Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

DB-YOLO: An Improved YOLOV8 Model Based on The Second Backbone for Small Defects on PCB Surface

Jun Liu, Yusen Gong, Hanli Zheng, Haocong Cai, Jinlin Yang

School of Automation, Guangdong Polytechnic Normal University.China.

Corresponding to: Jun Liu

Abstract: This study introduces the DB-YOLO method, targeting the existing PCB defect detection algorithms, which struggle to identify minor flaws in intricate, small-scale layouts with comparable backdrops, to enhance detection precision. Initially, the deepening of the network layer will lead to a large number of loss of detailed features of the detection target. Although the significant features of the target can be retained, this loss will make the information of the small target incomplete in the small target detection task. Therefore, we combined CBLinear and CBFuse modules in SOTAYOLOv9 to design the second backbone. By redistributing information at different levels, the micro features in the original information are strengthened, and the communication between channels is increased at the same time, which effectively improves the richness of feature information.; next, in order to shorten the forward propagation process and retain the original information to extract the features of small defects in the structure, we use partial convolution to design a new module named CSPC, which reduces the waste of computer resources caused by feature map redundancy. Furthermore, a refined version of the inverted residual multiscale attention module (iREMA) was introduced to augment the representation of features. In conclusion, our comparison of the DB-YOLO model with other current models reveals through experiments that our suggested model surpasses them, enhancing the mAP in the validation set by 3.6% relative to YOLOv8n.

Keywords: PCB Defect Detection; Computer Vision; Deep Learning

1 INTRODUCTION

Sophisticated industrial setups demand heightened efficiency and product excellence, with defect detection [1] serving as a technique to efficiently discard faulty items [2] and enhance the approval rate of products exiting the manufacturing process. With the continuous improvement of computer computing power and optimization of algorithm structure, Computer Vision (CV) -based detection technology has become more mature and efficient. For example, thresholding, edge detection, machine learning, and deep learning have all been successfully applied in the detection field. With the continuous progress of these methods, the disadvantages brought by traditional manual detection[3] are becoming more and more significant. As the electronic information industry [4] sees a steady rise in the need for high-quality PCB products and the advent of new information technology [5], the design and manufacturing of PCBs are advancing towards enhanced dependability, reduced dimensions, increased density, and so on. The mentioned brin The hurdles in detecting PCB defects using machine vision are significant, encompassing these aspects.

The creation and production of PCBs are intricate, frequently leading to a variety of defects. Each of these flaws significantly affects the stability of PCB component performance, necessitating the model to concurrently pinpoint multiple issues. precisely identify the type of defects and mark them as ple.

Owing to the exceptional production technique, the PCB component surface imperfections typically measure under 4500 pixels, with certain spur and mouse_bite flaws being merely around 300 pixels, representing a mere 0.005%-0.07% of the total 6,500,000 pixels. The PCB is captured in a high-resolution image, indicating a reduction in the number of features within the defects themselves, thereby complicating the detection process.

Present methods for detecting PCB defects in computer vision are typically categorized into two types: traditional target identification and deep learning-based detection [6]. Extracting and identifying minute flaws is challenging with the former, whereas CNN-based deep learning techniques excel in capturing minute defect characteristics, leading to outstanding outcomes in the def field. Identifying effects. Systems for detecting CNNS fall into two main categories: single-stage detectors and two-stage detectors [7]. Detectors with a dual-phase approach [8], like Faster-CNN, require the creation of regional recommendations to identify the target, thereby enhancing accuracy. Nonetheless, the application of this model to real-time detection remains challenging due to the increased computational demands and time expenditure of both stages, even though numerous academics have already enhanced it significantly. The single-stage detector based on the YOLO [9] uses the entire network as a channel, predicts the boundary boxes and probabilities of the entire image, and retrieves all objects in a one-step inference. Single-stage detectors have been widely used in industrial detection because of their good real-time performance and easy deployment. However, the model still has some shortcomings, such as excessive reliance on data quality for detecting minor defects, significant differences and instability in the performance of defect datasets for different variables, and the complexity of the model results in higher computational costs for training. To address these issues, many researchers have made different efforts. For example, to deal with this problem that the existing deep learning models are difficult to be deployed in practice, liu [10] et al. cleverly combined the lightweight characteristics of MobileNetV2 and the characteristics of feature pyramid network that are conducive to multi-scale feature fusion, and designed a new model, MobileNet-Yolo-Fast. Although the network did not realize industrial applications, it brought deep learning one step closer to practical deployment. Li [11] et al. proposed an integrated detection network that mixed classification tasks and detection tasks. Firstly, YOLOv2 screened quickly and roughly and generated a preliminary detection set with a smaller range, and then sent the data in the detection set to the ResNet-101 network for further classification. The proposed scheme has been applied in three real production lines, and the measured efficiency is improved by 33%. Niu [12] et al. used the k-Mean ++ method to improve the location matching of the original anchor box, and added the ECA-Net module, so that the model could capture the channel characteristics more effectively. These models have been verified in PCB surface defect detection tasks. However, due to the limitation and influence of the environment in the detection process, the current defect detection algorithms still need to improve the detection accuracy and detection speed.

Aiming at these problems and the physical characteristics of PCB defects, we strengthen the feature extraction ability, design a feature fusion link, and use the attention mechanism in the target detection stage to improve and optimize the whole model accordingly. This paper makes the following co-absorption in this detection field.

For sophisticated and complex circuit environments, defects often have similar structures to normal circuits. In order to better identify defects in these complex structures, we design a dual-backbone YOLO model (DB-YOLO). This model adds a second backbone to the existing YOLOv8n architecture. This design of a second backbone structure can improve feature extraction capabilities and the ability of networks to extract defect features from similarly colored PCB in complex contexts.

In order to complete the actual deployment more conveniently, it is necessary to minimize the number of model parameters. We designed the CSPC module, which effectively reduces the parameters, improves the detection speed, and improves the detection accuracy.

The inverted residual structure of iRMB module is used in EMAttention module, which can better retain the original features, reduce the use of training time and computing resources, and effectively improve the detection accuracy of structural missing types. This new structure is added to the C2f module in the detection part. The improved C2f module is called inverse multi-scale attention module

2 Related Work

2.1 *Vision-Based PCB Surface Defect Detection Method*

The use of computer vision is a watershed in the field of surface defect detection of industrial products, marking a more efficient era in the field of inspection. Although there are still some enterprises in the use of manual visual inspection of production and manufacturing. However, compared with computer vision methods, manual visual inspection is not only inefficient but also has a high miss rate, which seriously affects the qualification rate of the final product. In recent studies, the speed and accuracy of automatic optical inspection (AOI) have been significantly improved by continuously optimized algorithms and higher quality images. AOI is more suitable for large-scale production lines, so it has become a research hotspot. Liu [13] et al. completed the identification of PCB defects by image enhancement, image threshold segmentation and morphological segmentation, and provided a complete set of operation procedures for reference. However, this method relies too much on morphological geometric features and is easily affected by shadows, stains and other factors. Gaidhane [14] et al. completed the defect detection task by using the method of similarity measurement, which uses the rank of the matrix as a judgment threshold for similarity measurement, and higher than this threshold is denoted as a defect. This method has the advantages of simple calculation and fast detection speed, but it can not detect the type of defects, and has a good detection effect only for local defects. When the scope of defects becomes larger or the image has a certain offset deformation, the detection quality will be significantly decreased. Although the use of AOI has received good results and feedback, this approach has significant limitations, such as the problem of generalization and robustness. These two problems seriously limit the application of AOI technology in the field of industrial testing. In addition, compared with the convenience and improvement of production efficiency brought by AOI technology, the cost of equipment deployment is huge, so it is difficult to be widely promoted and applied in the field of detection. At the same time, many researchers have adopted traditional image processing methods mainly based on threshold segmentation[15], image alignment[16] and edge detection[17]. These traditional image processing methods are difficult to complete the defect detection

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

task with high efficiency and quality alone, so they have gradually become the auxiliary means of machine learning methods. Traditional machine learning methods are based on simple Back Propagation (BP) neural network and support vector machine (SVM). These methods have a certain feedback ability, which plays a limited role in the problem of poor robustness and low generalization ability. However, they are still trapped by the problems that the images are difficult to align perfectly, the images cannot be deeply understood, and the generalization ability is still not enough. Subsequent scholars have continuously proposed new structures and processing methods of rich information on the basis of machine learning, and gradually formed a network model for preliminary analysis of image information, and proposed the concept of deep learning.

In the field of detection, the method of deep learning has become a hot topic of research, and many excellent methods have emerged during this period, which greatly enriches the tools and methods we can use in the actual detection. However, in terms of the detection stage, all the existing detection algorithms can be roughly divided into one-stage detection algorithms and two-stage detection algorithms. Two-stage detection algorithm is represented by Faster R-CNN and its series of improved algorithms. The biggest structural feature of this kind of algorithm is that the detection is divided into two parts: prediction and detection, which can more effectively use computing resources to obtain more accurate detection results. The single-stage detection algorithm usually needs only one forward propagation to obtain the detection results, and its advantages are that the detection speed is fast, the model size is much smaller than that of the two-stage algorithm, and it is easy to be deployed in the actual environment. However, the improvement of speed and the reduction of model volume are at the expense of detection accuracy. So at present, most scholars' research direction for single-stage detection algorithm is to improve the accuracy or maximize its lightweight advantage. At present, the mainstream detection algorithms at that stage are YOLO series and the single-shot Multi-bracket detector[18] (SSD). As the peak of single-stage detection algorithm, YOLO can directly output the category and location information of the detected target. Compared with other algorithms, its biggest advantage is that in the case of limited computing resources, it achieves the balance between detection accuracy and detection speed to the greatest extent. While maintaining a fast detection speed, its detection accuracy can also meet most visual detection tasks. However, this is far from enough in the field of surface defect detection of industrial products. Due to the improvement of product accuracy and the limited hardware resources on the actual production line, we need algorithms with higher detection accuracy and more lightweight model structures. Recently, many scholars have devoted themselves to using the attention mechanism, improving the structure of the backbone network, or replacing the original module with a more lightweight module, trying to improve the detection accuracy while maintaining lightweight to meet the needs of actual production. For example, based on the YOLOv5 model, Yuan [19] et al. improved the original model in three parts: backbone, upsampling and detection head. Based on convolutional block attention module (CBAM), a multiple convolutional block attention module (MCBAM) for precise detection of multiple small targets is proposed. In the subsequent ablation experiments, the improvement of the model performance by these modules was verified. Based on the YOLOX model, Xuan [20]et al. used CSPDarknet to improve the backbone network part, and used the reverse residual structure and coordinated attention mechanism (CA) to optimize the network structure and improve the detection ability of the network. In addition, the author also uses data reinforcement on the original data set. Using the original data set and the enhanced data set to train the same network model, the results show that the enhanced data set can play a better training effect, which also reminds us that strengthening the data set in practical operation can also effectively improve the detection accuracy, so that the final detection algorithm meets the requirements of actual production.

2.2 PCB DEFECT DETECTION METHOD BASED ON YOLO

YOLOv8 series maintains the advantages of engineering simplicity and ease of use of YOLOv5 model structure. At the same time, the C2f module is proposed by branching more gradient streams in parallel with the idea of ELAN in YOLOv7, which ensures the lightweight and at the same time obtains richer gradient stream information. The Head part is the most changed, which is changed from the original coupling head to decoupling head, and changed from the Anchor-Based in YOLOv5 to Anchor-Free, which maintains a suitable balance between inference speed and detection accuracy. While the detection accuracy reaches the required level of precision electronic components, the detection speed also initially meets the needs of production. This model has high universality and can be flexibly improved to be suitable for different detection fields. Therefore, many scholars have used this model as a basis to conduct research in the field of PCB defect detection. For example, Yuan [21] et al. improved the loss function of the YOLOv8 model and lightweight the backbone network, realizing the fast and accurate detection of PCB surface defects. Chen [22] et al. introduced Wise IoU and proposed W-YOLOv8 model, and adopted a gradient gain allocation strategy with a dynamic non-monotonic focusing mechanism to make the model pay attention to the overall quality of the anchor frame, thereby improving the performance of the original model. Xia[23] et al. applied a high-resolution feature layer (P2) to obtain more details and location information of small objects, relying on this method to improve the detection of tiny defects. Yan[24] et al. added ShuffleAttention and BiFPN structures to the structure of YOLOv8, which enhanced the multi-scale feature fusion ability of the model and better adapted to small target detection. There have been many PCB defect detection algorithms developed based on YOLOv8, and these algorithms have been improved to varying degrees. However, in the detection of precision electronic components, this detection accuracy has not satisfied the researchers. And in the actual use process, we need to consider the limitation brought by the limited computing resources, so we need to consider the balance between the performance of the model and the size of the model. Although the above scheme improves the YOLOv8 model and achieves good results, there are still further optimization

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

schemes for the structure of the network model. In order to solve the shortage of computing power resources in the actual deployment and further improve the detection accuracy, a network model named DB-YOLO is designed. This model can detect small defects quickly and accurately on complex circuits.

YOLOv8 is positioned as an algorithmic framework rather than a specific algorithm, which implies that it is more easily adapted to different datasets and tasks, and this flexibility allows YOLOv8 to perform well in a variety of application scenarios. In addition, the YOLOv8 algorithm has been used in a wide range of applications in the field of surface defect recognition, including product quality inspection in a number of industries, including railroad tracks [25], wind turbines [26], and the interior of high-precision parts [27]. The efficacy and performance of the algorithm has been thoroughly validated. Therefore, in order to facilitate in model portability and improve generalizability, the mature YOLOv8 is chosen as the base framework in this paper.

3 PROPOSED METHOD

Various improved models based on YOLOv8 have achieved good detection results in most detection fields, and YOLOv8 model has also been widely recognized by many scholars. However, compared to natural inspection tasks that contain multiple sizes of inspection targets, the target sizes in PCB defect detection tasks are much smaller than the target sizes [28] in other inspection tasks. The color and shape of defective and non-defective images are difficult to distinguish, resulting in a high false alarm rate and a high leakage rate in the product quality inspection process. From the needs of today's production development, we need algorithms with higher detection accuracy and faster detection speed, and this algorithm also needs to have a small enough volume. To this end, we designed the DB-YOLOv8 network model, and the network structure diagram is shown in Figure 1. Where the red dashed box is the part of our model improvement.

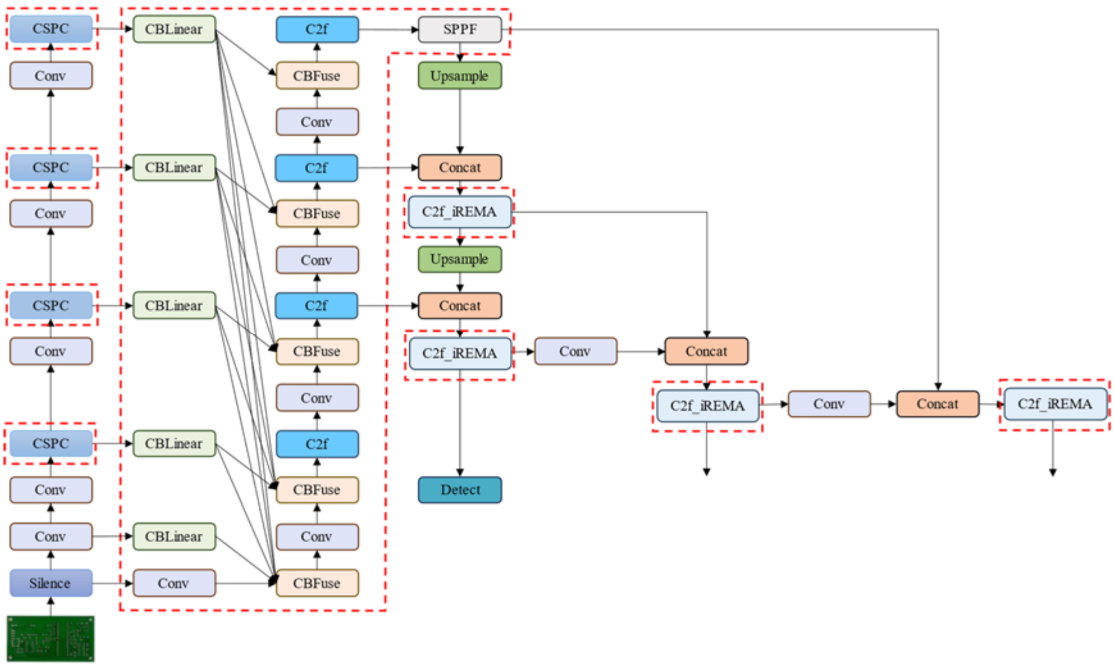

FIGURE 1 DB-YOLOV8 OVERALL STRUCTURE DIAGRAM

The sizes of different types of defects are obviously different. Therefore, in the face of defects of different sizes, some smaller effective features will be ignored, and these small features may not be very critical in other detection tasks. However, since PCB defects themselves carry very little information, these small features are particularly critical at this time. When deepening the network layers to extract high-level features, smaller defect features tend to lose a lot in this process, which seriously limits our improvement of the model. To solve this problem, a second backbone is added to the YOLOv8 model. Firstly, it input different features into CBLinear module to decompose them into more detailed information, and input these information into CBFuse module to recombine. After that, the fused information is used to extract new feature maps. This method can effectively deepen the information exchange between different channels, obtain feature maps that retain more information and high-level features, and can greatly alleviate the problem of information loss caused by network deepening. Since PCB defects are often distributed in precise and complex circuits and the defects themselves are smaller targets than circuits or components, the precise location of defects is strictly and finely required. Therefore, YOLOv8n may appear inadequate in detecting these tiny defects. Therefore, how to strengthen the feature extraction ability of YOLOv8n network so that it can avoid a large amount of interference information caused by background information is one of the problems currently faced.

Journal of
Research in Multidisciplinary Methods and Applications
http://www.satursonpublishing.com/
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

The convolutions used in YOLOv8 model are all traditional convolutions. This method of using a large number of traditional convolution stacks to extract features can greatly increase the generality of the model. However, for the PCB surface defect detection task, a large number of convolution operations will lead to the loss of key small features. At the same time, a large number of convolution operations will also lead to larger model size and a lot of waste of computing resources. How to improve the YOLOv8n network model and reduce its parameters to a certain extent is a major difficulty in object detection. In order to solve this problem, we use partial convolution (PConv) to construct a new module, named CSPC module, which uses the dynamic convolution kernel adjustment mechanism of PConv to solve the problem of feature map redundancy. It first checks the data points within that window. For those valid, non-missing data points, PConv applies a convolution kernel just like a regular convolution operation. However, for those missing or invalid data points, PConv will ignore them and not include them in the convolution calculation. This data processing method not only improves the robustness to missing data, but also extracts and utilizes the remaining effective information more effectively.

Since PCB defects are often distributed in precise and complex circuits and the defects themselves are smaller targets than circuits or components, the precise location of defects is strictly and finely required. Therefore, YOLOv8n may appear inadequate in detecting these tiny defects. Therefore, how to strengthen the feature extraction ability of YOLOv8n network so that it can avoid a large amount of interference information caused by background information is one of the problems currently faced. For this case, we apply inverse residue moving block (iRMB) to Efficient Multiscale Attention (EMA) to obtain a new inverted residual multi-scale attention mechanism, and successfully fuse this attention mechanism with the C2f module to obtain the C2f_iREMA module. This new module utilizes an inverted residual design to improve the handling of information flow and re-calibrate the channel weights in each parallel branch by encoding global information, capturing pixel-level relationships through cross-latitude interactions. This design enables the model to effectively use long-distance dependence with a small increase in parameters and model size, enhance the correlation between the defect part and the whole, and effectively reduce the influence of background on defect detection.

3.1 Design Of The Second Backbone

For Deep Learning, Feature Extraction Is Always The Basis And Key To Detection, And How To Obtain More Accurate Features Is Always One Of The Hot Research Directions In This Field. In The Defect Detection Of Pcb Products, Most Of The Defects Have The Characteristics Of Large Ratio Of Length To Width, Similar Structure To The Circuit, Or Large Difference In Scale Between Each Defect. These Characteristics Cause Great Trouble To Feature Extraction Mitigation. To Make Matters Worse, We Need To Perform Precise Positioning In Tight Circuits And Delicate Electronic Components, Which Requires That The Features We Extract Must Be Highly Accurate And Effective. A Big Problem With Neural Networks Is That As The Number Of Layers Of The Network Deetens, Some Information Will Be Gradually Lost In The Process Of Extracting Target Features. This Is An Unavoidable Problem In Feature Extraction, But We Can Add Some Information Interaction Structures Between Feature Information At Different Scales To Alleviate This Loss. Therefore Our Network Needs To Be Targeted To Select The Current Mission-Critical Information From A Large Amount Of Information, Mine Deeper Feature Information And Ensure That As Little Information As Possible Is Lost In The Process Of Information Transfer. For This Reason, We Have Designed The Second Backbone By Combining The Sotayolov9 Idea [29] With The Cblinear And Cbfuse Modules To Increase The Richness Of Information In Different Channels, And Strengthen The Key Details That Will Be Lost In The Process Of Forward Propagation While Completing Information Fusion. The Structure Of The Second Trunk Is Shown In Figure 2.

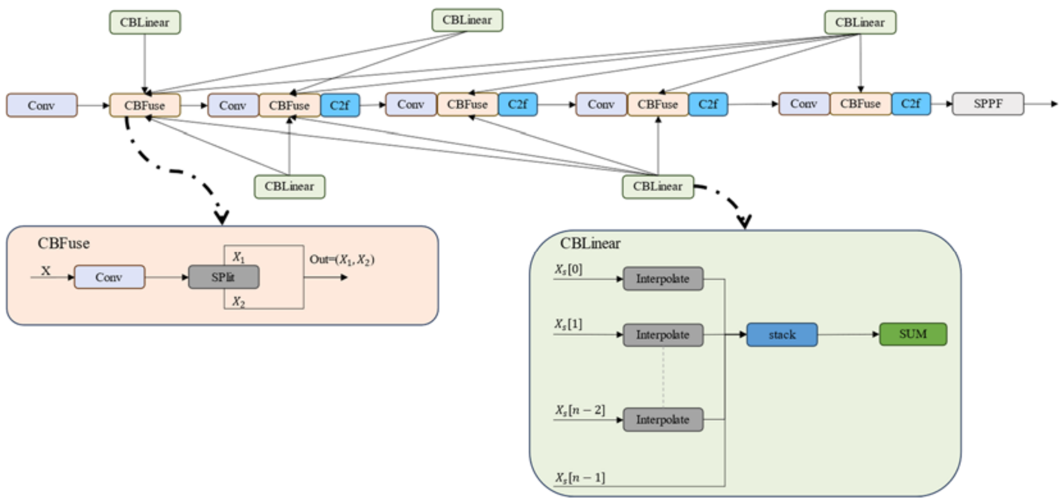

Figure 2 Structural Diagram Of The Second Backbone

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

3.2 CSPC MODULE

In real deployments with limited computing and storage resources, it is crucial to find the optimal configuration relationship between efficiency and performance. The CSPC module addresses this challenge by applying partial convolutions and a more concise structure, aiming replace the original C2f module in the model, which completes the lightweight improvement of the model. More importantly, unnecessary convolutions are reduced in the forward propagation process, the loss of effective features is reduced, and the final feature extraction is greatly improved, so that the model can complete the detection task faster and better. Their structure diagram is shown in FIG. 3. In the feature map of a convolutional neural network, there is a large amount of similar or duplicated information between different channels, which is referred to as feature map redundancy. In many cases, some channels of the feature map may contain features that are highly similar to other channels, which means that when performing forward propagation of the network, multiple processing of this information does not provide additional useful information, but rather increases the cost of computation and memory access. Chen [30] et al. proposed partial convolution (PConv) in 2023, which, by applying filters to only a few channels, the while keeping the other channels unchanged, resulting in fast and efficient computation. A key step in implementing PConv is to define a binary mask that can be used to distinguish the validity of data points. Specifically, for each data point, if it is valid, it is labeled as 1 at the corresponding mask location; If a data point is missing, it is labeled as 0. When performing the convolution operation, both the original data and the mask need to be convolved. The kernel will only be applied to locations where the mask value is 1. In addition, the convolution results need to be normalized to ensure that the convolution results of different regions are comparable and avoid the deviation of convolution results caused by missing data. The traditional feature extraction module generates feature maps by stacking convolution kernels, which often leads to effective information going through the same operation multiple times, wasting computing resources, and invalid information also wastes part of storage space. In contrast, we use partial convolution to design the CSPC module, which can obtain new features from the original image through limited convolution, and then fuse these new features with the original image. The new features have filtered out most of the invalid information, and the features we need on the original image are strengthened in the fusion process, which can ensure that the small key information are retained and strengthened. The specific structure of the module is shown in FIG. 3. It not only reduces the number of parameters, but also ensures the diversity and richness of the feature map to meet the requirements of accurate detection.

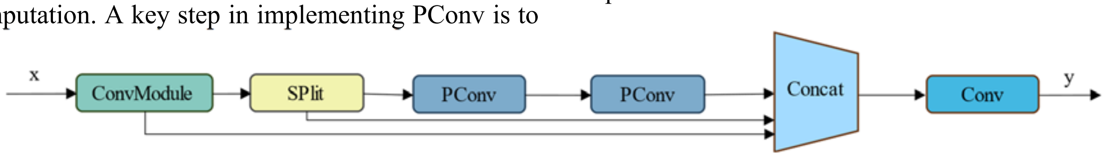

FIGURE 3 STRUCTURAL DIAGRAM OF CSPC MODULE

When the feature map is propagated in the CSCP module, it is first divided into multiple sub-feature maps after a convolution operation. These feature maps are concatenated with the original feature map without PConv in the channel dimension after two PConv operations, and finally the output result is output after a convolution. Stitching the feature map processed by PConv and the original feature map can fuse the feature map with more obvious details, avoiding a large loss of detail information in the convolution process. By shortening the forward propagation process and using lightweight convolution, this structure design reduces the training time and the use of computing resources, improves the training efficiency and performance, makes the model easier to deploy in the environment with limited computing resources, and brings new possibilities for the practical application of deep learning models. At the same time, the concat process fuses the overall feature map from before segmentation and the original sub-feature map after segmentation, so that the high-level features are obtained while the detailed features are retained on the new feature map.

After replacing the C2f module in the YOLOv8 model with the CSPC module, we test the amount of calculation and the number of parameters of the model, and compare the results with the YOLOv8 model. The specific data comparison is shown in FIG. 4. We use the data computed by the "calflops" tool. The number of model parameters is denoted by Mega (M). Model capacity is measured in giga (G).

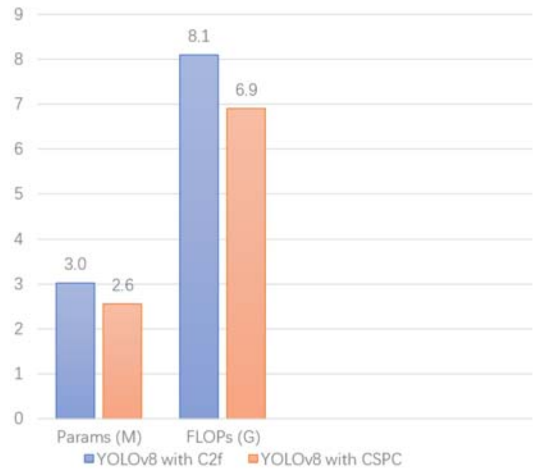

FIGURE 4 COMPARISON OF MODEL PARAMETERS AND FLOPS

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

3.3 C2F_IREMA MODULE

The attention mechanism can formulate different selection criteria to focus on specific information or deliberately ignore certain information, and directly invest the limited computing resources of the model into effective features or regions through the attention or inhibition of information. This method makes very effective use of computing resources, which not only strengthens the ability of the model to extract features, but also avoids invalid calculations and alleviates the pressure caused by limited computing power, so it has been widely used and recognized in the research and application of visual inspection. Attention mechanism has been a hot research topic since it was first proposed, and it is more widely used in the field of visual detection. Many scholars capture the features required by the corresponding detection tasks through different attention mechanisms, which significantly improve the detection effect of the model for small objects without increasing the computational burden and the size of the model. The squeeze-and-excitation block proposed by Hu et al. [31] has obtained better detection results in various detection fields, so it has received extensive attention. The module combines spatial information and channel information in the process of constructing feature information, and emphasizes the use of spatial information in all feature hierarchies to improve the detection ability. Another popular attention mechanism is the Convolutional Block Attention Module (CBAM) proposed by Woo et al. [32]. It combines channel attention and spatial attention mechanisms to highlight the relevance of channels and spatial locations. The channel attention module selectively emphasizes information channels, while the spatial attention module emphasizes important spatial locations. Through the combination of channel information and spatial information, the CBAM module deefies the model's understanding of image information, and greatly improves the model in various detection tasks, which fully proves the excellent universality of the module. Zhang [33] et al. combined the CNN-based IRB structure with the attention model and achieved good experimental results. In the subsequent research, they proposed the Single Residual Moving Block (MMB), which can effectively complete the lightweight process and simplify the model structure. A modern inverse residue moving block (iRMB) is derived based on simple but effective design guidelines, and its structure is shown in Fig. 5. IMB structure combines the lightweight nature of convolutional neural networks (CNNs) with the dynamic processing power of Transformer models to provide efficient performance in rings with limited computational power, and is therefore particularly suited for dense prediction tasks. Ouyang et al. [34] believe that the original features from the channel are more important, so his research team pays more attention to preserving the original information of the channel. In order to retain the original features to the maximum extent, they evenly distribute the spatial semantic features within each feature group to avoid a large loss of features, and thus propose a novel Efficient Multiscale Attention (EMA) module.

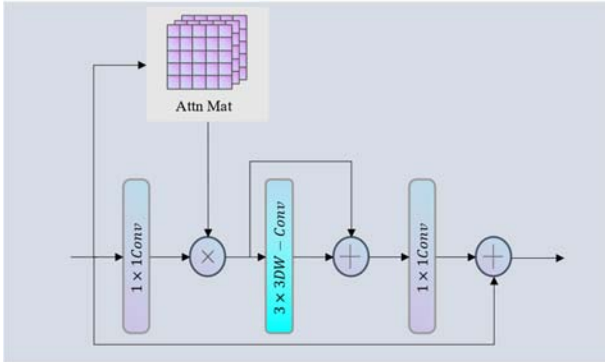

FIGURE 5 STRUCTURAL DIAGRAM OF IRMB

However, EMA module cannot establish the connection between the detailed features and the original features, and is easy to ignore the association between defects and the whole, resulting in misjudgment of similar structures. We draw lessons from iRMB module and EMAttention module and propose an iREMA module. This module can better help the model to complete the detection task when the discrimination between the detection target and the environment is not high and the detection target features are few. It is incorporated into the C2f module to design a new module C2f_iREMA, inverted residual multiscale attention module. Its structure is shown in Fig. 6. Firstly, an efficient multi-scale attention mechanism applies multiple parallel attention branch structures to focus on feature maps of different scales. After that, each branch calculates the attention weight through the convolution operation, so that the model pays more attention to the connection between the circuit part and the background. Finally, these feature maps are fused to obtain a multi-scale feature map representation. Then, the structure of the reverse residual module was applied to the module, and the convolution operation and batch normalization operation were put behind. Such a structure design can further compress the number of model parameters and improve the nonlinear expression ability. This module uses the inverse residual design of iRMB to improve the processing of information flow, and redistributes the weight of the channels in the parallel branch through the global information, and accurately obtains the pixel-level relationship in the cross-latitude information interaction.

This design allows the model to effectively distinguish the difference between part of the circuit and the overall background by using remote dependencies. By establishing this distinction, more attention can be paid to part of the circuit, and the model is lightweight.

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

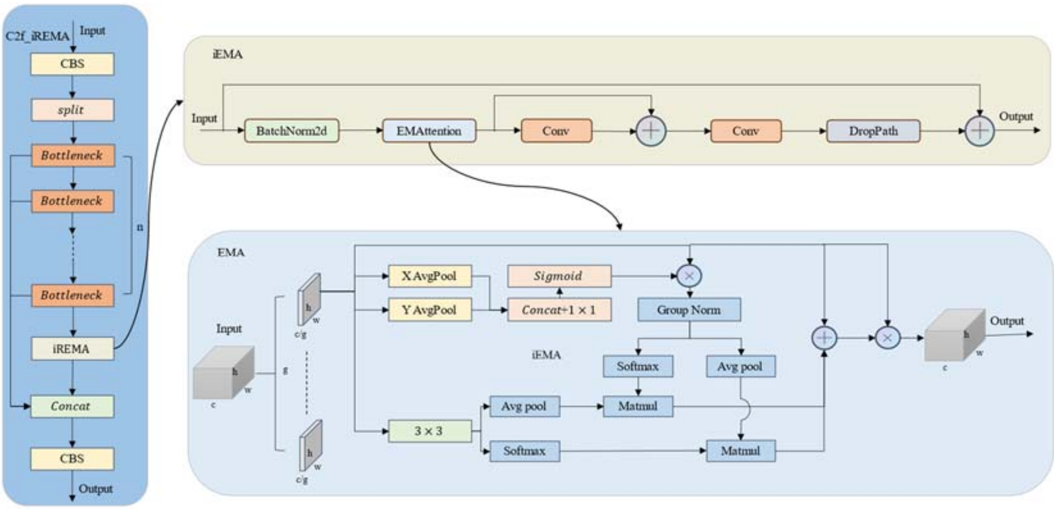

FIGURE 6 STRUCTURAL DIAGRAM OF IREMA MODULE

This design is a modular structure that can be integrated into various deep learning models. This combination allows the model to be more flexible to adapt to different scenarios, thus improving the adaptability and scalability of the model.

4 EXPERIMENTS AND RESULTS ANALYSIS

This paper introduces the work done in this paper, and compares our innovative method with some existing public methods, including the data source of PCB defects, experimental configuration, evaluation index, ablation experiment and comparison experiment.

4.1 DATA SET

The dataset used in the experiment is a public dataset released by Peking University, which contains 693 images and 6 defects with an average pixel size of 2777 × 2188 pixels. In this experiment, we set the partition ratio of training set to test set as 9:1. the defect types and numbers are shown in Table 1, and the defect samples are shown in Fig. 7.

TABLE 1 DEFECT SAMPLE STATISTICS

Type of defects	Number of images	Number of defects
Missing hole	115	497
Mouse bite	115	492
Open circuit	116	482
Short	116	491
Spur	115	488
Spurious copper	116	503
Total	693	2953

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

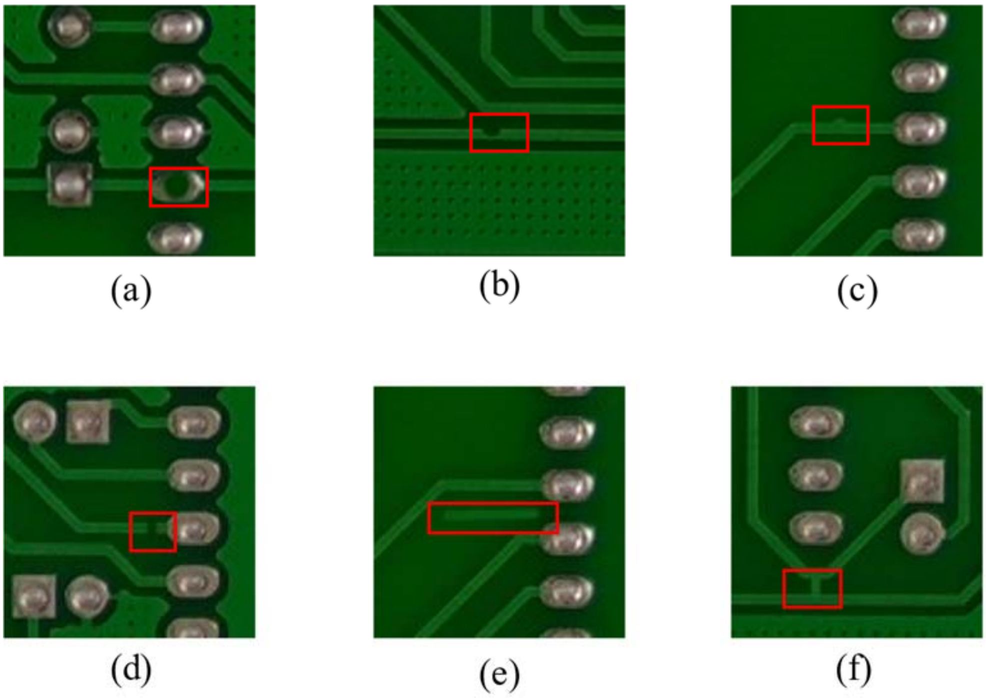

FIGURE 7 PCB DEFECT SAMPLE DIAGRAM (A)MISSING_HOLE (B)MOUSE_BITE (C)SPUR (D)OPEN_CIRCUIT (E)SPURIOUS_COPPER (F)SHORT

4.2 EXPERIMENTAL ENVIRONMENT

The hardware part of the experimental environment is under windows operating system and all the experiments are carried out under Pytorch framework with torch 2.1.1, CUDA 11.1 and programming language python 3.8. All the environment configurations for the experiments are shown in Table 2.

TABLE 2 EXPERIMENTAL ENVIRONMENT CONFIGURATION

Experimental environment	Value
Operating system	Windows 11
Deep learning framework	torch 2.1.1/cuda 11.1
Programming language	Python 3.8
CPU	Inter(R) Core(TM) i5-13400F 2.50 GHz
GPU	RTX 3060 12G
RAM	16G
image size	640×640
Batch size	4
Epoch	200
Initial learning rate	0.01
Optimizer	SGD

4.3 EVALUATION INDICATORS

There are many model performance metrics commonly used for object detection tasks, Here, we evaluate our model using three metrics commonly used by researchers: recall, precision, average precision, and mean average precision (mAP). Recall is the proportion of the actual number of positive samples out of all target samples detected. Precision is the number of positive samples detected as a proportion of all actual positive samples. These two metrics often interact with each other. When recall is high, it leads to lower detection accuracy and false detections. However, when precision is high, it leads to lower recall and missed detections.

The formulas for recall, precision, AP and mAP are as follows.

$$Recall = \frac{T_P}{T_P + F_N} \quad (1)$$

$$Precision = \frac{T_P}{T_P + F_P} \quad (2)$$

$$AP = \int_0^1 P(r)\,dr \quad (3)$$

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (4)$$

In the above equation, TP represents the number of positive samples predicted as positive by the model, FP represents the number of negative samples predicted as positive by the model, i.e., false detection, and FN represents the number of positive samples predicted as negative by the model, i.e., missed detection. n is the total number of detected categories, and N is set to 6 in this paper, i.e., leakage holes, mouse bites, open circuits, short circuits, stray circuits, and stray copper. ap is the area surrounded by PR curves, and mAP is the average value of detected categories of AP, and the larger mAP is, the better the algorithm performance. AP is the area surrounded by the PR curve, mAP is the average of the detected AP categories, the larger mAP represents the better performance of the algorithm. FPS represents the model detection speed. We expect the FPS of the model to be as large as possible.

4.4 *Experiments And Analysis Of Results*

In order to verify the enhancement effect of our proposed method on the model, we conduct an ablation comparison experiment, which improves the model by adding the second backbone network, the CSCP module, and the C2f_iREAM in turn while keeping all other conditions the same, and the specific experimental data and results are listed in Table 3

Table 3 Results Of Ablation Experiments Using YOLOv8 And Our Proposed Methodology

DB	CSPC	iREMA	mAP(0.5)	P	R	GFLOPS	Para	FPS
-	-	-	93.8%	96.7%	90.4%	8.1	3.0	181.8
√	-	-	95.7%	93.7%	93%	14.8	5.0	84.7
√	√	-	96.6%	97.1%	92.9%	13.5	4.5	78.7
√	√	√	97.4%	96.7%	93.9%	13.6	4.6	86.2

We add the modules one by one to the YOLOv8 model for experimentation. The experimental results are specifically listed in Table 3. According to the experimental results, it can be concluded that each module has an improvement effect on the model mAP, and there is no conflict between modules that makes the mAP drop; after adding the CSCP module, the mAP is improved by a small amount of 0.9%, while the GFLOPS is reduced by 8.878%, and the number of parameters is decreased by 10%, which proves that the CSPC module effectively reduces the number of parameters in the model; after adding the C2f_iBEMA module, the mAP is improved by 0.8%, and the FPS is improved by 7.5, which indicates that this module has some advantages in reducing the amount of calculation, and for practical applications, it improves the detection speed.

Table 4 Each Type Of Defect Detection Accuracy

DB	CSPC	iREMA	Missing hole	short	spur	Mouse bite	Spurious copper	Open circuit
-	-	-	0.995	0.995	0.844	0.919	0.939	0.936
√	-	-	0.995	0.995	0.875	0.946	0.948	0.983
√	√	-	0.995	0.995	0.931	0.949	0.960	0.969
√	√	√	0.995	0.995	0.920	0.963	0.995	0.976

As shown in Table 4, the detection accuracy has increased when the modules we designed are added in turn. Compared with the original model, the second backbone structure deeps the information interaction ability, and can obtain the relationship between more defect parts and the overall structure, so that it can effectively distinguish structurally similar defects, and also has a certain improvement effect on two types of defects: burr and mouse-bite. The CSCP module shorts the forward propagation process, effectively retains the original characteristics of the defect while reducing the amount of calculation, and reduces the loss of information in the propagation process. For smaller defects, such as burrs, it has a more obvious improvement, but also because the original characteristics are too much retained. These features may make some features of larger defects ineffective or even beneficial after information interaction. For example, after adding CSPC module, the detection accuracy of circuit open defects decreases greatly. Despite this, the introduction of this module makes the parameters of dual backbone model decrease by 10%, and the GFLOPS decrease by 8.878%. It has a good lightweight effect. The C2f_iREMA module uses the inverted residual structure and efficient multi-scale attention mechanism to reduce the information interference caused by background similarity. This design makes the model effectively use long-distance dependence and enhance the correlation between the defect part and the whole while increasing the parameters and model size by a small amount, which can effectively improve the detection accuracy of structurally similar defects and improve the detection speed to a certain extent.

Journal of http://www.satursonpublishing.com/
Research in Multidisciplinary Methods and Applications
ISSN 2957-3920 (Online) ISSN 3007-7060 (Print)
Volume 4, Issue 2, February 2025

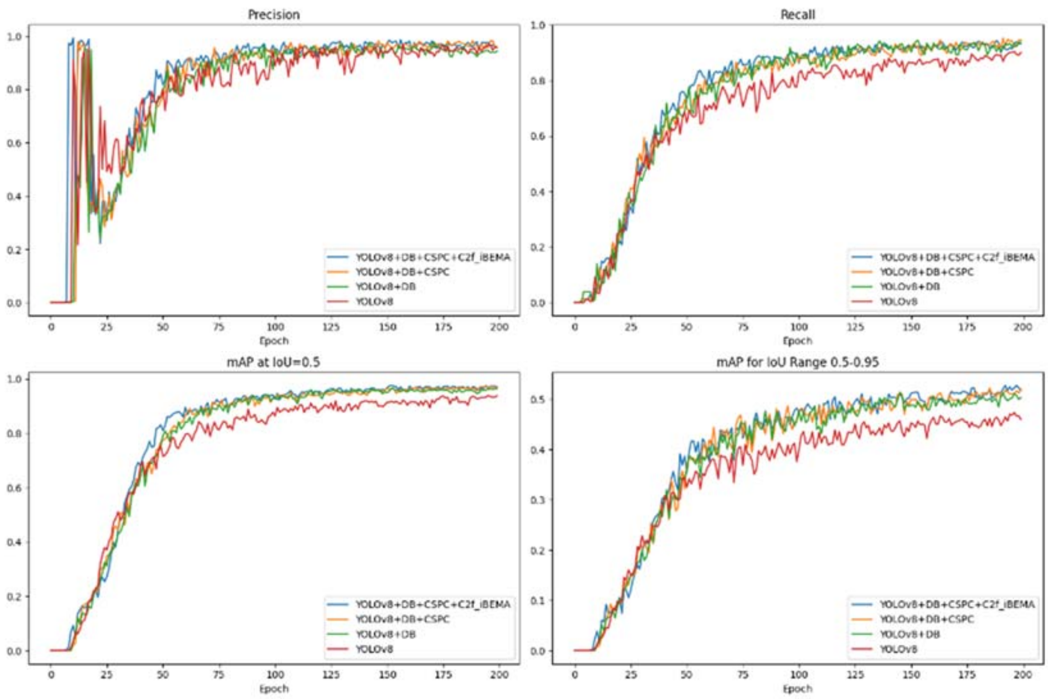

FIGURE 8 METRICS CURVES OF DIFFERENT MODULES

In FIG. 8, the loss change curve and the training process graph change curve of YOLOv8 and our proposed method are shown. According to the curve in the figure, it can be roughly determined that each index increases significantly with the increase of training times before 150 epochs, and the increase rate decreases after 150 epochs, and there is no significant increase of each index after 175 epochs. According to the change of the loss function curve, the loss function of the model is basically in the convergence state after 175 epochs, and the growth of the map change curve also reaches the growth limit at 175 epochs, which is stable at about 90%. After 175 epochs, the map value is basically smooth and no longer rises with the increase of training generations. It can be concluded that after 175 generations of training, the performance of the model reaches the optimal.

4.5 *COMPARATIVE EXPERIMENTS WITH DIFFERENT MODELS*

Under the same experimental conditions, we tested three models, YOLOv5-s, YOLOv9-s and YOLOv10-s, and obtained the relative comparison between the experimental data and the experimental results of DB-YOLO, as shown in Figure 4. The main detection data are accuracy, model size and detection speed.

TABLE 5 COMPARISON OF EXPERIMENTAL RESULTS

Model Name	Map	R	Model Volume/MB	FPS
YOLOv5-s	92.5%	72.2%	7.08	72.69
YOLOv9-s	95.1%	90.9%	9.79	70.40
YOLOv10-s	92.9%	86.3%	8.04	87.72
DB-YOLOv8	97.4%	93.9%	4.6	86.2

DB-YOLO model is compared with several other popular models in this field. Judging from the experimental data, our model has a lot of improvement in detection accuracy, and the model size is greatly reduced, which can better adapt to the environment with limited hardware conditions. Table 5 shows the specific results of the experiments performed. From the table, we can see that the best performance is the YOLOv9 model. The map value of the DB-YOLO model was 2.3 percentage points higher than that of the YOLOv9 model. It is 3 percentage points higher than the YOLOv9 model in R-value. In addition, the model size of DB-YOLO is much smaller than these models, which makes it easier to deploy in real environments. Compared with the first three indicators, the detection speed of DB-YOLO model is slightly slower than that of YOLOv10 model, but it is still much higher than the other two.

5 CONCLUSIONS

In this paper, we propose a new dual-backbone architecture consisting of CBLinear and CBFuse modules. The proposed structure can effectively retain the original information, better extract the effective features of the target, increase the information exchange ability between feature maps of different sizes, and effectively improve the detection accuracy. On the other hand, we propose a CSPC module that reduces the training time and the use of computational resources, improves the training efficiency and performance, and makes the model easier to be deployed in environments with limited computational resources. Finally, we combine the iRMB module and the EMAttention module to design an iREMA module that improves the information flow, which is integrated with the C2f module. This module improves the model's detection accuracy for missing parts of the circuit, such as mouse bites and open circuits.

FUNDING

Guangzhou Key Research and Development Plan - Artificial Intelligence Major Project, project number: SL2022B01J00019

REFERENCES

[1]Ren, Z., et al.: State of the Art in Defect Detection Based on Machine Vision. International Journal of Precision Engineering and Manufacturing-Green Technology. 9 (2), pp. 661-691(2022).http://dx.doi.org/10.1007/s40684-021-00343-6

[2]Sankar, V.U., et al.: A Review of Various Defects in Pcb. Journal of Electronic Testing. 38 (5), pp. 481-491(2022).http://dx.doi.org/10.1007/s10836-022-06026-7

[3]Zhou, Y., et al.: Review of Vision-Based Defect Detection Research and Its Perspectives for Printed Circuit Board. Journal of Manufacturing Systems. 70, pp. 557-578(2023).http://dx.doi.org/https://doi.org/10.1016/j.jmsy.2023.08.019

[4]Wang, F., et al.: Laser-Induced Thermography: An Effective Detection Approach for Multiple-Type Defects of Printed Circuit Boards (Pcbs) Multilayer Complex Structure. Measurement. 206, p. 112307(2023).http://dx.doi.org/https://doi.org/10.1016/j.measurement.2022.112307

[5]Bomgni, A.B., et al.: Multihop Uplink Communication Approach Based on Layer Clustering in Lora Networks for Emerging Iot Applications. Mobile Information Systems. 2023 (1), p. 5828671(2023).http://dx.doi.org/https://doi.org/10.1155/2023/5828671

[6]Liu, Y., et al.: A Survey of Real-Time Surface Defect Inspection Methods Based on Deep Learning. Artificial Intelligence Review. 56 (10), pp. 12131-12170(2023).http://dx.doi.org/10.1007/s10462-023-10475-7

[7]Wu, X., et al.: Recent Advances in Deep Learning for Object Detection. Neurocomputing. 396, pp. 39-64(2020).http://dx.doi.org/https://doi.org/10.1016/j.neucom.2020.01.085

[8]Ren, S., et al.: Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39 (6), pp. 1137-1149(2017).http://dx.doi.org/10.1109/TPAMI.2016.2577031

[9]Joseph, R., et al.: You Only Look Once: Unified, Real-Time Object Detection. arXiv - CS - Computer Vision and Pattern Recognition. (2015).http://dx.doi.org/arxiv-1506.02640

[10]Guohua, L., et al.: Printed Circuit Board Defect Detection Based on Mobilenet-Yolo-Fast. Journal of Electronic Imaging. 30 (4), p. 043004(2021).http://dx.doi.org/10.1117/1.JEI.30.4.043004

[11]Li, Y.T., et al.: Automatic Industry Pcb Board Dip Process Defect Detection System Based on Deep Ensemble Self-Adaption Method. IEEE Transactions on Components, Packaging and Manufacturing Technology. 11 (2), pp. 312-323(2021).http://dx.doi.org/10.1109/TCPMT.2020.3047089

[12]Niu, J., et al.: An Improved Yolov5 Network for Detection of Printed Circuit Board Defects. Journal of Sensors. (2023).http://dx.doi.org/10.1155/2023/7270093

[13]Liu, Z., et al.: Machine Vision Based Online Detection of Pcb Defect. Microprocessors and Microsystems. 82, p. 103807(2021).http://dx.doi.org/https://doi.org/10.1016/j.micpro.2020.103807

[14]Gaidhane, V.H., et al.: An Efficient Similarity Measure Approach for Pcb Surface Defect Detection. Pattern Analysis and Applications. 21 (1), pp. 277-289(2018).http://dx.doi.org/10.1007/s10044-017-0640-9

[15]Zhang, F., et al.: A Pcb Photoelectric Image Edge Information Detection Method. Optik. 144, pp. 642-646(2017).http://dx.doi.org/https://doi.org/10.1016/j.ijleo.2017.07.002

[16]Cao, G., et al.: Large-Complex-Surface Defect Detection by Hybrid Gradient Threshold Segmentation and Image Registration. IEEE Access. 6, pp. 36235-36246(2018).http://dx.doi.org/10.1109/ACCESS.2018.2842028

[17]Brian LiDarwin, A.I., et al.: Pcbdet: An Efficient Deep Neural Network Object Detection Architecture for Automatic Pcb Component Detection on the Edge. arXiv - CS - Computer Vision and Pattern Recognition. (2023).http://dx.doi.org/arxiv-2301.09268

[18]Wei, L., et al.: Ssd: Single Shot Multibox Detector. arXiv - CS - Computer Vision and Pattern Recognition. (2015).http://dx.doi.org/arxiv-1512.02325

[19]Yuan, M., et al.: Yolo-Hmc: An Improved Method for Pcb Surface Defect Detection. IEEE Transactions on Instrumentation and Measurement. 73, pp. 1-11(2024).http://dx.doi.org/10.1109/TIM.2024.3351241

[20]Xuan, W., et al.: A Lightweight Modified Yolox Network Using Coordinate Attention Mechanism for Pcb Surface Defect Detection. IEEE Sensors Journal. 22 (21), pp. 20910-20920(2022).http://dx.doi.org/10.1109/JSEN.2022.3208580

[21]Yuan, Z., et al., Lw-Yolo: Lightweight Deep Learning Model for Fast and Precise Defect Detection in Printed Circuit Boards', Symmetry, (2024)

[22]Chen, P., et al., A Machine Learning Approach for Automated Detection of Critical Pcb Flaws in Optical Sensing Systems', Photonics, (2023)

[23]Xia, K., et al.: Global Contextual Attention Augmented Yolo with Convmixer Prediction Heads for Pcb Surface Defect Detection. Scientific Reports. 13 (1), p. 9805(2023).http://dx.doi.org/10.1038/s41598-023-36854-2

[24]Long, Y.: Pcb Defect Detection Algorithm Based on Improved Yolov8. Academic Journal of Science and Technology. 7 (3), pp. 297-304(2023).http://dx.doi.org/10.54097/ajst.v7i3.13420

[25]Wang, Y., et al.: An Improved Yolov8 Algorithm for Rail Surface Defect Detection. IEEE Access. 12, pp. 44984-44997(2024).http://dx.doi.org/10.1109/ACCESS.2024.3380009

[26]Liu, L., et al.: A Wind Turbine Damage Detection Algorithm Designed Based on Yolov8. Applied Soft Computing. 154, p. 111364(2024).http://dx.doi.org/https://doi.org/10.1016/j.asoc.2024.111364

[27]Li, W., et al.: Rca: Yolov8-Based Surface Defects Detection on the Inner Wall of Cylindrical High-Precision Parts. Arabian Journal for Science and Engineering. (2024).http://dx.doi.org/10.1007/s13369-023-08483-4

[28]Tang, J., et al.: A Lightweight Surface Defect Detection Framework Combined with Dual-Domain Attention Mechanism. Expert Systems with Applications. 238, p. 121726(2024).http://dx.doi.org/https://doi.org/10.1016/j.eswa.2023.121726

[29]Chien-Yao, W., et al.: Yolov9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv - CS - Computer Vision and Pattern Recognition. (2024).http://dx.doi.org/arxiv-2402.13616

[30]Jierun, C., et al.: Run, Don't Walk: Chasing Higher Flops for Faster Neural Networks. arXiv - CS - Computer Vision and Pattern Recognition. (2023).http://dx.doi.org/arxiv-2303.03667

[31]Hu, J., et al.: Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 42 (8), pp. 2011-2023(2020).http://dx.doi.org/10.1109/TPAMI.2019.2913372

[32]Sanghyun, W., et al.: Cbam: Convolutional Block Attention Module. arXiv - CS - Computer Vision and Pattern Recognition. (2018).http://dx.doi.org/arxiv-1807.06521

[33]Jiangning, Z., et al.: Rethinking Mobile Block for Efficient Neural Models. arXiv - CS - Computer Vision and Pattern Recognition. (2023).http://dx.doi.org/arxiv-2301.01146

[34]Daliang, O., et al.: Efficient Multi-Scale Attention Module with Cross-Spatial Learning. arXiv - CS - Computer Vision and Pattern Recognition. (2023).http://dx.doi.org/arxiv-2305.13563