Surface Defect Detection Method of Strip Steel Based on Improved YOLOv5

Bin Wang, Ronaldo Juanatas, Jasmin Niguidula, and Hainan Luo

Abstract A strip steel surface defect detection method called YOLOv5-ABS is developed in order to address the issues of low detection accuracy, insufficient feature extraction ability, and insufficient feature fusion of YOLOv5. Firstly, in order to enhance the backbone network's capacity for feature extraction, the C3 module is swapped out for the SeC3 module with an attention mechanism. Secondly, in order to improve the network feature fusion capability, the bidirectional weighted feature pyramid network BiFPN is added in the Neck section. Finally, by introducing the SPPFCSPC spatial pyramid pooling structure, speed and accuracy are improved while keeping the receptive field unchanged. According to the experimental findings, the revised YOLOv5-ABS algorithm's mAP on the NEU-DET dataset is 78.6%, 3.8% larger compared to the initial YOLOv5s algorithm, and the detection speed is 142.8 FPS, enabling the quick and precise identification of strip steel surface defects.

Keywords Detected defects · Attention · YOLOv5 · BiFPN · SPPFCSPC

1 Introduction

Strip steel is an important industrial raw material in the field of mechanical manufacturing. Due to the influence of external factors such as the production process, raw material quality, and processing equipment, the surface of the product may produce various types of defects, such as rolled-in scale, scratches, and patches, during the strip steel production process. These flaws will not only degrade the product's look but will also reduce its steel strength, corrosion resistance, and wear resistance, and

B. Wang (✉) · R. Juanatas · J. Niguidula
Technological University of the Philippines, Manila, Philippines
e-mail: wangbin@ahcme.edu.cn

B. Wang
Anhui Technical College of Mechanical and Electrical Engineering, Wuhu 241002, Anhui, China

H. Luo
WuHu Hit Robot Technology Research Institute, Wuhu 241002, Anhui, China

W. Wang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 1032, https://doi.org/10.1007/978-981-99-7505-1_27

may even result in a safety mishap. As a result, one of the hottest topics in the world of mechanical manufacturing is how to quickly and precisely accomplish the identification of strip surface flaws.

Surface defect detection methods based on machine vision have received a lot of attention as machine vision technology has advanced. A typical machine vision defect detection system is made up of two parts: hardware and software. The hardware system mainly refers to the image acquisition device composed of an industrial camera, an industrial lens, a light source, and other equipment, which mainly completes the image acquisition work; the software system first performs noise reduction, segmentation, and morphological processing on the collected images, and then completes recognition and classification according to the extracted image feature information. Based on the shape template matching algorithm, some researchers have proposed an automatic detection method for shape defects of stamped parts, but this method is greatly affected by the light environment and is sensitive to noise [1]. Liu uses bilateral filtering and the Hilditch algorithm to improve the traditional Sobel operator and proposes a new edge detection algorithm for strip steel surface defect images. However, this method is not accurate for edge positioning, and the extracted edge lines are thicker [2]. A defect identification approach for solar cells based on SVM has been presented by some academics to address a number of common solar cell flaws, but it is challenging to train this system on huge amounts of data [3].

In the area of machine vision, deep learning has advanced quickly in recent years. Deep learning-based surface defect detection techniques have recently attracted a lot of attention in the target detection field. The deep learning-based approach may effectively address the complexity and unpredictability of manual feature extraction in conventional machine vision by automatically learning and extracting the input data's features [4]. To increase the effectiveness and accuracy of defect identification, academics domestically and internationally have applied deep learning to the detection of surface defects in a variety of products. A surface defect detection algorithm for a metal workpiece based on improved Faster RCNN is proposed by introducing multi-level ROI pooling layer structure and a bilinear interpolation method. This algorithm aims to address the issues of low accuracy and low speed of surface defect detection of metal materials [5]. Based on YOLOv3, Kou uses Anchor-Free to improve the speed of the model and designs dense blocks to extract richer feature information. The accuracy and robustness of the model have been improved [6]. Aiming at the problem of low accuracy of metal surface scratch detection, a method of metal surface scratch detection based on YOLOv4 is designed by introducing the small target detection idea, data enhancement, and adjusting anchor frame size [7]. Aiming at the problem of low detection efficiency of different scale defects on metal surfaces, an improved YOLOv5 detection network is designed. An adaptive anchor frame method is proposed, and a feature layer is added to the main components to enhance the useful feature information. In the prediction part, an effective loss function is used to solve the problem of data imbalance caused by small targets [8]. At present, there are many methods applied to the detection of strip steel surface defects, but these methods have problems such as low detection accuracy, low detection speed, and large model calculations.

In order to further improve the accuracy of strip steel surface defect detection, this study improves the YOLOv5s model and proposes the YOLOv5-ABS, which integrates an attention mechanism and a bidirectionally weighted feature pyramid structure. Firstly, a channel attention mechanism called SENet (squeeze and excitation network) is introduced into the backbone network to improve the network's feature extraction ability. Secondly, the BiFPN network is introduced to replace the feature fusion network PANET to enhance the fusion of shallow feature information and deep feature information. Finally, the spatial pyramid pooling structure SPPFCSPC is introduced to reduce the amount of calculation and improve the detection speed. Our experiments demonstrate that the updated YOLOv5-ABS network structure can achieve a maximum mAP of 78.6% on the NEU-DET dataset, which is 3.8% higher than the mAP of the previous YOLOv5s network. A better balance between detection speed and accuracy is achieved with a detection speed of 122 FPS.

2 Related Works

The YOLO series algorithm [9] is a one-stage target detection model based on deep learning and convolutional neural networks. Compared with previous versions, the YOLOv5 model has higher detection accuracy, faster detection speed, and a smaller model volume. According to the difference between network width and depth, the YOLOv5 model has five versions: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The detection accuracy of the five versions continues to increase, but as the model size and model parameters continue to increase, the detection speed decreases significantly. Considering the requirements of strip steel surface defect detection in terms of detection accuracy, detection speed, and real-time performance, YOLOv5s is selected for improved design in this study.

3 Methodology

In this section, we provide an improved model. Figure 1 represents the structure of the upgraded YOLOv5-ABS model, which is primarily made up of the Backbone, Neck, and Head.

Backbone: This component's primary job is to extract information about image features. It consists of the Conv module, the SeC3 module with channel attention mechanisms, and SPPFCSPC (Cross-Stage Partial Fast Spatial Pyramid Pooling). Conv performs convolution on the input features, BN prevents overfitting and facilitates model training, and the SiLU activation function is used to improve the expression ability of the model. The SeC3 module first uses the channel attention mechanism to enhance the feature extraction ability of the model, and then uses the C3 module to reduce the amount of calculation and memory. SPPFCSPC is used to fuse the feature

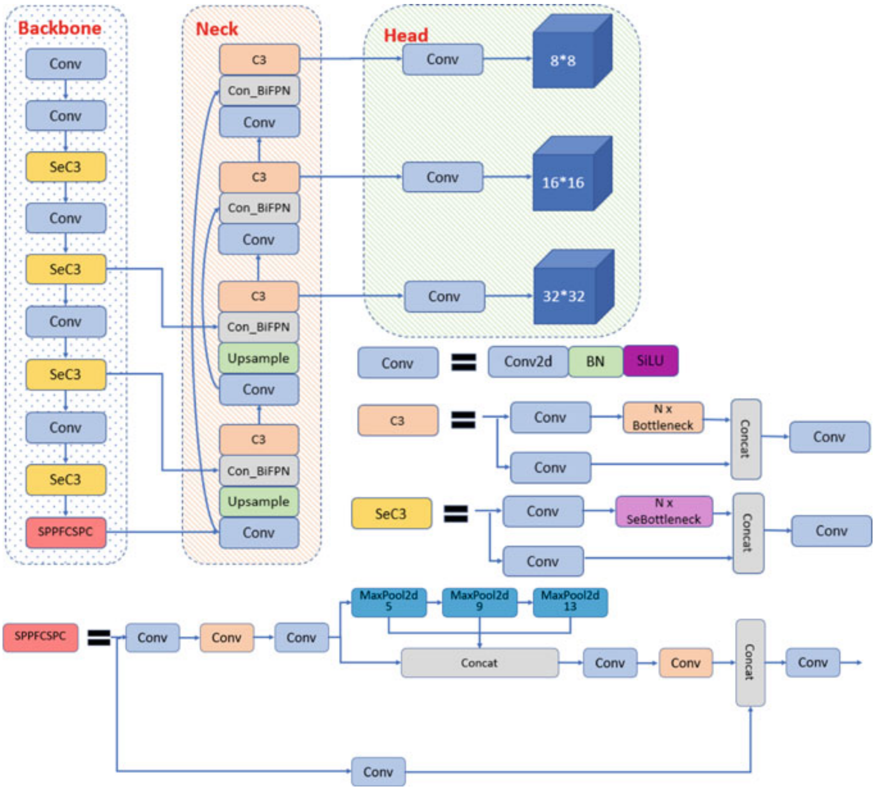

Fig. 1 YOLOv5-ABS model structure

maps of different receptive fields, enrich the feature expression ability, and improve the running speed.

Neck: Using BiFPN structure, the image feature information processed by the channel attention mechanism is integrated into the BiFPN network, which enhances the network's feature fusion ability and solves the problem of multi-scale feature fusion.

Head: the output of the model, which is mainly used to generate bounding boxes and predict defect categories. The loss function of the Bounding box is GIOU Loss, and the target box is filtered by NMS (Non-Maximum Suppression) to eliminate redundant boxes.

3.1 *SeC3 Module Integrated with SENet*

SENet (Squeeze-and-excitation networks) [10] is a channel attention mechanism network. The network mainly enhances the important features and suppresses the general features by obtaining the importance of each feature channel. Consequently, the model's capacity to extract target features and detection precision are enhanced.

The workflow of SE mainly includes three steps: Sequeeze, Excitation and Scale, as shown in Fig. 2. First, Sequeeze average pooling or maximum pooling the feature maps obtained by convolution to obtain a feature map of 1 × 1 × C. The Sequeeze formula is:

$$Z = Fsq(Xc) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} Xc(i, j) \tag{1}$$

In the formula (1), $H \times W$ is the size of channel space, X_c is the input feature map; (i, j) is the point with abscissa i and ordinate j on the feature map; $F_{sq}(X_c)$ represents the compression operation of the feature map; Z is the weight obtained by the compression channel.

Then, the relationship between the feature channels is grasped by the Excitation operation to generate weights for each feature channel. The Excitation formula is:

$$\mathrm{Sc} = Fex(Z, W) = Sigmoid[W2 \times \mathrm{Re}LU(W1, Z)] \tag{2}$$

In the formula (2), W is the dimension, S_c is the attention weight generated after F_{ex} operation.

Finally, the weight of the Excitation output is multiplied by the input feature to complete the recalibration of the original feature, so that the model can distinguish the features of each channel. The Scale formula is:

$$\widetilde{X} = F\mathrm{scale}(Xc, Sc) \otimes Sc \tag{3}$$

In the formula (3), $\otimes$ is the multiplication of elements, F_{scale} is a weight resetting operation, $\widetilde{X}$ is the result of SE channel attention.

The SeC3 module in this study was created by combining SE with the C3 unit. The SeC3 module is frequently utilized in the YOLOv5-ABS backbone, making it easier for the model to extract image features and increase detection accuracy.

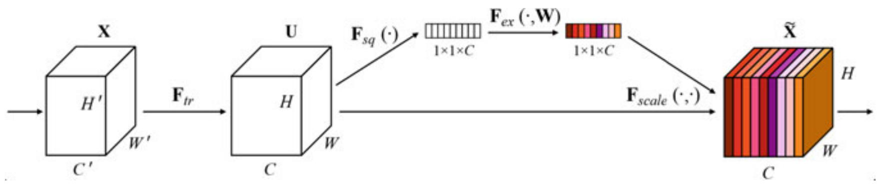

Fig. 2 SENet network structure

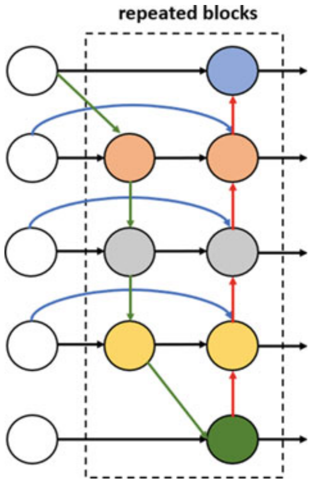

Fig. 3 BiFPN structure

3.2 Multi-scale Features Fusion Network BiFPN

Multi-scale feature fusion is frequently employed in target detection networks to enhance the model's detection performance. At present, the commonly used features of fusion networks are FPN [11], PANet [12], and BiFPN [13]. FPN establishes a top-down feature pyramid structure to achieve effective fusion of image semantic level and feature level, but it is limited by one-way feature information transmission. In order to make it simpler for the underlying information to be transferred to the top of the high-level, PANet integrates a bottom-up channel on top of FPN; however, this also makes the model more complex and decreases the efficiency of information transmission. BiFPN is a Bidirectional Feature Pyramid Network proposed by the Google team, as shown in Fig. 3. The network first deletes nodes with just one input edge; then, a new channel is established between the previous input point and the output point, allowing additional feature information to be fused without incurring excessive costs. In this study, BiFPN is added to the neck network to improve the network's capacity for feature fusion.

3.3 SPPFCSPC Module

According to the design scheme of SPPCSPC, this study combines the SPPF pyramid pooling structure and CSPNet to obtain the SPPFCSPC structure [14], which achieves speed improvement while keeping the receptive field unchanged. SPPF obtains different receptive fields through maximum pooling, which makes the model adapt to different resolution images. SPPF effectively solves the problem of repeated feature extraction, improves the generation speed of candidate boxes, and saves computing costs. CSPNet divides the features into two parts, only one of which is processed routinely. Finally, the two parts are combined, the calculation amount is reduced by half, and the speed and accuracy are improved.

4 Experimental Results and Analysis

4.1 Experimental Dataset

In this research, the strip steel surface defect dataset NEU-DET produced by Northeastern University was used, which contains six common types of surface defects: Rolled-in Scale (RS), Crazing (Cr), Inclusion (In), Patches (Pa), Scratches (Sc), Pitted Surface (PS). The dataset has a total of 1800 images, and each type of defect has 300 images. The grayscale image is 200 $\times$ 200 pixels in size in its original form. A total of 1440 images were chosen at random to make up the training-set; 72 images served as the validation-set; and 288 images served as the test-set. In order to increase the amount of training data and enrich the training scene, this study uses mosaic data enhancement technology to randomly crop four images and splice them into one image as training data. Considering the characteristics of the model and the original resolution, the input image's measurement is set to 256 $\times$ 256.

4.2 Experimental Environment and Evaluation Metrics

The Windows 11 operating system serves as the foundation for the study's experimental environment. The CPU is an Intel Core i7-12700H, the memory is 32 GB, and the GPU is a NVIDIA GeForce RTX 3070 Ti Laptop 8 GB. Using PyTorch 1.12.1 as the deep learning framework, the Python version is 3.8 and the CUDA version is 11.3. During the experiment, the epoch is set to 100 and the Batchsize is set to 128. The momentum is set to 0.937, the initial learning rate lr0 is set to 0.01, the cosine annealing algorithm is used to dynamically adjust the learning rate, and the weight _ decay is set to 0.0005.

This study uses Average Precision (*AP*) and Mean Average Precision (*mAP*) to evaluate the improved model. *AP* (formula 6) represents the mean value of Precision (*P*, formula 4) under different Recall (*R*, formula 5), that is, the *PR* curve obtained with *P* as the ordinate and *R* as the abscissa.

$$P = \frac{TP}{TP + FP} \times 100\% \tag{4}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{5}$$

$$AP = \int_{0}^{1} p(r)dr \tag{6}$$

where *TP* means True Positives, *FP* means False Positives, and *FN* means False Negative.

The mean *mAP* (formula 7) represents the average accuracy of all target detection categories.

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \tag{7}$$

where n represents the number of defect categories.

4.3 Ablation Experiment

The performance of the proposed method in strip surface defect detection is verified by ablation experiments. Based on the YOLOv5s network, the SeC3 module is used to replace the C3 module in the backbone named YOLOv5s-A; the introduction of the BiFPN module based on YOLOv5s-A is named YOLOv5s-AB; and based on YOLOv5s-AB, the SPPFCSPC module is used to replace the SPPF module named YOLOv5s-ABS. Table 1 displays the experimental outcomes.

According to the experimental results, after introducing the SE channel attention mechanism into the C3 module, the mAP of YOLOv5s-A is 75.8%, which is 1% higher than the original algorithm. This demonstrates that the SE attention mechanism effectively increases the model's ability to extract the image's global feature information and boosts detection accuracy. After introducing the BiFPN structure, the parameters and GFLOPs of YOLOv5s-AB are slightly improved compared with the original algorithm, but its mAP is further increased to 77.9%, which indicates that the use of BiFPN can effectively improve the detection accuracy of the network model. On this basis, the SPPFCSPC module is introduced. It can be found that the parameters and GFLOPs of YOLOv5s-ABS increase greatly, but the mAP is further improved to 78.6% and the detection speed is basically the same as the original algorithm, so it can still meet the real-time detection requirements.

Table 1 Results of ablation test

Model	Parameters	GFLOPs	FPS	mAP
YOLOv5s	6.29×10^6	15.0	147.4	74.8
YOLOv5s-A	6.06×10^6	13.8	145.1	75.8
YOLOv5s-AB	7.12×10^6	15.1	144.7	77.9
YOLOv5s-ABS	13.55×10^6	20.3	142.8	78.6

Figure 4 depicts the AP of each type of defect in YOLOv5s-ABS and YOLOv5s. The figure shows that the AP of YOLOv5s-ABS is enhanced as compared to the initial algorithm; the AP of RS, Cr, Pa, Sc, and Ps is increased by 1.6%, 19%, 1.4%, 0.7%, and 1.1%, respectively, except that the AP of In is reduced by 1.1%. The detection accuracy is substantially better than the previous algorithm, especially for Cr. It has been discovered that the application of the BiFPN multi-scale features fusion network, the SE channel attention mechanism, and the SPPFCSPC structure may significantly increase the detection accuracy of strip steel defects.

Figure 5 shows the test results for six kinds of defect samples. Different defects use different colors of Bounding-Box, and the upper right of each Bounding-Box has the defect name and confidence. The illustration demonstrates that for the same image, the enhanced algorithm has a better detection effect on all kinds of defects and can accurately detect the missing parts of the original algorithm.

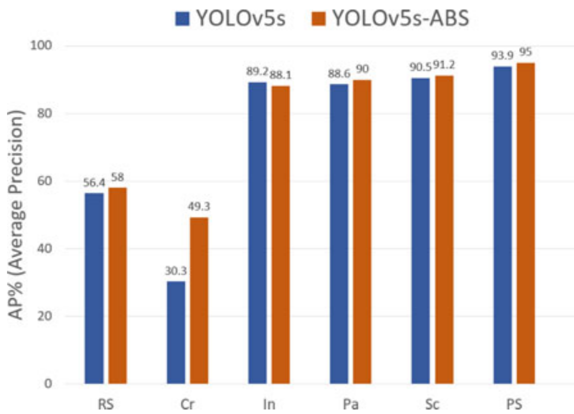

Fig. 4 Comparison of various types of AP between the two algorithms

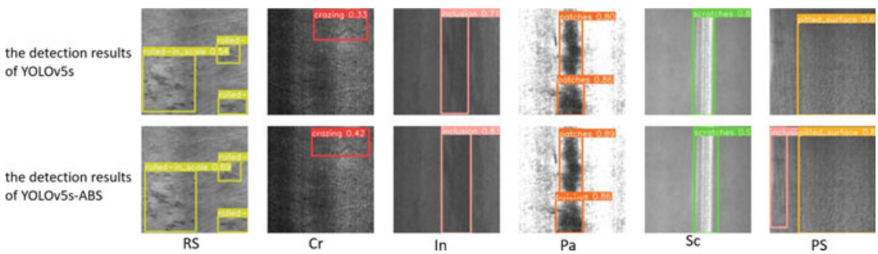

Fig. 5 Detection results of six kinds of defect samples

Table 2 Comparison of 5 different models on NEU-DET

Model	AP%						Weight/MB	mAP@0.5%
	RS	Cr	In	Pa	Sc	PS		
YOLOv3	53.1	36.7	62.8	77.4	72.2	68.5	235	61.8
YOLOv4	45.6	31.7	75.3	81.8	78.2	72.4	244	64.2
SSD	54	32.1	62.3	82.5	71.9	71.6	93.9	62.4
Faster-RCNN	57	37.3	83.6	89.7	90.6	90	108	74.7
YOLOv5s-ABS	58	49.3	88.1	90	91.2	95	26.7	78.6

4.4 Comparative Experiment

To further confirm the effectiveness of the improved YOLOv5s-ABS algorithm, this study selects four mainstream target detection algorithms YOLOv3, YOLOv4, SSD, and Faster-RCNN to compare with the improved YOLOv5s-ABS model on the NEU-DET dataset. Table 2 presents the outcomes.

According to Table 2, the mAP of the improved model YOLOv5s-ABS in this study reached 78.6%, which has a higher detection accuracy than the other four mainstream target detection models. Among them, the mAP of Faster-RCNN is 74.7%, which is close to the improved model. However, Faster-RCNN is a two-stage detection model with a complex training process, a large amount of calculation, and slow speed. The mAP of SSD is low, and it performs poorly in Cr defects but better in Pa defects. This is because Pa defects are mostly large targets, and Cr defects are mostly small targets, which reflects the shortcomings of SSD in small target defect detection. The detection accuracy of YOLOv3 and YOLOv4 models is low, the model volume is large, and the industrial deployment cost is high. In general, YOLOv5s-ABS has higher accuracy on the NEU-DET dataset, which can maintain a good balance between accuracy and speed, and a smaller model volume can be more convenient for industrial deployment.

5 Conclusions

This paper offers an improved YOLOv5s-ABS algorithm to solve the problem of strip steel surface defect detection and increase the accuracy and speed of strip steel surface defect detection. The algorithm first inserts the SENet channel attention mechanism into Backbone, which improves the feature extraction ability of the network. Secondly, the bidirectional weighted feature pyramid network BiFPN is introduced into Neck to strengthen the network features fusion ability. Finally, the SPPFCSPC structure is introduced to further improve speed and accuracy while keeping the receptive field unchanged. The algorithm's usefulness is demonstrated by experimental findings. Since the improved model has a low recognition rate for the two types of defects of Crazing and Rolled-in Scale, the next step will be to study the detection of these two types of defects and further optimize the detection accuracy of the model.

Acknowledgements This work was also supported by The Key project of natural science research in colleges and universities of Anhui Province (2022AH052365, KJ2020A1102) and Wuhu Engineering Technology Research Center (KJCXPT202204) and The Open Research Fund of Anhui Province Key Laboratory of Machine Vision Inspection.

References

1. Li YJ, Zhu PY, Sun XP (2017) Shape defect detection algorithm of stamping parts based on shape template matching. J Guangzhou Univ (Natural Science Edition) 16(5):62–66
2. Liu Y, Xia CL (2021) An image edge detection algorithm for strip steel surface defects based on Sobel opera-tor. Electronic Measurement Technol 44(3):138–143
3. Liu L, Wang C, Zhao SW (2018) Research on solar cells defect detection technology based on machine vision. J Electronic Measurement Instrument 32(10):47–52
4. Zhang T, Liu YT, Yang YN (2020) Review of surface defect detection based on machine vision. Sci Technol Eng 20(35):14366–14376
5. Dai XH, Chen HJ, Zhu CP (2020) Surface defect detection and realization of metal workpiece based on improved faster RCNN. Surface Technol 49(10):362–371
6. Kou XP, Liu SJ, Cheng KQ (2021) Development of a YOLO-V3-based model for detecting defects on steel strip surface. Measurement 182(2):109–119
7. Zhang BY, Leng YB (2022) Detection of metal surface scratch based on YOLOv4 network model. Acta Armamentarii 43(1):214–221
8. Wang K, Teng ZX, Zou TY (2022) Metal defect detection based on Yolov5. J Phys: Conf Ser 2218(1):120–128
9. Shao Y, Zhang D, Chu H (2022) A review of YOLO object detection based on deep learning. J Electron Inf Technol 44(2):1–12
10. Hu J, Shen L, Albanie S (2020) Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 42(8):2011–2023
11. Lin TY, Dollar P, Girshick R (2017) Feature pyramid networks for object detection. In: Computer vision and pattern recognition
12. Liu S, Qi L, Qin H (2018) Path aggregation network for instance segmentation
13. Tan M, Pang R, Le QV (2020) EfficientDet: scalable and efficient object detection. In: IEEE conference on computer vision and pattern recognition
14. He TJ, Li H (2023) The pavement damage detection model based on improved YOLOv5. Chin Civil Eng J 17(2):1–12