex Theory

4.2.1. Illumination Component Extraction

According to Retinex theory, an image $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y})$ can be expressed as the product of the illumination component $\boldsymbol{L}(\boldsymbol{x},\boldsymbol{y})$ and the reflectance component $\boldsymbol{R}(\boldsymbol{x},\boldsymbol{y})$:

$$\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{L}(\boldsymbol{x},\boldsymbol{y}) \cdot \boldsymbol{R}(\boldsymbol{x},\boldsymbol{y}) \quad (7)$$

Guided filtering is employed to estimate the illumination component due to its excellent edge-preserving properties:

$$\hat{\boldsymbol{L}} = \text{GuidedFilter}(\boldsymbol{I},\boldsymbol{I},\boldsymbol{r},\boldsymbol{\epsilon}) \quad (8)$$

4.2.2. Patch-based Adaptive Gamma Correction

After extracting the illumination component, it needs to be corrected to improve the brightness distribution. This paper proposes a patch-based adaptive gamma correction that dynamically adjusts the gamma parameter according to the local

luminance:

$$\gamma_{i,j} = \alpha \cdot \left(\frac{\mu_{\text{global}}}{\mu_{i,j} + \epsilon} \right)^{\beta} \quad (9)$$

The corrected illumination component is then:

$$L_{\text{enh}}(x, y) = L(x, y)^{\gamma_{i,j}} \quad (10)$$

4.2.3. Defect Enhancement Based on Top-Hat Transform

After illumination correction, a multi-scale Top-Hat transform is used to enhance defect regions:

$$T_{\text{multi}}(I) = \sum_{k=1}^{K} w_k \cdot T_{\text{hat}}^{(k)}(I) \quad (11)$$

where $T_{\text{hat}}^{(k)}(I) = I - (I \circ B_k)$ is the Top-Hat transform corresponding to the structuring element B_k at the k-th scale.

4.2.4. Image Fusion

The illumination-corrected image is fused with the defect-enhanced image:

$$I_{\text{enh}} = L_{\text{enh}} \cdot R + \lambda \cdot T_{\text{multi}}(I) \quad (12)$$

In this paper, the image enhancement module is designed as a differentiable network layer, embedded before the detection network to enable end-to-end joint optimization. The parameters of the enhancement module are updated according to the detection loss, learning the enhancement strategy most conducive to defect detection.

4.3. Experiments and Results Analysis

4.3.1. Experimental Setup

Dataset: PCB images captured under various lighting conditions were collected to construct a complex illumination test set.

Evaluation Metrics: Information entropy, contrast, peak signal-to-noise ratio; detection accuracy mAP, recall.

Comparison Methods: Histogram equalization, CLAHE, MSRCR.

4.3.2. Experimental Results

To validate the effectiveness of the proposed image enhancement method, comparative experiments were conducted on the complex illumination test set. Four typical image enhancement methods were selected for comparison: Histogram Equalization, Contrast Limited Adaptive Histogram Equalization (CLAHE),

Multi-Scale Retinex with Color Restoration (MSRCR), and the proposed adaptive Retinex-based method. Evaluation metrics included objective image quality indicators (information entropy, contrast) and downstream detection task indicators (mAP50, recall).

From the perspective of image quality, the original images had an information entropy of 6.82 and a contrast of 42.3, reflecting information loss due to uneven illumination. After processing with the proposed method, the information entropy increased to 7.68 and the contrast to 68.9, both outperforming the other comparison methods. This indicates that the proposed method effectively enhances the discrimination between defects and the background while preserving image details.

From the perspective of detection performance, running the baseline detection algorithm directly on the original images yielded an mAP50 of 92.5% and a recall of 88.7%. After enhancement with the proposed method, mAP50 increased to 95.1% and recall to 92.3%, representing improvements of 2.6 and 3.6 percentage points, respectively. In comparison, CLAHE and MSRCR achieved mAP50 of 93.8% and 94.2%, and recalls of 90.2% and 91.0%, respectively. The experimental results demonstrate that the proposed method yields the most significant improvement in detection performance, especially in enhancing the detection rate of defects in highlight and shadow areas.

In summary, the proposed method improves brightness distribution through patch-based adaptive gamma correction and enhances defect regions using multi-scale Top-Hat transform, effectively improving image quality and detection performance under complex lighting conditions, thereby providing higher-quality input for subsequent detection algorithms.

5. Conclusion and Future Work

5.1. Conclusion

This paper systematically investigates the key technologies for PCB micro-defect detection in complex industrial environments. Addressing challenges such as strong background interference, small defect scales, complex lighting conditions, and high real-time requirements in practical industrial scenarios, the work is carried out at three levels: core detection algorithm, image enhancement preprocessing, and lightweight model deployment. A relatively complete technical system for PCB micro-defect detection has been developed.

5.2. Future Work

While this paper has achieved certain results in researching key technologies for PCB micro-defect detection, several directions warrant further exploration due to limitations in research time and experimental conditions:

Few-shot and Zero-shot Defect Detection Research

Defect samples on industrial production lines are low-probability events, and constructing large-scale annotated defect datasets is costly. This research relies on labeled defect datasets, but in practical scenarios, new or rare defects may be difficult to collect in advance. Future work could explore the application of few-shot learning, meta-learning, or zero-shot learning techniques in PCB defect detection, enabling models to adapt quickly to new defect types with few or even no samples.

Further Balancing Model Efficiency and Accuracy

The lightweight model proposed in this paper achieves 42 FPS on a Jetson Nano, meeting the requirements of conventional production lines. However, as PCBs evolve towards higher density and finer line widths, the demands for detection accuracy will continue to increase. Further research is needed to optimize the model structure for real-time detection on more resource-constrained MCU-level devices, or to achieve higher precision real-time detection on high-performance computing platforms.

(3) Multi-modal Information Fusion Detection

This research is based solely on visible light images for defect detection. However, certain types of defects (e.g., rough hole walls, delamination) are not clearly visible in 2D images. Future work could consider integrating multi-modal information such as 3D structured light, thermal imaging, and laser point clouds to perceive defect characteristics from more dimensions, further enhancing the comprehensiveness and accuracy of detection.

(4) Research on Explainability of Detection Algorithms

Current deep learning models perform excellently in defect detection, but their decision-making processes lack explainability. In high-reliability fields such as aerospace and medical electronics, this limitation hinders practical application. Future work could explore combining the physical causes of defects and process knowledge with deep learning models to make detection results explainable, thereby increasing industrial users' trust in the algorithms.

(5) Construction of Edge-Cloud Collaborative Detection Systems

With the development of the Industrial Internet, the synergy between edge computing and cloud computing is becoming a trend. Future research could investigate an edge-cloud collaborative architecture for PCB defect detection: the edge side handles real-time preliminary detection and rapid response, while the cloud side is responsible for secondary verification of complex defects, continuous model training, and knowledge sharing across production lines. Through edge-cloud collaboration, the breadth and depth of detection capabilities can be continuously improved.

In summary, PCB micro-defect detection technology still has broad research prospects in terms of algorithm accuracy, environmental adaptability, and deployment efficiency. With the continuous development of deep learning theory and edge computing hardware, it is believed that more innovative achievements will

emerge in this field, providing stronger technical support for quality control in the electronics manufacturing industry.

References

[1] Xu Haida, Huang Yun, Wang Pengyu, et al. Improved Faster R-CNN Algorithm for PCB Defect Detection[J/OL]. Mechanical Design and Manufacturing, 1-5[2026-03-01].

[2] Liu Chunjuan, Zhao Haoran, Zhang Mingxuan, et al. PCB Defect Detection Model Integrating Wavelet Transform Convolution and Knowledge Distillation[J/OL]. Journal of Instrumentation, 1-13[2026-03-01].

[3] Du Qian, Su Yingying, Peng Jie, et al. Lightweight PCB Defect Detection Algorithm Based on Improved YOLOv11n[J/OL]. Journal of Chongqing University of Science and Technology (Natural Science Edition), 1-10[2026-03-01].

[4] Cao Pengbin, Li Jianke, Wang Qiang, et al. Research on Defect Detection of Laser Soldering Joints with Small Samples Based on Generative Adversarial Network Data Augmentation and Transfer Learning[J/OL]. Journal of Welding, 1-15[2026-03-01].

[5] Liu Yang, Guo Peiying, Ma Wenlong, et al. Research on PCB Defect Detection Method Based on Improved DETR Model[J]. Inner Mongolia Petrochemical Industry, 2025, 51(12): 14-16+28.

[6] Zheng Jiali, Dong Ling, Li Heng, et al. Lightweight PCB Defect Detection Method Based on PMAC-YOLO[J/OL]. Journal of Electronic Measurement and Instrumentation, 1-16[2026-03-01].

[7] Guo Yuru. Research on Circuit Board Defect Detection Based on Multi-Scale Features and YOLOv8[J]. Laboratory Testing, 2025, 3(24): 12-14.

[8] Qiu Yongfeng, Luo Kaixi, Wang Zhibing, et al. Adaptive Global-Local Attention PCB Defect Detection Algorithm AGS-YOLO[J]. Semiconductor Technology, 2026, 51(02): 190-199.

[9] He Yongjiao, Wang Lin, Zhang Xiaoshuang, et al. A Small Sample Surface Defect Detection Model for Industrial Products Based on BP-DFO Faster R-CNN[J]. Modern Manufacturing Engineering, 2025, (12): 1-10+26.