International Core Journal of Engineering Volume 12 Issue 3, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

YOLO-RSH: A PCB Surface Component Detection Model

Changjuan Bai

School of Information Science, Yunnan Normal University, Kunming, China

Abstract

Printed Circuit Boards (PCBs) are an essential component of modern electronic products. The surface components of PCBs exhibit characteristics such as multi-scale, dense distribution, high similarity, and small size, which can lead to missed or false detections during inspection. This paper proposes a novel YOLO-RSH detection model based on YOLOv11. First, a new C2SCSA attention module is introduced to replace the C2PSA module in the backbone network, which enhances the network's feature extraction ability. Next, the newly proposed C3k2 convolution module in YOLOv11 is augmented with the RepVGG mechanism, forming the C3k2-RepVGG convolution, which improves the backbone's focus on important features. Finally, a hierarchical feature fusion block (HFFB) is added before the detection head, which better integrates multi-scale target features. Experimental results show that compared to the baseline model, the YOLO-RSH model achieves an average precision improvement of 1.8 percentage points.

Keywords

PCB; YOLOv11; Small Object Detection; Feature Extraction; Feature Fusion.

1. Introduction

Printed Circuit Board (PCB) is the backbone of modern electronic information industries, carrying electronic components like integrated circuits, resistors, capacitors, etc. The rapid development of electronic technology has led to the miniaturization, integration, and diversification of PCBs [1]. It is expected that in the future, PCBs will become increasingly smaller, more complex, and specialized. Surface Mount Technology (SMT) has enabled the high-density and high-speed automatic assembly of components. During the manufacturing and assembly process, the rapid and accurate identification and positioning of PCB components are becoming increasingly important [2]. With advancements in technology and the increasing complexity of PCBs, the limitations of manual inspection are becoming more evident. No matter how skilled the human operator is, it is easy to make mistakes when faced with the intricate designs, miniaturized parts, and densely packed components on the circuit boards, which also leads to reduced efficiency. Therefore, the ability to automatically detect different electronic components on the PCB to achieve automated inspection holds significant value. Component detection can be performed after the components have been mounted by comparing the actual position and category of the components with a standard template, allowing for the precise identification of category misplacement or position errors. Furthermore, identifying components on the PCB helps in recycling undamaged components, which is meaningful for environmental protection and resource conservation.

There are two main types of detection methods for electronic components: traditional machine learning algorithms and deep learning-based algorithms. In traditional machine learning algorithms, Su Mingming [3] proposed a PCB electronic component detection method based on digital image processing, starting from pattern recognition and image processing. Crispin [4] and others used genetic algorithms and gray model fitting to generate a set of general templates for accurate component localization and identification. Yin [5] used a multi-level image matching approach for

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

PCB component detection. First, PCB images are matched, then coarse matching is used to find areas similar to the standard components, and finally, fine matching is done to determine if the target component is present.

With the rapid development of neural networks, the application of deep learning algorithms in PCB component detection or defect detection has gradually become the mainstream approach. Deep learning-based object detection algorithms are mainly divided into single-stage algorithms and two-stage algorithms. The single-stage algorithms, such as the YOLO series [6-17] (You Only Look Once) and SSD [18], calculate the confidence and position coordinates of candidate boxes while creating them. A single detection is sufficient to obtain the results. On the other hand, two-stage algorithms separate region detection from object detection. Typically, they first use algorithms for region detection on images and then perform object detection in the proposed areas. Representative two-stage detection algorithms include R-CNN [19], FAST-RCNN [20], etc., which are used to generate candidate boxes and then classify the generated candidate boxes.

In summary, many research achievements have been made in the field of PCB component detection. However, the currently widely used object detection algorithms often miss small targets or misidentify components with high similarity during detection. Therefore, this paper is based on the YOLO11 model and improves the original network architecture to make it more suitable for the field of PCB component detection. The innovations in this paper include:

(1)Proposing a novel C2SCSA attention module: A new attention mechanism module is designed to focus more attention on the target area, effectively reducing noise interference, thus improving the network's detection capability for small targets.

(2)Hierarchical Feature Fusion Module (HFFB): A hierarchical feature fusion module is added before the detection head to better fuse information from different layers.

(3)Proposing a new C3k2-RepVGG module: This module is able to focus on important spatial features while reducing the number of channels, thus enhancing the model's detection capabilities.

2. Methods

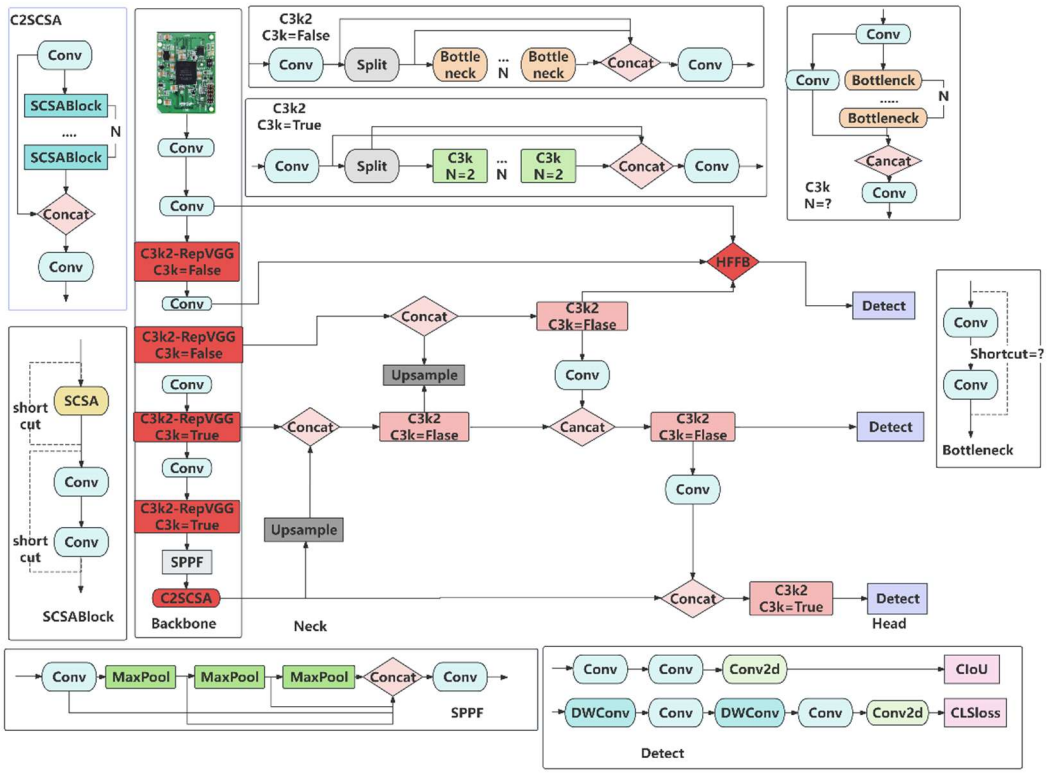

Fig. 1 YOLO-RSH

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

The architecture diagram of the YOLO-RSH network is shown in Fig. 1.the red module is the newly added module.

2.1 C2SCSA

The C2PSA module proposed in YOLO11n is a general and efficient attention module, designed to balance computational cost and performance for general object detection tasks. However, in PCB component recognition, the C2PSA module has a fixed receptive field, which makes it less adaptable to components with large scale variations (such as large ICs and tiny resistors coexisting on the PCB). Additionally, the components on the PCB are densely arranged and highly similar, so the C2PSA module is prone to misclassifying densely packed small components as background noise. To reduce noise interference and address the issue of insufficient receptive fields or incomplete information in multi-scale component detection, this paper integrates the SCSA module with the C2PSA module proposed in YOLO11 to introduce a new module called C2SCSA, as shown in Fig 2.

The C2SCSA module combines the efficient CSP structure with the SCSA [21] attention mechanism. The CSP structure splits the input channels, with one channel retaining the original feature information to ensure direct gradient backpropagation, while the other channel uses the SCSA module for feature extraction. This design alleviates issues of gradient vanishing and feature redundancy. Spatial and Channel Spatial Attention (SCSA) is an efficient attention mechanism, consisting of two parts: the Shared Multi-semantic Spatial Attention (SMSA) and the Progressive Channel Word Attention (PCSA). The SMSA utilizes multi-scale depth-shared 1D convolutions to capture multi-semantic spatial information, enhancing both local and global feature representations. The PCSA refines channel features using self-attention based on the input features from SMSA, effectively mitigating semantic differences and ensuring robustness across channels.

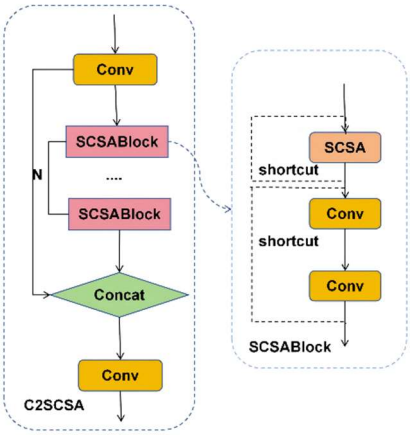

Fig. 2 The C2SCSA module

The SMSA module processes the input feature map in three steps. First, the feature map is averaged along both the height and width dimensions, then these pooled features are divided into four independent sub-features. Next, 1D depth convolutions with kernels of size 3, 5, 7, and 9 are used to capture features at different scales. These sub-features from the two sequences are aligned through shared convolution kernels. Finally, group normalization (GN) is applied to normalize features at different levels, and the multi-semantic spatial attention information is generated through a Sigmoid activation function. The structure diagrams can be found in Figs 3 and 4.

In the SMSA phase, multi-semantic spatial information is extracted from each feature to provide precise spatial priors for the subsequent PSCA phase. The PSCA phase refines the semantic understanding of local sub-features using the overall feature map, alleviating semantic discrepancies introduced by multi-scale convolutions in the former phase. The SCSA attention mechanism strikes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

a good balance between computational cost and accuracy. The SMSA phase in the SCSA module helps suppress interference from the complex background of the PCB, while also enabling the extraction of multi-scale spatial information. The subsequent PCSA phase enhances the channel feature representation ability, improving the distinction and consistency of key features and boosting the robustness of the attention mechanism.

With directionally decomposed multi-scale spatial attention and spatial-prior-based channel semantic refinement, the C2SCSA module can achieve more precise feature focusing and background suppression for small, multi-scale objects with high similarity on the PCB. This enhances the model's accuracy in recognizing PCB components.

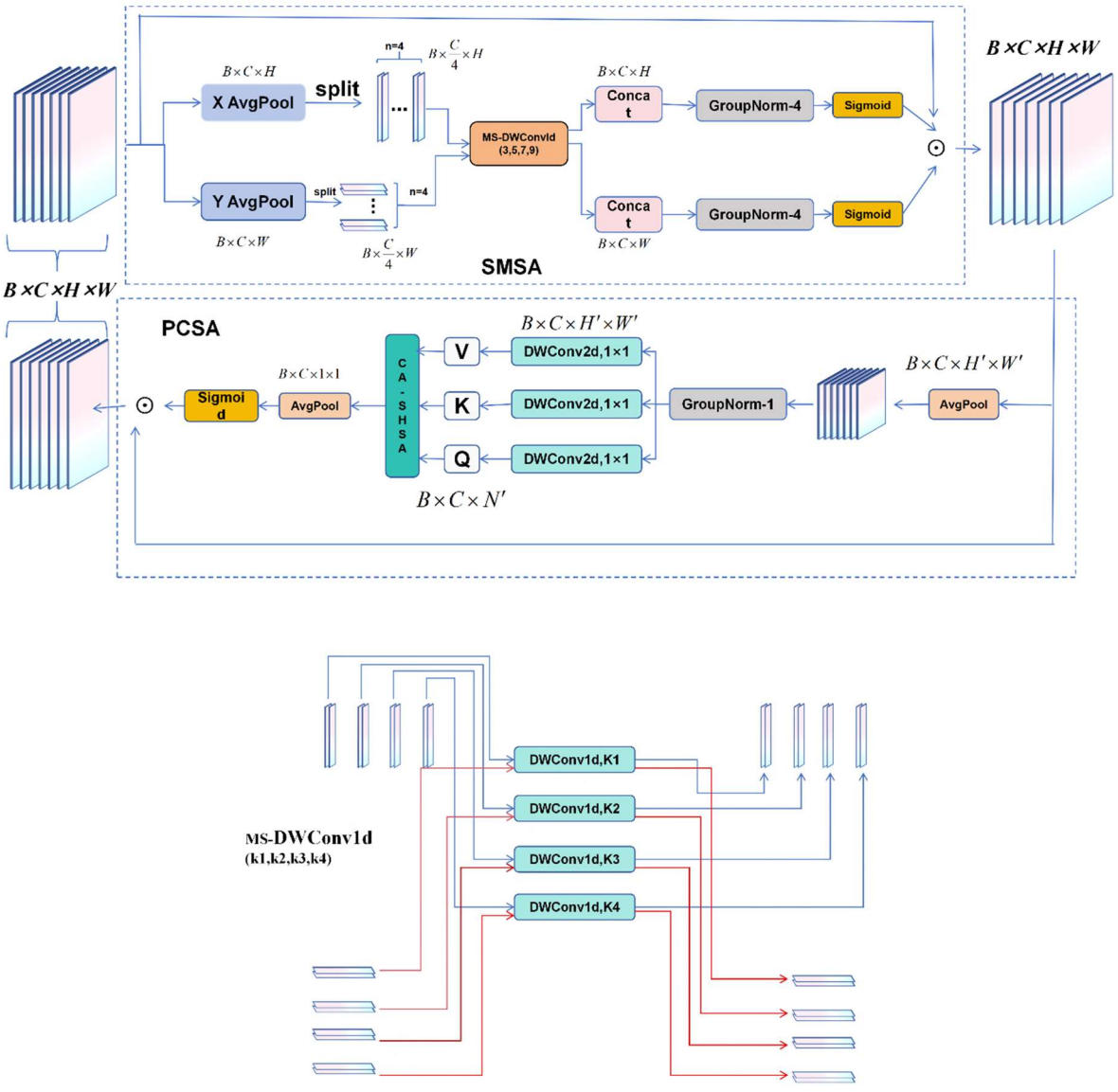

Fig. 3 and **4.** The SCSA module

2.2 HFFB

YOLO11n uses PANet as the neck network, where the feature fusion method is relatively fixed, and its adaptive capability for features is weak. During multiple up-sampling and down-sampling processes, small target information may be lost, resulting in insufficient reconstruction of small target details.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

In the YOLO-RSH network, the HFFB [22] module is introduced. The inputs of HFFB come from the 1st, 3rd, and 16th layers of the network. Among these, the features extracted from the 1st layer convolution represent the most primitive features of the image, the 3rd layer represents low-level features extracted through the C3k2-RepVGG convolution module, and the 16th layer contains small target features with high-level semantic information. Using the unique function of this module, the features extracted from these three different layers are efficiently fused. This structure allows the model to understand image features from coarse to fine, from global to local, and then fuse them with features from the previous stage. This enables the model to capture both macro features and focus on the subtle differences of small targets. It provides stronger recognition ability for densely arranged small targets on PCBs, such as resistors, capacitors, and other components.

The HFFB module is an adaptive hierarchical feature fusion block, capable of integrating features obtained from different layers. The adaptive hierarchical feature fusion block can dynamically merge local features, global representations, and fused semantic information from previous stages based on the input features, as shown in Fig 5. Here, Gi represents the feature matrix output by the global feature block, Li represents the feature matrix output by the local feature block, Fi−1 represents the feature matrix output by the previous stage HFFB, and Fi represents the feature matrix fused by the current stage HFFB. The self-attention mechanism in the global feature block can capture both spatial and temporal global information. To some extent, the HFFB module feeds the incoming global features into the channel attention (CA) mechanism. This mechanism improves the representation of specific semantic features by utilizing the interdependence between channel mappings. The local features are input into the spatial attention (SA) mechanism to enhance local details and suppress irrelevant regions. Finally, the results generated by each attention mechanism are fused with the feature fusion path and connected to the residual inverse MLP (IRMLP). This helps avoid issues like gradient vanishing, explosion, and network degradation, effectively capturing both global and local feature information from each layer.

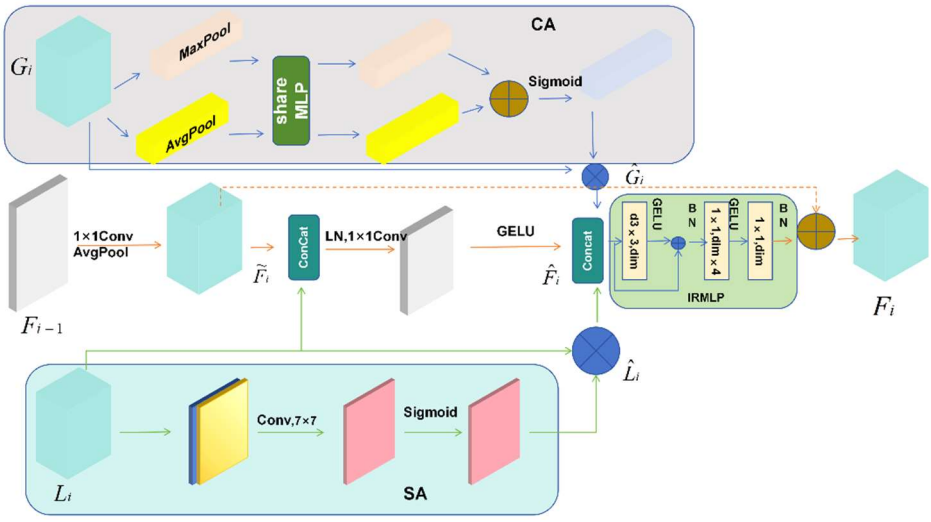

Fig. 5 The HFFB module

2.3 C3k2-RepVGG

The C3k2 module in the YOLO11n model is the core convolutional component of the network. It is a static module, which may not fully capture the fine features of PCB components or micro-components.

In YOLO11-RSH, a new C3k2-RepVGG convolution component is introduced. By employing a multi-branch structure during training, it provides stronger representational power and optimization flexibility, allowing the model to learn finer and more robust features of PCB components.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

Additionally, through re-parameterization techniques, the model's accuracy can be improved without adding computational delays, as shown in Fig 6.

The C3k2-RepVGG module is an efficient feature extraction component that combines cross-stage partial connections of C3k2 with the RepVGG[23] reparameterized structure. This module adopts a dual-branch design: one branch serves as an identity mapping, while the other consists of multiple cascaded R-Bottleneck or C3k sub-modules. Finally, features are fused through concatenation and convolution. By introducing the RepVGG structure, the multi-branch topology enhances feature expression during the training phase, while in the inference phase, the structure is re-parameterized into a single 3×3 convolution, balancing performance and efficiency.

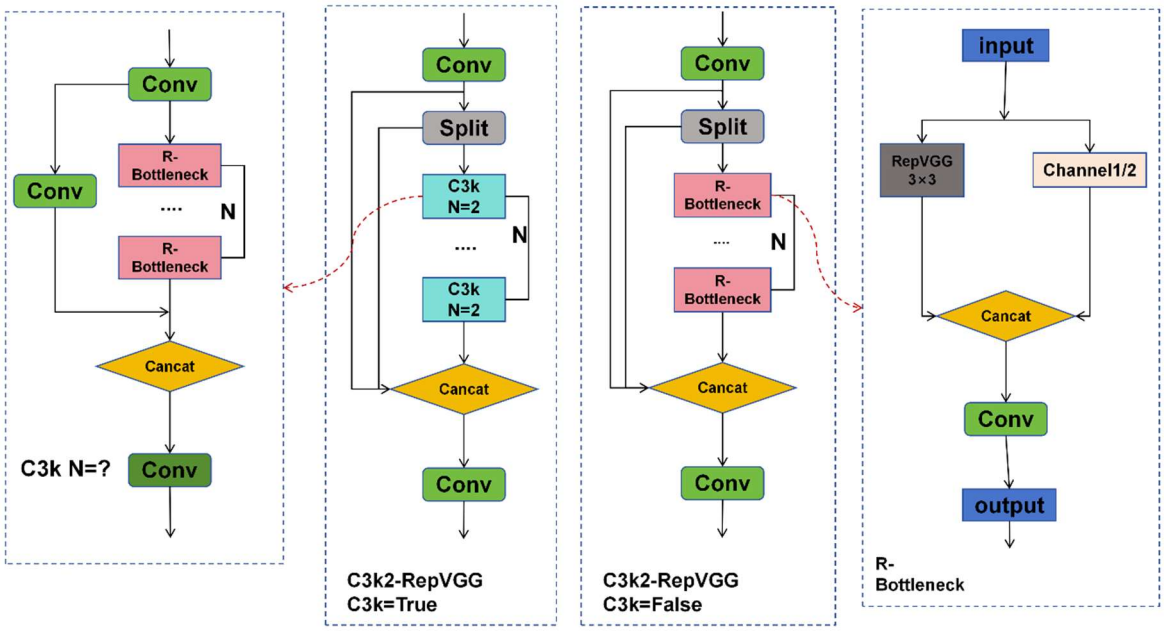

Fig. 6 The C3k2-RepVGG module

3. Experiments and Results

3.1 Dataset Details

Currently, there are several publicly available PCB datasets that can be used for PCB component detection. Chia-Wen Kuo and others not only provided new ideas for the development of PCB research but also introduced a public dataset, pcb_wacv_2019, which contains 47 PCB images. Some images are sourced from the internet, while others are captured by industrial cameras or DSLR cameras. The images are annotated with various PCB components, including resistors, ICs, capacitors, etc. The PCB_METAL dataset [24] was created by C. Pramerdorfer and others, using a DSLR camera under different lighting conditions. This dataset consists of 984 images from 123 PCB boards, recording the locations of various component types, including integrated circuits, capacitors, resistors, and inductors. However, the distribution of components in this dataset is limited, with most images containing fewer than 20 integrated circuits and capacitors. The PCB_DSLR dataset [25] records the locations of integrated circuits in 165 PCBs and annotates components in some images. Due to resolution limitations, smaller components such as resistors and capacitors are not included. Although the aforementioned datasets have large data scales, they cannot truly simulate the wide variations that could challenge PCB detection performance in the real world.

After reviewing numerous papers, it was found that the FICS_PCB dataset [26] is currently one of the most authoritative datasets for PCB component detection. Therefore, the FICS_PCB dataset was selected for training and evaluating the model. This dataset consists of 9,912 images from 31 PCB

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

samples and includes 77,347 annotated components. The images were captured with a DSLR camera and under a microscope. It contains four color variations (green, red, blue, black) and includes 77,347 PCB components (ICs, capacitors, resistors, inductors, transistors, diodes). The dataset contains annotations for six component types, as shown in Fig 7. The four PCB sample datasets are listed in Table 1. During training, the dataset is divided into training, validation, and test sets in a ratio of 7.5:1.5:1.5.

Table 1. Sample size in each dataset

Dataset	IC	Capacitor	Inductor	Transistor	Diode	Resistor
Pcb_wacv	3686	27110	717	935	756	21406
PCB_METAL	5844	3175	541	—	—	2670
PCB_DSLR	9313	—	—	—	—	—
FICS-PCB	3243	36639	1292	1398	1593	33182

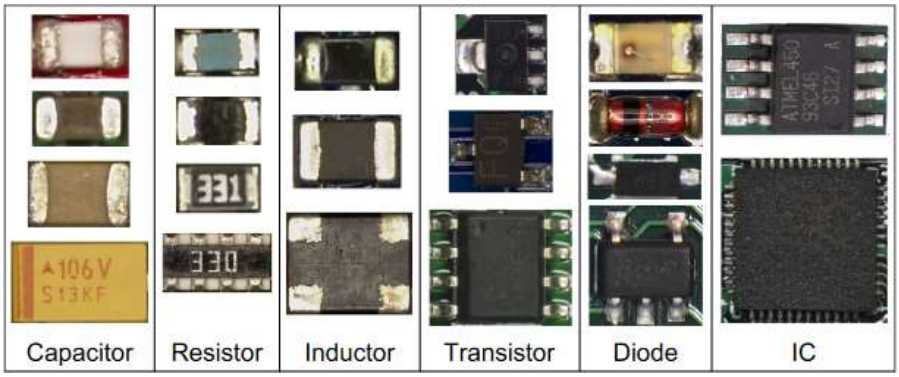

Fig. 7 Dataset example

3.2 Evaluation Metrics

In this model, we use Precision, Recall, and Average Precision as evaluation metrics. The detected targets are defined as positive samples, and non-detected targets are defined as negative samples.

1) Precision Formula:

$$Precision = \frac{TP}{TP + FP}$$

Where TP represents the number of positive samples correctly detected, and FP represents the number of negative samples incorrectly detected as positive.

2) Recall Formula:

$$Recall = \frac{TP}{TP+FN}$$

Where TP represents the number of positive samples correctly detected, and FN represents the number of positive samples incorrectly predicted as negative.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

3) Average Precision Formula:

$$Average\ Precision=\frac{1}{N}\sum_{i=1}^{N}AP$$

Where N is the number of detection classes, and AP represents the Average Precision for each class. This metric reflects the model's precision and stability for each class.

3.3 Results

Table 2. Ablation Experiment Comparison of YOLO-RSH Model Modules

Methods	Modules	Precision(%)	Recall(%)	mAP50(%)	mAP95(%)	parameters	flops
0	YOLO11	0.828	0.811	0.858	0.701	2583322	6.3
1	YOLO11+C2SCSA	0.873	0.806	0.862	0.699	2535450	6.3
2	YOLO11+C2SCSA +HFFB	0.898	0.803	0.867	0.703	3079662	13.5
3	YOLO11+C2SCSA +HFFB+C3k2-RepVGG	0.863	0.808	0.876	0.708	3111302	13.5

Table 3. Comparison of YOLO-RSH Model with Other Models

Modules	Precision(%)	Recall(%)	mAP50(%)	mAP95(%)	parameters	flops
YOLOv5n	0.895	0.752	0.839	0.685	2182835	5.8
YOLOv6n	0.842	0.752	0.825	0.672	4155618	11.4
YOLOv8n	0.872	0.711	0.838	0.674	2685538	6.8
YOLOv10n	0.879	0.73	0.817	0.663	2696756	8.2
YOLO11n	0.828	0.811	0.858	0.701	2583322	6.3
YOLO12n	0.867	0.768	0.846	0.691	2509514	5.8
YOLO-RSH	0.863	0.808	0.876	0.708	3111302	13.5

From the ablation experiments in Table 2, it can be seen that after adding different modules to the original backbone network, the network performance showed significant improvement. Compared to the original YOLO11 network, after adding the C2SCSA attention module, the precision increased by 4.5%, although the recall slightly decreased, the mAP0.5 increased by 0.4%, and the number of parameters and FLOPs remained almost the same as YOLO11. After adding the HFFB module, the precision increased by 7%, the mAP0.5 increased by 0.9%, and the mAP0.95 increased by 0.4%. After further adding the C3k2-RepVGG module, compared to the previous two modules, the recall increased by 0.5%, while the precision slightly decreased, but the average precision increased by 0.9%, and the mAP0.95 increased by 0.5%. This shows that the added modules have positively impacted the network's detection performance and effectively improved the detection capabilities of the network.

To compare the detection performance of the YOLO-RSH model with other original general-purpose object detection algorithms in PCB component detection, a series of comparative experiments were conducted, and the detailed results are shown in Table 3.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

Among all the compared models, YOLO-RSH exhibited the best performance. Compared to the baseline model YOLO11, the mAP0.5 improved by 1.8%; compared to YOLOv5, YOLOv6, YOLOv8, and YOLOv10, the mAP0.5 improved by 3.7%, 5.1%, 3.8%, and 5.9%, respectively. This demonstrates that the YOLO-RSH model has certain improvements in detection accuracy compared to other basic detection models.

4. Conclusion

To address the issues of missed detections and false detections of components on PCBs, this paper proposes the YOLO-RSH model. By introducing the C2SCSA module, the model's feature extraction capability is enhanced. Then, the HFFB module is added to the network head to further promote the fusion of features at different scales and strengthen the detection ability for small targets. Finally, the C3k2-RepVGG module replaces the original C3k2 module to improve the model's average detection accuracy.

Compared to general-purpose YOLO series models, the YOLO-RSH model demonstrates higher detection accuracy for PCB component detection, meeting the requirements for industrial component inspection. The next step will focus on researching ways to streamline the model, reducing both the number of parameters and computational load, in order to improve the detection speed and make the model better suited for real-world applications.

References

[1] Y .Shi, Z. Xin, P. C. Loh, and F. Blaabjerg, ‘‘A review of traditional helical to recent miniaturized printed circuit board Rogowski coils for powerelectronic applications,’’ IEEE Trans. Power Electron., vol. 35, no. 11, pp. 12207–12222, Nov. 2020.

[2] F. R. Leta and F. F. Feliciano, ‘‘Computational system to detect defects in mounted and bare PCB based on connectivity and image correlation,’’ in Proc. 15th Int. Conf. Syst., Signals Image Process., Jun. 2008, pp. 331–334.

[3] Xu Mingming. Detection and Recognition of PCB Components [D]. Dalian University of Technology, 2006.

[4] Crispin, A.J., Rankov, V. Automated inspection of PCB components using a genetic algorithm template-matching approach. Int J Adv Manuf Technol 35, 293–300 (2007). https://doi.org/10.1007/s00170-006-0730-0

[5] YIN H M. A template-matching-based fast algorithm for PCB components detection[J]. Advanced materials research, 2013, 690/691/692/693:3205-3208.

[6] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016

[7] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[8] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement, 2018.

[9] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection, 2020.

[10] Jocher Glenn. Yolov5 release v7.0. https: // github. com/ ultralytics/ yolov5/ tree/ v7. 0 , 2022.

[11] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. Cspnet: A new backbone that can enhance learning capability of cnn. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pages 390–391, 2020.

[12] Joseph Redmon. Darknet: Open source neural networks in c. http://pjreddie.com/ darknet/, 2013–2016.

[13] Chuyi Li, Lulu Li, Yifei Geng, Hongliang Jiang, Meng Cheng, Bo Zhang, Zaidan Ke, Xiaoming Xu, and Xiangxiang Chu. Yolov6 v3.0: A full-scale reloading. arXiv preprint arXiv:2301.05586, 2023.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202603_12(3).0013

[14] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-offreebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7464–7475, 2023.

[15] Jocher Glenn. Yolov8. https: // github. com/ ultralytics/ ultralytics/ tree/ main , 2023.

[16] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616, 2024.

[17] Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-time end-to-end object detection. arXiv 2024, arXiv:2405.14458.

[18] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.

[19] RENS,HEK, GIRSHICKR,etal. Faster r_cnn: Towards real-time object detection with regionproposal networks[J].Advances in neural information process-ing systems,2015,28.

[20] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1137–1149.

[21] Si, Yunzhong et al. “SCSA: Exploring the Synergistic Effects Between Spatial and Channel Attention.” Neurocomputing 634 (2024): 129866.

[22] Xiangzuo Huo, Gang Sun, Shengwei Tian, Yan Wang, Long Yu, Jun Long, Wendong Zhang, Aolun Li,HiFuse: Hierarchical multi-scale feature fusion network for medical image classification,Biomedical Signal Processing and Control,Volume 87, Part A,2024,105534,ISSN 1746-8094,https://doi.org/10.1016/j.bspc.2023.105534.

[23] Ding, Xiaohan et al. “RepVGG: Making VGG-style ConvNets Great Again.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 13728-13737.

[24] C. Pramerdorfer and M. Kampel, “A dataset for computer-vision-based pcb analysis,” in 2015 14th IAPR International Conference on Machine Vision Applications (MVA), pp. 378–381, May 2015.

[25] G. Mahalingam, K. M. Gay, and K. Ricanek, “Pcb-metal: A pcb image dataset for advanced computer vision machine learning component analysis,” in 2019 16th International Conference on Machine Vision Applications (MVA), pp. 1–5, 2019.

[26] Hangwei Lu, Dhwani Mehta, Olivia Paradis, Navid Asadizanjani, Mark Tehranipoor, Damon L. Woodard. FICS-PCB: A Multi-Modal Image Dataset for Automated Printed Circuit Board Visual Inspection. Cryptology ePrint Archive, Paper 2020/366, 2020. [Online]. Available: https://eprint.iacr.org/2020/366.