IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 21 July 2022, accepted 1 September 2022, date of publication 7 September 2022, date of current version 13 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204683

RESEARCH ARTICLE

A Defect Detection Method for a Boiler Inner Wall Based on an Improved YOLO-v5 Network and Data Augmentation Technologies

XIAOMING SUN[1,2], XINCHUN JIA[1], (Member, IEEE), YUQIAN LIANG[1,2], MEIGANG WANG[1], AND XIAOBO CHI[2]

[1]School of Automation and Software Engineering, Shanxi University, Taiyuan 030013, China
[2]School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China

Corresponding author: Xinchun Jia (xchjia@sxu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61973201 and Grant U1610116, and in part by the Key Research and Development Projects of Shanxi Province under Grant 201903D121145.

ABSTRACT During the long-term operation of a coal-fired boiler, some defects of its inner wall are unavoidable. The traditional manual detecting method is time-consuming and not safe for maintenance engineers. In this paper, we propose an automatic detection method to deal with inner wall defects based on an improved YOLO-v5 network and data augmentation technologies. Specifically, some shallow features and original deep features are fused on the basis of the original YOLO-v5 network for the small objects. Meanwhile, a squeeze-excitation (SE) attention module is added behind the network's backbone to improve the feature extraction efficiency of the network, and a varifocal loss function is adopted to make it easier for the network to detect those dense objects. Moreover, 176 images including four types of typical inner wall defects (castables falling off, anti-wear layer damage, perforation and bruise) are collected from a power plant boiler, and five data augmentation technologies are introduced to increase the number of samples. The experimental results demonstrate that the proposed method can effectively detect various defects of a boiler inner wall with a satisfactory accuracy, and bring a great facilitation to the maintenance of a power plant.

INDEX TERMS Boiler inner wall defects, object detection, improved YOLO-v5 network, data augmentation.

I. INTRODUCTION

Owing to the complex structure and harsh environment in a boiler of coal-fired power plants, some defects are inevitable during the long-term uninterrupted operation process, especially on the inner wall of a boiler. Although a boiler can run steadily at most of time, once unexpected defects occur, it must be shut down for maintenance. As a result, the cost of the power plant should be increased and the safety of the boiler's operation will be threatened [1], [2], [3]. Therefore, it is of great significance to accurately and automatically detect the defects of a boiler inner wall.

Traditionally, the manual maintenance method requires maintenance engineers of power plants to carry numerous detection instruments (e.g., endoscopes) into a boiler inner wall for inspecting the steel pipes one by one [4]. This inspecting process costs great workload and its effects depend largely on the maintenance engineers' experiences. Clearly, such a traditional method is not effective, especially for the safety of engineers. Therefore, an effective automatic detection technology for inner wall defects of a boiler is required instead.

The development of deep learning technologies makes it possible to solve the problem aforementioned. Among them, convolutional neural networks are widely used in image classification, object detection, semantic segmentation and other tasks due to their powerful image feature extraction ability [5], [6], [7], [8]. In object detection tasks, the mainstream two-stage network is regions with convolutional neural networks features (R-CNN) [9], [10], [11],

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

Castables falling off   Anti-wear layer damage   Perforation   Bruise

FIGURE 1. Four main types of inner wall defects in a boiler.

[12], and the mainstream one-stage networks are you only look once (YOLO) [13], [14], [15], [16], [17] and single shot multibox detector (SSD) [18], [19]. In reference [20], Fang *et al.* developed a hybrid network, which adopted a Faster R-CNN to robustly detect the object parts, and used a model-driven clustering algorithm to group the related partial detections and suppress false detections. Based on YOLO-v3 network, Kou *et al.* added several dense convolutional blocks to make the network more efficient in extracting image features and improve the network's performance. Some experimental results in [21] showed that the proposed object detection network can detect various defects on the surface of steel bars well, which is superior to the original YOLO-v3 network. Noting the shortcomings of unmanned aerial vehicles in detecting pedestrians near the ground in low illumination environment, Wang *et al.* proposed a pedestrian detection method based on image fusion and YOLO-v3 network, and introduced a convolutional block attention module to improve the network's performance [22]. In reference [23], Zeng *et al.* added an adversarial occlusion network to the standard Faster R-CNN detection network, and this improved network showed a fine accuracy. As for the insensitivity problem of existing object detection algorithms for large or medium defect targets on bearing covers, Zheng *et al.* developed an improved YOLO-v3 network to detect defects in real-time [24].

The above results show some effective applications of excellent object detection networks in practice. Nevertheless, few of them take the density and size of the objects to be detected into account. Moreover, the training efficiency of a detection network is of great importance. Thus, it is significant to improve the performance of an object detection network from the perspective of training efficiency.

During a training process, the quality and quantity of a dataset play crucial roles in a network's performance, and the sample number of a dataset used for object detection is insufficient in many cases. As an effective technique to increase the number of samples, data augmentation has attracted more and more attention of researchers [25], [26], [27]. In reference [28], Zhang *et al.* input the augmented dataset into a convolutional neural network to classify the images, and their data augmentation technology showed a satisfactory effect. To enhcance the accuracy of image classification, Takahashi *et al.* applied a novel data augmentation technology, and achieved a new state-of-the-art test error of 2.19% on CIFAR-10 [29]. These research results indicate that data augmentation technologies can effectively increase the network's capacity to extract features from samples.

Given the above-mentioned problems, in this paper, three aspects are improved based on the original YOLO-v5 network. First, some shallow features are fused with the original deep features to enhance the network's ability to extract features of small objects. Then a squeeze-excitation (SE) attention module is added at the end of the YOLO-v5 network's backbone to make the network more efficient for feature extraction. Third, the detection network's loss function is replaced with a varifocal loss to improve the detection ability of the network for those dense objects. Moreover, to deal with the shortage of some defect images in the original dataset, five data augmentation technologies are utilized to augment the original inner wall defects dataset of a boiler. Finally, experimental results demonstrate that the improved YOLO-v5 network can detect the defects effectively while having a satisfactory accuracy, and is superior to the original YOLO-v5 network and some commonly used networks. Furthermore, the data augmentation technologies adopted in this paper are also shown to be effective.

The rest of this paper is organized as follows. The inner wall defects of a boiler under study are described in Sec. 2. Sec. 3 introduces the research methods involved in this paper. And the detailed experiment results are represented in Sec. 4. At last, remarkable conclusions are drawn and future works are given in Sec. 5.

II. RESEARCH OBJECT DESCRIPTION

As recommended by some skilled power plant engineers, there are four main types of common defects on the inner wall of a boiler, namely castables falling off, anti-wear layer damage, perforation, and bruise, as shown in Fig. 1. Specifically, the castables wrapped in the inner wall of a boiler can effectively prevent the inner wall from being affected by harsh environments such as high temperature, high pressure, and corrosion, thereby prolonging the service life of the boiler. Once the castables fall off somewhere, it is likely to crack

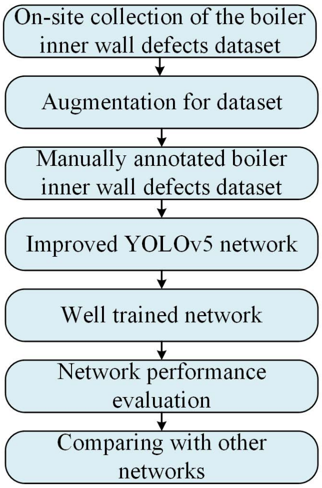

FIGURE 2. Flow chart of the proposed defects detection process.

there and its nearby areas, this situation will greatly endanger the boiler's safe operation. An anti-wear layer plays a significant role during the safe operation of a boiler, and its main function is to improve the fire resistance of its components and prevent damage to these components. Besides, some perforations and bruises are found on the inner wall during the process of collecting dataset on site, and they are mostly caused by the damage and maintenance of certain parts in the boiler.

The above defects affect the boiler's stable operation to various degrees. However, as represented in Sec. I, manual detection is time-consuming, inefficient, and not conducive to the safety of maintenance engineers. Consequently, the research object of this paper is to fully use a large amount of data collected on site in the boiler to train a network that automatically detects various defects on the inner wall of the boiler, so as to facilitate the maintenance of the boiler, and reduce the cost of the power plant, more importantly, ensure the safe operation of the boiler. Specifically, the flow chart of detection process of inner wall defects in a boiler is shown in Fig. 2.

III. RESEARCH METHODS

A. Data AUGMENTATION

In the field of computer vision, data augmentation is a powerful technology to improve a network's performance [30]. Similar to a human learning process, if a network fully learns various characteristics of the objects in a training process, it will show an excellent performance in a test set. Benefical from some data augmentation technologies, a network can learn the objects' characteristics in different lighting, angles, and scales, leading to an improvement over a network's reasoning ability.

Among data augmentation technologies, image rotation and flipping are the most common. Scale transformation, adding noise, and illumination processing are also useful methods [31], [32]. Besides, generative adversarial networks also perform well in data augmentation tasks [33], [34], [35].

TABLE 1. Number of samples for each type of defect before and after data augmentation.

Type of defects	Number of samples	
	Before augmentation	After augmentation
Castables falling off	30	267
Anti-wear layer damage	42	378
Perforation	64	525
Bruise	40	331
Total	176	1501

According to the different characteristics of various defects in the original dataset, five technologies are adopted to augment the original dataset, namely rotating, blurring, darkening, random clipping, and flipping. The above technologies improve the original dataset's quality from different aspects, and the number of samples is increased from 176 to 1501. The augmented images of four types of defects are shown in Fig. 3. And the number of image samples for various types of defects before and after augmentation is shown in Table 1.

Clearly, the defect images under different angles, light intensity, scale, and noise intensity are added after the five data augmentation technologies. When an object detection network fully learns these features, it will have a strong robustness. Moreover, the effectiveness of these data augmentation technologies on the network's performance is verified in Sec. 4.2.

B. YOLO-v5 NETWORK

Similar to YOLO-v4 network, the overall structure of YOLO v5 network changes little, but its detection speed and accuracy are greatly improved. Specifically, at the input side of the YOLO-v5 network, four images from the original dataset are randomly cropped and spliced to one image by a mosaic method. At the same time, some black edges are adaptively added to images with different length-to-width ratios so as to accelerate network reasoning. During each training process, the prediction anchors are output according to the initial set anchor values, and the best anchor values suitable for different training data are continuously adjusted. The network's backbone is composed of a focus module and some CSP modules, in which the image with the size of $608 * 608 * 3$ is sliced and convoluted to obtain a feature map with the size of $304 * 304 * 32$ in the focus module, and the CSP module is mainly composed of convolution, batch normalization, mish, residual and concat operations, as shown in the blue box in Fig. 4. At the neck side of the network, various features at different levels are fused together to improve the detection ability of the network for different objects. Besides, a focal loss is chosen as the loss function, which considers the length-to-width ratios of both prediction anchors and object anchors, and a non-maximum suppression operation is applied to improve the prediction anchors' accuracy.

C. IMPROVED YOLO-v5 NETWORK

The original YOLO-v5 network has a high accuracy in the detection of large objects, such as pedestrian detection, vehicle detection, etc. However, it does not perform pretty well for

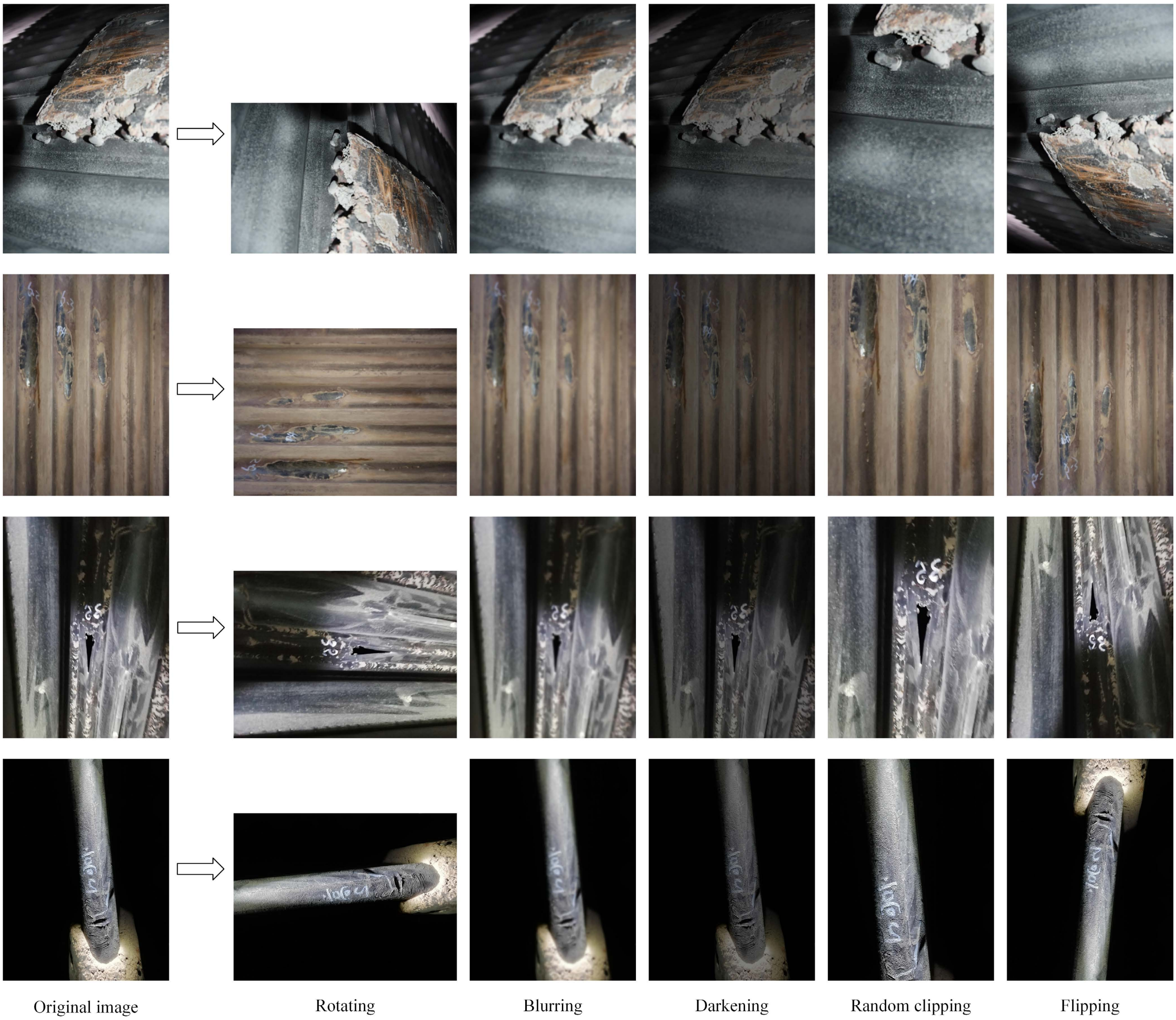

FIGURE 3. Defect images before and after data augmentation.

objects that are relatively small or dense, such as inner wall defects of a boiler. Therefore, the following three improvements are made to the original YOLO-v5 network.

1) FUSION OF MULTILEVEL FEATURES

In object detection tasks, pedestrians, buildings, etc. are usually considered as larger objects, and in this paper, the four defects to be detected are relatively medium or small. Moreover, from Fig. 1, it can be seen that there are obvious differences among the four types of defects. Specifically, the defect objects of anti-wear layer damage, perforations and bruise are mostly small, while those of castables falling off are mostly large. During a network training process, the essence of convolutional layers is the layer-by-layer extraction of image features. So, the former convolutional layers can form some larger feature maps to capture those small objects, while the latter ones can form some smaller feature maps to capture those large objects. For the four types of defects in the dataset, the feature levels required for each type of defect are different. However, the original YOLO-v5 network has only some deep-level feature extraction modules, but not a shallow one [36]. As a result, those deep-level feature extraction modules may not be enough to comprehensively extract the features of the small objects.

In response to the above problem, a shallow-level feature extraction module for small objects is introduced to the original YOLO-v5 network, as shown in Fig. 4. An up-sampling operation is taken behind the 19th block so as to further expand the feature map. After the 20th block, the obtained feature map is concatenated with that of the second

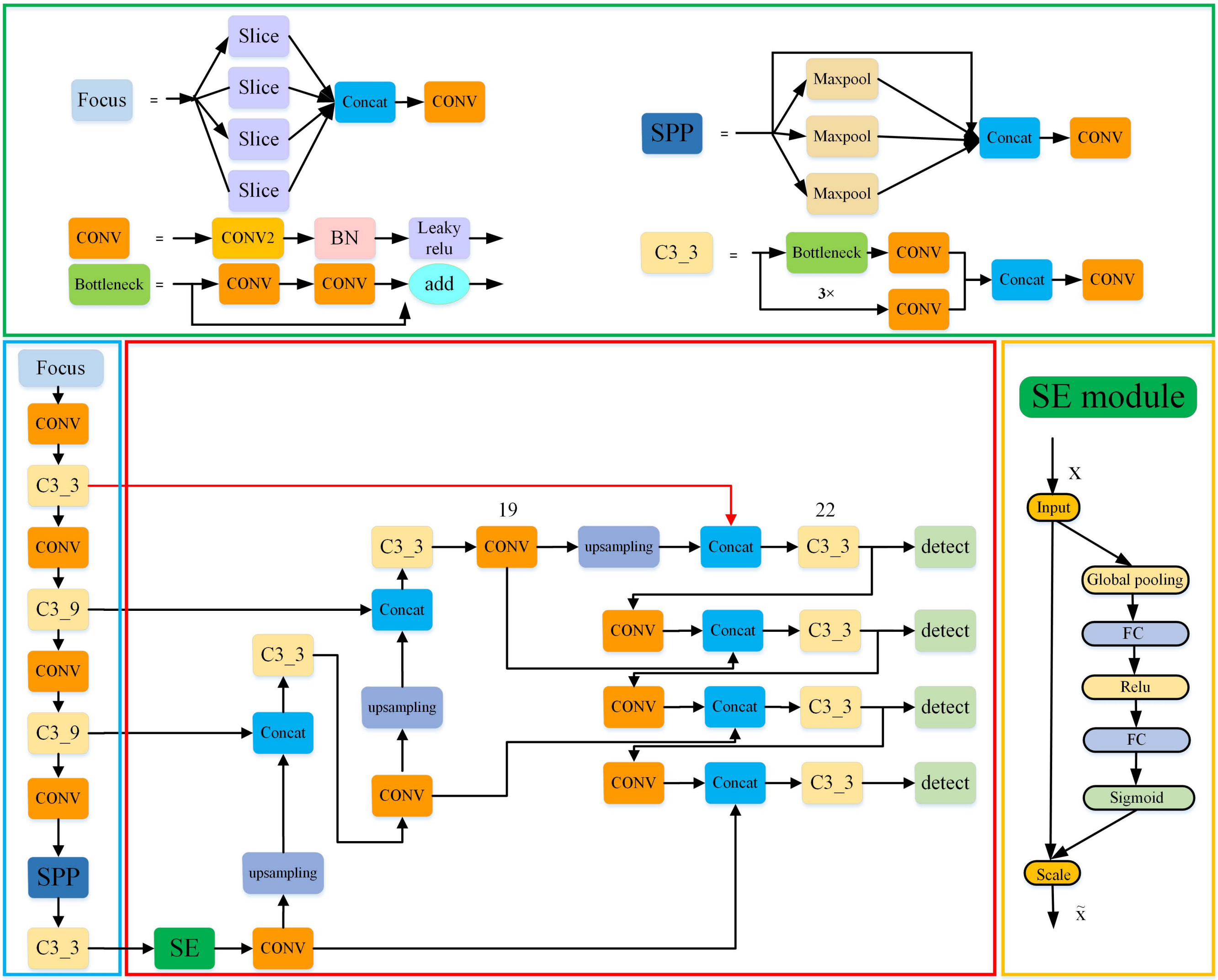

FIGURE 4. **Detailed structure of the improved YOLO-v5 network.**

convolutional block in the network's backbone, see the red connecting line in Fig. 4. Meanwhile, the detection of small objects is added to the detection side. Through such changes, multiple levels of features are thoroughly fused, hence the detection network may have a more powerful capacity of detecting the relatively small defects, such as anti-wear layer damage and bruise.

2) SE ATTENTION MODULE

In an object detection task, different features of objects are extracted by different convolutional channels, and these features have different importance to the object detection task. Once the network invests too many training resources into those less important features, its training efficiency will be affected, so is the network's accuracy.

For this reason, an SE attention module is added behind the feature extraction layer of the original YOLO-v5 network, as shown in Fig. 4. The calculation of this module mainly consists of two stages: feature compression and feature excitation. Firstly, in the feature compression stage, the feature map ($h * w * c$) extracted in the previous step outputs the compressed global features ($1 * 1 * c$) through a global average pooling operation. Then in the feature excitation stage, the compressed global feature passes through two fully connected layers of two activation functions, relu and sigmoid, and outputs the weights (importance) of each convolutional channel. Finally, the original feature map is multiplied by the different weights of the corresponding channels [37], see the orange box in Fig. 4.

After some multi-layer convolutional operations, the network has extracted different features of the objects. And it gives different weights to different convolutional channels by adding this SE attention module. The more important the feature, the larger the weight, and vice versa, thereby improving the network's feature extraction efficiency as well as the performance.

3) VARIFOCAL LOSS

The loss function is an important factor affecting the network's convergence. During a training process, the network aims at minimizing the loss function and continuously performs gradient descent. Sometimes, however, the objects to be detected is relatively dense, making it difficult for the network to accurately detect, so the optimization direction of the network gradually deviates from what we want, as described in [38].

For alleviating this problem, a varifocal loss is adopted as the loss function, which is given by

$$VFL(p, q) = \begin{cases} -q(qlog(p) + (1 - q)log(1 - p)), q > 0 \\ -\alpha p^{\gamma} log(1 - p), q = 0 \end{cases} \tag{1}$$

where p is the predicted IoU-aware classification score; q represents the object's IoU score; a IoU score describes the ratio of the intersection and union of the predicted bounding box and the ground truth bounding box. Specifically, the q values of negative samples are 0, while for positive samples, they are their IoU scores. Besides, α and γ are the weight coefficients, and $VFL(\cdot)$ is the varifocal loss.

With this setting, the network uses q to increase the weight of the positive sample loss with high IoU, so as to focus the training on high-quality samples, hereby enhancing the network's performance.

The above three operations improve the original YOLO-v5 network from three aspects of small objects detection, feature extraction efficiency and dense objects detection. And the detailed structure of the improved YOLO-v5 network is shown in Fig. 4. Furthermore, the effectiveness of the proposed improved YOLO-v5 network is verified in Sec. 4.3.

IV. EVALUATION INDICATORS AND EXPERIMENT RESULTS

In this section, the network performance evaluation indicators are introduced and the effectiveness of the research methods used in this paper is comprehensively verified through several experiments.

Some hyper-parameters of the proposed improved YOLO-v5 network are set as follows: the size of each training batch is 4, the network is optimized by using an adam optimizer and runs 150 epochs for training. 80% of the augmented dataset are chosen for training the network while the rest are tested. And the settings of other hyper-parameters are the same as the original YOLO-v5 network.

A. EVALUATION INDICATORS

MAP is adopted to evaluate the network performance, and its formulas are as follows:

$$IOU = \frac{area(B_p \bigcap B_{gt})}{area(B_p \bigcup B_{gt})} \tag{2}$$

$$P(r) = \frac{TP}{TP + FP} \tag{3}$$

$$AP = \int_0^1 P(r)dr \tag{4}$$

TABLE 2. Comparison of the network performance before and after data augmentation.

MAP values	Before augmentation	After augmentation
Original yolo-v5	77.0%	**88.9%**
Improved yolo-v5	84.3%	**94.9%**

$$MAP = \frac{\sum(AP)}{N_{class}} \tag{5}$$

where B_p and B_{gt} are the predicted anchor and actual anchor of an object, respectively, and IOU is the degree of overlap between the two anchors. Besides, TP is the number of detection anchors with $IoU > 0.5$, while FP is the number of detection anchors with other cases. AP is the precision of a certain type of defect. And MAP is the mean AP of four types of defects.

B. EFFECTIVENESS OF THE DATA AUGMENTATION TECHNOLOGIES

The datasets before and after augmentation are input into the original and improved YOLO-v5 network, respectively, so as to fully test the effectiveness of the data augmentation technologies adopted in this paper.

Specifically, for the original YOLO-v5 network, the *MAP* values of the data before and after augmentation are 77.0% and 88.9%, respectively, that is, the data augmentation technologies utilized in this paper increase the *MAP* value of the original YOLO-v5 network by 11.9%. And for the improved YOLO-v5 network proposed in this paper, using the data before and after augmentation, the *MAP* values are 84.3% and 94.9%, respectively. The data augmentation technologies increase the *MAP* value by 10.6%. A detailed comparison of the network performance before and after data augmentation is listed in Table 2.

In conclusion, the data augmentation technologies utilized in this paper can effectively enhance the quantity and quality of the original dataset, and then improve the network's performance.

Furthermore, the data augmentation technologies used in this paper can also be extended to other scenarios, such as wall defects, surface defects of solar panels, surface defects of parts in industrial production, surface defects of high-risk buildings, etc.

C. EFFECTIVENESS OF THE IMPROVED YOLO-v5 NETWORK

To thoroughly verify the effectiveness of the improved YOLO-v5 network proposed in this paper, the augmented dataset is used to compare the performance of the network in five experiments: (1) Original YOLO-v5 network; (2) Replacing the loss function in the original YOLO-v5 network with a varifocal loss function; (3) Adding an SE attention module to the original YOLO-v5 network; (4) Adding a shallow-level feature extraction module to the original YOLO-v5 network; (5) Replacing the loss function of the original YOLO-v5 network with a varifocal loss function,

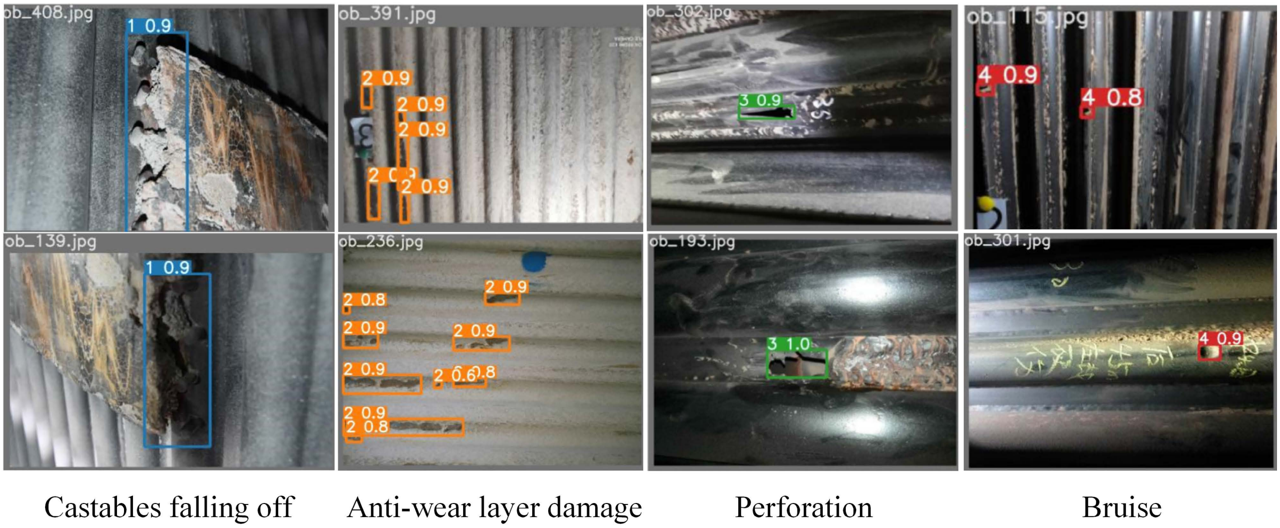

FIGURE 5. Detection results of the improved YOLO-v5 network.

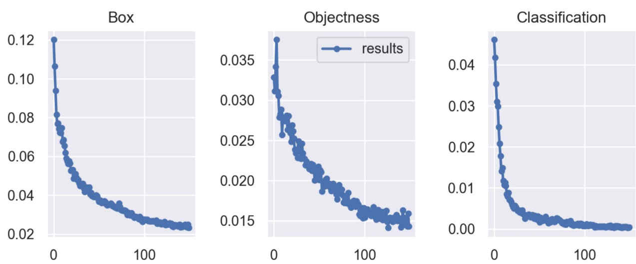

FIGURE 6. Box, objectiveness and classification loss during network training.

adding an SE attention module and a shallow-level feature extraction module.

The detection results of the improved YOLO-v5 network in the test set are shown in Fig. 5, and Fig. 6 reveals the box, objectiveness, and classification loss during the training process, respectively. It can be seen from Fig. 6 that the network's training loss first decreases gradually, and then tends to be stable. Specifically, the *MAP* value of the network in experiment 1 is 88.9%, while those of experiments 2, 3, 4, and 5 are 89.7%, 93.0%, 94.0% and 94.9%, respectively. Evidently, the methods used in this paper improve the performance of the original YOLO-v5 network in varying degrees from three aspects: loss function, shallow-level feature extraction, and feature extraction efficiency. And Table 3 summarizes the MAP and Frames Per Second (FPS) values of the five experiments. Besides, in experiment 1, the *AP* values of the four types of defects are 93.9%, 84.2%, 87.7% and 89.8%, respectively. While in experiment 4, they are 94.4%, 93.7%, 91.5% and 96.4%, respectively. The fusion of multi-level features in this paper is shown to be effective.

D. COMPARISON WITH OTHER NETWORKS

To comprehensively verify the performance of the improved-YOLO-v5 network, we compared it with several commonly used object detection networks: Faster R-CNN, SSD, and Retinanet. And Table 4 summarizes the MAP and FPS values of the four networks.

To sum up, the data augmentation technologies used in this paper can effectively ameliorate the quality of the original dataset, and then improve the performance of the object detection network. More importantly, our improvements over the original YOLO-v5 network have also achieved a satisfactory result.

TABLE 3. Comparison of the five experiments' detection results.

Experiments	MAP values	FPS
Experiment 1	88.9%	65.2
Experiment 2	89.7%	66.8
Experiment 3	93.0%	64.9
Experiment 4	94.0%	63.4
Experiment 5	**94.9%**	64.5

TABLE 4. Comparison of detection results for the four networks.

Experiments	MAP values	FPS
Improved YOLO-v5	**94.9%**	64.5
Faster R-CNN	83.5%	43.8
SSD	85.4%	37.6
RetinaNet	73.2%	29.0

V. CONCLUSION

To deal with the low efficiency and safety problems of the manual maintenance method in inner wall defects of a boiler, in this paper, some deep learning technologies are applied to the automatic detection of inner wall defects in a boiler. The main contributions can be summarized as follows:

(1) The original YOLO-v5 network has been improved from three aspects of shallow-level feature extraction, feature extraction efficiency, and loss function. And the effectiveness

of these three improvements has been verified through our experiments.

(2) The improved YOLO-v5 network in this paper is superior to other commonly used networks and has a satisfactory performance.

(3) Five data augmentation technologies adapted to different defects of a boiler inner wall have been utilized to increase the number of samples, thereby improving the performance of the object detection network. Moreover, these data augmentation technologies can be extended to other similar fields.

With the improved YOLO-v5 network proposed in this paper, the maintenance engineers of a power plant only need to use wall climbing robots or other tools to capture the images of each height and position of a boiler inner wall and input the images into the fine trained improved YOLO-v5 network, so as to automatically detect and locate the defects of a boiler inner wall, and then repair the defects at the specified position in time, which avoids the faults in boiler operation and protects the personal safety of power plant engineers.

Nevertheless, the augmentation of the original dataset inevitably increases the object detection network's training burden. At the same time, the SE attention module and shallow-level feature extraction module added in the network can increase the parameters to a certain extent, which makes the training more difficult. How to reduce the training burden and simplify the training as much as possible while maintaining the network performance is one of important issues to be considered in our follow-up researches. Besides, using semantic tags on images along with an ontology to describe the meaning of the tags are also beneficial to our methods, and the detection effect of the proposed improved YOLO-v5 network in other scenarios is also our future experiment content.

REFERENCES

[1] X. Song, X. Wu, and Y. Kang, "An inspection robot for boiler tube using magnetic flux leakage and ultrasonic methods," *Insight, Non-Destructive Test. Condition Monitor.*, vol. 46, no. 5, pp. 275–278, 2004.

[2] S.-Y. Jing, Q. Liang, H. Zeng, X.-M. Liu, and D.-F. Wan, "Analysis and inspection on linear defects in inner surface of boiler," *Electr. Power Construct.*, vol. 31, no. 11, pp. 84–88, Nov. 2010.

[3] Z.-B. Chen and Y.-P. Wang, "Defects in boiler supervision inspection of 2×600mw power units and treatment," *Electr. Power*, vol. 41, no. 5, pp. 19–22, May 2008.

[4] X. Gao, Z. Jiang, J. Gao, D. Xu, Y. Wang, and H. Pan, "Boiler maintenance robot with multi-operational schema," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, Aug. 2008, pp. 610–615.

[5] Y. D. Li, Z. B. Hao, and H. Lei, "Survey of convolutional neural network," *J. Comput. Appl.*, vol. 36, no. 9, pp. 2508–2515 and 2565, Sep. 2016.

[6] B. Zhao, M. Dai, P. Li, R. Xue, and X. Ma, "Defect detection method for electric multiple units key components based on deep learning," *IEEE Access*, vol. 8, pp. 136808–136818, 2020.

[7] 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.

[8] Z. Li, J. Li, and W. Dai, "A two-stage multiscale residual attention network for light guide plate defect detection," *IEEE Access*, vol. 9, pp. 2780–2792, 2021.

[9] 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.

[10] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1440–1448.

[11] 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. 1–10.

[12] T. Bai, J. Yang, G. Xu, and D. Yao, "An optimized railway fastener detection method based on modified faster R-CNN," *Measurement*, vol. 182, Sep. 2021, Art. no. 109742.

[13] 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.

[14] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

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

[16] Z. Ying, Z. Lin, Z. Wu, K. Liang, and X. Hu, "A modified-YOLOv5s model for detection of wire braided hose defects," *Measurement*, vol. 190, Feb. 2022, Art. no. 110683.

[17] Y. Guo, Y. Zeng, F. Gao, Y. Qiu, X. Zhou, L. Zhong, and C. Zhan, "Improved YOLOv4-CSP algorithm for detection of bamboo surface sliver defects with extreme aspect ratio," *IEEE Access*, vol. 10, pp. 29810–29820, 2022.

[18] 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.* Berlin, Germany: Springer, 2016, pp. 21–37.

[19] F. Liu, K. Mao, H. Qi, and S. Liu, "Real-time long-term correlation tracking by single-shot multibox detection," *Opt. Eng.*, vol. 58, no. 1, 2019, Art. no. 013105.

[20] F. Fang, L. Li, H. Zhu, and J.-H. Lim, "Combining faster R-CNN and model-driven clustering for elongated object detection," *IEEE Trans. Image Process.*, vol. 29, pp. 2052–2065, 2020.

[21] X. Kou, S. Liu, K. Cheng, and Y. Qian, "Development of a YOLO-V3-based model for detecting defects on steel strip surface," *Measurement*, vol. 182, Sep. 2021, Art. no. 109454.

[22] C. Wang, D. Luo, Y. Liu, B. Xu, and Y. Zhou, "Near-surface pedestrian detection method based on deep learning for UAVs in low illumination environments," *Opt. Eng.*, vol. 61, no. 2, Feb. 2022, Art. no. 023103.

[23] L. Zeng, B. Sun, and D. Zhu, "Underwater target detection based on faster R-CNN and adversarial occlusion network," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104190.

[24] Z. Zheng, J. Zhao, and Y. Li, "Research on detecting bearing-cover defects based on improved yolov3," *IEEE Access*, vol. 9, pp. 10304–10315, 2021.

[25] J. Niu, Y. Chen, X. Yu, Z. Li, and H. Gao, "Data augmentation on defect detection of sanitary ceramics," in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2020, pp. 5317–5322.

[26] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple copy-paste is a strong data augmentation method for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2918–2928.

[27] H. Naveed, "Survey: Image mixing and deleting for data augmentation," 2021, *arXiv:2106.07085*.

[28] K. Zhang, Z. Cao, and J. Wu, "Circular shift: An effective data augmentation method for convolutional neural network on image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1676–1680.

[29] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep CNNs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 2917–2931, Sep. 2020.

[30] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122.

[31] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[32] M. Saran, F. Nar, and A. N. Saran, "Perlin random erasing for data augmentation," in *Proc. 29th Signal Process. Commun. Appl. Conf. (SIU)*, Jun. 2021, pp. 13001–13008.

[33] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process.*, vol. 35, no. 1, pp. 53–65, Jan. 2017.

[34] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," 2017, *arXiv:1711.04340*.

[35] N. Jain, A. Olmo, S. Sengupta, L. Manikonda, and S. Kambhampati, "Imperfect ImaGANation: Implications of GANs exacerbating biases on facial data augmentation and snapchat face lenses," *Artif. Intell.*, vol. 304, Mar. 2022, Art. no. 103652.

[36] S. Tan, G. Lu, Z. Jiang, and L. Huang, "Improved YOLOv5 network model and application in safety helmet detection," in *Proc. IEEE Int. Conf. Intell. Saf. for Robot. (ISR)*, Mar. 2021, pp. 330–333.

[37] 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.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

XIAOMING SUN received the B.S. degree in measurement and control technology and instruments and the M.S. degree in control engineering from Shanxi University, Taiyuan, China, in 2019 and 2022, respectively.

His current research interests include machine learning, deep learning, and intelligent control.

XINCHUN JIA (Member, IEEE) received the B.Sc. degree in mathematics from Shanxi University, Taiyuan, China, in 1985, the M.Sc. degree from the Institute of Systems Science, Chinese Academy of Sciences, Beijing, China, in 1988, and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2003.

From 2009 to 2010, he was a Visiting Professor with the Centre for Intelligent and Networked Systems, Central Queensland University, Australia. From 2002 to 2018, he was a Professor and the Associate Dean of the School of Mathematical Sciences, Shanxi University, where he has been a Professor and the Dean of the School of Automation and Software Engineering, since 2018. He has authored more than 100 articles and two inventions. His current research interests include networked control systems, multi-agent control, and intelligent control.

YUQIAN LIANG received the B.S. degree in measurement and control technology and instruments from Putian University, Putian, China, and the M.S. degree in control engineering from Shanxi University, Taiyuan, China, in 2018 and 2022, respectively.

Her current research interests include deep learning, computer vision, and image processing.

MEIGANG WANG received the M.S. degree in automation from the Taiyuan University of Technology, Taiyuan, China, in 2006.

His current research interests include neural networks, computer vision, image processing, and multi-agent systems.

XIAOBO CHI received the B.Sc. degree in information and computing sciences and the Ph.D. degree in mathematics from Shanxi University, Taiyuan, China, in 2007 and 2014, respectively.

He joined the School of Mathematical Sciences, Shanxi University, in 2014. From 2013 to 2014, he was a Visiting Ph.D. Student with the Centre for Intelligent and Networked Systems, Central Queensland University, Australia. From 2016 to 2017, he was a Visiting Scholar with the College of Engineering and Applied Science, University of Wyoming, WY, USA. He has authored more than 20 articles and two inventions. His current research interests include networked control systems, fuzzy filtering, sensor networks, and applications. He was a recipient of the Outstanding Young Talents of Shanxi Province's "Three Jin Talents" Support Program.