IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 November 2022, accepted 13 December 2022, date of publication 20 December 2022,
date of current version 29 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3230894

METHODS

Efficient Detection Model of Steel Strip Surface Defects Based on YOLO-V7

YANG WANG, HONGYUAN WANG, AND ZIHAO XIN
School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213000, China

Corresponding author: Hongyuan Wang (hywang@cczu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976028, and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant SJCX22_1431.

ABSTRACT During the production process of steel, there are often some defects on the surface of the product. Therefore, detecting defects is the key to produce high-quality products. At the same time, the defects of the steel have caused huge losses to the high-tech industry. A steel surface defect detection algorithm based on improved YOLO-V7 is proposed to address the problems of low detection speed and low detection accuracy of traditional steel surface defect detection methods. First, we use the de-weighted BiFPN structure to make full use of the feature information to strengthen feature fusion, reduce the loss of feature information during the convolution process, and improve the detection accuracy. Secondly, the ECA attention mechanism is combined in the backbone part to strengthen the important feature channels. Finally, the original bounding box loss function is replaced by the SIoU loss function, where the penalty term is redefined by taking the vector angle between the required regressions into account. The experimental results show that the improved model proposed in this paper has higher performance compared with other comparison models. Based on our experiments, the proposed method yields 80.2% mAP and 81.9% on the GC10-DET dataset and NEU-DET dataset with high speed, which is better than other existing models.

INDEX TERMS Machine vision, object detection, deep learning, feature extraction.

I. INTRODUCTION

As an important industrial product, steel is an indispensable raw material in daily life, machinery manufacturing and defense industries. With the continuous development of the industrial level, great progress has been made in steel production technology. The market also attaches great importance to the appearance and quality of products. In the production process of steel, for some reasons such as the quality of raw materials, manufacturing equipment and production conditions, different types of defects will exist on the surface of the product. As shown in Figure 1, different types of defects have different sizes and characteristics. These defects will reduce the strength, performance and wear resistance of steel, affect normal use, and may even cause serious consequences. Therefore, aiming to guarantee the quality of steel, it is required to inspect the surface of steel for defects. Detection of steel surface defects of steel in production line is of the utmost importance. [1].

Traditional steel surface defect detection methods include manual detection method and stroboscopic flash detection method, both of which require labor. Inspectors need to perform a lot of repetitive work, which is prone to visual fatigue, resulting in missed or false detections [2]. Machine vision method has been widely used to detect defects [3]. The YOLO series are a target detection algorithm based on deep learning and convolutional neural network. Its advantages include fast speed, high detection accuracy and real-time monitoring [4]. YOLO-V7 exhibits excellent performance in defect detection [5]. Based on the task of steel defect detection, this paper proposes an improved YOLO-V7 algorithm, which aims to improve the accuracy and speed of the steel surface defects detection. The performance of the algorithm is verified on the public datasets GC10-DET and NEU-DET.

The main contributions of this study are:

(1) A de-weighted bidirectional feature pyramid network (BiFPN) is combined into the original model, which strengthens the feature fusion between different parts, reduces the loss of feature information during the convolution process and improves the detection accuracy of the algorithm.

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson.

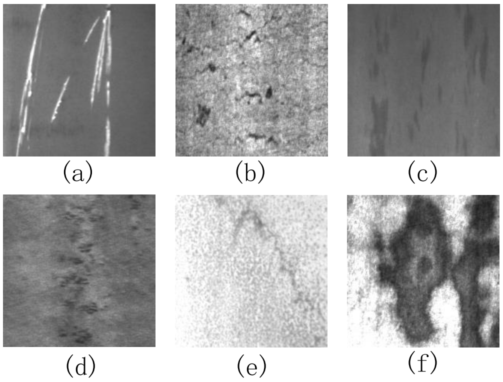

FIGURE 1. Some common steel defects images. (a) scratches, (b) crazing, (c) inclusion, (d) rolled-in scale, (e) pitted surface, (f) patches.

(2) Aiming to improve the feature extraction ability of the YOLO-V7 model in steel surface defects detection, the ECA attention mechanism was combined. Through experimental exploration, the ECA attention mechanism was embedded into the backbone of YOLO-V7, which improves the algorithm's feature learning ability and makes the algorithm pay more attention to useful information.

(3) In this paper, SIoU loss function is adopted to replace the original YOLO-V7 bounding box loss function and redefine the penalty term.

II. RELATED WORK

Machine vision technology has developed rapidly and has been widely used in the fields of image classification, face recognition, industrial manufacturing, and object detection in recent years [6]. Machine vision technology has the advantages of stability and efficiency [7]. Some scholars apply this technology to defect detection, which improves the detection efficiency [8]. L.Qiu et al. proposed an effective algorithm for pixel surface defect based on deep learning [9]. JEON et al. proposed a filtering scheme combined with illumination method for steel surface defect detection [10]. SON et al. proposed a method to determine the surface corrosion area of steel bridges [11]. HE et al. proposed classification priority network (CPN) and multi-group convolutional neural network (MG-CNN), which can make the detection accuracy of hot-rolled surface defects more than 94%, and the classification rate exceeds 96% [12]. Zhao et al. proposed a spherical multi-output Gaussian process (S-MOGP) method to model and monitor 3D surfaces [13]. XU et al. used a machine vision method to detect metal surface defects, which can detect cracks, dents and even other defects under uneven lighting conditions [14].

In defect detection based on deep learning, the accuracy and speed of existing detection methods are expected to be improved. The detection of large targets is gradually improving, but there is still a lot of work to be done for the detection of small targets [15]. From an optimization point, we offer a solution to detect metal surface defect [16], the improved YOLO-V7 shows better performance than other existing methods.

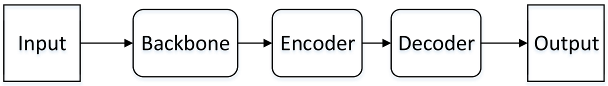

FIGURE 2. General detection pipline.

III. METHODS

A. YOLO-V7 MODEL

YOLO-V7 refers to the concept of YOLO series and its speed and accuracy exceed all known object detectors between 5 FPS to 160 FPS. YOLO-V7 balances speed and accuracy perfectly which makes itself favored by the industry. As shown in Figure 2, its general detection pipeline consists three parts: backbone, encoder, and decoder [17]. The structure of YOLO-V7 mainly consists three parts: input part, backbone feature extraction network part, strengthen feature extraction network and predictions part.

B. IMPROVED NETWORK STRUCTURE BASED ON YOLO-V7 MODEL

This method takes YOLO-V7 as the baseline, and proposes a new steel strip surface defect detection algorithm. The structure of YOLO-V7 is shown in Figure 3. On the whole, YOLO-V7 first resizes the input image to 640*640, then inputs it to the backbone network, and then outputs three layers of feature maps of different sizes through the head network, and then outputs the prediction result through RepConv [18]. The backbone part of YOLO-V7 mainly uses ELAN, MP structures, and Silu activation function. The ELAN structure can learn and converge efficiently by controlling gradient paths, deeper networks. ELAN-W is also similar. The network structure of ELAN and ELAN-W are shown in Figure 4. The MP structure is used for downsampling. The MP structure is shown in Figure 5. At the bottom of the backbone, we added the ECA attention mechanism [19]. The structure of ECA is shown in Figure 6. This module follows part of the SE attention mechanism [20]. The main improvement over SE attention mechanism is that the size of the one-dimensional convolution kernel is adaptively selected, and the dimension is maintained during local cross-channel interaction, reducing network complexity and improving model performance [21]. In ECA, the original feature image is input first, and all channels of the original image are globally averaged and pooled, and then a fast one-dimensional convolution with a size of Q is used to generate channel weights, and the corresponding probabilities of different channels are calculated and then compared with the original image. The input features are multiplied together as the input to the next layer. This method determines the Q value through function adaptation, and its value is proportional to the channel dimension C, as shown in formula (1) (2):

$$C = \phi\left(Q\right) = 2^{(\lambda Q - b)} \tag{1}$$

$$Q = \psi\left(C\right) = |\frac{\log_2\left(C\right)}{\lambda} + \frac{b}{\lambda}|odd \tag{2}$$

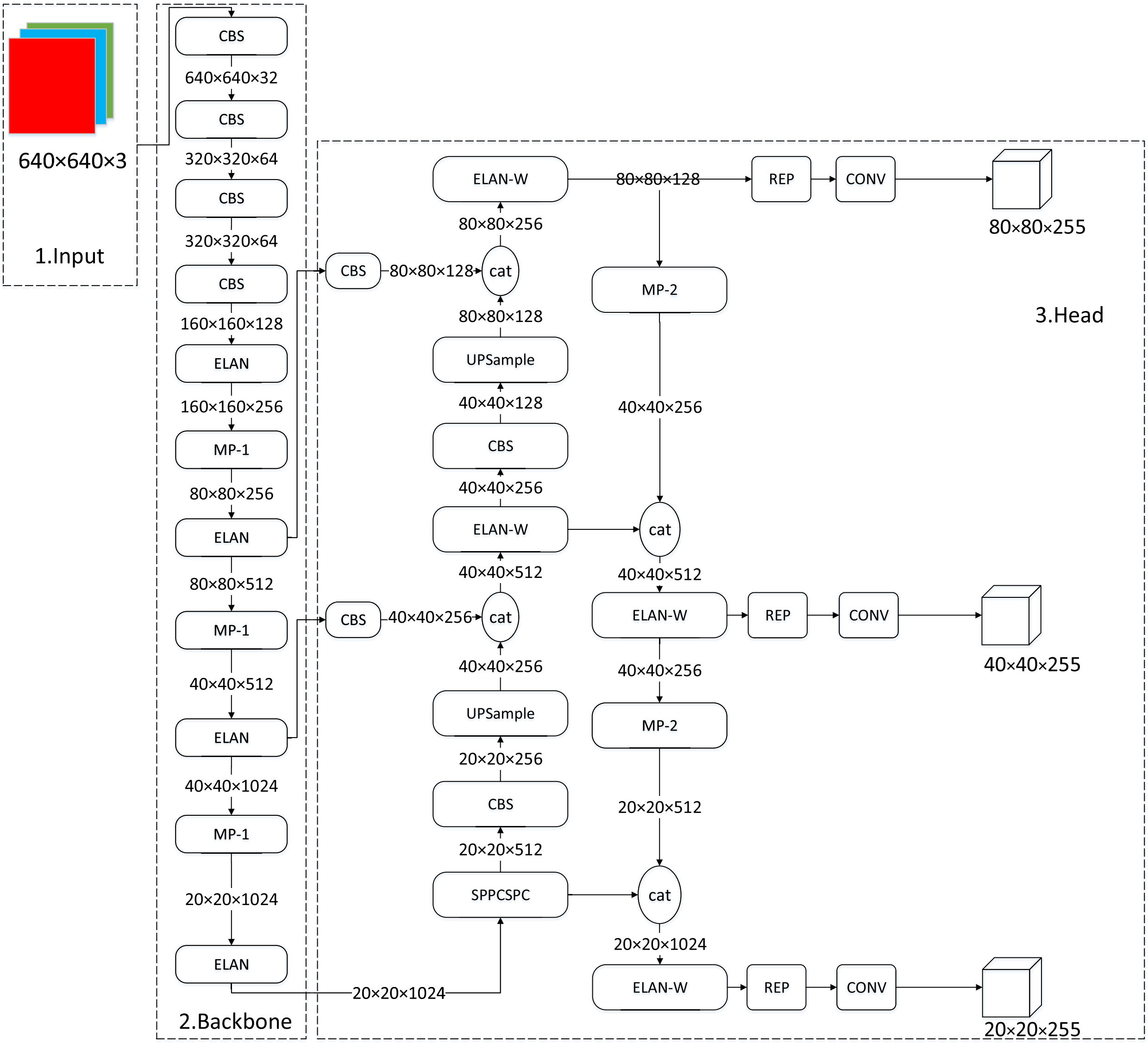

FIGURE 3. Architecture of YOLO-V7.

where λ=2, b=1, Q takes the nearest odd number. ECA has a flexible and lightweight structure, which can adaptively select one-dimensional convolution kernels, avoid dimensionality reduction and directly conduct cross-channel communication, enhance useful semantic information in feature maps, propose redundant and invalid information, and improve the effective extraction of steel surface defect features. ECA improves the efficiency of YOLO-V7 and is suitable for the datasets in this article. The backbone part after adding the ECA attention mechanism is shown in Figure 7.

The head part of YOLO-V7 integrates the neck part and the head part of YOLO-V4 [22] and YOLO-V5, and is a FPN + PAN structure [5]. FPN [23] can transfer the stronger semantic information possessed by the deep feature layer to the deep feature layer [24]. FPN is combined with PAN to perform parameter aggregation for different detection layers from different backbone layers. Although this combination effectively improves the feature fusion ability of the network, it will also lead to a problem, that is, the input of the PAN structure is all the feature information processed by the FPN structure, and a part of the original feature information of the backbone feature extraction network part is missing. Lack of original information to participate in learning can easily lead to deviations in training learning and affect detection accuracy. Aiming to solve this problem, we used an improved bidirectional feature pyramid network [25] to improve the original YOLO-V7 head part. The original BiFPN network constructs a bidirectional channel, proposes a cross-scale connection method, adds an extra edge, and fuses the features in the feature extraction network directly with the relative size features in the bottom-up path. Therefore, the network retains more shallow semantic information without losing too much deep semantic information. The original BiFPN sets different weights according to the importance of different input

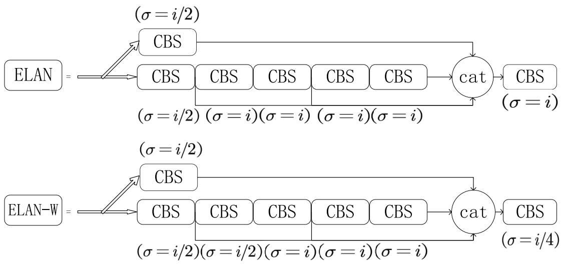

FIGURE 4. Architecture of ELAN and ELAN-W.

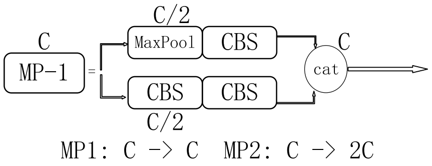

FIGURE 5. Architecture of MP.

features, and meanwhile, this structure is used repeatedly to strengthen feature fusion. However, we found that the result after the introduction of weighted BiFPN in YOLO-V7 is not ideal. We thought the reason is that weighting the input feature layer is very similar to the mechanism of adding an attention mechanism. Therefore, we removed the weight part of BiFPN and introduced a de-weighted BiFPN. Our de-weighted BiFPN structure is shown in Figure 8.

C. LOSS FUNCTION

The loss function of YOLO-V7 consists three parts: the bounding box loss function, the objectness loss function and the class loss function. The bounding box loss function is used to measure the error of the prediction box for the coordinate positioning error. The objectness loss function reflects the confidence error of the prediction box. The class loss function reflects the error caused by the prediction error of the prediction box for the target category. The objectness loss function and class loss function of YOLO-V7 are BCEWithLogitsLoss. The objectness loss function is CIoU loss. CIoU considers the distance between the ground truth box and the prediction box, the overlap rate, the box scale and the penalty term, which makes the bounding box regression more stable, as shown in the Figure 9. CIoU loss is defined by Equation (3).

$$\mathrm{Loss}_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{3}$$

where $\rho^2\left(b, b^{gt}\right)$ represents the Euclidean distance between the center point of the prediction box and the center point of the ground truth box and it is denoted by d in Figure 9. Where c represents the diagonal distance of the minimum closure rectangle containing both the ground truth box and the prediction box. α is defined by Equation (4). v is defined by Equation (5).

$$\alpha = \frac{v}{1 - \mathrm{IoU} + v} \tag{4}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{5}$$

where w^{gt} represents the width of the ground truth box. Where h^{gt} represents the height of the ground truth box. Where w represents the width of the prediction box. Where h represents the height of the prediction box. Regarding the bounding box loss function, such as CIoU, the direction between the ground truth box and the prediction box is not considered, causing a slow convergence speed. For this, SIoU introduces the vector angle between the ground truth box and the prediction box to redefine the correlation [26]. As a result, we replaced the CIoU loss function with the SIoU loss function. The SIoU loss function contains four parts: the angle loss, the distance loss, the shape loss, and the IoU loss.

1) ANGLE COST

The angle cost is defined by Equation (6). Its diagram is shown in Figure 10.

$$\begin{aligned} \Lambda &= 1 - 2 \times \sin^2\left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right) \\ &= \cos\left(2 \times \left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right)\right) \end{aligned} \tag{6}$$

where c_h is the height distance between the center point of the ground truth box and the prediction box. Where σ is the distance of the center point between the ground truth box and the prediction box.

$$c_h = \max\left(b_{c_y}^{gt}, b_{c_y}\right) - \min\left(b_{c_y}^{gt}, b_{c_y}\right) \tag{7}$$

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x}\right)^2 + \left(b_{cy}^{gt} - b_{c_y}\right)^2} \tag{8}$$

where $\left(b_{c_x}^{gt}, b_{c_y}^{gt}\right)$ is the barycentric coordinate of the ground truth box. Where $\left(b_{c_x}, b_{c_y}\right)$ is the center coordinate of the prediction box.

2) DISTANCE COST

The distance cost is defined by Equation (9). Its diagram is shown in Figure 11.

$$\Delta = \sum_{t=x,y}\left(1 - e^{-\gamma\rho_t}\right) = 2 - e^{-\gamma\rho_x} - e^{-\gamma\rho_y} \tag{9}$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2, \quad \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2 \tag{10}$$

$$\gamma = 2 - \Lambda \tag{11}$$

where (c_w, c_h) is the width and height of the minimum circumscribed matrix of the ground truth box and the prediction box.

3) SHAPE COST

The shape cost is defined by Equation (12).

$$\Omega = \sum_{t=w,h}\left(1 - e^{-w_t}\right)^\theta = \left(1 - e^{-w_w}\right)^\theta + \left(1 - e^{-w_h}\right)^\theta \tag{12}$$

$$w_w = \frac{\left|w - w^{gt}\right|}{\max\left(w, w^{gt}\right)}, \quad w_h = \frac{\left|h - h^{gt}\right|}{\max\left(h, h^{gt}\right)} \tag{13}$$

where (w,h) is the width and height of the prediction box. Where $\left(w^{gt}, h^{gt}\right)$ is the width and height of the ground truth

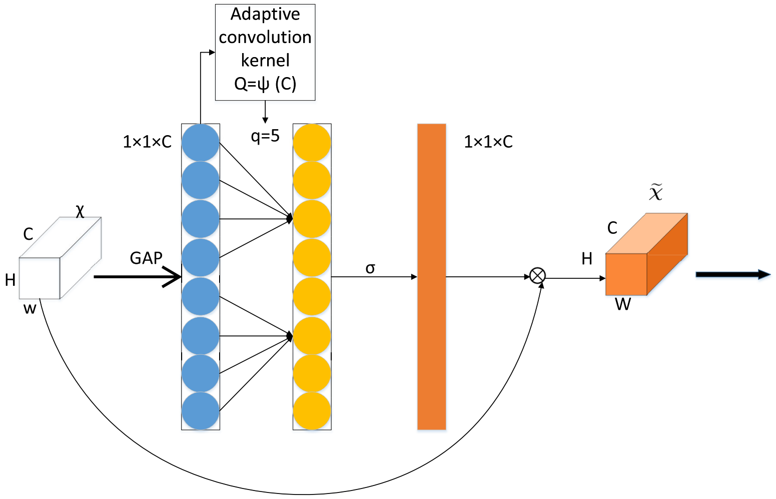

FIGURE 6. Architecture of ECA-Net.

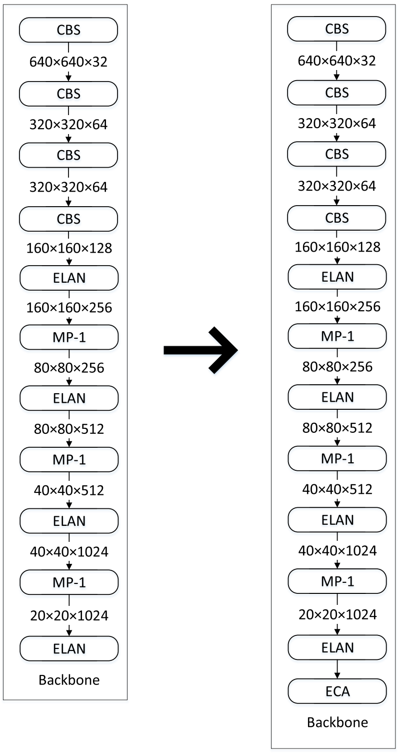

FIGURE 7. YOLO-V7 incorporating ECA Modules.

box. Where θ controls how much attention is paid to shape loss.

4) IoU COST

The IoU cost is defined by Equation (14).

$$\mathrm{IoU} = \frac{A}{B} \tag{14}$$

where A represents the intersection of the ground truth box and the prediction box. Where B represents the union of the ground truth box and the prediction box.

5) SIoU LOSS

In conclusion, the SIoU loss is defined by Equation (15).

$$\mathrm{Loss}_{\mathrm{SIoU}} = 1 - \mathrm{IoU} + \frac{\Delta + \Omega}{2} \tag{15}$$

IV. EXPERIMENTS

To complete the experiment, we used the PyTorch framework [27]. The experimental condition is : Ubuntu 16.04 LTS operating system, Python 3.8, Pytorch 1.9.0 and Nvidia GTX2080Ti GPU with 11GB memory.

A. DATASETS

In our experiment, we used two popular public datasets to verify the practicality of the proposed method, namely, GC10-DET and NEU-DET (see Figure 12).

1) GC10-DET

The GC10-DET dataset contains the detection images of steel surface defects in actual industrial production, with a total of 2257 images. It contains 10 types of defects including Pu (punching), Wl (weld line), Cg (crescent gap), Ws (water spot), Os (oil spot), Ss (silk spot), In (inclusion), Rp (rolled pit), Cr (crease), Wf (waist folding). The resolution of image is 4096*1000. The training set, validation set and test set were divided into 6:1:3 ratio.

2) NEU-DET

The NEU-DET dataset was produced by the team at Northeastern University [28]. It contains six types of defects: Rs (rolled-in scale), Pa (patches), Cr (crazing), Ps (pitted surface), In (inclusion) and Sc (scratches). There are 1800 images in the dataset in total. The resolution of image is 200*200. All images are at grayscale. The training set, validation set and test set were divided into 6:1:3 ratio.

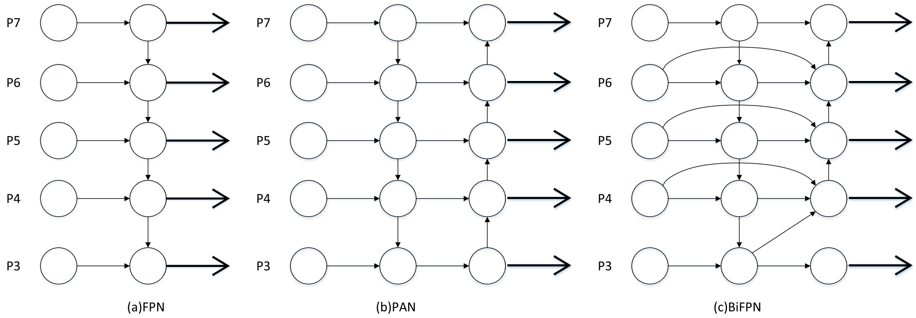

FIGURE 8. Feature network design (a) FPN (b) PAN (c) BiFPN.

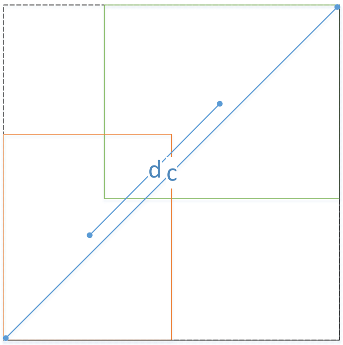

FIGURE 9. CIoU.

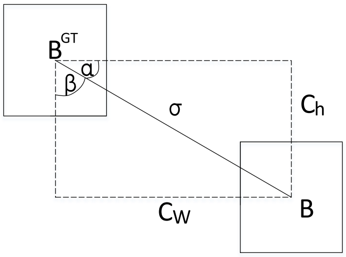

FIGURE 10. Angle cost.

B. PERFORMANCE EVALUATION

In industrial production, the accuracy and speed of defect detection are the two most concerned indicators. If the inspection result of the type or the location of the defect is wrong, it may cause the machine to misjudge. If the inspection speed is too slow, it will greatly reduce the efficiency of defect detection, and may even lead to accidents. So as to solve the above problems, three measurements: AP, mAP, and FPS are used to evaluate the strip defect detection model. AP represents the average precision of each defect, mAP represents the mean average precision of all classes, and FPS represents the frames per second. We use the above indicators to determine whether the model can meet the requirements of real-time monitoring.

C. EXPERIMENTS

The total parameters of the improved YOLO-V7 model are 38,665,321. The proposed method performs defect detection at a fast speed and high accuracy. Figure 13 shows part of the steel strip defect detection results. Various defects are of different types and sizes.

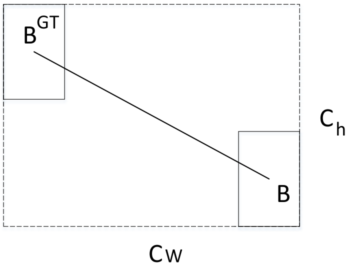

FIGURE 11. Distance cost.

1) ABLATION STUDY

Our experiments are performed on the GC10-DET dataset and the NEU-DET dataset. We demonstrate the effectiveness of each part through ablation study. This can demonstrate the improved accuracy of our proposed surface defect detection method. To verify the effectiveness of each improvement part on the YOLO-V7 network model, combined experiments are performed on each improvement strategy to control the variables. The results of GC10-DET dataset and NEU-DET dataset for ablation experiments are listed in Table 1 and Table 2.

We can learn from Table 1 and 2 that after using de-weighted BiFPN, ECA attention mechanism, and SIoU loss in the YOLO-V7 network, the mAP of the proposed method on GC10-DET is 2.8% higher than the original one, and the mAP on NEU-DET is 4.6% higher than the original one. From the data in the table, it can be learned that the model proposed in this study has a great improvement over the original model. The feature pyramid is improved for multi-scale fusion, more original features are fused. The attention mechanism enables the network to automatically learn the importance of each feature channel and assign more weights to more useful features. The improved loss function is also beneficial for small-scale defects. As a result, the detection ability of the proposed method for small-scale defects such as Pu, Ss, In, Cr, Rs, Sc is significantly improved.

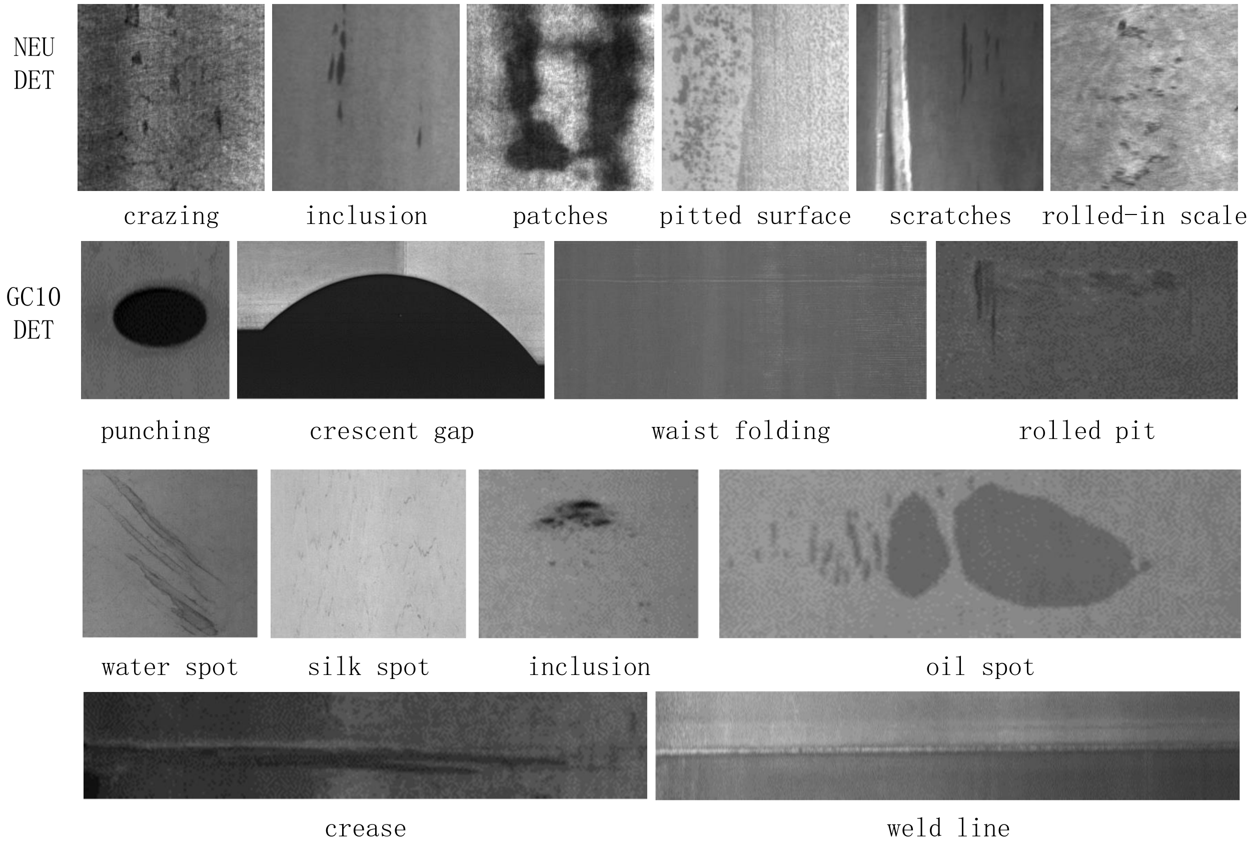

FIGURE 12. Two datasets with different resolutions.

TABLE 1. Ablation experiments on GC10-DET.

scheme	Pu	Wl	Cg	Ws	Os	Ss	In	Rp	Cr	Wf	mAP	P	R	FPS
YOLO-V7	0.930	0.681	0.986	0.905	0.883	0.891	0.782	0.848	0.830	0.000	0.774	0.790	0.718	111.1
YOLO-V7+BiFPN	0.928	0.827	0.992	0.947	0.906	0.883	0.848	0.815	0.852	0.000	0.800	0.799	0.725	96.2
YOLO-V7+ECA	0.964	0.743	0.973	0.910	0.833	0.888	0.797	0.748	0.902	0.000	0.776	0.772	0.772	109.9
YOLO-V7+SIoU	0.968	0.734	0.986	0.921	0.882	0.891	0.831	0.787	0.880	0.000	0.788	0.805	0.711	105.3
Proposed method	**0.969**	0.798	0.991	0.908	0.885	**0.910**	**0.869**	0.826	0.868	0.000	**0.802**	0.821	0.703	105.2

TABLE 2. Ablation experiments on NEU-DET.

scheme	Cr	In	Pa	Ps	Rs	Sc	mAP	P	R	FPS
YOLO-V7	0.517	0.861	0.932	0.811	0.640	0.874	0.773	0.662	0.755	58.8
YOLO-V7+BiFPN	0.547	0.888	0.943	0.841	0.724	0.867	0.802	0.826	0.697	57.5
YOLO-V7+ECA	0.454	0.875	0.920	0.805	0.722	0.884	0.776	0.736	0.753	54.9
YOLO-V7+SIoU	0.477	0.877	0.943	0.885	0.667	0.850	0.783	0.797	0.707	59.9
Proposed method	**0.690**	0.853	0.907	0.843	**0.727**	**0.896**	**0.819**	0.792	0.765	55.56

2) COMPARATIVE EXPERIMENT

In order to verify the effectiveness of the proposed method, this paper studies the currently widely used object detection network. SSD [29] and Faster-RCNN [30] represent the classic one-stage and two-stage object detection networks, respectively. SSD and Faster-RCNN represent the network uses ResNet50. YOLO-V5 is the same detection network of YOLO series as YOLO-V7.

Table 3 and Table 4 present a comparison of AP, mAP and FPS for the defects on the GC10-DET dataset and NEU-DET dataset. On the GC10-DET dataset, the detection method based on YOLO-V7 is much higher than other models in detection accuracy and detection speed. Among them, the improved YOLO-V7 model has the highest AP in the seven types of defects including Pu, Cg, Ws, Os, Ss, In, and Cr. In addition, the detection speed is significantly faster than other models, although slightly lower than the original YOLO-V7 model, but it can be ignored at high speeds of FPS > 100. But on defects with lighter color such as Wf, the YOLO-V7 model performs very poorly. To further verify the

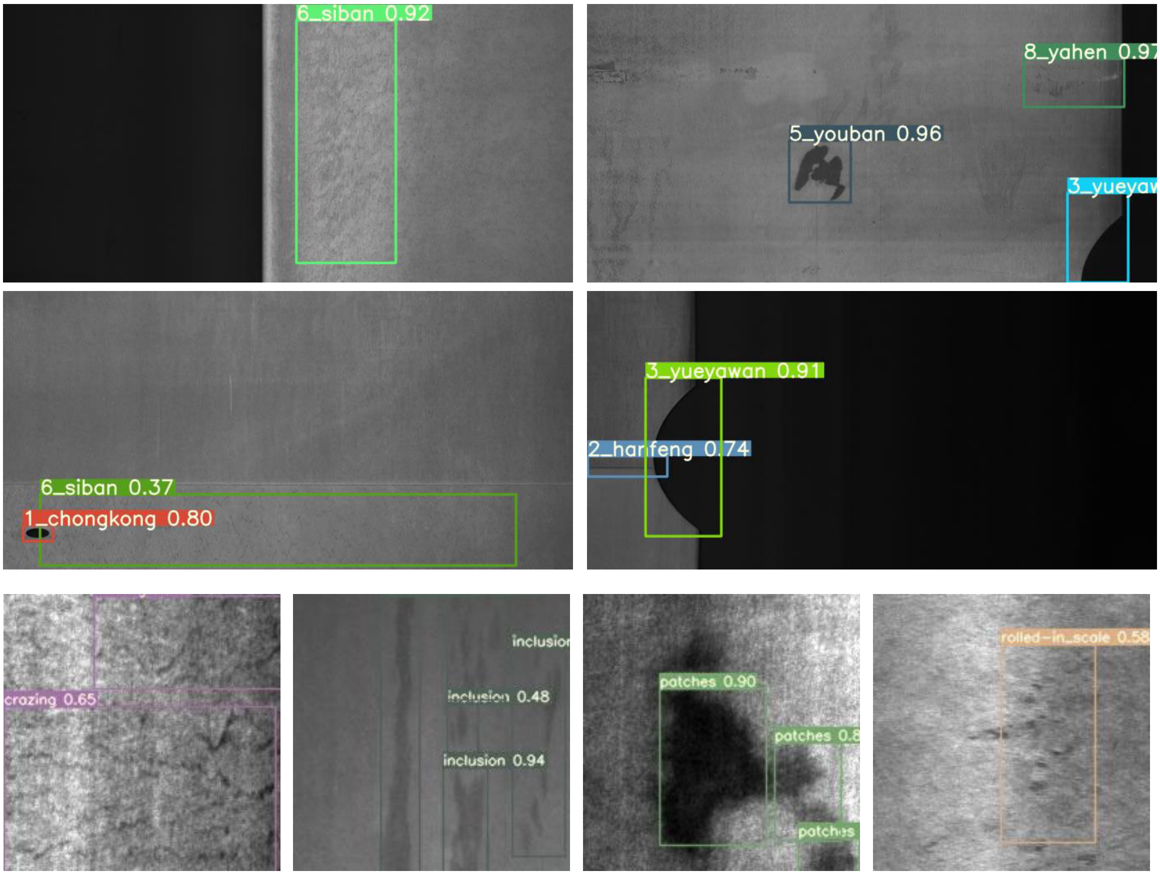

FIGURE 13. Part of steel strip defect detection results.

TABLE 3. AP results of different models on GC10-DET.

Types	SSD300	SSD512	Faster-RCNN	YOLO-V5	YOLO-V7	Proposed Method
Pu	0.860	0891	0.899	0.953	0.93	**0.969**
Wl	**0.974**	0.885	0.554	0.877	0.681	0.798
Cg	0.861	0.848	0.872	0.96	0.986	**0.991**
Ws	0.552	0.558	0.599	0.76	0.905	**0.908**
Os	0.612	0.662	0.653	0.627	0.883	**0.885**
Ss	0.689	0.650	0.579	0.687	0.891	**0.91**
In	0.168	0.256	0.194	0.367	0.782	**0.869**
Rp	0.105	0.364	0.364	0.539	**0.848**	0.826
Cr	0.527	0.521	0.736	0.427	0.83	**0.868**
Wf	**1.000**	0.919	0.818	0.822	0	0
mAP	0.635	0.651	0.627	0.702	0.774	**0.802**
FPS	29.3	27.1	23.6	-	**111.1**	105.2

robustness of the method, comparative experiments are conducted on the NEU-DET dataset. On the NEU-DET dataset, the improved YOLO-V7 model has the highest recognition rates for Cr, Rs, and Sc defects. In terms of detection speed, it is also slightly lower than the original YOLO-V7 model and significantly faster than other models. From the experimental results, it can be learned that the model proposed in this study is superior to other models in accuracy and speed in steel surface defect detection.

TABLE 4. AP results of different models on NEU-DET.

Types	SSD300	SSD512	Faster-RCNN	YOLO-V5	YOLO-V7	Proposed Method
Cr	0.411	0.417	0.374	0.509	0.517	**0.69**
In	0.796	0.763	0.794	0.855	**0.861**	0.853
Pa	0.839	0.863	0.853	**0.948**	0.932	0.907
Ps	0.839	**0.851**	**0.851**	0.829	0.811	0.843
Rs	0.621	0.581	0.545	0.617	0.64	**0.727**
Sc	0.836	0.856	0.882	0.874	0.874	**0.896**
mAP	0.714	0.724	0.711	0.772	0.773	**0.819**
FPS	37.6	29.0	23.8	-	**58.8**	55.6

V. CONCLUSION

This research improves the feature pyramid based on the YOLO-V7 network, integrates more features without increasing the cost, adds the ECA attention mechanism to help the algorithm make better use of the feature information, which is helpful for small target detection. The target detection loss function relies on the aggregation problem of bounding box regression metrics and SIoU bounding box loss function is adopted. The proposed method is evaluated by two steel strip defect datasets with different resolutions. The experimental results prove that the proposed method has high detection accuracy and speed, which meets the requirements of market. However, its detection effect on defects with lighter colors such as Wf in the GC10-DET dataset is not ideal. It is our future research direction to try to improve the detection accuracy of the model on such defects.

ACKNOWLEDGMENT

The authors would like to thank Prof. Wang Hongyuan for his key guidance on this paper, and Dr. Ding Zongyuan for his writing guidance. Finally, thanks to everyone who helped with this research.

REFERENCES

[1] S. Kim, W. Kim, Y.-K. Noh, and F. C. Park, "Transfer learning for automated optical inspection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2517–2524.

[2] Z. Li, J. Zhang, T. Zhuang, and Q. Wang, "Metal surface defect detection based on MATLAB," in *Proc. IEEE 3rd Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Oct. 2018, pp. 2365–2371.

[3] X. Chen, J. Lv, Y. Fang, and S. Du, "Online detection of surface defects based on improved YOLOV3," *Sensors*, vol. 22, no. 3, p. 817, Jan. 2022.

[4] 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.

[5] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

[6] X. Zihao, W. Hongyuan, Q. Pengyu, D. Weidong, Z. Ji, and C. Fuhua, "Printed surface defect detection model based on positive samples," *Comput., Mater. Continua*, vol. 72, no. 3, pp. 5925–5938, 2022.

[7] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020.

[8] P. K. R. Maddikunta, Q.-V. Pham, P. B, N. Deepa, K. Dev, T. R. Gadekallu, R. Ruby, and M. Liyanage, "Industry 5.0: A survey on enabling technologies and potential applications," *J. Ind. Inf. Integr.*, vol. 26, Mar. 2022, Art. no. 100257.

[9] L. Qiu, X. Wu, and Z. Yu, "A high-efficiency fully convolutional networks for pixel-wise surface defect detection," *IEEE Access*, vol. 7, pp. 15884–15893, 2019.

[10] Y.-J. Jeon, D.-C. Choi, S. J. Lee, J. P. Yun, and S. W. Kim, "Steel-surface defect detection using a switching-lighting scheme," *Appl. Opt.*, vol. 55, no. 1, pp. 47–57, Jan. 2016.

[11] H. Son, N. Hwang, C. Kim, and C. Kim, "Rapid and automated determination of rusted surface areas of a steel bridge for robotic maintenance systems," *Autom. Construct.*, vol. 42, pp. 13–24, Jun. 2014.

[12] D. He, K. Xu, and P. Zhou, "Defect detection of hot rolled steels with a new object detection framework called classification priority network," *Comput. Ind. Eng.*, vol. 128, pp. 290–297, Feb. 2019.

[13] C. Zhao, J. Lv, and S. Du, "Geometrical deviation modeling and monitoring of 3D surface based on multi-output Gaussian process," *Measurement*, vol. 199, Aug. 2022, Art. no. 111569.

[14] K. Xu, S. Liu, and Y. Ai, "Application of Shearlet transform to classification of surface defects for metals," *Image Vis. Comput.*, vol. 35, pp. 23–30, Mar. 2015.

[15] G. Li, S. Du, B. Wang, J. Lv, and Y. Deng, "High definition metrology-based quality improvement of surface texture in face milling of workpieces with discontinuous surfaces," *J. Manuf. Sci. Eng.*, vol. 144, no. 3, Mar. 2022, Art. no. 031001.

[16] Y. Xu, K. Zhang, and L. Wang, "Metal surface defect detection using modified Yolo," *Algorithms*, vol. 14, no. 9, p. 257, Aug. 2021.

[17] K. Wang, Z. Teng, and T. Zou, "Metal defect detection based on Yolov5," *J. Phys., Conf. Ser.*, vol. 2218, no. 1, Mar. 2022, Art. no. 012050.

[18] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13733–13742.

[19] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13–19.

[20] 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.

[21] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13–19.

[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[23] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[24] Q. Chen, Y. Wang, T. Yang, X. Zhang, J. Cheng, and J. Sun, "You only look one-level feature," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13039–13048.

[25] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10781–10790.

[26] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.

[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.

[28] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.

[29] 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.

[30] 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.*, vol. 28, 2015, pp. 1–14.

YANG WANG received the B.E. degree from the Qingdao University of Technology, in 2020. He is currently pursuing the M.E. degree with Changzhou University. His main research interests include computer vision and defect detection.

HONGYUAN WANG received the Ph.D. degree in computer science from the Nanjing University of Science and Technology. He is currently a Professor with Changzhou University. His research interests include pattern recognition, intelligence systems, and pedestrian trajectory discovery in intelligent video sureillance.

ZIHAO XIN received the B.E. degree from Changzhou University, in 2019, where he is currently pursuing the M.E. degree in computer science. His main research interests include computer vision and defect inspection.