Computer Life
ISSN: 1819-4818 | Vol. 13, No. 2, 2025

Surface defect detection algorithm for special-shaped electronic components based on improved RT-DETR

Shugang Liu [a], Shang Shi [b], Jie Wu [c]

School of Computer Science, NORTH CHINA ELECTRIC POWER UNIVERSITY, Baoding 071003, China
[a] lsg69@qq.com, [b] 1367348920@qq.com, [c] 2973395257@qq.com

Abstract: The detection of surface defects of special-shaped electronic components is a key link to improve the level of insertion technology of special-shaped plug-in machines. Traditional manual detection is susceptible to human subjectivity, the efficiency of template matching algorithm is low, and in the case of insufficient sample data, the existing deep learning technology has problems such as low accuracy and insufficient real-time performance in defect detection. In order to improve the accuracy and real-time performance of special-shaped component detection, this paper improves the object detection model RT-DETR and proposes a real-time multi-dimensional feature adaptive network (RT-MDAFNet): Firstly, the Adaptive Fusion Pyramid Network (AFPN) is designed at the feature fusion layer of the model, and the adaptability and feature extraction ability of the model to multi-scale targets are improved through the dynamic channel attention mechanism and selective feature fusion mechanism. Then, the adaptive channel-spatial aggregation network module (SASE-RepNet) was designed to improve the detection accuracy and efficiency in complex backgrounds by combining multi-level feature aggregation, channel adaptive weight allocation and spatial selectivity enhancement mechanism. In the absence of existing datasets, a dataset of special-shaped electronic components was constructed, and the RT-MDAFNet model was compared with 8 baseline models, including DETR, Faster R-CNN, YOLO series, etc.

Keywords: Detection of irregular-shaped electronic components; RT-DETR; Object detection; Data augmentation.

1. Introduction

In the modern electronics manufacturing industry, special-shaped electronic components are widely used in high-density circuit board assembly, and their quality directly affects the performance and reliability of electronic products. However, in industrial production, the detection of surface defects is a challenging task due to the complex structure and different shapes of special-shaped electronic components. The surface defect detection of traditional special-shaped electronic components mainly relies on manual inspection, which is time-consuming and subjective, which is difficult to adapt to the needs of industrial large-scale production, and can easily lead to false detection and missed detection. In this context, the template matching algorithm based on machine vision has been introduced into the research in this field. This kind of algorithm realizes the identification of surface defects of special-shaped electronic components by calculating the similarity characteristics between the test image and the standard template.

2. Object detection algorithm based on improved RT-DETR

2.1. Construction of the RT-MDAFNet model

Backbone Network Selection and Model Structure Different convolutional neural networks, such as ResNet[1] and HGNet [2], can be used in the RT-DETR backbone network to adapt to different application scenarios and needs. As a classical deep residual network, ResNet has strong feature extraction capabilities, especially when dealing with image classification and object detection tasks in complex scenes [3]. However, in order to achieve high detection accuracy, ResNet usually requires a deeper network structure, which leads to a significant increase in the number of parameters and computations, and it is difficult to meet the needs of scenarios with high requirements for real-time performance and lightweight [4]. In contrast, HGNet is a network structure designed for lightweight, with significantly reduced parameter quantity and computational complexity, making it ideal for embedded devices or industrial application scenarios that require high real-time performance and low power consumption. Therefore, HGNet is chosen as the backbone network for RT-DETR (Figure 1).

2.2. The RT-DETR model with HGNet as the backbone network

has a fast convergence speed and good performance in detection accuracy, but there is still some room for improvement: First, the model has room for improvement in detection accuracy [8]. Second, in terms of the running speed of the model and the number of model parameters, further weight reduction can still be made. Considering the accuracy and lightweight requirements of special-shaped component object detection in industrial applications, this paper improves the RT-DETR model [4], designs the Adaptive Fusion Pyramid Network Module (AFPN) and the Adaptive Channel-Spatial Aggregation Network Module (SASE-RepNet), and proposes the RT-MDAFNet model, the structure of which is shown in Figure 2.

Adaptive Fusion Pyramid Network Module (AFPN) In order to improve the adaptability and feature extraction ability of the model to multi-scale targets, this paper proposes an adaptive fusion pyramid network module (AFPN) based on the HS-FPN module to improve the channel attention mechanism and feature fusion method. In terms of channel attention mechanism, the AFPN module adopts the Dynamic Channel Attention (DCA) module. The DCA module not only obtains the importance weights of the channel based on the global average pooling and global maximum pooling, but also

dynamically generates the attention weights based on the input characteristics. Specifically, the DCA module first performs global average pooling and global maximum pooling operations on the input features, and extracts the global mean and peak characteristics of each channel respectively to ensure the representativeness and significance of the features. Then, these pooled features are processed through the learnable Fully Connected Layer (FC), and the adaptive channel weights are dynamically generated, so that the AFPN module can flexibly adjust the weights according to the changes of the input features, and enhance the adaptability to different scale targets, so as to highlight the important channels and suppress the secondary channels, so as to make adaptive weighting adjustment to the original features. Match. Finally, the weight values are multiplied by the original features to obtain the filtered features. In terms of feature fusion, the AFPN module adopts the Selective Feature Fusion (SFF) mechanism to fuse the selected low-level features and high-level features, and adds an upsampling operation in the feature fusion process. The high-level features are scaled to match the low-level features by an upsampling strategy that combines transpose convolution and bilinear interpolation. Specifically, advanced features are first extended by transposing convolutions to increase the resolution of the features. Then, bilinear interpolation is used to further adjust its size to match the scale of the low-level features, so as to enhance the fusion effect of the features of different levels. This upsampling strategy can effectively improve the resolution of high-level features, so that the AFPN module can better capture detailed information and adapt to object detection at different scales. The feature fusion module of the AFPN module uses the high-level features generated by the dynamic channel attention mechanism as weights to screen out the most representative information among the low-level features, thereby improving the expression ability of the model. The formula for calculating the scale-adjusted high-level features and the fused multi-scale features in the AFPN module is as follows:

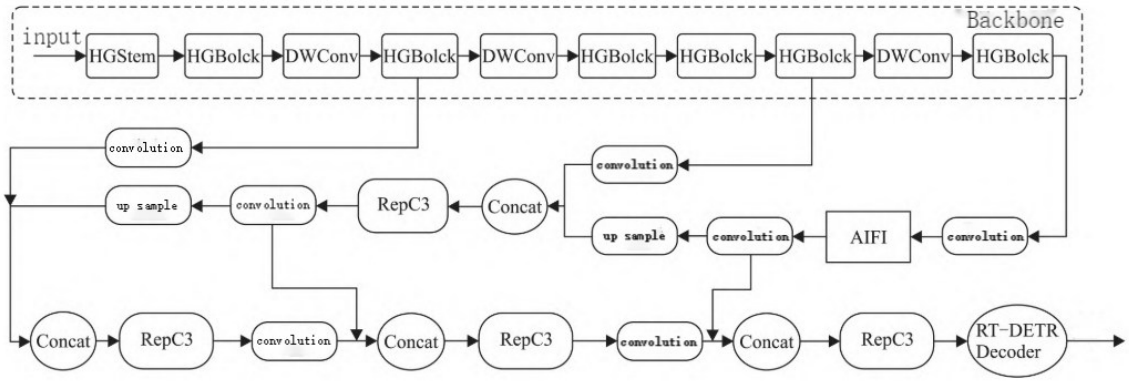

Figure 1. Structure diagram of RT-DETR model

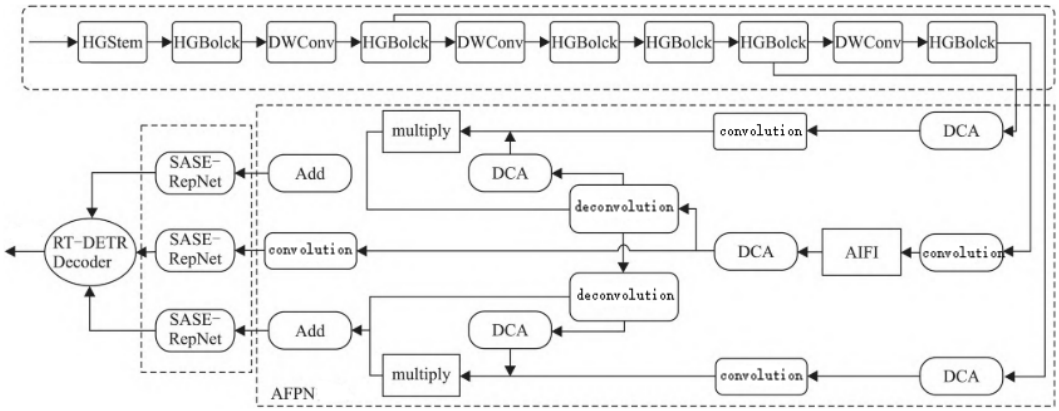

Figure 2. Structure diagram of RT-MDAFNet model

$$f_{\text{att}} = \text{BL}\left(T - \text{Conv}\left(f_{\text{high}}\right)\right), \quad (1)$$

$$f_{\text{out}} = f_{\text{low}} * \alpha\left(f_{\text{att}}\right) + f_{\text{att}}, \quad (2)$$

where fhigh is the high-level feature; flow is a low-level feature; FATT is a scale-adjusted high-level feature; BL(·) is a bilinear interpolation operation; T is the global feature obtained by global average pooling, which is used to guide feature enhancement; fout is the result of the weighted superposition of low-level features and high-level features, representing the multi-scale features after fusion. * is a convolution operation; α (·) is the operation of the DCA module. The feature fusion process in the AFPN module is as follows: First, the high-level features fhigh and the low-level feature flow are input, and the high-level features are extended by transposing convolutions with a step size of 2 and a kernel size of 3×3 to obtain the enlarged features. The high-level features are then scaled using bilinear interpolation to match the dimensions of the lower-level features. Then, the DCA module is used to convert the weights of high-level features into corresponding attention weights, which are used to filter useful information from low-level features. Finally, the filtered low-level features are combined with the high-level features to generate multi-scale feature fout. 1.2. 3 Adaptive Channel-Spatial Aggregation Network Module (SASE - RepNet) Traditional convolutional feature extraction methods cannot fully cope with the problems of diverse forms, different scales, and easy confusion with complex backgrounds when dealing with special-shaped electronic

components. In order to solve the complex feature extraction problem in the detection of surface defects of special-shaped electronic components, the Adaptive Channel-Spatial Aggregation Network Module (SASE-RepNet) is designed to extend and fuse the multi-level feature aggregation mechanism, the channel adaptive weight allocation mechanism and the spatial selectivity enhancement mechanism, in order to effectively improve the accuracy and efficiency of detection in complex backgrounds. The weight index of the deep convolution kernel, where Xi is the input feature of the ith channel, Li is the deep convolution kernel of the ith channel, and C is the total number of channels. In the point-by-point convolution stage, Xdepthwise is subjected to point-by-point convolution to aggregate the information of different channels and improve the expressive ability of features, and the calculation formula is as follows:

$$Xpointwise = Lp * Xdepthwise, \quad (3)$$

where Xpointwise is the final extracted feature, p is the weight index of the point-by-point convolution kernel, and Lp is the point-by-point convolution kernel of the pth channel. The deep separable convolution structure can reduce the amount of computation while maintaining the effectiveness of feature extraction, making the module suitable for embedded environments. Then, through the channel block operation, the feature is divided into two parts and passed to two parallel submodules, Conv2 and Conv3, respectively. Both sub-modules contain a layer aggregation mechanism that generates detailed feature representations through multiple convolution and aggregation layers, ensuring that feature information of different scales and complex structures can be efficiently transmitted, and maintaining feature integrity at multiple scales, so as to accommodate different sizes and shapes of special-shaped electronic components. In terms of the aggregated features, the channel adaptive weighting mechanism is further used to distribute the attention of each channel. Specifically, the channel adaptive weighting mechanism extracts global information through global average pooling, and dynamically generates the adaptive weight of each channel through two fully connected layers, calculated as follows:

$$z = \sigma(W2 (W1 \cdot GAP(X)))), \quad (4)$$

where GAP(X) represents the global average pooling operation on feature X, W1 and W2 are the learned weight matrices, and σ (·) is the activation function Sigmoid. By adjusting the weights between channels, the model pays more attention to the key channel features, making it better at detecting small and complex components.Then, in order to enhance the positioning ability of the model spatially, a spatial selectivity sensitivity mechanism was introduced: firstly, the average and maximum values of features in the spatial dimension were calculated. Then, the two are spliced together and generated by convolution operation. The formula for calculating the spatial weight map is as follows:

$$S=\sigma(fconv (concat(avg(X) ,max(X))))), \quad (5)$$

where avg(X) and max(X are the average pooling and maximum pooling of feature X in the channel dimension, respectively. By generating a spatial weight map, the mechanism weights important areas to make the positioning of the model more accurate in complex backgrounds. Finally, the fused features are further optimized to generate the final output features through a depth-separable convolution Conv4, so that the channel and space are finely fused to provide high-quality feature representation for the detection task.

3. Experimental results

Comparative experiments The RT-MDAFNet model was compared with eight existing baseline models, including the single-stage model (YOLO series), the two-stage model (Faster R-CNN), and the transformer-based model (DINO, DETR, DN-DETR). Each baseline model is described as follows.From the experimental results (Table 1), it can be seen that: (1) Although the YOLO series model has certain advantages in terms of frame rate, parameter quantity, GFLOPs and other indicators, its detection accuracy is lower than that of the RT-MDAFNet model. Among them, the mAP50 and mAP50-95 values of the YOLOv11-s model with the highest detection accuracy in the YOLO series were 76.89% and 45.89%, respectively, which were significantly lower than the mAP50 and mAP50-95 values of the RT-MDAFNet model (80.87% and 50.43%). Moreover, the detection accuracy of the Faster R-CNN, DINO, DETR, and DN-DETR models is lower than that of the RT-MDAFNet models [5]. The results show that the AFPN module and SASE-RepNet module in the RT-MDAFNet model are conducive to capturing the surface defect characteristics of special-shaped electronic components. (2) Among the eight baseline models, the mAP50 and mAP50-95 values of the DINO model were 77.56% and 46.97%, respectively, second only to the RT-MDAFNet model, but its model complexity was higher and the number of parameters was larger. The DETR model had the lowest number of parameters (41.0 M), but it was significantly higher than that of the RT-MDAFNet model. The Faster R-CNN model had the lowest GFLOPs (76.4) and was also higher than the RT-MDAFNet model. The frame rate of the MDAFNet model is 41.5 FPS, which is significantly higher than that of the Faster R-CNN model. These results show that compared with the two-stage model (Faster R-CNN) and the model based on Transformer architecture (DINO, DETR, DN-DETR), the RT-MDAFNet model not only has higher detection accuracy, but also maintains lower model complexity and parameter quantity [6].

Table 1. Results of contrast experiment

MODEL	mAP50/%	mAP50-95/%	M	GFLOPs	FPS
YOLOV5-s [26]	72.43	42.94	7.10	15.8	63.2
YOLOV8-s [27]	75.69	44.98	11.13	28.4	128.4
YOLOV10-s [28]	74.87	44.45	8.04	24.5	76.5
YOLOV11-s [29]	76.89	45.89	9.41	21.3	100.0
Faster R-CNN [14]	77.45	46.21	41.38	76.4	12.1
DINO [30]	77.56	46.97	47.43	279.2	—
DETR [31]	75.32	46.96	41.00	187.3	—
DN-DETR [32]	75.12	47.57	48.30	195.8	—
RT-MDAFNet	80.87	50.43	24.31	75.3	41.5

On the whole, the RT-MDAFNet model achieves a good balance between accuracy, parameter quantity, computational complexity and frame rate, effectively controls the amount of computation and parameter scale under the premise of high accuracy requirements, and improves the frame rate, which is suitable for application scenarios with limited computing resources and high-precision and efficient detection [7].

4. Conclusion

In order to solve the problems of low efficiency, poor real-time performance and low accuracy of defect detection of existing special-shaped electronic components, a real-time multi-dimensional feature adaptive network model (RT-MADFNet) is proposed on the basis of the RT-DETR model. The model includes two key modules: AFPN and SASE-RepNet MADFNet: firstly, the feature extraction of multi-scale special-shaped electronic components is carried out through the dynamic channel attention mechanism and the selective feature fusion mechanism. Then, by combining multi-level feature aggregation, channel adaptive weight allocation and spatial selectivity enhancement mechanism, the detection efficiency and accuracy of special-shaped electronic components in the complex detection background were improved. On the self-built dataset, the RT-MDAFNet model was compared with the existing 8 baseline models. The experimental results show that the mAP50 and mAP50-95 values of the RT-MADFNet model are 80.87% and 50.43%, respectively, the frame rate is 41.5 FPS, and the GFLOPs are 75.3, which is better than that of the baseline models. The ablation experiment verified the effectiveness of the AFPN module and the SASE-RepNet module in the RT-MDAFNet model. Although the RT-MDAFNet model has achieved good experimental results, there is still room for improvement in the robustness of the model under some extreme conditions, such as uneven lighting or high background noise. In the follow-up research, we can consider combining self-supervised learning or generative adversarial network (GAN) to improve the robustness of the model under harsh conditions, so as to further improve the comprehensive performance of the detection algorithm.

References

[1] Wang Xueqi. Research on visual inspection and positioning technology of special-shaped electronic component plug-in machine[D]. Hunan: Hunan University, 2019. WANG X R Q.esearch on visual inspection and positioning technology of alien electronic component plug-in machine[D]. Hunan: Hunan University, 2019.

[2] Wang Qianqian. Research on insertion and positioning technology of special-shaped components based on machine vision[D]. Harbin: Harbin University of Science and Technology, 2019. WANG Q Q. Research on installation and location technology of special-shaped components based on machine vision[D]. Harbin: Harbin University of Science and Technology, 2019.

[3] Zhou Yang. Research on visual inspection and positioning technology of special-shaped component plug-in system[D]. Wuhan: Wuhan Textile University, 2023. ZHOU Y. Research on visual inspection and positioning technology of special-shaped component plug-in system [D]. Wuhan: Wuhan Textile University, 2023.

[4] Wang Yuan. Research on key technologies of visual inspection and positioning of special-shaped plug-in system [D]. Mianyang: Southwest University of Science and Technology, 2017. WANG Y. Vision inspection and positioning key techniques research of irregular component plug-in machine [D].Mianyang: Southwest University of Science and Technology, 2017.

[5] Gao Zong. Research on quality inspection technology of electronic components based on machine vision[D]. Guiyang: Guizhou University, 2018. GAO Z. Research on quality inspection technology of electronic components based on machine vision[D]. Guiyang: Guizhou University, 2018.

[6] Zeng Yao, Gao Faqin. Surface Defect Detection Algorithm for Electronic Components Based on Improved YOLOv5[J]. Journal of ZheJiang University (Engineering Science) ,2023,57 (3) : 455-465. ZENG Y, GAO F Q. Surface defect detection algorithm of electronic components based on improved YOLOv5[J]. Journal of Zhejiang University (Engineering Science) , 2003, 57(3) : 455-465.

[7] HAN K, WANG Y H, TIAN Q, et al. Ghostnet: more features from cheap operations[C]// Proceedings of the IEEE /CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE, 2020: 1580-1589.

[8] Wang Haonan. Research on electronic component detection and grasping system based on machine vision[D]. Taiyuan: Taiyuan University of Science and Technology, 2023. WANG H N. Research on the detection and gripping system of electronic components based on machine vision [D]. Taiyuan: Taiyuan University of Science and Technology, 2023.