

Received 19 November 2022, accepted 6 December 2022, date of publication 9 December 2022, date of current version 15 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228206

APPLIED RESEARCH

PCB Defect Detection Method Based on Transformer-YOLO

WEI CHEN^{ID}, ZHONGTIAN HUANG^{ID}, QIAN MU, AND YI SUN

School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, China
Xi'an Key Laboratory of Network Convergence Communication, Xi'an, Shaanxi 710054, China

Corresponding author: Zhongtian Huang (20207223098@stu.xust.edu.cn)

ABSTRACT In order to solve the problem of low accuracy and efficiency in printed circuit board(PCB) defect detection using reference methods, a Transformer-YOLO network detection model is proposed. Firstly, an improved clustering algorithm is used to generate the anchor box suitable for the PCB defect data set of this paper. Secondly, abandoning the traditional idea of using convolutional neural network to extract image feature, Swin Transformer is used as the feature extraction network, which can effectively establish the dependency between image features. Finally, to modify the order of the channels in the feature map and enable the network to more effectively focus on the information with greater value, the convolution and attention mechanism module is added to the feature detection network component. Comparing the network model proposed in this paper with Faster R-CNN, SSD, YOLOv3, YOLOv4 and YOLOv5, the experimental results show that the proposed model improves the accuracy by 23.90%, 15.51%, 10.70%, 7.83% and 6.12% respectively, which is better than other most mainstream target detection models and has relatively small volume.

INDEX TERMS PCB defects, clustering algorithm, convolutional neural network, Swin Transformer, attention mechanism.

I. INTRODUCTION

Printed circuit board is a key component in most electronic devices [1]. With the advent of the fourth industrial revolution, there is an increasing demand for electronic products such as smartphones and laptops in our life. As the bottom operating platform of its electronic products and devices [2], the future prosperity of the PCB industry is beyond unquestionable. At the same time, with the increasing demand for PCBs and the rising productivity, PCBs manufacture tends to be more complex and miniaturized [4], which brings a rather harsh challenge to the detection of defects in circuit boards. If there are certain defects in PCBs, it will lead to great changes in performance, indirectly affect the quality and sales of later products, and cause serious economic losses. Therefore, it is necessary to carry out defect detection during the production of PCBs, which can reduce the production

cost of the product and effectively improve the product qualification rate [5].

In the industrial production process, PCB defects will be caused by welding failure, improper operation mode and storage mode. The current PCB defects are mainly divided into six categories: short, open circuit, spur, spurious copper, mouse bite and missing hole [7]. PCB defect detection methods mainly include manual inspection, performance testing, reference comparison method and non reference method [8]. Manual inspection mostly relies on the technicians' subjective assessment of whether defects are visible to the naked eye on the PCB. The efficiency of this method is low and it is simple to miss the inspection. Performance testing is mainly functional testing of the PCB to determine whether there are defects in the PCB, but this method will cause some damage to the PCB. Reference comparison method [9] is currently widely used in the industrial field, which mainly exploited a technical mean based on template matching. The PCB to be tested is compared with the standard template to determine whether there are certain defects, but

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan^{ID}.

the main drawback of the method is the need for accurate alignment in space, otherwise the false detection rate is high. In recent years, Automatic optical inspection (AOI) has become the main method used in industrial inspection, compared to other methods, the accuracy of the detection rate have been improved. However, due to the strict parameter setting requirements and high sensitivity of the AOI system, it is particularly prone to over-screening and therefore requires subsequent manual secondary screening [10]. Compared with the above mentioned traditional defect detection methods, the deep learning-based defect detection methods have better performance and can effectively avoid the low detection accuracy and efficiency. Therefore, deep learning-based defect detection methods have naturally become a hot topic of research in recent years.

Currently, the technical development of PCB defect detection using deep learning algorithms has been relatively mature. The relevant algorithms can be divided into two categories: One is one-stage target detection algorithms represented by SSD [11], YOLO series [12], etc. And the other is two-stage target detection algorithms represented by R-CNN [16], Fast R-CNN [17], Faster R-CNN [18], etc. Meanwhile, myriad scholars, regarding the task of PCB defect detection, have done a large sum of research and achieved some research results. Xin et al. [19] proposed an improved network model for YOLOv4, by analyzing the backbone architecture CSPDarknet53 and appropriately changing the settings of the model hyperparameters, thus increasing the accuracy rate to 96.88%. Moreover, Li et al. [20] achieved 93.07% detection accuracy by applying to a joint approach of real PCB images and virtual PCB images to preprocess the data, while changing the three-layer predicted output layer of YOLOv3 to a four-layer predicted output layer. Additionally, Hu et al. [21] completed the detection of common PCB defect types by improving the Faster RCNN and utilizing ResNet50 with image pyramids as a feature extraction network. Besides, Zeng et al. [22] proposed an enhanced multi-scale feature fusion method for the task of small target detection, as well as the effectiveness of the method was verified on the PCB data set, which also obtained fabulous results.

Based on the above-mentioned related literature research results, and in order to better balance the detection accuracy, detection speed and the volume of the network model for the task of PCB defect detection, the paper makes related improvements based on the YOLOv5 model and proposes a new network model. The main contributions of this paper are as follows:

- The original clustering algorithm is boosted in order to make the resulting anchor boxes more appropriate for the PCB defect data set, consequently improving the model's detection accuracy.
- Adopt Swin Transformer as the feature detection network instead of the traditional convolutional neural network, because Swin Transformer has a strong global interaction mechanism that can efficiently establish the dependencies

between image features, and there is no problem with limiting the receptive field in convolutional neural network.

- The attention mechanism module is added to the interactive part of the feature extraction network and the feature detection network, and the arrangement order of channels in the feature map is adjusted so that the network model can more efficiently focus on the channel information with high importance. Finally, the task of detecting and identifying common PCB defect types is completed.

II. METHOD

YOLOv5 is a widely used target detection model. Compared with YOLOv3 and YOLOv4, the detection accuracy of this model has been further improved. The prediction results of the network output mainly include the type information and location information of the target object. The overall architecture of the network is mostly consistent with YOLOv3 and YOLOv4. As demonstrated in Figure 1, it mainly includes four parts: The first part is the input section of the network, it mostly consists of the pre-processing of the input picture, including data augmentation and automatically learning anchor boxes, etc. The second part is the backbone network and YOLOv5 adopts CSPDarknet as its network for extracting picture features. The third part is the feature detection network, which fuses several layers of feature information top-down and bottom-up. The fourth part is the output network that displays the predictive information of the kind, confidence level, and position of the target object.

III. RELATED WORKS

This paper mainly carries out relevant improvements based on YOLOv5 network, however the model remains the same architecture, namely input, image feature extraction network, feature detection network, and output prediction results. Firstly, the original clustering technique in this paper was improved to make the resulting anchor boxes more appropriate for the target data set. Secondly, Swin Transformer is used for image feature extraction. Finally, the connection between the feature extraction network and the feature detection network now includes convolution and SE Block module. The overall architecture of its network is shown in Figure 2.

A. IMPROVED CLUSTERING ALGORITHM

The anchor boxes in the original YOLOv5 model is primarily formed by the k-means clustering algorithm and the genetic algorithm, and the anchor boxes have a direct impact on the network's detection accuracy. There are two major problems in the original model using K-means algorithm: One problem is to determine all the initial cluster center point by random selection in the clustering process, and the other problem is that Euclidean Distance is selected as the standard of similarity clustering between samples. In conclusion, these two problems will have an impact on the generated anchor boxes as well as the model's detection

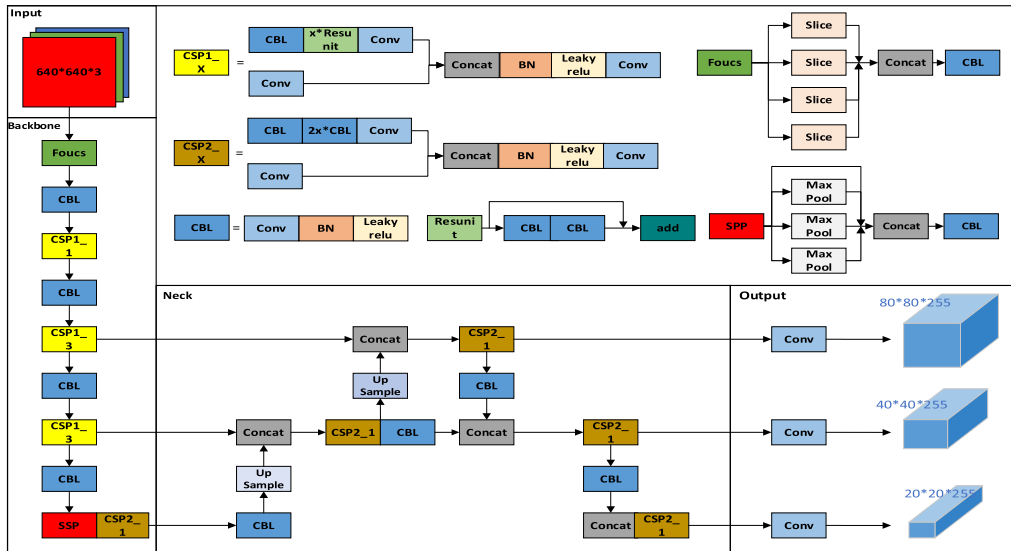


FIGURE 1. Architecture of YOLOv5.

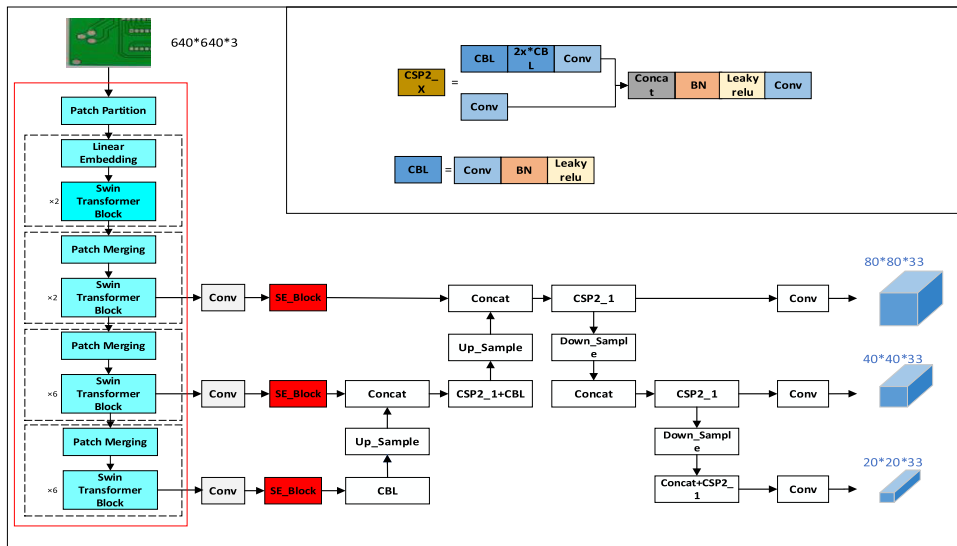


FIGURE 2. Network architecture of this paper.

accuracy. Therefore, in this paper, K-means++ is selected to cluster data sets. By randomly selecting a clustering center iteratively until all the clustering centers are selected. And secondly the main purpose of setting the anchor boxes is to maximize the overlap area between the prediction box and the ground truth box, so the clustering metric in the clustering algorithm is redefined and formula 1 is used to remeasure the distance formula between samples in the data set. The final result of its clustering is shown in Table 1, and the main process of the improved clustering algorithm is as follows:

Step1: Choose one of the PCB defect data set samples to serve as the initialization cluster center.

Step2: Calculate the distance between all the remaining samples in the data set and the initialization cluster center according to formula 1.

TABLE 1. Different clustering algorithms.

Algorithm	Original clustering algorithm	Clustering algorithm in our model
anchor boxes	(18,18),(24,23)(19,32) (33,20),(30,30)(25,49) (50,25),(40,39)(55,55)	(19,19),(23,23)(20,33) (33,20),(30,30),(24,49) (50,24),(38,38)(47,47)

Step3: Calculate the probability P(x) of each sample being selected as the next cluster center according to formula 2.

Step4: To choose all nine cluster center points, repeat Steps 2 and 3, where 9 is the number of anchor boxes that will be produced by the network model in this paper.

Step5: For each sample x in the data set, calculate the distance between the sample and the 9 cluster centers selected above according to formula 1, and divide the sample into the category corresponding to the nearest cluster center.

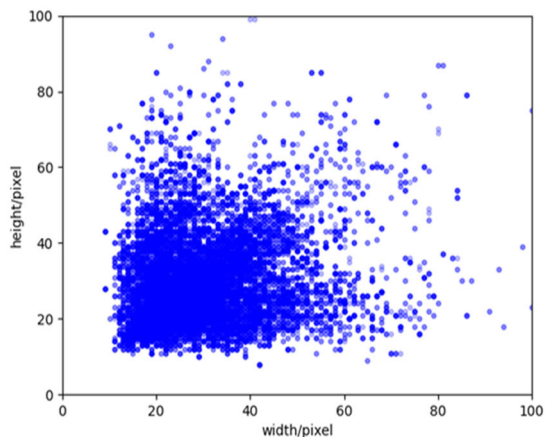


FIGURE 3. Data set distribution.

Step6: For each cluster category, recalculate the cluster center according to formula 3, and repeat Step5 until the location of the cluster center does not change.

$$D(x, c) = 1 - IOU(x, c) \tag{1}$$

$$P(x) = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \tag{2}$$

$$C_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{3}$$

B. IMPROVED THE FEATURE DETECTION NETWORK

Defects on PCBs are typically characterized by a relatively small area. The point is best illustrated by the data set distribution shown in Figure 3 in this paper. The defect’s pixel size is mostly less than 60, and it represents a very modest part of the entire image in the photograph. At present, the convolutional neural network, which is made up of a series of convolution kernels of various sizes, is still the primary source of support for the feature extraction network of one-stage and two-stage target detection models. As an illustration, take the YOLOv5 model. The model employs CSPDarknet as the feature extraction network, which is mostly made up of Focus and CSP [23] architecture, and its network comprises a significant amount of convolution kernels. Although the depth of convolutional neural networks can be adjusted freely, if only a small number of convolution kernels are used in the network design process, the model can only extract a part of the shallow feature information of the image due to the lack of depth. On the contrary, if it uses a lot of convolution kernels in the process of network design, more deep semantic information can be extracted. But at the same time, a new problem arises: because the network architecture is excessively complex, the feature transfer effect across layers is weakened, which reduces the capacity to extract local fine features. The convolutional neural network also has poor capacity to learn global aspects of the picture and pays more attention to local feature information of the image, neglecting the correlation degree of contextual

feature information, as it iterates between layers, eventually expanding the receptive field.

Since Google proposed the model Transformer [24] in 2017, it has profoundly accelerated the development process in the field of Natural Language Processing. In the past two years, many researchers tried to introduce Transformer model into the field of Computer Vision, and achieved relatively remarkable results. The case that Carion et al. [25] applied Transformer to target detection for the first time and proposed a new network framework DETR in 2020 is a prime example. In October of the same year, Dosovitskiy et al [26] proposed the variant network architecture ViT model of Transformer, an image classification network based on pure attention mechanism. During the development of this field of Computer Vision, convolutional neural network model has occupied a dominant position and played an irreplaceable role and its core is mainly the convolution kernel module. And the disadvantages of the convolution neural network model are discussed in detail in the previous paper. As opposed to the convolutional neural network, the Transformer is mostly made up of the attention mechanism module construction, which is entirely distinct from the former model. The Transformer model can quickly determine the degree of connection between image features based on its attention mechanism module, fully utilize upper and lower information, and accomplish the modeling of both global and local image features with a superior global interaction mechanism.

In this paper, the feature extraction network is mainly improved, and the Swin Transformer [18] is used to replace the convolutional neural network. The model is mainly improved on the basis of the Transformer, which is composed of Patch Partition, Linear Embedding, Swin Transformer Block, etc. And the core of the network is the Swin Transformer Block module. In this module, the computation process of the network can be parallelized by using the window attention mechanism; the problem of limited receptive field in convolutional neural networks is skillfully solved by using sliding window multi-head attention mechanism. According to the number of stacks in the Swin Transformer Block and heads used in the attention mechanism, the Swin Transformer Block can be divided into four models with different parameter configurations. In this paper, Swin Transformer Tiny is used as the feature extraction network, the specific network architecture, parameter configuration and shape changes of feature map shown in Figure 4.

The network model Swin Transformer Tiny of Figure 4 is explained here with a color image of input size 640 × 640: Firstly, the network will use Patch Partition to divide the input image into 160 × 160 size according to the non-overlapping size of 4 × 4 pixels, with the number of channels unchanged, and a total of 16 patches will be obtained. The 16 patches will be superposition and flatten on the channels so as to obtain the feature map of 160 × 160 × 48. Secondly, the network implement cascading stacking of Stage1, Stage2, Stage3, and Stage4 modules. The Linear Embedding architecture

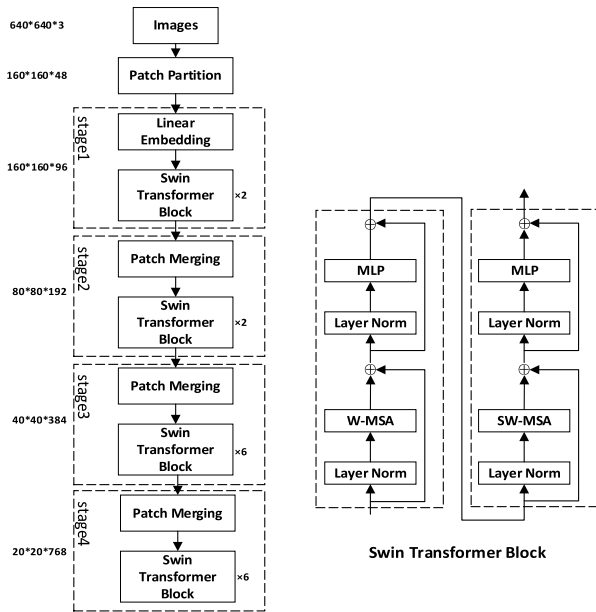


FIGURE 4. Swin transformer tiny.

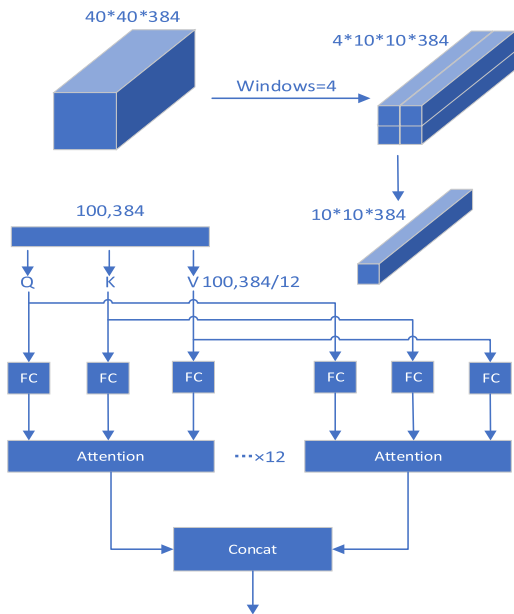


FIGURE 5. Window attention mechanism.

in Stage1 only adjusts the number of channels without changing the size of the feature map. Simultaneously, the Patch Merging architecture of Stage2 through 4 changed the size of the feature map and the number of channels. However, the Swin Transformer Block module, does not alter the geometry of the feature map prior to or following its input and output. The module is mainly composed of window attention mechanism, sliding window attention mechanism, layer normalization and multilayer perception mechanism, which need to be used in pairs, the former layer uses window attention mechanism, and the latter requires a sliding window attention mechanism. The calculation process of

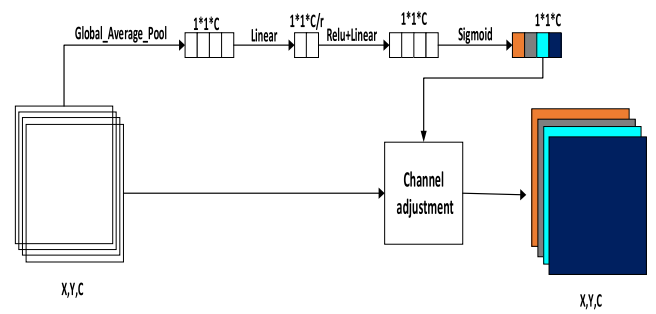


FIGURE 6. SE block.

window attention mechanism is revealed in Figure 5, and the data in the figure mainly correspond to the Swin Transformer Block module in Stage3. The window attention mechanism in the module mainly divides the input feature map according to the specified window size, and uses the way of parallelization to process. The sliding window mechanism mainly enables feature information interaction between different windows, which can not only rapidly increase the size of receptive field, but also pay attention to local feature information and global feature information at the same time.

C. IMPROVE THE FEATURE DETECTION NETWORK

Because the original network model pays attention to the feature information contained in all channels in the feature map indiscriminately, this method is obviously unreasonable and ineffective. For the information of different channels in the feature map, their contributions to network detection are different. Thus, in this paper, considering the importance of network training in Transfer Learning and not destroying the architecture of the feature extraction network, our model choose to add SE Block [28] to the interaction part between the feature extraction network and the feature detection network. The main purpose is to learn a set of weights to represent the importance of the feature information contained in different channels in the feature map, and reorder the channels of the feature map according to the importance of the feature information contained in the channels. Therefore, the network can focus on the channel information with higher importance and ignore a part of the channel information with relatively lower significance. The overall architecture of SE Block is clarified in Figure 6. It main process is as follow: Firstly, after the shape of the given input feature map being (X, Y, C) , according to formula 4, the information contained in different channels in the feature map is transformed into an objective data one by one, thereafter combined to obtain a vector of $1 \times 1 \times C$. Secondly, dimension reduction and dimension increase are carried out through two fully connected layers, and the Sigmoid function is used to fix the output value of the vector between 0 and 1, plus the output value represents the importance of feature information contained in each channel number in the original feature map. Ultimately, the vector is multiplied with the original feature map, and the order of the channels in the feature map is

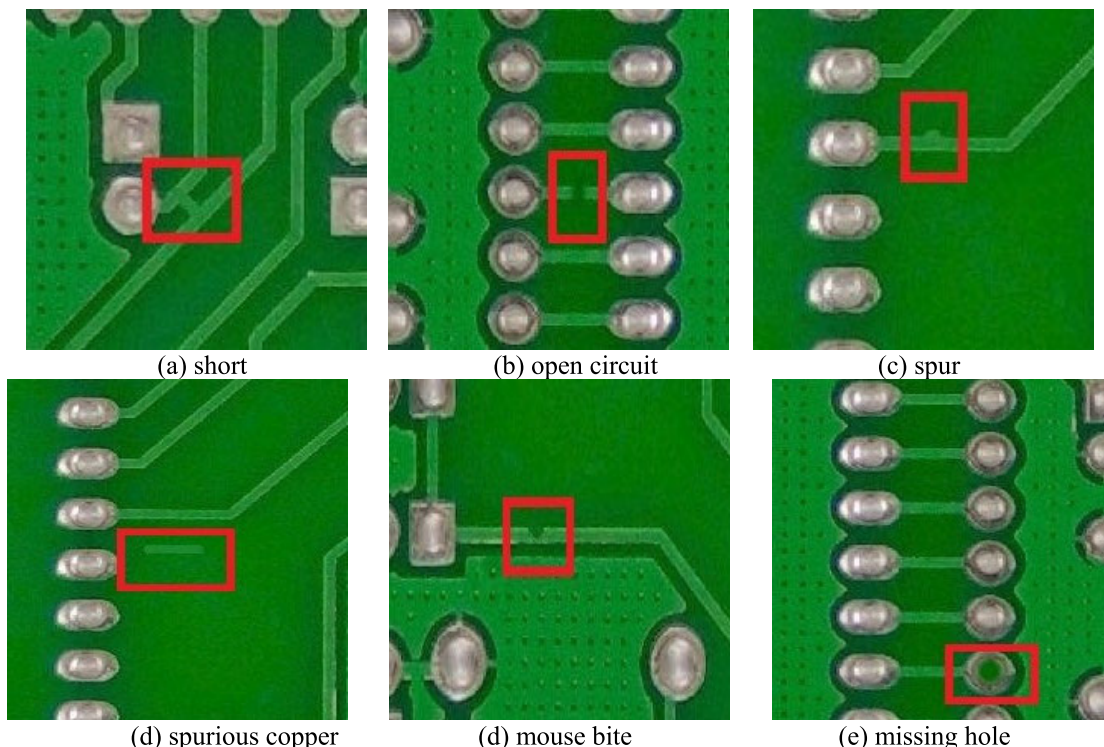


FIGURE 7. Defects Of PCB.

TABLE 2. PCB dataset.

Defect types	Number f original data distribution	Expand the number of data set distributions
short	116	1732
open circuit	116	1740
spur	115	1752
spurious copper	116	1760
mouse bite	115	1852
missing hole	115	1832
Total number	693	10668

readjusted according to the importance of the channels, a new feature map is obtained consequently.

$$Z_c = \frac{1}{X \times Y} \sum_{i=1}^X \sum_{j=1}^Y U_c(i, j) \tag{4}$$

IV. EXPERIMENT

A. EXPERIMENTAL PLATFORM

Operating system: Windows10, CPU: Inter(R) Core(TM) I7-9700FCPU, GPU: NVIDIA GeForce RTX 2070, deep learning framework: PyTorch.

B. EXPERIMENTAL DATA

The data set comes from the PCB defect data set published by Intelligent Robotics Open Laboratory of Peking University. The defect types mainly include short, open circuit, spur, spurious copper, mouse bite and missing hole. In order to prevent network overfitting, we expanded the original 693 data sets by random rotation, random cropping, brightness adjustment, noise addition, etc. The final number of data sets reached

10,668. The distribution of the number of various defects is shown in Table 2, and the six defects are shown in Figure 7.

C. NETWORK TRAINING

All the experiments involved in this paper adjusted and set the parameters on the basis of the pre-training weight. There were 300 rounds of network training. The batch size was 8 for the first 50 rounds of frozen network training, and the batch size was 4 for the last 250 rounds of unfrozen network training. Data set division: (training set + validation set): test set = 9:1, in which the ratio of training set and validation set is still 9:1, and the final training set is divided into 8640 pieces, the validation set into 961 pieces, and the test set into 1067 pieces.

D. EVALUATION INDEX

In this paper, the mean average precision (MAP), detection speed (FPS) and model size (MB) are the main reference indexes to evaluate the advantages and disadvantages of the

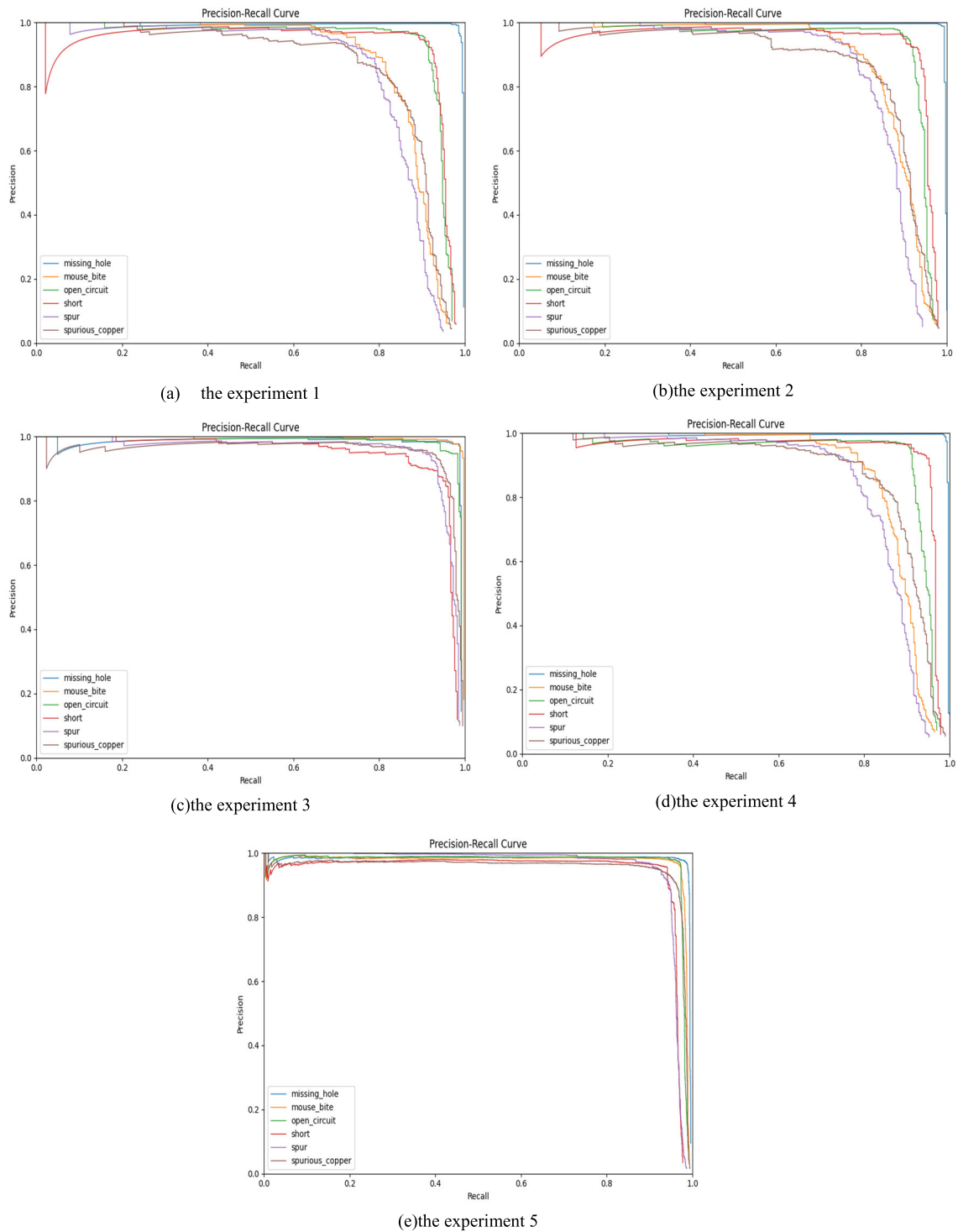


FIGURE 8. Precision-recall curve.

model.

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$AP = \int_0^1 P(R)dR \tag{7}$$

$$MAP = \frac{\sum_{i=0}^n AP(i)}{n} \tag{8}$$

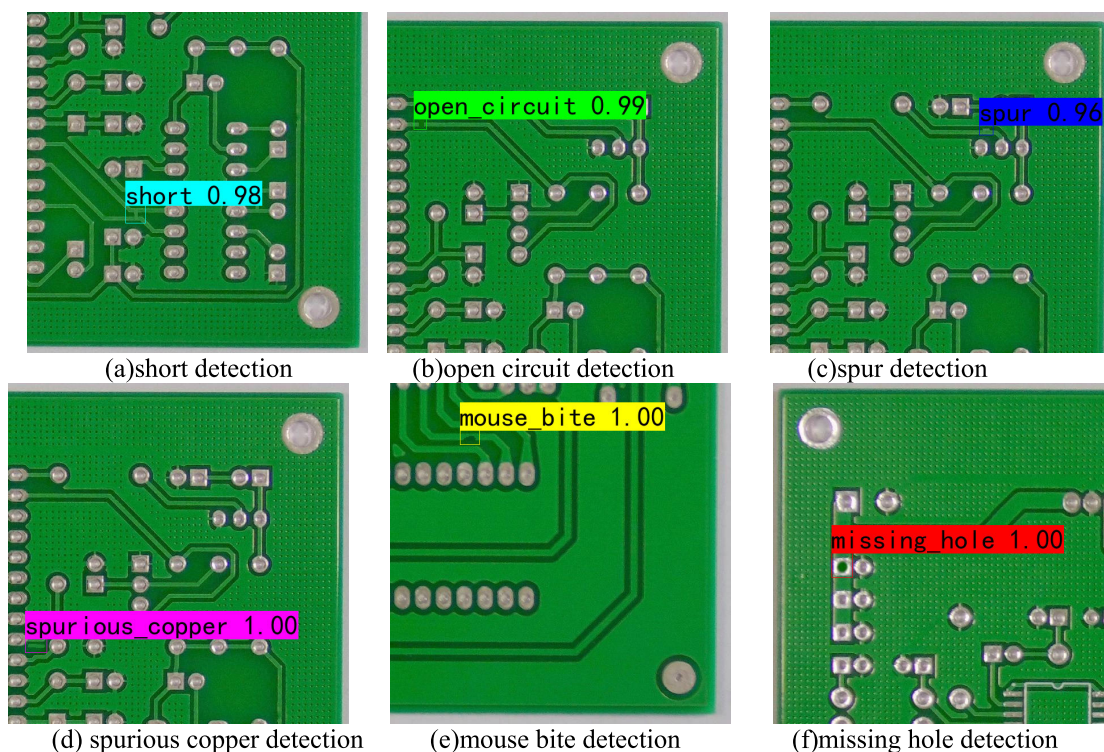


FIGURE 9. Defect detection effect.

In the above Figure, P and R represent precision and recall respectively. P can be understood as the accuracy of the network that does not mistake the background for the target. R can be understood as the sensitivity of the network that does not mistake the target for the background.

Here, TP, FP and FN are explained in combination with the data set in this paper: TP represents the actual PCB defect, and the model predicts that it is PCB defect. FP indicates that the actual is not a PCB defect, and the model predicts that it is a PCB defect. FN is actually a PCB defect, and the model prediction is not a PCB defect.

V. RESULTS AND ANALYSIS

A. ABLATION EXPERIMENTS

Ablation experiments were designed to verify the three improvements mentioned above one by one. There are mainly five groups of experiments: The experiment 1 represents the original YOLOv5 model. The experiment 2 represents the modification of clustering algorithm based on the original model. The experiment 3 represents the replacement of feature extraction network based on the original model. The experiment 4 represents the addition of attention mechanism architecture based on the original model, and the experiment 5 represents the proposed model in this paper. Meanwhile, for the models used in the five groups of experiments, the PR curves of the six types of defects are shown in Figure 8. And the MAP and AP of the six types of defects are shown in Table 3.

From the analysis of the results in Figure 8, it can be seen that the area enclosed by the PR curve and the coordinate axis represents the detection accuracy of each defect category. Here we can roughly see that the area enclosed by the PR curve of the six defect categories is obviously larger for the proposed model in this paper, that is, the PR curve of 8(e), compared to the PR curve of 8(a) experiment 1.

From the analysis of the results in Table 3, the experiment was analyzed as follows: For experiment 1, the MAP of six types of defects reached 90.92%, and the overall accuracy was not very low. However, from the AP value, it can be found that the original model had a poor detection effect for three types of defects, mouse bite, spurious copper and spur, which was lower than the MAP, while the detection performance for missing hole was extremely good. For experiment 2, only the clustering algorithm of the original algorithm was improved. It can be seen from the results in Table 3 that the MAP was increased by 0.62%, and the detection accuracy of the six types of defects was basically improved to a certain extent by improving the clustering algorithm. For experiment 3, Swin Transformer is used as the image feature extraction network. It can be seen from the results in Table 3 that the MAP value is improved by 5.33%, and the improvement effect is obvious. Meanwhile, compared with the AP index of six types of defects in experiment 1, the AP value in experiment 3 is significantly improved. In particular, the detection accuracy of three kinds of defects, including spur, spurious copper and mouse bite, is improved remarkably, and the detection accuracy of six kinds of defects is not

TABLE 3. Ablation experiments.

experiments	MAP(%)	AP(%)					
		missing hole	short	open circuit	mouse bite	spurious copper	spur
1	90.92	99	93	93	88	86	85
2	91.54	100	94	93	89	87	86
3	96.25	98	97	97	95	95	94
4	91.52	100	95	93	89	88	85
5	97.04	98	98	97	96	95	95

TABLE 4. Comparison of the effect of common algorithms.

Model	MAP(%)	FPS(s)	Volume(MB)
Faster R-CNN(VGG16)	73.14	6	522.98
SSD	81.53	14	100.27
YOLOV3	86.34	23	236.32
YOLOV4	89.21	20	245.53
YOLOv5	90.92	34	31.76
Our model	97.04	21	93.95

different. For experiment 4, the attention mechanism module was mainly added to the interaction between image feature extraction network and feature detection network. As can be seen from Table 3, MAP was improved by 0.60%, and the detection accuracy of most of the six types of defects was improved to a certain extent. For experiment 5, the proposed network model in this paper, MAP is improved by 6.12% to 97.04%. Among the five groups of experiments, MAP index of the proposed model in this paper has the best performance. And it also maintains good detection accuracy for six types of defects.

B. COMPARISON OF ALGORITHMS

The network model proposed in this paper is compared with other current mainstream object detection models. The results are shown in Table 4. As can be seen from the Table 4, compared with the Faster R-CNN network, the proposed model has certain advantages in model volume and detection speed, and the detection accuracy is improved by 23.9%. Compared with SSD network, the detection speed of the proposed algorithm is improved by 15.51% even though the network model volume and reasoning speed are slightly superior. Compared with the YOLO series algorithm, taking YOLOv3 and YOLOv4 as an example, the network model in this paper is relatively small under the condition of little difference in detection speed, and the detection accuracy is improved by 10.70% and 7.83%. Compared to the YOLOv5 algorithm, the speed of detection is increased by 6.12%, despite sacrificing some detection speed and model size. The model achieved 97.04% detection accuracy and the model of this paper is better than the current mainstream target detection algorithm. Finally, the model in this paper is applied to PCB defect detection. The detection effects of six common defects are shown in Figure 9.

VI. CONCLUSION

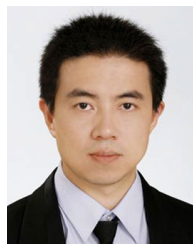
In contrast to the prevalent target detection network model currently in use, this research suggests a novel hybrid network model for the detection of PCB defects. First and foremost,

the problems with the current clustering algorithm are fixed in order to produce a more appropriate anchor boxes for the PCB defect dataset in this paper. After that, the attention mechanism network is utilized to extract image features, and the interaction between the image feature extraction network and the feature detection network is enhanced by the addition of the attention mechanism module. Briefly said, it is confirmed that the network model developed in this study achieves optimum detection accuracy on PCB public datasets and is better than the present mainstream techniques, which can essentially match the demands of industrial applications.

REFERENCES

- [1] A. Angelopoulos, E. T. Michailidis, N. Nomikos, P. Trakadas, A. Hatziefremidis, S. Voliotis, and T. Zahariadis, "Tackling faults in the industry 4.0 era—A survey of machine-learning solutions and key aspects," *Sensors*, vol. 20, no. 1, p. 109, Dec. 2019.
- [2] D. Powell, M. C. Magnanini, M. Colledani, and O. Myklebust, "Advancing zero defect manufacturing: A state-of-the-art perspective and future research directions," *Comput. Ind.*, vol. 136, Apr. 2022, Art. no. 103596.
- [3] V. A. Adibhatla, H.-C. Chih, C.-C. Hsu, J. Cheng, M. F. Abbod, and J.-S. Shieh, "Defect detection in printed circuit boards using you-only-look-once convolutional neural networks," *Electronics*, vol. 9, no. 9, p. 1547, Sep. 2020.
- [4] F. Psarommatis, G. May, P.-A. Dreyfus, and D. Kiritsis, "Zero defect manufacturing: State-of-the-art review, shortcomings and future directions in research," *Int. J. Prod. Res.*, vol. 58, no. 1, pp. 1–17, Jan. 2020.
- [5] K. Goto, K. Kato, T. Saito, and H. Aizawa, "Adversarial autoencoder for detecting anomalies in soldered joints on printed circuit boards," *Proc. SPIE*, vol. 29, no. 4, 2020, Art. no. 041013.
- [6] L. Huang, S. Shen, F. Xie, J. Zhao, J. Han, and K. Feng, "A novel multi-pattern solder joint simultaneous segmentation algorithm for PCB selective packaging systems," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 13, Dec. 2019, Art. no. 2058005.
- [7] R. Ding, L. Dai, G. Li, and H. Liu, "TDD-Net: A tiny defect detection network for printed circuit boards," *CAAI Trans. Intell. Technol.*, vol. 4, no. 2, pp. 110–116, Jun. 2019.
- [8] D.-M. Tsai and Y.-H. Chou, "Fast and precise positioning in PCBs using deep neural network regression," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4692–4701, Jul. 2020.
- [9] R. Ye, M. Chang, C.-S. Pan, C. A. Chiang, and J. L. Gabayno, "High-resolution optical inspection system for fast detection and classification of surface defects," *Int. J. Optomechatron.*, vol. 12, no. 1, pp. 1–10, Jan. 2018.
- [10] Y.-S. Deng, A.-C. Luo, and M.-J. Dai, "Building an automatic defect verification system using deep neural network for PCB defect classification," in *Proc. 4th Int. Conf. Frontiers Signal Process. (ICFSP)*, Sep. 2018, pp. 145–149.
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2016, pp. 21–37.
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [13] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.
- [14] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

- [15] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [17] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [19] H. Xin, Z. Chen, and B. Wang, "PCB electronic component defect detection method based on improved YOLOv4 algorithm," *J. Phys., Conf. Ser.*, vol. 1827, no. 1, Mar. 2021, Art. no. 012167.
- [20] J. Li, J. Gu, Z. Huang, and J. Wen, "Application research of improved YOLO v3 algorithm in PCB electronic component detection," *Appl. Sci.*, vol. 9, no. 18, p. 3750, Sep. 2019.
- [21] B. Hu and J. Wang, "Detection of PCB surface defects with improved faster-RCNN and feature pyramid network," *IEEE Access*, vol. 8, pp. 108335–108345, 2020.
- [22] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.
- [23] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.
- [25] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 213–229.
- [26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [27] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted Windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
- [28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.



WEI CHEN received the Ph.D. degree. He is currently an Associate Professor and a Master's Tutor. His research interests include computer vision, artificial intelligence, image processing, and optoelectric detection.



ZHONGTIAN HUANG is currently pursuing the M.S. degree in electronic information with the Xi'an University of Science and Technology, Xi'an, China. His research interests include computer vision, artificial intelligence, image processing, and optoelectric detection.



QIAN MU is currently pursuing the M.S. degree in electronic information with the Xi'an University of Science and Technology, Xi'an, China. Her research interests include computer vision, artificial intelligence, image processing, and optoelectric detection.



YI SUN received the Ph.D. degree. He is currently an Associate Professor and a Master Tutor. His current research interests include blockchain, game theory, swarm robotics, and distributed systems.

• • •