Research on PCB defect detection algorithm based on mixed attention mechanism

Zhenhua Li [1, a], Lei Zhang [1, b], Yu Wang [2, c]

[1]Jiangsu Institute of Technology, School of Electrical and Information Engineering, Changzhou, China

[2]Southeast University, National Mobile Communications Research Laboratory, Nangjing, China

[a]1640973805@qq.com, [b] zhlei@jsut.edu.cn, [c]yuwang_edina@jsut.edu.cn

Abstract. The production process of PCB is complex and cumbersome, which leads to the complex and diverse defects of the PCB and seriously affects the productivity. Therefore, timely and effective detection of its surface defects is particularly momentous. For this article, a new mixed attention module C_Efficient Channel Attention (C_ECA) is presented by connecting spatial attention with effective Channel attention. A one-dimensional convolution can be substituted for the fully connected layer in the channel attention module of C_ECA; for the spatial attention module of C_ECA, we can use 3*3 and 7*7 convolution kernel in place of the 7*7 convolution kernel to use multiple scale information to consider the significance of diverse regional trait. We integrated the module into YOLOv3 and tested it on PCB data set. Finding in the laboratory indicate that the network performance can be greatly improved and PCB defects can be accurately identified.

Keywords: Deep learning, Attention mechanism, ECANet.

1. Introduction

As the core component of electronic products, the quality detection of PCB is the key to meet the increasing quality demands of the electronic manufacturing industry. In the production process, affected by equipment, environment, manual operation, etc. PCB will produce defects such as short circuit, open circuit, spur, mouse bite, missing hole and spurious copper. There are many types of defects which the size and shape are complex and changeable, and the defect characteristic area is small. Therefore, fast and accurate inspection of disfigurement in PCB circuit board has become a research hotspot in recent years.

PCB disfigurement inspection falls into two main sorts: traditional PCB defect inspection and PCB defect inspection on account of deep learning. The conventional PCB defect inspection has a lot of issues such as long detection time, high false detection rate and great influence by human subjective factors, which greatly reduces the production efficiency of enterprises and can no longer up to grade of modern circuit board inspection demand. Deep learning target inspection algorithms can solve this problem. In this research, YOLOv3 is used as the main network. By improving the network, the feature description ability is further enhanced. Therefore, It can solve the problem of small target defect inspection.

2. Related work

Target detection is to detect the desired target and then judge its category. In recent years, Target inspection algorithms on account of deep learning falls into two main sorts: two-stage target detection algorithms based on region and one-stage target detection algorithms based on regression. The representative algorithms of the former include r-cnn [3], fast r-cnn [4] and fast r-cnn [5], and the latter is represented by SSD [8] and Yolo [9-12] series algorithms. Hu et al. [6] proposed an improved feature pyramid in Faster RCNN, which effectively make use of the semantic message of deep feature layer, but does not fully utilize the information of shallow feature maps, resulting in PCB defect location is not accurate.

In last few years, Attention module is widely used in computer vision. We add the attention mechanism to the neural network, so the performance of the network becomes significantly higher. Hu et al. proposed the SE (Squeeze Excitation) module [15], which learns the correlation between channels in the feature map thus generates channel attention. So it makes the network raise more and more concern about information-rich channels thereby bringing significant performance improvements to CNNs. The CBAM (Convolutional Block Attention Module) module [14] is further extended on the basis of the SE module. This module pools the feature map in space globally to obtain spatial attention. BAM (Bottleneck Attention Module) [16] integrates spatial attention and channel attention in parallel. The SK (Selective Kernel) [17] construction presented by Li et al. combines the idea of SE with the residual network, which enables the network to dynamically select different receptive fields on the basis of different scales of the feature layer. please close this file and download the Microsoft Word, Letter file.

3. This paper method

3.1 Yolov3

Yolov3 is a network presented by Redmon et al. It is a single-stage target inspection algorithm based on the yolov2 algorithm, which combines residual network and feature pyramid network. The Yolov3 network is mainly composed of three pieces, including the backbone feature extraction network, the feature pyramid network and the yolohead. The network structure is shown in Figure 1.Its backbone feature extraction network is DarkNet53, which draws on the idea of residual network and alleviates the problem of gradient disappearance in deep network by stacking 5 residual blocks. The CBL unit is the basic unit of yolov3, which is composed of convolution, batch normalization (BN) and Leaky ReLu activation function. Through the backbone feature extraction network, three feature layers of different sizes can be obtained, namely 13*13, 26*26 and 52*52 effective feature layers.

The feature pyramid FPN mainly Aggregates three feature layers obtained by Backbone. FPN first performs 5 convolutions on the effective feature layer of size 13*13, then upsampling and stacks it with the effective feature layer of size 26*26. Then it passes the stacked feature layer through 5 convolutions, and then performs Upsampling. Then it is stacked with the effective feature layer with the size of 52*52, and finally three output feature layers are generated respectively.

YOLOv3 bring in the thought of using anchor boxes in object position prediction, and predicts 3 anchor boxes on each feature map. For an input image, the YOLOv3 algorithm divides it into 13×13 blocks and predicts the bounding boxes of 3 object on each small block. YOLOv3 introduces a multi-scale fusion method which Make predictions for the bounding box of the object at 3 scales,thereby greatly improving the accuracy of small target inspection.

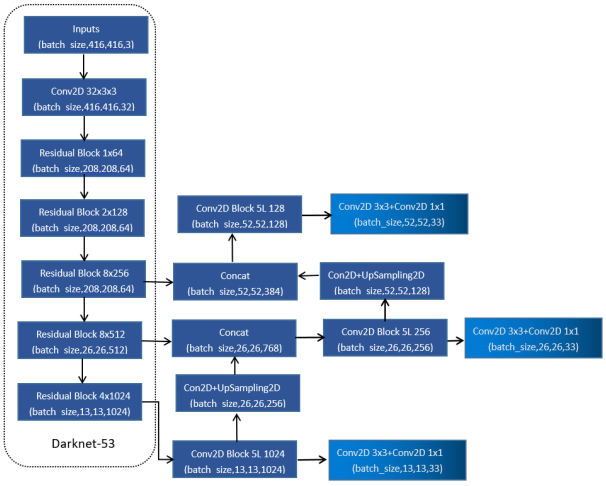


Figure 1. Table Type Styles

3.2 Mixed attention mechanism

We combine spatial attention module and channel attention module to form mixed attention. There are many mixed attention models in the field of computer vision. The imparity is mainly divided into two aspects: First, the combination mode of the two is different, and second, the calculation mode of the two is different. This paper is a series of channel attention and spatial attention. First focusing on what a given image is, and then focusing on where the useful information of the image is.

Given a feature layer as input, C_ECA will in turn derive the one-dimensional channel attention weight and the two-dimensional spatial attention weight, as shown in Figure 2, the entire calculation course can be sum up:

$$F' = M_c(F) \otimes F + F \tag{1}$$

$$F'' = M_s(F') \otimes F' \tag{2}$$

where $\otimes$ represents element-wise multiplication. F'' for the final output. Figures 3 and 4 describe the computation process of each attention map. Details of each attention module are described below.

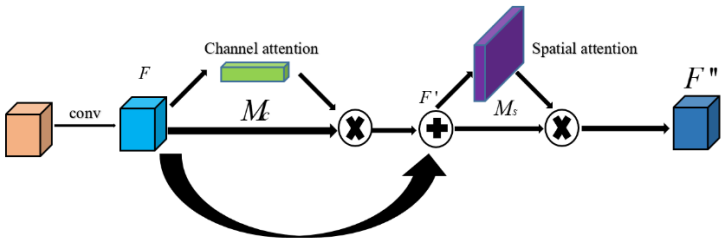


Figure 2. C_ECA attention module structure

channel attention. In order to efficiently compute channel attention, the average pooling and max pooling operations are firstly used to aggregate the spatial information of the feature maps to generate two different spatial descriptor F^c_{avg} and $F^c_{\max}$, which represent the average pooled features and the max pooled features respectively; Based on the idea of reference [18], one-dimensional convolution with convolution kernel length k is used for channel feature aggregation. The two convoluted features are added according to the elements, and our channel attention $Mc(F) \in \mathbb{R}^{C\times1\times1}$ is generated through the operation of sigmoid function. The channel attention is extended along the two dimensions of space to $\mathbb{R}^{C\times H\times W}$, and then multiplied with the input feature map according to the corresponding elements to obtain the feature map after injecting channel attention. The channel attention calculation process can be expressed as follows:

$$M_c(F) = \sigma\left(f^k_{1D}\left(F^c_{avg}\right) + f^k_{1D}\left(F^c_{\max}\right)\right) = \sigma\left(f^k_{1D}\left(AvgPool(F)\right) + f^k_{1D}\left(MaxPool(F)\right)\right) \tag{3}$$

Among them, σ represents the Sigmoid function, and f^k_{1D} represents the one-dimensional convolution operation with a convolution kernel size of k. The size of k is adaptively determined by the equation in [18]:

$$k = \left|\frac{\log_2(C)}{2} + \frac{1}{2}\right|_{odd} \tag{4}$$

C represents the number of channels of the input feature map, and $|t|_{odd}$ represents the odd number closest to t.

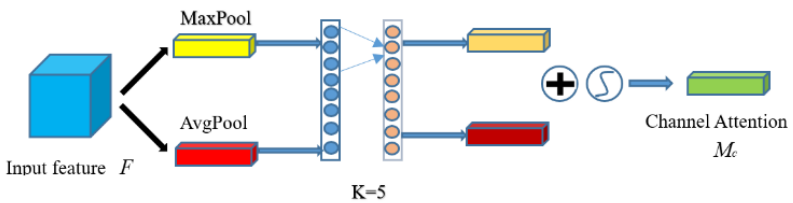


Figure 3. Channel Attention Module

Spatial attention. We aggregate the channel information of the input feature maps through two pooling operations to generate two 2D feature maps $F^s_{avg} \in \mathbb{R}^{1\times H\times W}$ and $F^s_{\max} \in \mathbb{R}^{1\times H\times W}$. Each represents a max-pooled feature and an average-pooled feature in a channel. Then the two feature maps are spliced. The convolution kernels of different sizes are used to enhance the adaptability of the model to different scale features, and they are convoluted by two standard convolutions of different sizes. The two convolutional results are added element-wise and operated by the Sigmoid function to generate two-dimensional spatial attention. In short, spatial attention is computed as:

$$\begin{aligned} M_s(F) &= \sigma(f^{7\times7}([AvgPool(F);MaxPool(F)]) + f^{3\times3}([AvgPool(F);MaxPool(F)])) \\ &= \sigma(f^{7\times7}([F^s_{avg};F^s_{\max}]) + f^{3\times3}([F^s_{avg};F^s_{\max}])) \end{aligned} \tag{5}$$

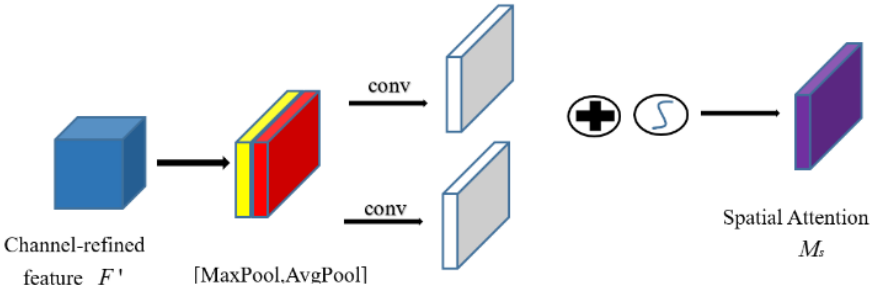


Figure 4. Spatial Attention Module

3.3 YOLOv3 with C_ECA

The mixed attention C_ECA proposed in this paper is a general-purpose CNNs that can be inserted anywhere in the network. This paper integrates the attention model into YOLOv3 and mainly follows the following two principles:

Does not significantly increase the network complexity. Because YOLOv3 itself is a one-stage network, designed for real-time object detection, it is not necessary to add attention to every position in the entire network.

In the deep convolutional neural network, the features of the shallow layer are more general, and it conforms to the general features of the image. while the features of the deep layer are more abstract and complex, and their representational ability is more unique, which is more suitable for attention.

According to the above two principles, this paper does not modify the structure of YOLOv3's backbone network Darknet53. As shown in Figure 5, we choose to introduce attention to the three prediction layer branches of the network. So it can increase the network's ability to select and capture features.

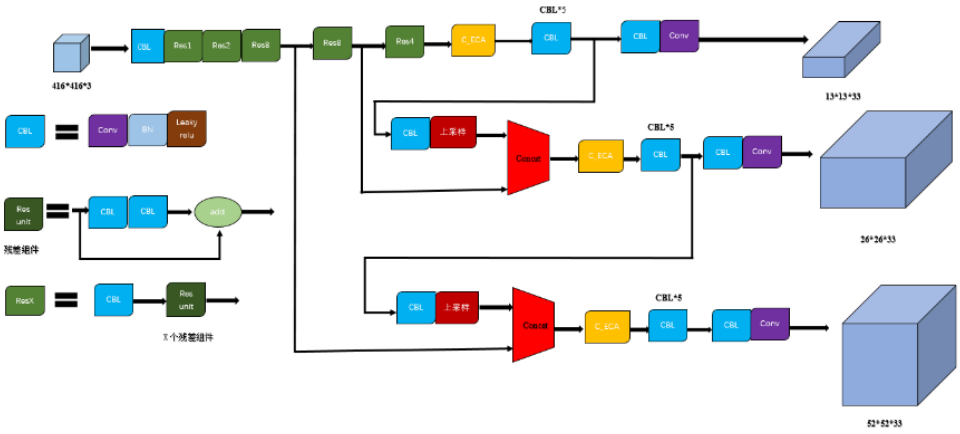


Figure 5. Improved YOLOv3 network structure

4. Experiment

4.1 Evaluation indicators

In order to evaluate the detection performance of the model, this paper uses the average precision (AP) of various defects, the average precision (mAP) of all categories, and FPS as the evaluation indicators. The formula for calculating AP is as follows:

$$AP = \int_0^1 P(R) d_R \tag{6}$$

In formula (6), P is precision, R is Recall, and the specific calculation formulas of precision and Recall are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{8}$$

In formula (7)(8), TP (True Positive) represents the number of positive samples predicted by the model as positive, FP (False Positive) represents the number of negative samples misjudged as positive by the model, and FN (false negative) indicates the number of positive samples misjudged as negative by the model. Through the above formula, we can take the precision and recall obtained as the ordinate and abscissa of the coordinate axis respectively. and the P-R curve is drawn by selecting different thresholds corresponding to different Precision and Recall values. The various AP values obtained by calculating the area of the P-R curve are averaged to obtain the mAP value. The calculation formula is as follows:

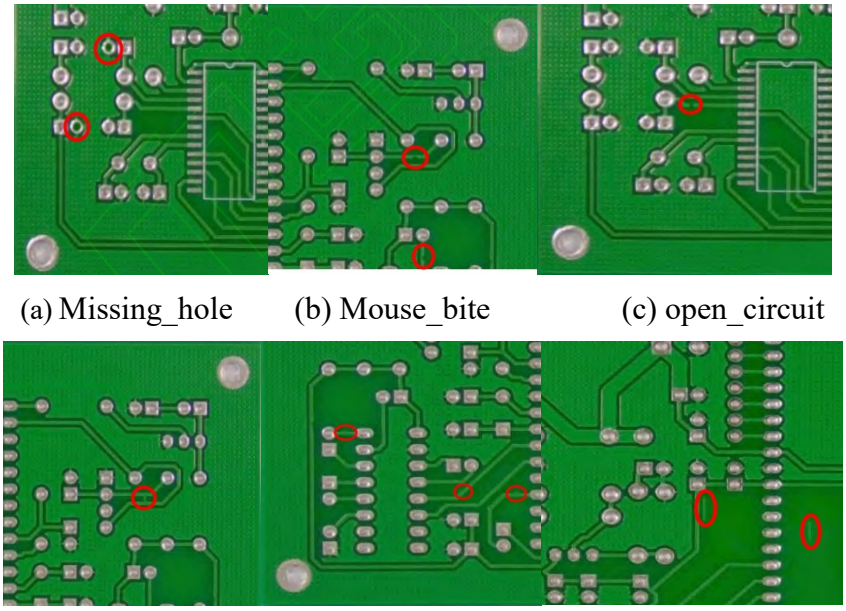
$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{9}$$

N in the formula (9) is the number of defect categories. In the process of calculating mAP, we judge whether the bounding box of target prediction is correct according to the positioning accuracy of the measured object. We use intersection over Union (IOU) to measure the error between the prediction frame and object real frame. A pre-defined IoU threshold, if the IoU of the predicted box and the ground-truth box is greater than the threshold, the object will be considered successfully detected, otherwise it will be identified as false detection. When calculating mAP in this experiment, the IoU threshold is 0.5. FPS is a measure of the model detection speed. That is the number of images processed per second.

4.2 Experimental environment and data

The operating system used in the experiment is Windows 10, and the hardware configuration used in the experiment is: Central Processing Unit (CPU): Inter(R) Core(TM) i9-9900KCPU @3.60GHz; Graphics Processing Unit (GPU): NVIDIA GeForce RTX 3080. The software environment is CUDA 11.1, cuDNN v8.0.5.39, and the deep learning framework used in this article is pytorch1.8.0.

Due to the fact that there are few defective PCB bare boards produced by the factory, it is difficult to obtain first-hand data pictures. There are few defect datasets, and it is difficult to sample the original image data of the data set. Therefore, the data set of the Intelligent Robot Open Laboratory of Peking University is used. The defect types in the data set include PCB open circuit, short, spur, mouse bite, spurious copper and missing hole. There are a total of 693 original images. Later, we expanded the number of pictures to 10668 by using the data expansion method of random horizontal flip and brightness adjustment. The 6 defect types of the dataset are shown in Figure 6.



(a) Missing_hole (b) Mouse_bite (c) open_circuit

(d) Short (e) Spur (f) spurious copper

Figure 6. PCB defect image

4.3 Experimental environment and data

Experimental environment and data

In order to verify the advantages and disadvantages of the algorithm in this paper compared with other object detection algorithms, a comparative experiment was carried out, and the results are shown in Table 1. Faster R-CNN is a typical two-stage detection algorithm. Efficientdet, SSD and YOLOv4 are all commonly used one-stage detection algorithms. Comparing the performance of each algorithm on the dataset, it can be found that the SSD and YOLOv4 algorithms have better performance, while the Faster R-CNN and Efficientdet algorithms have poor performance. Faster R-CNN is a two-stage algorithm, which is much slower than the one-stage algorithm. The poor detection effect on PCB data set is mainly due to the large network receptive field, which is not conducive to thedetection of small object. The poor performance of Efficientdet is because the downsampling multiple of the backbone network is too large, and there is too much useless information at the bottom, which is not conducive to the detection of PCB defects. The algorithm in this paper has better accuracy in PCB defect detection, with mAP reaching 99.42%. Compared with other algorithms, the AP of various defects has been improved. In terms of detection speed, the FPS of this algorithm is 40.38, which is far slower than SSD, but it can meet the requirements of industrial detection speed.

Analysis of the experimental results of the algorithm in this paper

In this paper, the training dataset and the test dataset are set according to the ratio of 9:1. During the training process, We choose Adam optimizer to optimize the network parameters, set β_1 to 0.9, set β_2 to 0.999, and set the training batch size to 8. We load the pre-training model of yolov3 backbone network during training. first freeze the parameters of the backbone network to train 20 epoch, the learning rate is set to 1E-3, and then the entire network is unfrozen to train the global parameters for 80 epochs, and the learning rate is set to 1E-4. In order to verify the detection effect of different attention modules embedded in the neural network, five groups of experiments were set up to evaluate the detection effect. The results of the ablation experiments are shown in Table 2.

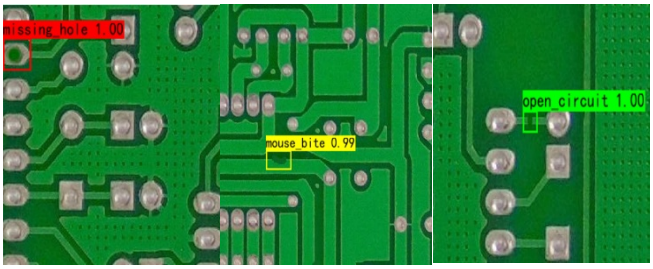
It can be seen from Table 2 that the neural network embedding channel and spatial attention have the best detection performance. Although the speed is not as fast as the single embedding channel or spatial attention, the performance is improved. The training results of the improved YOLOv3 algorithm proposed in this paper are applied to PCB defect detection, and the detection effect is shown in Figure 7. It can be seen from Figure 7 that the method proposed in this paper can accurately realize the defect detection of PCB. Comparison of detection performance of different algorithms

Table 1. Comparison of detection performance of different algorithms

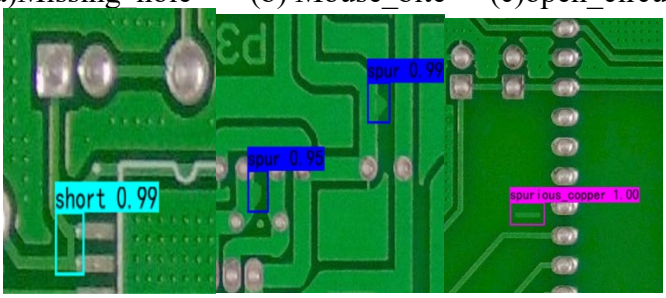
Methods	*FPS*	*mAP/%*	*AP/%*					
			missing hole	*mouse bite*	*open circuit*	*short*	*spur*	*Spurious copper*
Faster R-CNN	1.02	71.55	88.90	81.51	77.91	69.40	51.13	60.45
Efficientdet(D0)	25.59	71.93	96.84	67.86	54.30	85.84	61.26	65.45
SSD	53.39	91.28	96.14	90.93	79.84	94.41	92.23	94.13
YOLOv4	33.49	97.11	98.65	97.28	97.00	96.39	96.68	96.64
The algorithm of this paper	40.38	99.42	99.69	99.82	99.54	99.46	98.36	99.64

Table 2. Ablation experiments

Methods	*FPS*	*mAP/%*	*AP/%*					
			missing hole	*mouse bite*	*open circuit*	*short*	*spur*	*spurious copper*
YOLOv3	47.65	99.23	99.99	99.54	98.79	99.04	98.51	99.51
YOLOv3+spatial	41.76	99.28	99.95	99.60	99.11	99.39	98.46	99.13
YOLOv3+channel	43.35	99.40	99.68	99.86	99.27	99.24	98.81	99.57
YOLOv3+spatial+channel	38.88	99.38	99.70	99.61	99.58	99.27	98.89	99.22
The algorithm of this paper	40.38	99.42	99.69	99.82	99.54	99.46	98.36	99.64



(a)Missing_hole (b) Mouse_bite (c)open_circuit



(d) Short (e) Spur (f) spurious copper

Fig. 7 Application of improved YOLOv3 algorithm on PCB defect detection

5. Conclusion

By integrating the attention mechanism into the YOLOv3 network model, the network can pay more attention to the defect feature information in the feature map, and can also suppress irrelevant background and other non-defect secondary information. In this way, we can effectively improve the performance of the network model.

Acknowledgments

Supported by National Natural Science Foundation of China (No. 61701202, No. 61901196) and the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2021D14).

References

[1] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks [C] //Advances in neural information processing systems. 2015: 2017-2025. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Fu J, Zheng H, Mei T. Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition[C]// IEEE Conference on Computer Vision &Pattern Recognition. IEEE, 2017: 4438- 4446.

[3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic 7 segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580-587.

[4] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[5] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

[6] Hu B, Wang J H. Detection of PCB surface defects with improved Faster -RCNN and feature pyramid network[J]. IEEE Access,2020,8:108335-108345.

[7] Tsai D M,Chou Y H. Fast and precise positioning in PCBs using deep neural network regression[J].IEEE Transactions on Instrum entation and Measurement,2020,69(7): 4692-4701.

[8] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector [C] //European conference on computer vision. Springer, Cham, 2016: 21-37.

[9] Redmon J, Divvala S, Girshick R, et al. You Only look once: unified, real time object detection[C]//Computer Vision and Pattern Recognition. 2017: 6517-6525.

[10] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C] //IEEE conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.

[11] Redmon J, Farhadi A. YOLOv3: An incremental improvement [C] //IEEE conference on Computer Vision and Pattern Recognition, 2018, arXiv: 1804.0276.

[12] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.

[13] Li Y T, Kuo P, Guo J N. Automatic industry PCB board DIP process defect detection with deep ensemble method[C]// 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE). Delft, Netherlands: IEEE, 2020: 453-459.

[14] WOO S, PARK J, LEE J Y, et al. CBAM, convolutional block attention module[J]. arXiv, 1807.06521, 2018.

[15] HU J, SHEN L, SUN G, et al. Squeeze- and- excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8), 2011-2023.

[16] PARK J, WOO S, LEE J Y, et al. A simple and lightweight attention module for convolutional neural networks[J]. International Journal of Computer Vision, 2020, 128(9), 783-798.

[17] LI X, WANG W, HU X, et al. Selective kernel networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[18] WANG Q, WU B, ZHU P, et al. ECA-Net, efficient channel attention for deep convolutional neural networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.