Article citation info:
Li Y, Wu F. Micro defect detection method for electronic components based on blending attention and full-scale feature enhancement. Diagnostyka. 2026;27(1):2026114. https://doi.org/10.29354/diag/219906.

DIAGNOSTYKA, 2026, Vol. 27, No. 1

e-ISSN 2449-5220
DOI: 10.29354/diag/219906

MICRO DEFECT DETECTION METHOD FOR ELECTRONIC COMPONENTS BASED ON BLENDING ATTENTION AND FULL-SCALE FEATURE ENHANCEMENT

Yanfang LI [1], Fangzhou WU [2,*]

[1] College of Digital Technology and Engineering, Ningbo University of Finance & Economics, China
[2] Hangzhou Research Institute, Huawei Technologies Co., Ltd., China
[*] Corresponding author, e-mail: yfli468@163.com

Abstract

In response to the low efficiency and insufficient accuracy of traditional detection methods when facing micro defects, a micro defect detection method for electronic components on the basis of blending attention and Full-Scale Feature Enhancement Module (FFEM) is proposed. This method proposes a blending attention by combining spatial and channel domain attention, and incorporates a bidirectional feature pyramid structure to achieve multi-scale feature fusion, thereby improving the efficiency of feature extraction. The test results on the PCB and NEU-DET datasets showed that the Weighted Blending Attention Module (WBAM) achieved the average accuracy by about 10%, 15%, and 20% respectively at thresholds of (Mean Average Precision, mAP)@0.5, mAP@0.75, and mAP@0.9. The FFEM also significantly improved in precision and recall, with an average precision increase of 12% and a recall increase of 18%. In addition, the model that combined WBAM and FFEM further improves its performance indicators, with an F1-score increase of about 4%, accuracy increase of about 3%, and recall increase of about 5%. The research provides efficient and accurate new technological solutions for detecting micro defects in the electronic industry, which is of great significance for improving production quality and reducing production costs. The proposed module in this method is highly versatile. Its feature enhancement mechanism for detecting minor defects provides a transferable technical solution for similar fine detection problems in industrial vision.

Keywords: WBAM, FFEM, micro defect detection, weighted fusion, bidirectional feature pyramid

List of Symbols/Acronyms

WBAM: Weighted Blending Attention Module
FFEM: Full-Scale Feature Enhancement Module
mAP: Mean Average Precision
SAN: Spatial Attention Net
CAM: Channel Attention Mechanism
SE Block: Squeeze-and-Excitation Block
CNN: Convolutional Neural Network
MLP: Multi-Layer Perceptron
FCL: Fully Connected Layer
ReLU: Rectified Linear Unit
Sigmoid: Sigmoid Activation Function
IoU: Intersection over Union
UAV: Unmanned Aerial Vehicle
IC: Integrated Circuit
PCB: Printed Circuit Board
U-Net: U-shaped Network
CBAM: Convolutional Block Attention Module
DETR: Detection Transformer
RT-DETR: Real-Time Detection Transformer

1. INTRODUCTION

With the updating and iteration of electronic industry technology, electronic components, as the core of modern electronic devices, directly affect product performance and safety. During the production process, micro defects such as cracks and holes often occur. If they cannot be accurately detected, it will cause huge losses. However, traditional detection methods are inefficient and inaccurate, which limits their ability to meet the current demand for high-precision detection of electronic components. In this context, Liu et al. proposed a Swin converter architecture to address the imbalanced Integrated Circuit (IC) defect samples and difficulties in defect detection and segmentation in electronic component welding defect detection. The results showed that the method successfully solved the sample imbalance, improved the Mean Average Precision (mAP) in IC pin welding defect detection, and increased the detection speed [1]. Wu et al. built a deep learning model method relying on generative adversarial networks

Received 2025-12-09; Accepted 2026-03-23; Available online 2026-03-23

and Convolutional Neural Networks (CNN) to address the sample imbalance in electronic product detection. This method could increase the number of electronic component sample images required for training, thereby improving detection accuracy [2]. Weiss et al. proposed a Cybord AI detection method to address the device failure caused by the exposure of electronic component pins to adverse environments during the assembly process. The results indicated that the artificial intelligence detection system could effectively detect and prevent pollution problems in electronic assembly [3]. Wang et al. built a YOLOv5 detection model for automatic visual inspection of defects in smart grid transmission lines and electrical equipment. The accuracy and recall reached 96.8% and 93.3%, significantly improving the detection performance of key targets and defects [4]. Dong et al. built an improved method to address the severe impact of scaling on image components for unmanned aerial vehicle (UAV) detection of transmission lines. The results indicated that the detection accuracy of this method was significantly improved on the CPLID dataset. The mAP50 and mAP for relatively small and medium targets were improved compared with the baseline [5].

Driven by deep learning technology and CNN architecture, Weighted Blending Attention Module (WBAM) and Full-Scale Feature Enhancement Module (FFEM) have become a hot topic. WBAM and FFEM technologies are largely applied in image classification, object detection, and image segmentation, and are also extensively researched in computer vision. Guan et al. proposed a new "plug and play" blending attention module to address the the difficulty of traditional CNN in obtaining global information and detecting partial occlusions in object detection. The results indicated that the proposed model could better detect obstructions and outperformed other lightweight models [6]. Zhang et al. proposed a remote sensing scene classification model on the basis of multi-scale features and blending attention mechanism to address the target scale difference, complex spatial distribution, and class similarity in high-resolution remote sensing image classification. The experiment showed that the classification accuracy was as high as 97.29%, which was superior to other advanced models on classification performance and computational efficiency [7]. Ni et al. proposed a novel blending attention attack method to address the insufficient research on the robustness of deep learning translation and the challenge of adversarial attack generation. This method achieved the highest attack performance when attacking three high-performance translation models [8]. Jiang M built an attention guided full-scale feature aggregation network to address the insufficient multi-scale feature fusion and information redundancy leading to poor model fitting in high-resolution remote sensing images. The F1-score and Intersection over Union (IoU) were improved on both public datasets, outperforming other mainstream methods [9]. Cao et al. designed a CNN relying on dual-attention mechanism and multi-scale feature fusion to address the image occlusion and human scale changes in human pose estimation. The network demonstrated effectiveness in benchmark testing [10].

In summary, existing research has made excellent progress in the field of electronic component micro defect detection, blending attention, and FFEM. However, there is still a long tail distribution defect problem in the balance between detection accuracy and speed. Therefore, a method for detecting micro defects in electronic components based on blending attention and FFEM is proposed to optimize the detection speed, recognition range, and accuracy of micro defects in electronic components. The innovation of the research lies in utilizing the Spatial Attention Net (SAN) mechanism to focus on the graphic position and learn positional weights to highlight important regions, thereby improving the efficiency of feature extraction. Combining SAN and Channel Attention Mechanism (CAM), WBAM is proposed to improve the speed of detecting micro defects. In addition, the model offers a more efficient solution for detecting micro defects in electronic components.

2. METHODS AND MATERIALS

The study first combines SAN and CAM for weighted fusion to generate a blending attention model. Secondly, the study constructs an FFEM based on a bidirectional feature pyramid and incorporates a full-scale deep supervision mechanism [11-12]. Finally, a micro defect detection model on the basis of blending attention and FFEM is constructed.

2.1. Construction of micro defect detection model based on blending attention

Existing detection methods often fail to accurately identify micro defects in electronic components due to their extremely small size. The attention mechanism can improve the detection accuracy by guiding the model to emphasize key areas in the input data. Therefore, a WBAM model is proposed based on the spatial and channel domains. This model utilizes the attention weights of spatial and channel domains through weighted superposition, improving the efficiency and accuracy of feature extraction [13-14]. SAN emphasizes the positional information of each pixel, highlighting important regions by learning the weights of each position, as presented in Figure 1.

Figure 1 presents the SAN. The structure first sums up each feature point on the channel. Afterwards, the dimension of the feature layer is changed from H×W×1 to 1× (H×W). Then, the summed feature layer is activated using the Softmax function, and the Softmax output is rescaled to H×W×1. Finally, the reshaped weights are multiplied by the original feature layer to get the

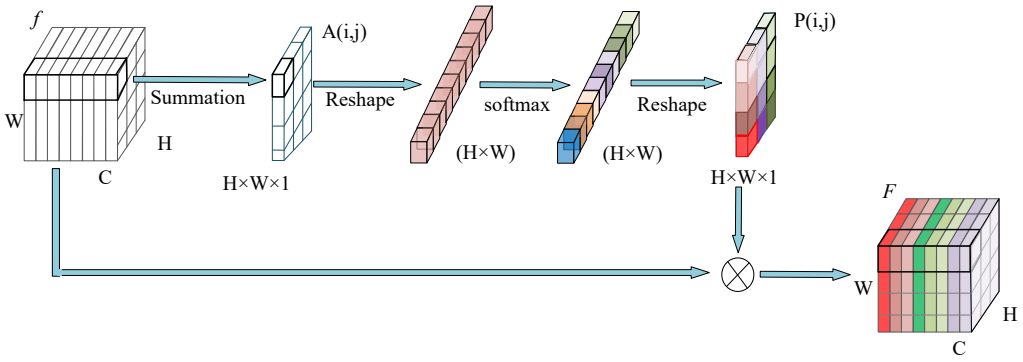


Fig. 1. Schematic diagram of SAN structure (Source from: Author's self drawn)

output result. Before performing weighted fusion, attention coefficients need to be calculated. Firstly, feature map F is subjected to two different types of pooling, namely, global average and global maximum, in order to generate spatial statistics, as shown in equation (1).

$$\begin{cases} F_{gap}(i,j) = \frac{1}{C}\sum_{c=1}^{C} F_c(i,j) \\ F_{gmp}(i,j) = MAX_C F_c(i,j) \end{cases} \quad (1)$$

In equation (1), $F_c(i,j)$ represents the eigenvalue of the C -th channel at spatial position (i,j). $F_{gap}(i,j)$ represents performing a global average pooling operation on the eigenvalues of all C channels at position (i,j). $F_{gmp}(i,j)$ represents performing a global max pooling operation on the eigenvalues of all C channels at position (i,j). F_{gap} and F_{gmp} are concatenated along the channel dimension and spatial features are extracted through convolutional layers, as shown in equation (2).

$$F_{spatial} = Conv(F_{gap} \oplus F_{gmp}) \quad (2)$$

In equation (2), $Conv$ represents the convolution operation. $F_{spatial}$ represents the final spatial feature map. $\oplus$ represents channel dimension concatenation. Finally, the spatial attention weight matrix is produced by activating the Sigmoid function, as shown in equation (3).

$$G_t = \sigma(F_{spatial}) \quad (3)$$

In equation (3), σ represents the activation function. G_t represents the spatial attention coefficient matrix. The SAN coefficient matrix is calculated. CAM models the dependency relationships between channels to determine the importance of feature channels. The current typical CAM algorithm structure is the Squeeze-and-Excitation Block (SE Block) [15], as shown in Figure 2.

Figure 2 shows the SE Block structure diagram. The first step is to start from a single image and extract image features. Then, the dimensions of feature map [H,W] are averaged or maximally pooled, and the weights of each channel are obtained through a Multi-Layer Perceptron (MLP) network. Finally, the weight is applied to feature map U[C,H,W]. The input F undergoes global average pooling, a process that extracts global information from each channel, as shown in equation (4).

$$F_{gap} = \frac{1}{H \cdot W}\sum_{h=1}^{H}\sum_{w=1}^{W} J(:,h,w) \in R^C \quad (4)$$

In equation (4), J represents a three-dimensional array with dimensions of C×W×H. H signifies the height size of the three-dimensional array J. W signifies the width size of the three-dimensional array J. C signifies the channel dimension size of the three-dimensional array J. $J(:,h,w)$ represents the C -dimensional vector of the three-dimensional array J at the height and width positions. Two Fully Connected Layers (FCLs) are used to perform nonlinear transformation on channel information. The output of the first FCL is shown in equation 5.

$$F_{fc1} = FC1(F_{gap}, W_{fc1}) + b_{fc1} \in R^{C/2} \quad (5)$$

In equation (5), the output of the FCL F_{fc1} signifies an $C/2$-dimensional vector. $FC1$ signifies the first FCL. W_{fc1} signifies the weight matrix of the $FC1$. b_{fc1} represents the bias vector of the $FC1$. Afterwards, the model is activated with the ReLU

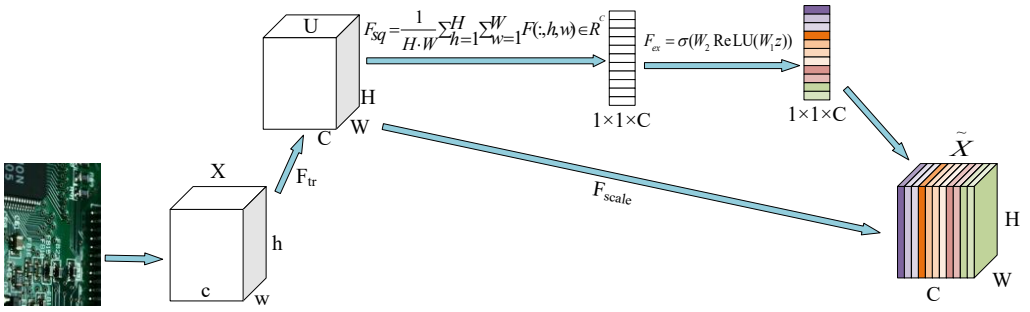


Fig. 2. SE Block structure diagram (The physical image in the picture is sourced from: https://colorhub.me/photos/Zwoxk)

function and the second FCL is output, as shown in equation (6).

$$F_{fc2} = FC2((\mathrm{ReLU}(F_{fc1}) \in R^{\frac{C}{2}}), W_{fc2}) + b_{fc2} \in R^{C} \quad (6)$$

In equation (6), $FC2$ represents the second FCL. ReLU represents the modified linear unit activation function. W_{fc2} signifies the weight matrix of the $FC2$. b_{fc2} represents the bias vector of the $FC2$, with a dimension of C. F_{fc2} represents the output of the $FC2$. Finally, the CAM weight vector is generated through the activation function, as shown in equation 7.

$$G_d = \sigma(F_{fc2}) \in R^{C} \quad (7)$$

In equation (7), σ represents the Sigmoid activation function. G_d represents the final output CAM weight vector. The attention coefficients and their weights of the obtained spatial and channel domains are weighted and fused to obtain the WBAM coefficient G_{cmf}, as shown in equation (8).

$$G_{cmf} = \frac{\gamma_d \cdot G_d + \gamma_t \cdot G_t}{\gamma_d + \gamma_t + \in} \quad (8)$$

In equation (8), γ_d represents the weight scalar of CAM, used to measure their respective importance. γ_t represents the weight scalar of SAN, used to measure their respective importance. $\in$ represents the constant used to prevent situations where the denominator is zero. G_{cmf} represents the WBAM coefficient. The WBAM coefficient G_{cmf} is element wise multiplied with the original F to get the final blending attention feature map, as shown in equation (9).

$$F_{wbam} = F \odot G_{cmf} \quad (9)$$

In equation (9), $\odot$ represents element level multiplication. F represents the original feature map. F_{wbam} represents a blending attention feature map. The novelty of the WBAM theory lies in the proposal of an adaptive weighted fusion mechanism. It dynamically balances the contributions of channel and spatial attention through learnable weight parameters, rather than simply adding or concatenating. This enables the model to adaptively adjust the attention focus based on the input features, enhancing the flexibility of feature selection. The module flowchart of WBAM is shown in Figure 3.

The complete weighted fusion attention module is shown in Figure 3. In the upper part of the figure, the input feature map receives global average pooling operation, and then sequentially passes through two FCLs. The nonlinearity is introduced using the ReLU activation function and channel weight maps are obtained. The original feature map in the lower part of the figure first goes through a convolutional layer, uses the ReLU activation function, and then goes through the convolutional layer again. Finally, the spatial weight map is generated using the Sigmoid. The spatial weight map is multiplied element by element with the F, and the result is weighted and summed with the F. Finally, it is input into the Sigmoid to further modify the features and obtain the final output feature map.

2.2. Construction of micro defect detection model for electronic components based on blending attention and FFEM

The defect detection model relying on blending attention can focus on key areas in the feature map, highlight important information, and has high detection accuracy and good feature extraction ability. However, the model still has limitations on computational resource consumption and robustness. Therefore, the study further introduces FFEM to optimize the detection model. FFEM can extract and fuse features, improving the model's feature extraction and classification capabilities, making it more robust in complex backgrounds and detecting micro defects. FFEM constructs rich feature representations by extracting multi-scale features from input features [16]. It mainly has feature extraction, feature fusion, and feature enhancement. The architectural innovation of FFEM lies in the integration of bidirectional feature pyramids and full-scale deep supervision. Through the bidirectional path, it achieves complementary enhancement of details and semantic information, and combines multi-layer auxiliary loss functions to strengthen multi-scale feature learning, thereby improving the feature representation ability for minor defects.

During feature extraction, three parallel convolutional layers are taken to extract feature maps, as shown in equation (10).

$$\begin{cases} F_{small} = Conv_{1\times1}(F) \\ F_{medium} = Conv_{3\times3}(F) \\ F_{large} = Conv_{5\times5}(F) \end{cases} \quad (10)$$

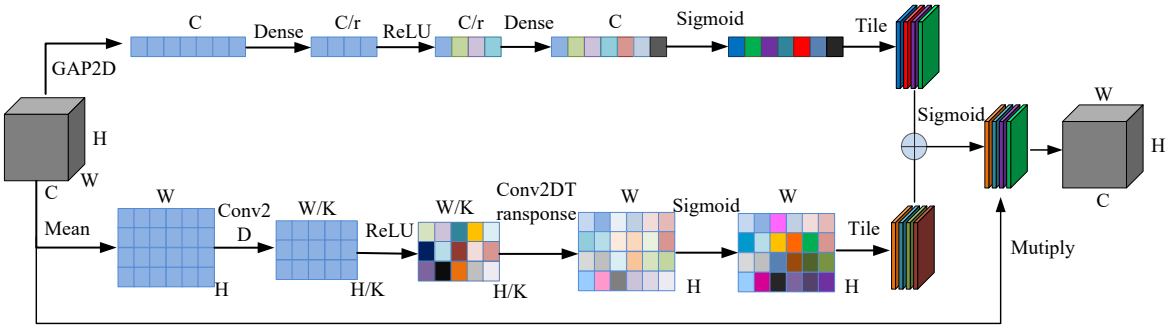


Fig. 3. Weighted fusion attention frame diagram (Source from: Author's self drawn)

In equation (10), $Conv_{1\times1}$ signifies the convolution operation using a 1×1 convolution kernel. $Conv_{3\times3}$ represents the convolution operation using a 3×3. $Conv_{5\times5}$ signifies the convolution operation using a 5×5. Cross-scale feature fusion takes a bidirectional feature pyramid structure to achieve cross-scale information exchange. This structure can extract multi-scale features and achieve cross-scale information exchange, enhance feature expression ability, and effectively solve the incomplete single scale feature information. The expression is shown in equation (11).

$$F_i^{top-down} = \text{Upsample}(F_{i+1}^{top-down}) \oplus F_i \quad (11)$$

In equation (11), Upsample represents the upsampling operation. $F_{i+1}^{top-down}$ represents the top-down feature of the $i+1$-th layer. F_i represents the original features of the i -th layer. $F_i^{top-down}$ represents the top-down feature of the i -th layer. The bottom-up path is used to convey detailed information, as shown in equation (12).

$$F_i^{bottom-up} = Conv_{3\times3}(Concat(F_i^{top-down}, F_{i-1}^{bottom-up})) \quad (12)$$

In equation (12), $F_i^{bottom-up}$ represents the bottom-up feature fusion result of the i -th layer. $F_{i-1}^{bottom-up}$ represents the bottom-up feature of the $i-1$ -th layer. $Concat$ represents the feature concatenation operation. The fused feature enhancement is enhanced through the input blending attention module, as shown in equation (13).

$$F_{enhanced} = \text{WHA}(Concat(F_{top-down}, F_{bottom-up})) \quad (13)$$

In equation (13), $Concat$ represents the feature quantity obtained through enhancement. WHA represents the blending attention module constructed in the previous text. Figure 4 displays the FFEM.

Figure 4 is an FFEM based on a bidirectional feature pyramid structure. The input feature map is processed through a multi-scale feature extraction layer, and features F1, F2, and F3 of different scales are extracted through different convolution kernels. The extracted features are input into top-down and bottom-up feature paths, and upsampling, feature fusion, concatenation, and convolution operations are performed, respectively. Finally, the features of the bidirectional path are fused to obtain the enhanced feature Fenhanced, achieving complementarity and enhancement of details and semantic information. A full-scale deep supervised network is added to the model, and auxiliary classifiers are introduced at multiple levels of the network to alleviate gradient vanishing and enhance shallow feature discriminative ability [17-18], as shown in equation (14).

$$L_{aux}^k = \text{CrossEntropy}(y, \text{Classifier}(F_{enhanced}^k)) \quad (14)$$

In equation (14), L_{aux}^k represents the auxiliary loss of the k -th layer. CrossEntropy signifies the cross-entropy loss function. Classifier represents the classifier. $F_{enhanced}^k$ signifies the enhanced features of the k -th layer. y signifies the true label. The research incorporates loss functions to enhance balance in classification [19-20], as shown in equation (15).

$$L_{total} = L_{main} + \lambda \sum_{k=1}^{3} L_{aux}^k \quad (15)$$

In equation (15), L_{main} represents the loss of the main output layer. L_{main} represents the total output layer loss. λ represents the weight coefficient used to balance the contributions of the main loss and auxiliary loss in the total loss.

The specific value of the balance coefficient λ was determined through ablation experiments on the validation set, and the final selected value was the one that resulted in the best model performance. The structure diagram of the full-scale deep supervision network is shown in Figure 5.

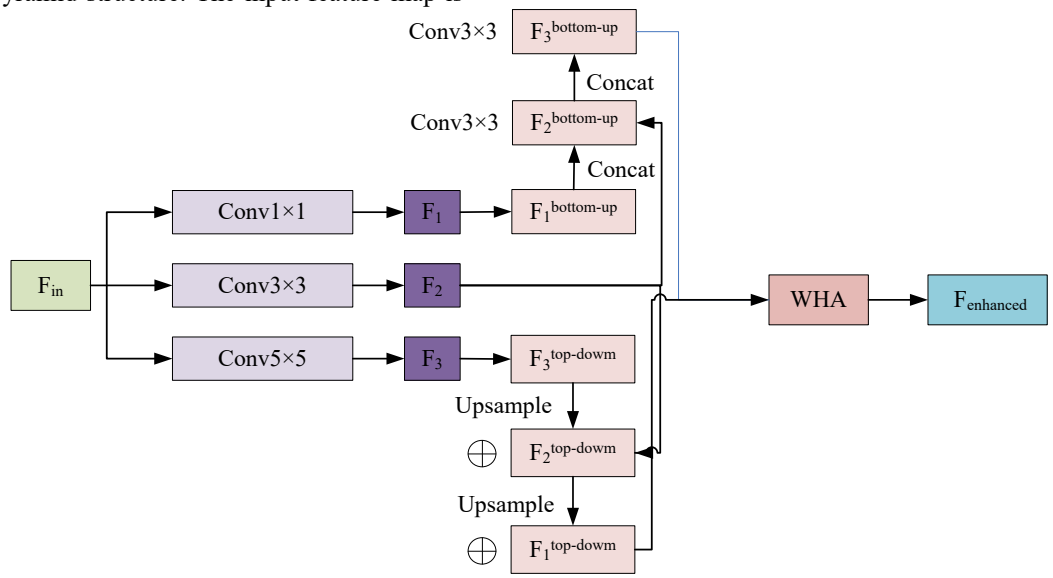


Fig. 4. Overall framework diagram of the FFEM (Source from: Author's self drawn)

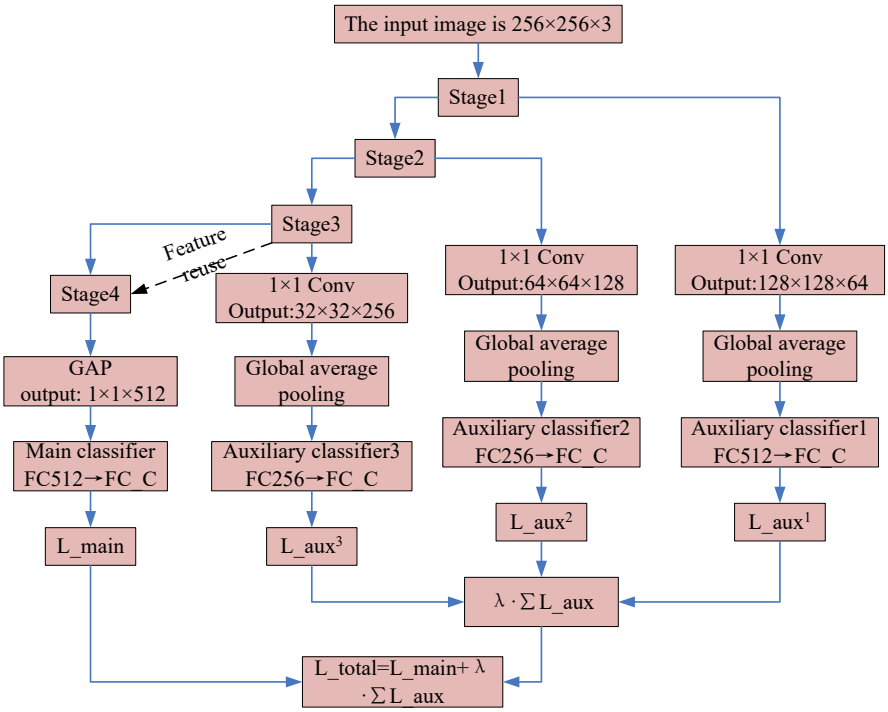


Fig. 5. Full-scale deep supervision network architecture diagram (Source from: Author's self drawn)

In Figure 5, the architecture takes a 256×256×3 image as input, undergoes convolution and pooling processing from Stage 1 to Stage 4, and performs feature reuse between Stage 3 and Stage 4. After each stage, the dimensionality is adjusted through 1×1 convolution. After global average pooling, the main classifier and auxiliary classifier calculate the L_main and L_aux1-3 losses, respectively. The final total loss L_total is composed of the main loss and weighted auxiliary loss, achieving multi-scale feature utilization and auxiliary supervision. Finally, a micro defect detection model for electronic components combining WBAM and FFEM is constructed. The model structure diagram is shown in Figure 6.

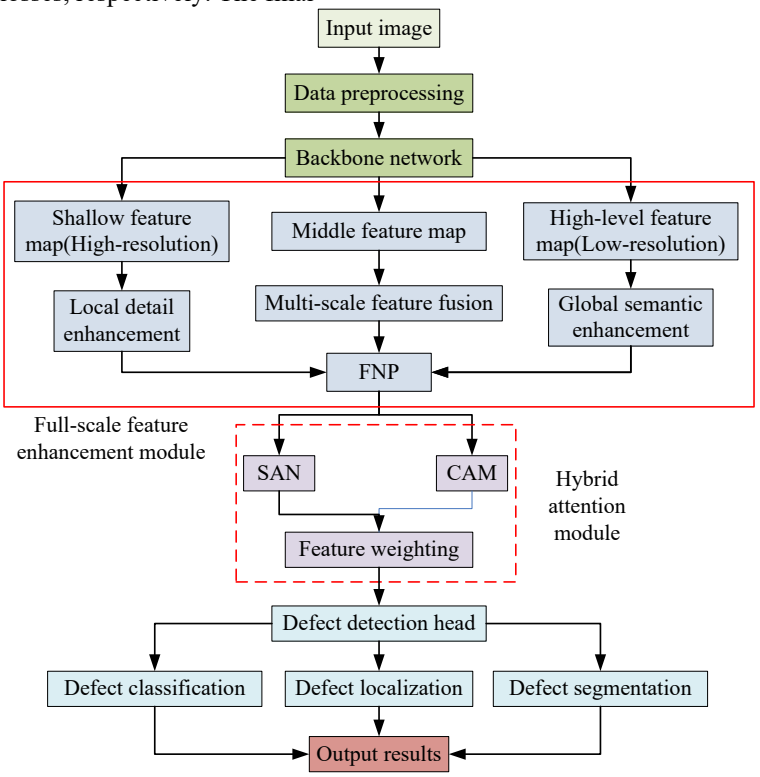


Fig. 6. Structure diagram of a micro defect detection model combining blending attention and FFEM (Source from: Author's self drawn)

In Figure 6, the input image is preprocessed, and shallow, middle, and deep feature maps are extracted through a backbone network, and then subjected to partial detail enhancement, cross-scale feature fusion, and global semantic enhancement operations. The processed models are combined and a feature pyramid is constructed. The pyramid features are weighted by a blending attention module and then aggregated at multiple scales to enter the defect detection head. Detection classification, localization, and segmentation are implemented at the detection head. Finally, the detection results are output. The structure integrates WBAM and FFEM (WBAM-FFEM), and improves defect detection accuracy through cross-scale feature interaction and feature weighting, ensuring accurate and reliable classification, localization, and segmentation results.

3. RESULTS

The study first validates key indicators such as detection accuracy and recall of the electronic component micro defect detection model. Secondly, the same indicators for the FFEM-based electronic component micro defect detection model are validated. Finally, the performance metrics and resource consumption of the WBAM and FFEM models are compared with other models.

3.1. Performance verification of micro defect detection model for electronic component based on blending attention

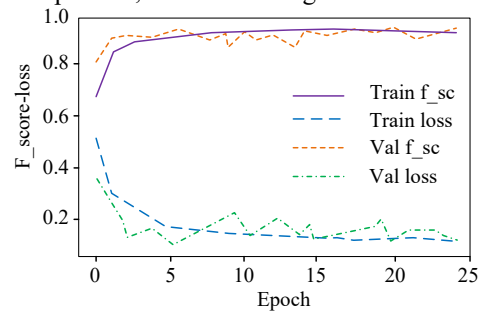
To verify the actual performance, some comparative experiments are performed. This study selected widely adopted and performance-stable classic models such as YOLOv5 and Faster R-CNN as the main baselines, aiming to conduct a fair and reproducible comparison with the recognized benchmarks in the field, thereby clearly verifying the effectiveness of the proposed WBAM and FFEM modules themselves. Before training and testing, the experimental environment and dataset are configured. The hardware system configuration and software system parameters required for the research are presented in Table 1.

According to Table 1, the environment required for experimental validation needs to be configured, and a test dataset needs to be prepared for training and testing the model. The publicly available PCB defect dataset prepared for the experiment includes typical defects of six electronic components: missing holes, mouse bites, open circuits, short circuits, stray, and pseudo copper. It is suitable for analyzing the generalization of micro defects. The surface defects such as scratches and inclusions in NEU-DET steel dataset are highly similar to the surface damage of electronic components and have micro target characteristics, which are often transferred to microelectronic defect detection tasks. This study used the publicly available PCB and NEU-DET datasets for validation. The main considerations were as follows: Firstly, the types of defects contained in these datasets are highly similar in morphology, scale, and appearance to the common micro-defects in the actual production of electronic components, providing an appropriate benchmark for the verification of the core capabilities of this method. Secondly, testing on public benchmarks is conducive to conducting fair and reproducible performance comparisons with existing mainstream methods, and establishing the relative advantages of the method. Although industrial site data is usually more complex, by validating the model's capabilities in small target detection, multi-scale feature fusion, and subtle feature discrimination on such challenging public datasets, it can provide strong support for the potential generalization of the method in real scenarios. Before delving into the performance, the convergence during the training and validation process is analysed. The F1-score and loss value are key indicators for evaluating the learning effectiveness of a model, which can intuitively reflect the performance in different iteration processes.

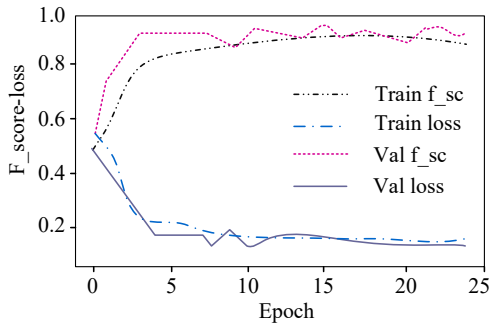
Table 1. Hardware and software configuration parameters table

Category	Item	Model/Version
Hardware	Computer host	Intel Core i7-10700K, 16GB DDR4 3200MHz RAM, 256GB SSD, 1TB HDD
	GPU accelerator card	NVIDIA RTX 3080, 10GB+ VRAM
	Camera	1920×1080 resolution, 30fps frame rate, autofocus
	Image acquisition card	Compatible with camera, supports multiple resolutions and formats
	Electronic component stage	Adjustable angle and position, with lighting equipment
	Server	Multi-GPU configuration, scalable based on data and computation needs
Software	Operating system	Linux Ubuntu 20.04
	Deep learning framework	PyTorch 1.9.x
	Data processing tools	Python 3.8+, NumPy, Pandas, OpenCV
	Image annotation tool	labelImg
	Model evaluation tools	Scikit-learn, TensorBoard
	Visualization tools	Matplotlib, Seaborn
	Database management system	MySQL

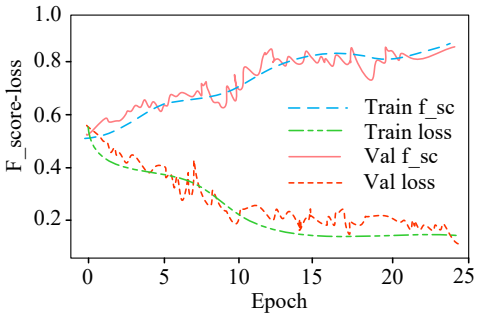
CNN for Biomedical Image Segmentation (U-Net) is a type of CNN used for biomedical image segmentation, which can effectively handle image segmentation tasks. DeepLab Version 3 Plus (DeepLab v3+) is a model for semantic image segmentation, suitable for semantic segmentation tasks in complex scenes. U-Net with Convolutional Block Attention Module (U-Net+CBAM) introduces CBAM on the basis of U-Net to optimize the ability to obtain key features. The following compares the F1-score and loss values of four models during the iteration process, as shown in Figure 7.



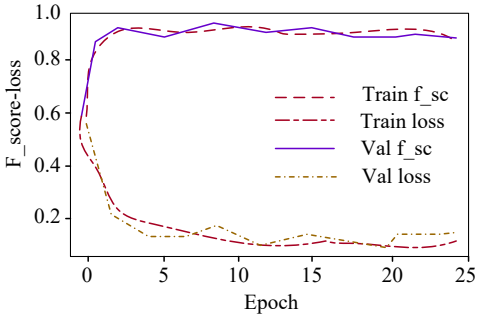
(a) U-Net+CBAM iterative process



(b) U-Net iterative process



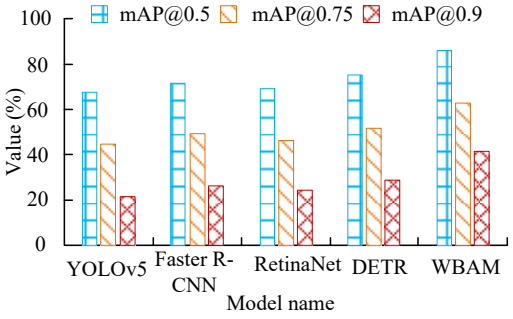
(c) DeepLab v3+ iterative process



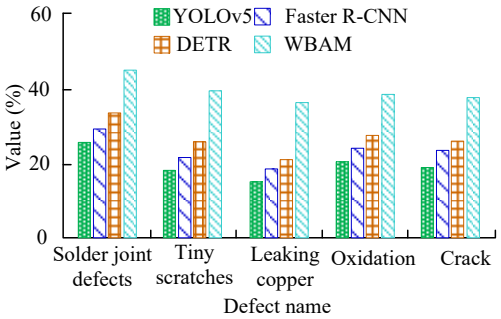
(d) WBAM iterative process

Fig. 7. Diagram of the iterative process of network results (Source from: Author's self drawn)

From Figures 7 (a), 7 (b), 7 (c), and 7 (d), WBAM exhibited faster convergence speed and higher F1-score during training and validation compared with other models. Specifically, the F1-score of WBAM during final convergence increased by about 15% compared with UNet+CBAM, about 20% compared with U-Net, and about 18% compared with DeepLabV3+. Meanwhile, WBAM had the lowest loss value. These data indicate that WBAM has significantly improved both training efficiency and detection accuracy. mAP is widely used as a performance evaluation metric for object detection tasks, which can fully reflect the detection precision and recall. The study takes multiple mAP thresholds for evaluation, comprehensively measuring the detection ability. The specific comparison is shown in Figure 8.



(a) Three-threshold mAP comparison chart

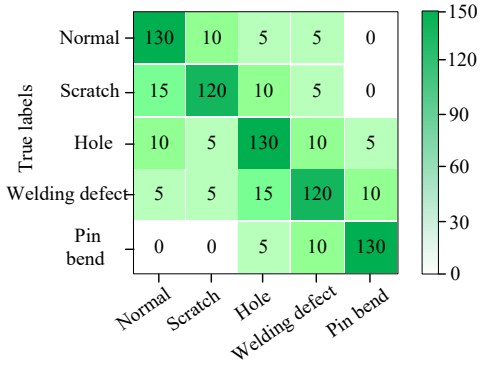


(b) Performance of different defect types at mAP@0.9

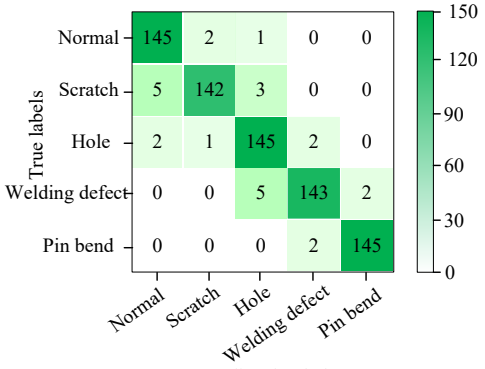
Fig. 8. Comparison of defect detection performance for electronic components under multi-model and multi-threshold (Source from: Author's self drawn)

Figure 8 (a) compares the overall detection performance of You Only Look Once Version 5 (YOLOv5), Faster Regions with CNN (Faster-R-CNN), Focal Loss for Dense Object Detection (RetinaNet), Detection Transforms (DETR), and WBAM at three different mAP thresholds of 0.5, 0.75, and 0.9. WBAM had an average accuracy improvement of about 10% at mAP@0.5, about 15% at mAP@0.75, and about 20% at mAP@0.9. Figure 8 (b) shows the detection performance of WBAM for different types of defects at mAP@0.9. The detection accuracy of WBAM was improved by about 25% in solder joint defects, about 30% in tiny scratches, about 20% in pin bend, about 25% in oxidation, and about 20% in cracks compared with other models. These data indicate that WBAM exhibits higher detection accuracy and robustness

compared with other models. The confusion matrix can visually display the performance differences of different attention models in multi-class defect classification tasks, as shown in Figure 9.



(a) Channel attention model



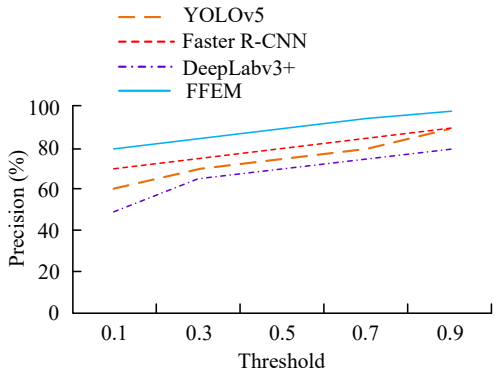
(b) Hybrid attention model

Fig. 9. Structure diagram of a micro defect detection model combining blending attention and FFEM (Source from: Author's self drawn)

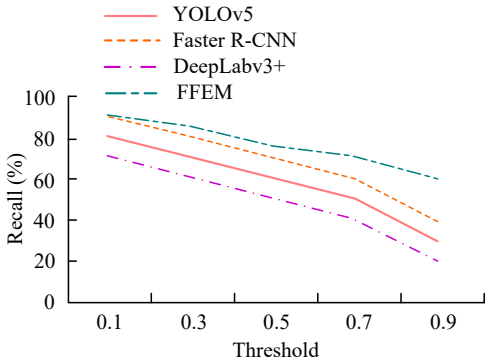
The classification results of CAM and WBAM for five types of defects are shown in Figure 9. WBAM improved accuracy by 4.6% compared with SAM in the "normal" category, 18% in the "scratch" category, 2.2% in the "hole" category, 2.4% in the "welding defect" category, and 15.4% in the "pin bend" category. This indicates that the blending attention model has stronger capabilities in feature extraction and classification, which can more accurately identify and distinguish different types of defects.

3.2. Performance verification of electronic component micro defect detection model based on FFEM

The above experimental research has verified the excellent performance of the WBAM-based electronic component micro defect detection model in improving detection accuracy and robustness. Based on FFEM, the micro defects in electronic components is further detected by, taking precision and recall as key indicators, as shown in Figure 10.



(a) Comparison of accuracy of different models



(b) Comparison of recall rates of different models

Fig. 10. Comparison of precision and recall of different models (Source from: Author's self drawn)

Figures 10 (a) and 10 (b) show the precision and recall. Between the threshold range of 0.1-0.9, FFEM improved the average precision by about 12% compared with YOLOv5, about 10% compared with Faster R-CNN, and about 15% compared with DeepLabv3+. The average recall of FFEM within the same threshold range increased by about 18% compared with YOLOv5, about 13% compared with Faster R-CNN, and about 22% compared with DeepLabv3+. The data results indicate that FFEM is suitable for defect detection tasks that require high precision and high recall. Further research is conducted to visually demonstrate the classification performance of each model on different categories through a confusion matrix, as shown in Figure 11.

According to Figures 11 (a), 11 (b), 11 (c), and 11 (d), in normal category detection, the accuracy of FFEM was improved by 10%, 3%, and 10% compared with YOLOv5, Faster R-CNN, and DeepLabv3+, respectively. In scratch category detection, the accuracy of FFEM was increased by 13%, 5%, and 13% compared with YOLOv5, Faster R-CNN, and DeepLabv3+. For hole category detection, the accuracy of FFEM was improved by 10%, 7%, and 3% compared with YOLOv5, Faster R-CNN, and DeepLabv3+. The classification accuracy of FFEM was significantly higher than other models in all categories, indicating that the research model can identify and distinguish different types of defects.

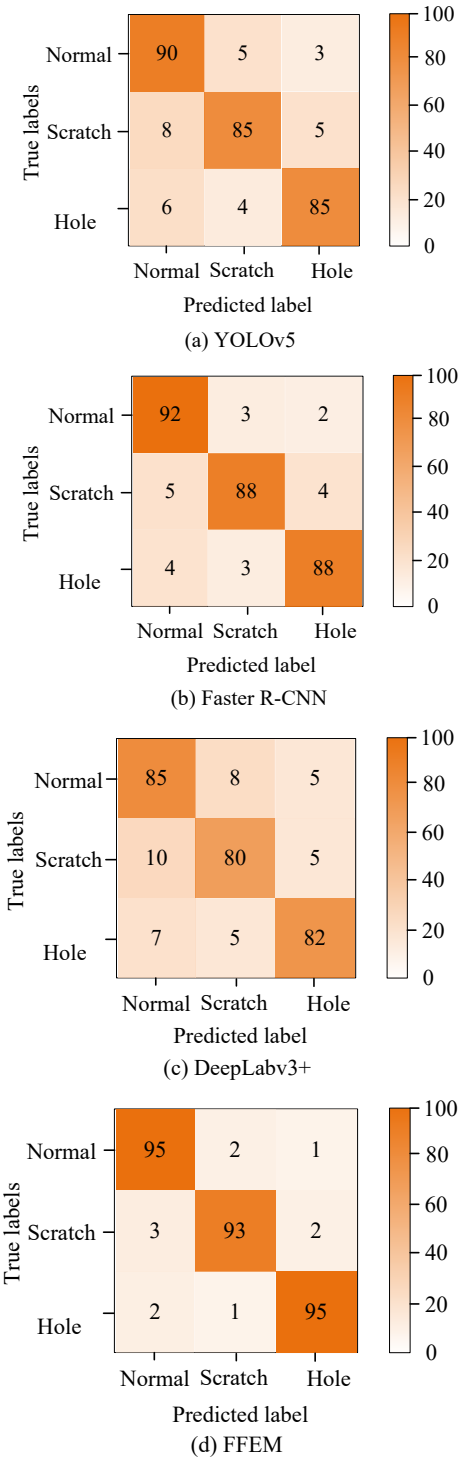


Fig. 11. Comparison of confusion matrices for different models (Source from: Author's self drawn)

3.3. Performance verification of micro defect detection model for electronic component based on blending attention and FFEM

The performance verification of the electronic component micro defect detection model based on WBAM-FFEM is conducted in the above content, achieving excellent performance results. The next step is to validate the WBAM-FFEM in electronic component micro detection, verifying direct performance indicators like precision, recall, F1-score, inference time, and CPU usage. The specific comparison is shown in Figure 12.

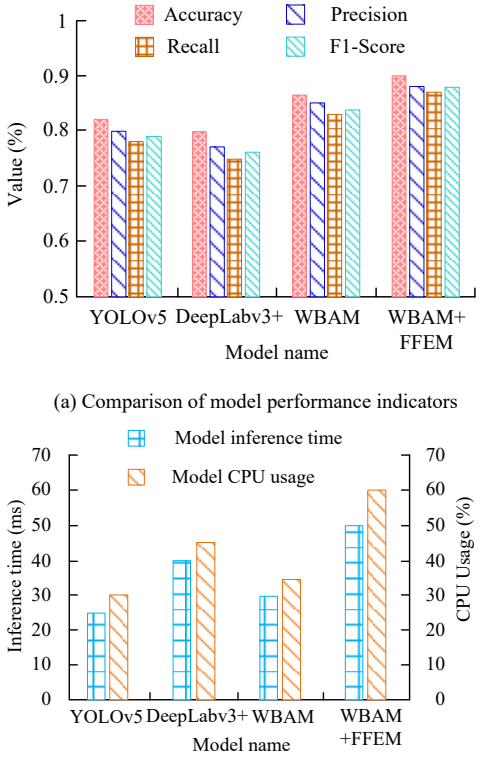


Fig. 12. Comprehensive comparison and analysis of performance indicators and resource consumption of different models (Source from: Author's self drawn)

In Figure 12 (a), WBAM-FFEM showed an average improvement of about 4% in accuracy compared with other models, with an average improvement of about 3%, 5%, and 4% in precision, recall, and F1-score. In terms of resource consumption in Figure 12 (b), the inference time of WBAM-FFEM was improved to varying degrees compared with other models, with the maximum improvement being 100%. The CPU usage increased by 100% compared with YOLOv5, 33% compared with DeepLabv3+, and 71% compared with WBAM. WBAM-FFEM has significantly improved performance, but the required computing resources have increased. To assess the model's sensitivity to the loss balance coefficient λ, this study tested the performance of the WBAM-FFEM model with λ at different values on the validation set. The results are shown in Table 2.

The experiment shows that when λ = 0.5, the model achieves the best performance. When λ varies between 0.3 and 0.7, the fluctuation range of mAP is less than ±0.8%, indicating that the model is not sensitive to changes in the λ value within this range and has good robustness. However, when λ deviates too much from this range, the performance significantly declines, suggesting that an appropriate amount of auxiliary supervision is crucial for the stability of training.

Table 2. Model performance under different loss balance coefficients λ

λ	mAP@0.5:0.95 (%)	Precision (%)	Recall rate (%)
0	68.2	85.1	72.3
0.1	70.5	86.4	74.8
0.2	72.8	87.9	76.5
0.3	74.9	89.2	78
0.4	75.8	90	78.9
0.5	76.1	90.2	79.2
0.6	75.9	89.9	79
0.7	75	89.3	78.3
0.8	73.1	88	76.8
0.9	71	86.5	75
1	69.3	85.3	73.4

4. DISCUSSION AND CONCLUSION

Driven by electronics industry, the quality detection of electronic components is facing higher requirements. Traditional detection methods suffer from low efficiency and insufficient accuracy in detecting micro defects, which makes it hard to meet the production needs. Therefore, the research proposes a micro defect detection method on the basis of WBAM-FFEM to optimize the detection accuracy and efficiency through WBAM and FFEM. The research method outperformed existing methods in key indicators such as model convergence speed, detection accuracy, and recall. The F1-score of WBAM during final convergence increased by 15%, 20%, and 18% compared with UNet+CBAM, U-Net, and DeepLabV3+, respectively. The average accuracy of WBAM improved by about 10%, 15%, and 20% respectively at thresholds mAP@0.5, 0.75, and 0.9. In defect classification, WBAM improved the average accuracy by 8.5% compared with SAM. Compared with other models, FFEM improved the average accuracy by 12% and the recall by 18%. FFEM improved the average accuracy of category detection by 8.6% compared with other models. In addition, WBAM-FFEM further improved its performance indicators, with F1-score, precision, and recall increasing by approximately 4%, 3%, and 5%, respectively. The experimental results show that this method performs exceptionally well in addressing key challenges such as dealing with tiny targets, complex backgrounds, and the differentiation of multiple types of defects. These capabilities are core requirements in industrial inspection scenarios. The design of the WBAM and FFEM modules focuses on enhancing the model's feature extraction and discrimination abilities rather than overfitting to specific datasets. This fundamentally enhances its potential to adapt to data from different sources. However, this study acknowledges that to ultimately prove its robustness in specific production line environments, it will still need to collaborate with industrial partners in future work, conducting end-to-end verification and fine-tuning on proprietary large-scale datasets that include more noise, lighting variations, and defect variants. Although excellent results have been made, the increased computational resource consumption, long inference time, and robustness in handling extremely complex backgrounds or micro defects still need further validation. It should be noted that although the computational cost of the current model has increased, during industrial deployment, through optimization techniques such as model lightweighting, hardware-specific acceleration, and model pruning-distillation, the inference speed can be significantly improved while ensuring the detection accuracy, thus meeting the strict real-time requirements of the production line. The final feasibility of the model is the engineering optimization result achieved through the trade-off between accuracy and speed. Future research will emphasize the computational efficiency of the model and exploring its potential applications in detecting more defect types.

Source of funding: *This research received no external funding.*

Author contributions: *research concept and design, Y.L.; Collection and/or assembly of data, Y.L.; Data analysis and interpretation, Y.L.; Writing the article, F.W.; Critical revision of the article, F.W.; Final approval of the article, F.W.*

Declaration of competing interest: *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.*

REFERENCES

1. Amin S, Shivakumara P, Jun T, Chong K, Zan D, Ramachandra R. An augmented reality-based approach for designing interactive food menu of restaurant using android. Artificial Intelligence and Applications. 2023;1(1):26-34. https://doi.org/10.47852/bonviewAIA2202354.
2. Wu H, Lv Q, Yang J, Yan X, Xu X. Electronic component detection based on image sample generation. Soldering & Surface Mount Technology. 2022;34(1):1-7. https://doi.org/10.1108/SSMT-08-2020-0036.
3. Weiss E. Revealing hidden defects in electronic components with an ai-based inspection method: A corrosion case study. IEEE Transactions on

Components, Packaging and Manufacturing Technology. 2023;13(7):1078-1080. https://doi.org/10.1109/TCPMT.2023.3293005.
4. Wang S, Tan W, Yang T, Zeng L, Hou W, Zhou Q. High-voltage transmission line foreign object and power component defect detection based on improved YOLOv5. Journal of Electrical Engineering & Technology. 2024;19(1):851-866. https://doi.org/10.1007/s42835-023-01625-6.
5. Dong C, Zhang K, Xie Z, Shi C. An improved cascade RCNN detection method for key components and defects of transmission lines. IET Generation, Transmission & Distribution. 2023;17(19):4277-4292. https://doi.org/10.1049/gtd2.12948.
6. Guan R, Man KL, Zhao H, Zhang R, Yao S, Smith J, Lim EG, Yue Y. MAN and CAT: Mix attention to nn and concatenate attention to YOLO. The Journal of Supercomputing. 2023;79(2):2108-2136. https://doi.org/10.1007/s11227-022-04726-7.
7. Zhang Z, Xu C. A Lie group-based model for remote scene classification with multi-scale feature fusion and mixed attention mechanisms. International Journal of Remote Sensing. 2025;46(10):3800-3830. https://doi.org/10.1080/01431161.2025.2491818.
8. Ni M, Wang C, Zhu T, Yu S, Liu W. Attacking neural machine translations via hybrid attention learning. Machine Learning. 2022;111(11):3977-4002. https://doi.org/10.1007/s10994-022-06249-x.
9. Jiang M, Zhang X, Sun Y, Feng W, Gan Q, Ruan Y. AFSNet: Attention-guided full-scale feature aggregation network for high-resolution remote sensing image change detection. GIScience & Remote Sensing. 2022;59(1):1882-1900. https://doi.org/10.1080/15481603.2022.2142626.
10. Cao D, Liu W, Xing W, Wei X. Human pose estimation based on feature enhancement and multi-scale feature fusion. Signal, Image and Video Processing. 2023;17(3):643-650. https://doi.org/10.1007/s11760-022-02271-7.
11. Li HC, Ma H, Che YB, Yang ZD. A two-way dense feature pyramid networks for object detection of remote sensing images. Knowledge and Information Systems. 2023;65(11):4847-4871. https://doi.org/10.1007/s10115-023-01916-4.
12. Mishra S, Zhang YZ, Chen DZ, Hu XS. Data-driven deep supervision for medical image segmentation. IEEE Transactions on Medical Imaging. 2022;41(6): 1560-1574. https://doi.org/10.1109/TMI.2022.3143371.
13. Zhai WZ, Li QL, Zhou Y, Li XS, Pan JF, Zou GF, Gao ML. DA^{2}Net: A dual attention-aware network for robust crowd counting. Multimedia Systems. 2023; 29(5):3027-3040. https://doi.org/10.1007/s00530-021-00877-4.
14. Hu J, Liu Y, Wu K. Neural network pruning based on channel attention mechanism. Connection Science 2022;34(1):2201-2218. https://doi.org/10.1080/09540091.2022.2111405.
15. Liu CJ, Liu XQ, Chen C, Wang QK. Soft thresholding squeeze-and-excitation network for pose-invariant facial expression recognition. The Visual Computer. 2023; 39(7): 2637-2652. https://doi.org/10.1007/s00371-022-02483-5.
16. Hong YS, Neu R, Berto F, Nakamura T, Palin-Luc T, Reed O. Editorial: Renewal of FFEMS editorial board. Fatigue & Fracture of Engineering Materials & Structures. 2022;45(2):331. https://doi.org/10.1111/ffe.13637.
17. Rodríguez-Puigvert J, Martínez-Cantín R, Civera J. Bayesian deep neural networks for supervised learning of single-view depth. IEEE Robotics and Automation Letters. 2022;7(2):2565-2572. https://doi.org/10.1109/LRA.2022.3142915.
18. Losoi P, Konttinen J, Santala V. Substantial gradient mitigation in simulated large-scale bioreactors by optimally placed multiple feed points. Biotechnology and Bioengineering. 2022;119(12):3549-3566. https://doi.org/10.1002/bit.28232.
19. Nie J, Zhong S. Loss functions for finite sets. Computational Optimization and Applications. 2023; 84(2):421-447. https://doi.org/10.1007/s10589-022-00420-9.
20. Liu F, Wang J, Chen D, Shen C, Xu F. Asymmetric exponential loss function for crack segmentation. Multimedia Systems. 2023; 29(2): 539-552. https://doi.org/10.1007/s00530-022-00944-4.

Yanfang LI Obtained his BE in Process and Equipment of Machinery Manufacturing from Guilin University of Electronic Technology in 1994. He obtained his ME in computer application technology from Shanghai Jiao Tong University in 2004. Presently, he is working as an associate professor in College of Digital Technology and Engineering, Ningbo University of Finance & Economics. His areas of interest are machine vision, mechanical structure design and optimization, engineering education.
Contact: liyanfang@nbufe.edu.cn

Fangzhou WU Obtained his BE in Electrical & Information Engineering from Southwest Jiaotong University in 2020. He obtained his ME in Signal Processing and Communications from the University of Edinburgh in 2021. Presently, he is working as a programmer at Huawei's Hangzhou Research Institute. His areas of interest are intelligent perception, software architecture, edge computing.
Contact: yfli468@163.com