IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 August 2023, accepted 28 September 2023, date of publication 6 October 2023, date of current version 19 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3322444

RESEARCH ARTICLE

An Improved Lightweight YOLOv5 Algorithm for Detecting Railway Catenary Hanging String

SHUO ZHANG, YUJIAN CHANG, SHUOHE WANG, YUESONG LI, AND TANGQI GU
Hebei Provincial Collaborative Innovation Center of Transportation Power Grid Intelligent Integration Technology and Equipment, Shijiazhuang Tiedao University, Shijiazhuang, Hebei 050043, China
School of Electrical and Electronics Engineering, Shijiazhuang Tiedao University, Shijiazhuang, Hebei 050043, China

Corresponding author: Yujian Chang (changyj@stdu.edu.cn)

This work was supported in part by the Natural Science Foundation of Hebei Province under Grant A2022210024, in part by the Natural Science Foundation of China under Grant 12072205, and in part by the Graduate Innovation Fund of Shijiazhuang Tiedao University under Grant YC2023027.

ABSTRACT Aiming at the problems of small target and low recognition accuracy of high-speed railway contact network hanging chord defects, this paper proposes a target detection algorithm for hanging chord defects based on YOLOv5. To enhance the original YOLOv5 algorithm, the MobielNetv3 module was used as the efficient and lightweight backbone feature extraction network. Depth-separable convolution was adopted instead of standard convolution, reducing the number of network parameters by 2×10^6 and increasing detection speed by 23%. Introducing BiFPN feature pyramid structure with fusion of different feature layers in neck network improves detection accuracy by 0.4%. Adding CBAM attention mechanism at the prediction end improves the feature extraction ability of the model for small target images, which further improves the detection accuracy by 0.5%. The loss function CIoU was improved to Focal EIoU in order to solve the problems of unbalanced sample datasets and vanishing IoU gradients during the training process. The experimental results exhibit that the improved algorithm achieves an average accuracy of 98.5% on the dataset, a 39% enchancment in model detection speed and a 28% reduction in model parameters, verifying that the algorithm has the advantages of high recognition accuracy and fast detection speed. It can effectively solve the technical difficulties in the detection of defects in the existing contact network suspension chords, and provides a new way of thinking for intelligent railway inspection.

INDEX TERMS YOLOv5, lightweight network, image detection, railway, deep learning.

I. INTRODUCTION

In the electrified railway industry, the pantograph chord is an important part of the high-speed railway contact network system, mainly serving as a suspension and current-carrying function [1]. However, during railway operation, the complex mechanical and electrical interaction between the contact lines in the electrical traction network and the pantographs on the railway vehicles inevitably leads to potential and difficult to detect defects such as loose strands and broken strands in the suspension strings, thus posing a threat to the safety of high-speed railways [2], [30]. In the past, manual inspections were inefficient and time-consuming, while traditional image processing methods were based on matching templates [3], which were not robust enough to meet the inspection requirements due to the varying images captured. It is therefore particularly important to propose an intelligent and efficient algorithm for the detection of hanging string defects.

In recent years, domestic and international research institutes and enterprises have achieved fruitful results in railway contact network inspection [33]. Currently, target detection algorithms based on hanging thread defects can be divided into two main categories. One class is based on traditional image processing and machine learning algorithms, which can be specifically divided into feature matching methods and statistical pattern methods. The other category is based on deep learning algorithms, and the core idea is to automatically extract and represent the features of the data through a multi-level neural network model. Compared to traditional algorithms, it is more adaptable and flexible to the needs

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

of different tasks and less susceptible to interference from environmental factors, enabling continuous training and convenient parameter adjustment. Representative network models include target detection algorithms such as SSD, YOLO, SPP-Net and Faster R-CNN [4], [5], [6]. Han et al. [7] proposed a target detection method based on an improved CNN model, which achieved component identification for 16 types of substation equipment based on the difference of regional RTDs in the equipment. Zhang et al. [8] improved the detection accuracy of insulators in the image based on the proportion of the faulty area on the insulator string to the whole image and by introducing the feature pyramid network into the YOLOv3 model. Lei et al. [9] proposed a deep convolutional neural network approach based on the Faster R-CNN algorithm to locate bird nests, which achieved intelligent fault detection on high-voltage lines. Wang et al. [10] accomplished the task of diagnosing the location of defects in transmission line images by comparing different models of VGG16, VGG19, ResNet50 and ResNet101 with multi-scale training and horizontal mirroring of training samples.

Although some good results have been achieved with the above methods, the large size of the models, the high number of model parameters and the slow speed of model detection do not meet the needs of actual field workers very well. In order to solve the above problems, this paper proposes a method to improve YOLOv5s for the detection of contact network suspension string defects, the innovation points of the method mainly include:

(1) Replacement of the YOLOv5 backbone network with the MobileNetv3 network to reduce the parameters of the network and increase the speed of detection.
(2) The introduction of BiFPN feature pyramid networks in the neck section to achieve bi-directional fusion of top-down and bottom-up deep and shallow feature maps, fusing more features through the flow of feature information from the same layer without incurring excessive computational costs.
(3) The addition of the CBAM attention mechanism on the prediction side allows the model to focus more on important feature information and suppress minor feature information.
(4) Replacing the CIoU border regression loss function with Focal-EIou, which not only solves the problem of sample imbalance but also speeds up convergence.

II. YOLOv5 ALGORITHM PRINCIPLE

The network structure of YOLOv5s [11], [12] algorithm is shown in Figure 1, which consists of four main components: input, backbone, neck and prediction.

(1) Inputs: The YOLOv5s algorithm uses mosaic data enhancement operations on the input images to enrich the background for target detection. It also uses adaptive anchor frame calculation with adaptive image scaling to reduce computational complexity during training.
(2) Backbone network: The backbone feature extraction network uses CSPDarknet53 for the initial extraction of feature images. Three feature maps of sizes (80, 80, 256), (40, 40, 512), and (20, 20, 1024) can be obtained for later classification and regression prediction. The Focus slicing operation is introduced to take a value every other pixel before the image enters the backbone, stacking the four independent feature layers obtained, at which point the information in the width dimension is converted to the channel dimension and the input channel is expanded by four times, and then the feature extraction is carried out, which can effectively reduce the calculation of parameters and memory usage.
(3) Neck network: The FPN+PAN structure is used as a feature fusion network in the neck, which improves detection accuracy by carrying information about target features of different sizes.
(4) Prediction side: GIoU [13] was chosen as the bounding box loss function on the prediction side, and NMS maximum value suppression was used to solve the problem of redundant generated boxes.

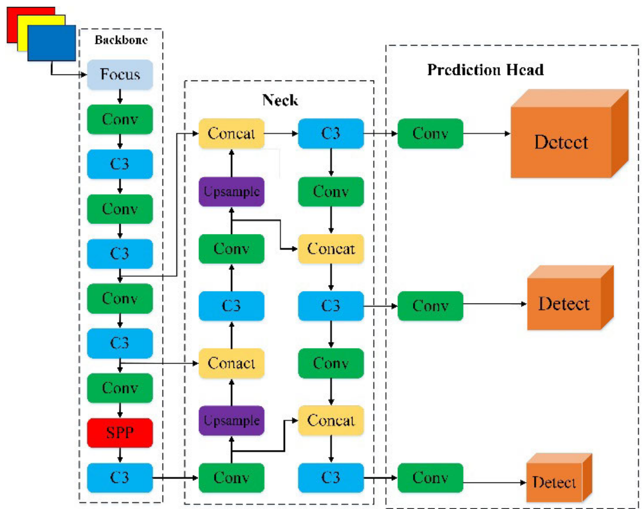


FIGURE 1. The original YOLOv5 network structure.

III. IMPROVEMENT OF THE YOLOv5 NETWORK ARCHITECTURE DESIGN

A. IMPROVEMENTS OF THE BACKBONE NETWORK

MobileNet network is a lightweight CNN proposed by Google in 2017, whose main contribution is to replace ordinary convolution with depth-separable convolution, which significantly reduces the model parameters while ensuring accuracy [14], [32]. Depth-separable convolution requires separate convolution of each channel with different convolution kernels, and then upscaling and downscaling by point-by-point convolution to obtain the corresponding feature maps. The difference between depth-separable convolution and standard convolution is shown in Figure 2.

Assume that the input image resolution is of size $D_K \times D_K \times D_K$, and that the output feature size remains the same

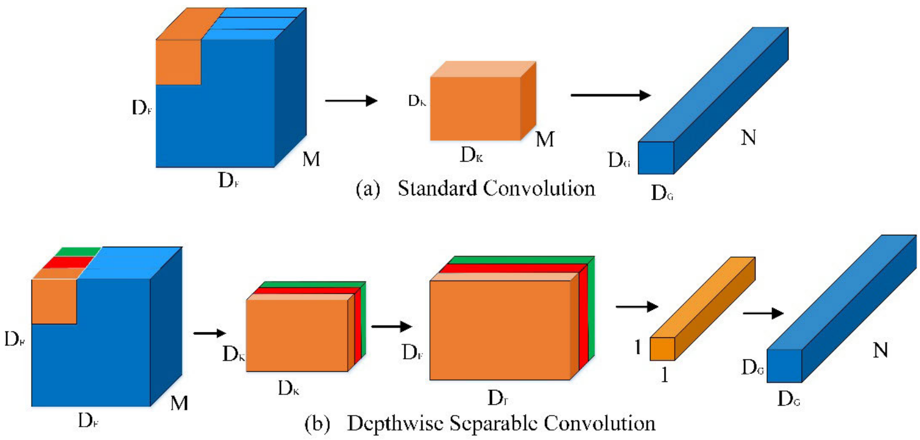


FIGURE 2. **Comparision between conventional convolution and depth separable convolution.**

after passing through a convolution kernel of size $D_F \times D_F$ with a step size of 1 and a number of channels N. Then the ratio of computational effort between the depth-separable convolution and the standard convolution is

$$\frac{D_K \times D_K \times M \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \tag{1}$$

When the convolution kernel size D_F is 3, it is known that the depth-separable computation is reduced to approximately $\frac{1}{8} - \frac{1}{9}$ of the original standard convolution, which can greatly reduce the computational cost of the model and improve the detection speed.

In order to solve the problem of convolutional kernel failure in the deep convolutional part of the network model that tends to occur during training, the MobileNetv2 network [15] introduces an inverse residual structure based on MobileNetv1, as shown in Figure 3.

The feature map is first up-dimensioned by 1×1 point-by-point convolution, then extracted by 3×3 deep convolution, and finally down-dimensioned by 1×1 point-by-point convolution, presenting a structure with a large middle and small ends. And in the bottleneck layer, the linear activation function ReLU6 is used instead of the traditional ReLU function. This provides greater robustness while avoiding the loss of feature information.

The biggest highlight of MobileNetv3 is the inclusion of the SE module, which consists of two parts, Squeeze and Excitation [16], [17], as shown in Figure 4. This module performs global average pooling on the input feature maps to obtain the output vectors. In this process, two fully connected layers are set up to reduce the number of channels and lower the number of parameters by setting different numbers of neurons.

The SE module can assign a weighting relationship to each input channel according to its importance, so that it focuses more on the key information in the image while ignoring irrelevant information.

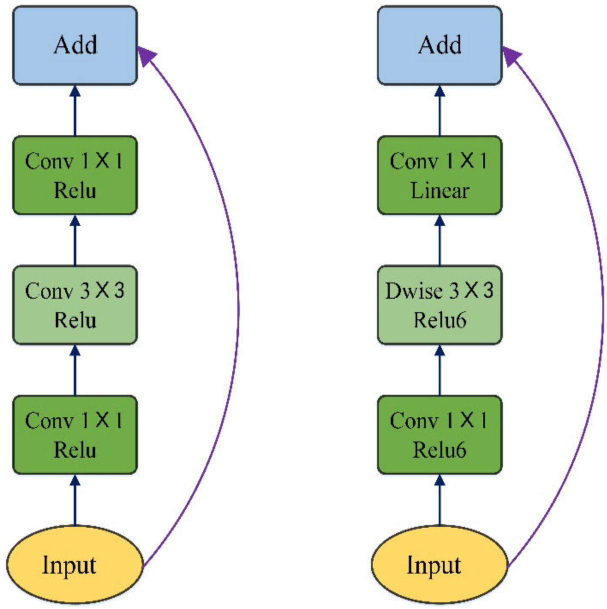


FIGURE 3. **Residual structure and Inverse Residual structure.**

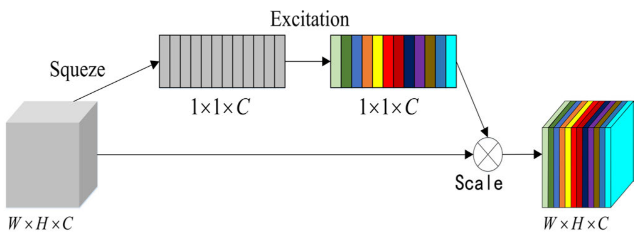


FIGURE 4. **Structure of the SE module.**

YOLOv5 is a regression-based single-stage target detection algorithm that can classify targets while detecting them [28], [29]. This paper describes the replacement of the YOLOv5 network's original CSPDarknet53 network with MobileNetv3 structure. The replacement leads to a

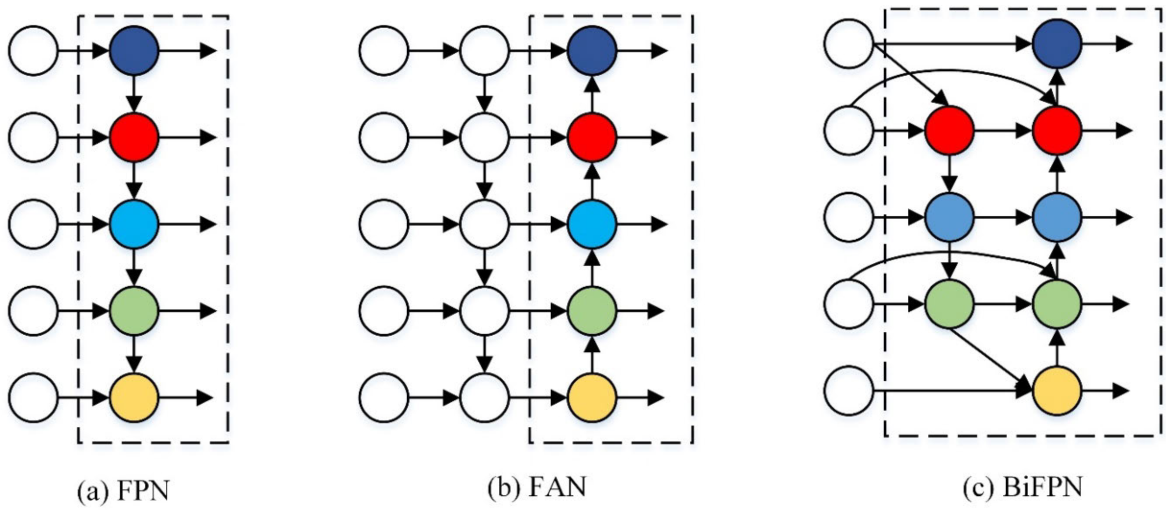


FIGURE 5. FPN, FAN and BiFPN network structure.

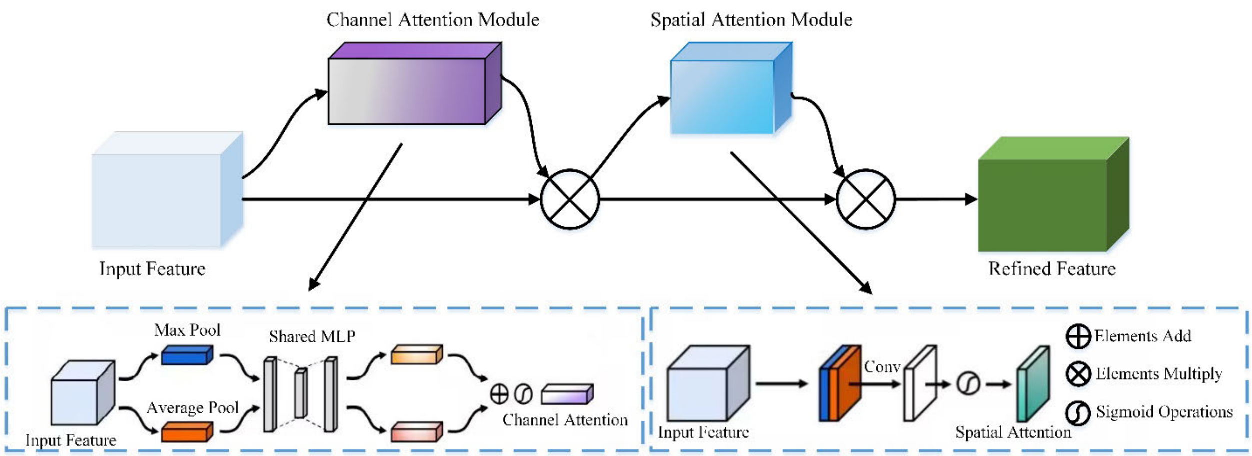


FIGURE 6. Structure diagram of CBAM attention module.

23% improvement in the model's computation rate and a 0.7% increase in detection accuracy, due to the efficient feature extraction technique and low parameter count of MobileNetv3.

B. IMPROVEMENTS OF THE NECK NETWORK

The input images need to be processed by the neck network after feature extraction by the YOLOv5s backbone network and input to the prediction side, while the original YOLOv5s feature fusion uses the FPN+PAN network structure as the neck network [18], as shown in Figure 5(a), (b). The PANet network structure [19] is a bottom-up fusion mechanism that can convey feature maps with strong localization feature information at the bottom, but relatively weak semantic feature information; the FPN network structure is a top-down fusion mechanism that can convey structure maps with strong semantic feature information at the top, but weak localization feature information. Although the original neck network can achieve the fusion of localisation feature information and semantic feature information through bidirectional feature fusion, the two parts use a direct summation operation, which may lead to important feature information being ignored.

To solve the above problems, this algorithm adds the more powerful BiFPN [20] module to the neck part of YOLOv5 to improve the accuracy, robustness and computational efficiency of target detection, making the model more powerful and efficient. As shown in Figure 5(c), the main improvement points compared to the traditional PANet network structure are the following three aspects:

(1) To simplify the network structure, BiFPN removes intermediate nodes with only one input edge from the original characteristic pyramid network.
(2) To fuse more feature information without incurring more computational cost, BiFPN adds a feature fusion path between input and output nodes at the same level.

(3) The BiFPN structure adds extra weight to each input feature map, allowing the network to gradually learn the importance of each input feature during the training process, and to achieve a fusion of higher-level features by stacking them multiple times.

C. IMPROVEMENTS OF THE PREDICTION HEAD

In order to refine the feature maps extracted by the model and improve the classification effect of the model, the algorithm in this paper introduces the CBAM attention module in the prediction head network [21]. The model's detection accuracy improved by 0.5% by weighting different parts of the input data and multiplying them with corresponding feature maps.

The CBAM attention module is shown in Figure 6, which mainly includes the channel attention module and the spatial attention module. First, the input feature image is fed into the channel attention module, and the n-dimensional feature vector is obtained after the maximum pooling and average pooling operations, and then the weight coefficients are obtained after processing by the fully connected layer and sigmoid function. Then, the results obtained in the previous step are fed into the spatial attention module, and after the same operation, the obtained vectors are stacked by connection. Finally, the spatial attention Ms is generated by the output of the convolution and sigmoid operation.

The CBAM attention module process equation is

$$\begin{cases} \dfrac{F' = M_c(F) \otimes F}{F'' = M_S(F') \otimes F'} \end{cases} \tag{2}$$

D. IMPROVEMENTS OF THE LOSS FUNCTION

YOLOv5 uses CIoU Loss (Complete Intersection over Union Loss) as the loss function, although it takes into account the overlap area, centroid distance, and width-height of the bounding box regression. However, the $\boldsymbol{\nu}$ in its formula indicates the difference in width and height, rather than the true difference between width and height respectively and the confidence level, which may lead to slow convergence and inaccurate regression. To address this problem, this paper improves the CIoU loss by introducing the Focal EIoU loss as a loss function [31], which not only solves the sample imbalance problem in the bounding box regression task, but also makes the newly generated loss function capable of obtaining more accurate prediction frames and more precise target detection results. As a result of this, there was a 1.3% increase in accuracy. The formula for CIoU Loss is shown in (3): where ω^{gt}, h^{gt}, b^{gt}, ω, h and b present the width, height and centre point of the real frame and the predicted frame respectively; ω^c, h^c epresents the width, height and Euclidean distance of the smallest outer rectangle, ρ represents b and b^{gt}.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(\omega^c)^2 + (h^c)^2} + \alpha\nu \tag{3}$$

where

$$\alpha = \frac{\nu}{(1 - IoU) + \nu} \tag{4}$$

$$\nu = \frac{4}{\pi^2}(\arctan\theta\frac{\omega^{gt}}{h^{gt}} - \arctan\theta\frac{\omega}{h})^2 \tag{5}$$

The Focal EIoU loss function is shown in (6), with IoU representing the cross-merge ratio, L_{IoU}, L_{asp} and L_{dis} representing the IoU loss, the width-height loss and the distance loss respectively, and γ being a hyperparameter controlling the arc of the curve.

$$L_{Focal\mathrm{EIoU}} = IoU^{\gamma} L_{EIoU} \tag{6}$$

where

$$\begin{aligned} L_{EIoU} &= L_{IoU} + L_{dis} + L_{asp} \\ &= 1 - IoU + \frac{\rho^2(b, b^{gt})}{(\omega^c)^2} + \frac{\rho^2(\omega, \omega^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \end{aligned} \tag{7}$$

E. OVERALL IMPROVEMENT IDEAS

The improved YOLOv5 network structure is shown in Figure 7, and the overall improvement steps are as follows:

a) The lightweight MobileNetv3 network structure model is used as the backbone feature extraction network, and the standard convolution in the network model is replaced with a depth-separable convolution, and the SE attention mechanism is introduced on the basis of the inverse residual structure, so that its feature extraction capability is further enhanced.
b) By introducing the BiFPN module in the Neck part instead of the original FPN and PAN structure, the model can better capture the target information of different scales during the training process and improve the detection accuracy. At the same time, the propagation path of the features can be dynamically adjusted to improve the robustness of the model.
c) By adding a CBAM module to each of the three branches on the prediction side of YOLOv5, the importance of each channel in the feature map can be adaptively learned and the detailed features of the target can be enhanced. And a large performance improvement is achieved without adding too much extra computational overhead.
d) Improving the border regression loss function CIoU to Focal EIoU reduces the width and height difference between the target and anchor frames, while speeding up convergence and improving localisation.

IV. EXPERIMENTAL SETTINGS

A. INTRODUCTION TO THE DATASET

The dataset for this paper is derived from field shots taken by inspection drones(M300RTK) along the railway line, most of which are samples of normal hanging chords. In order to increase the diversity of image samples and alleviate the data imbalance problem, in this paper, the faulty hanging

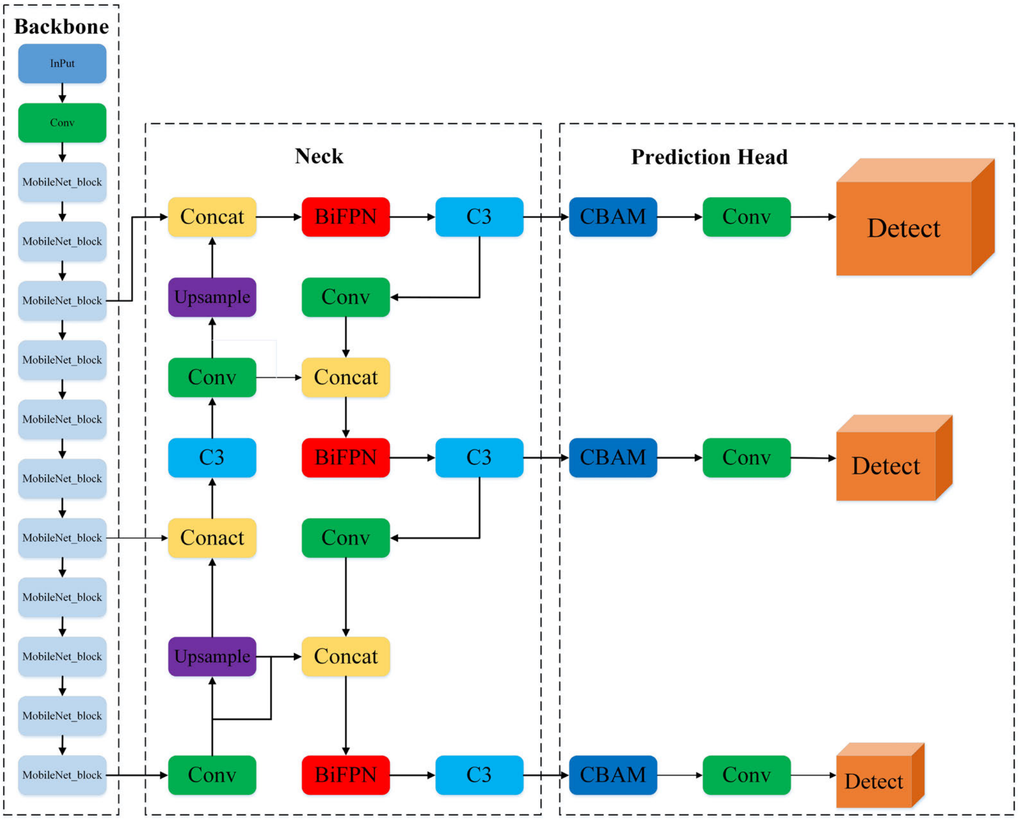


FIGURE 7. Improved YOLOv5s network structure.

chord can be augmented by 17 times of the data set through the data augmentation algorithm by rotating the chord by 90 degrees/180 degrees, flipping it, blurring it, changing the luminance, increasing the noise, and combining the two by two with each other, and other operations. The hanging string categories are divided into three categories: normal, loose strands and broken strands, and some images of the data set are shown in Figure 8.

FIGURE 8. Dataset images.

The final hanging string dataset has 2589 images, 1263 normal samples, 639 loose strand samples and 687 broken strand samples. And 60% were randomly selected as the training set, 20% as the validation set and 20% as the test set, the specific quantities are shown in Table 1.

TABLE 1. Distribution of the number of dataset categories.

Sample type	Total	Training set	Validation set	Test Set
Normal	1263	759	252	252
Loose	639	383	128	128
Broken	687	412	137	138

B. EXPERIMENTAL ENVIRONMENT

To have an objective and fair evaluation of the improved algorithm proposed in this paper, all the experiments were conducted in the same experimental environment. The experimental environment uses Window10 operating system, CPU with Inter(R) Core (TM)i5-7500@3.40GHz, GPU with NVIDIA GeForce RTX 2070, size 8GB. the CUDA version is 12.0, Pytorch version is 1.9.0, Python version is 3.8.3.

C. EVALUATION INDICATORS

In order to evaluate the target detection effect objectively and fairly, this paper selects precision rate, recall rate, mAP@0.5, mAP@0.5:0.95, FPS (Frames Per Second), and number of parameters as evaluation indexes. Precision and recall were calculated by confusion matrix Table 2.

TABLE 2. Confusion matrix.

True Result	Predicted Result	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

The precision rate is the proportion of true positive cases in the sample with positive prediction results, calculated as in (8) as:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

where TP indicates that the correct category is predicted to be the correct category and FP indicates that the incorrect category is predicted to be the correct category. AP is the area enclosed by the P-R curve and the coordinate axis, which can reflect the detection effect of the target model. When IOU is set to 0.5, the expression formula of mAP@0.5 is as follows:

$$mAP@0.5 = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{9}$$

where AP_i denotes the average precision of target detection in category i and K denotes the category.

V. ANALYSIS OF EXPERIMENTAL RESULTS

A. ANALYSIS OF ABLATION EXPERIMENT

In order to verify the accuracy and superiority of this algorithm, the YOLOv5 algorithm as well as the algorithm proposed in this paper are compared and verified experimentally under the same experimental conditions, respectively. The ablation experiment table is shown in Table 3, where "✓" indicates that the strategy is used and "✕" indicates that the strategy is not used. The training process are set Batchesize is 32, the number of iterations is 150 epochs, and the initial learning rate is 0.0001. The evaluation indexes are P, R, mAP@0.5, mAP@0.5:0.95, and the comparison graph of each evaluation index is shown in Figure 9-12.

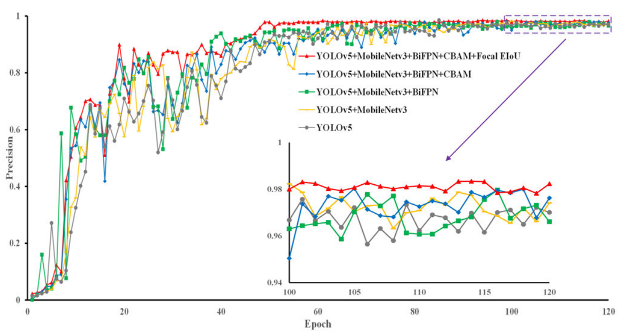


FIGURE 9. Precision comparision chart.

From the ablation experimental data of each model in Table 3, it can be seen that the original YOLOv5 model evaluation metrics P, R, mAP@0.5, and mAP@0.5:0.95 have

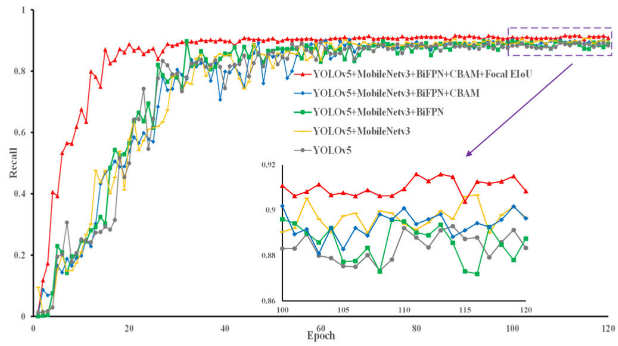


FIGURE 10. Recall cmparision chart.

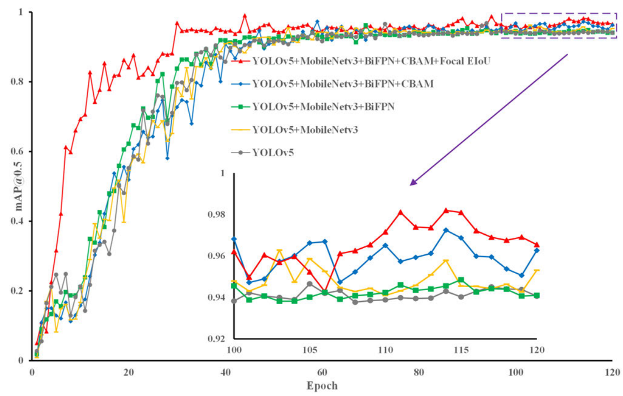


FIGURE 11. mAP@0.5 comparision chart.

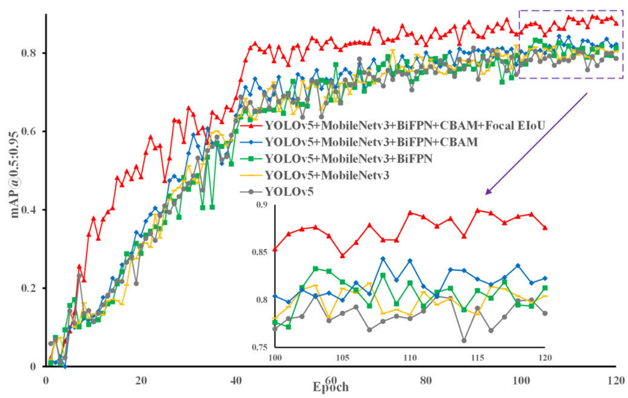
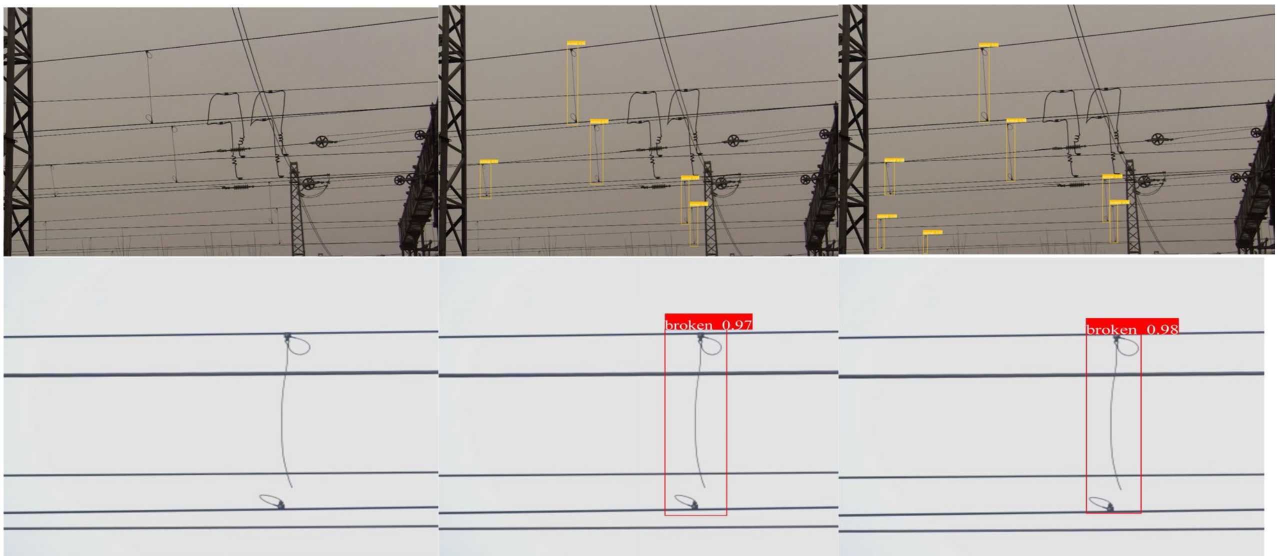


FIGURE 12. mAP@0.5:0.95 comparision chart.

values of 97.2%, 88.6%, 93.7%, and 77.5%, respectively, in the training of the dangling string dataset. Replacing the model backbone with the MobileNetv3 module resulted in a reduction in the amount of parameters of the model by 2×10^6, while the values of the evaluation metrics increased by 0.7%, 1.8%, 1.1%, and 1.3%, respectively. The introduction of BiFPN feature pyramid in the neck accelerates the fusion of feature information in the same layer of images. It makes the evaluation index of mAP0.5:0.95 improve by 2.1% while ensuring the precision rate and recall rate remain basically unchanged. The addition of the CBAM attention mechanism on the prediction side significantly improves the

TABLE 3. Ablation experiment.

Model	MobileNetv3	BiFPN	CBAM	Focal EIoU	Precision	Recall	Map@0.5	Map@0.5:0.95	Params	FPS
1	×	×	×	×	97.2	88.6	93.7	77.5	7.03×10^6	178
2	✓	×	×	×	97.9	90.4	94.8	78.8	5.03×10^6	312
3	✓	✓	×	×	97.6	89.5	94.3	80.9	5.04×10^6	295
4	✓	✓	✓	×	98.1	90.2	96.1	82.1	5.06×10^6	295
5	✓	✓	✓	✓	98.5	91.5	97.3	85.3	5.06×10^6	290



(a) The Original image (b) Image of YOLOv5s detection (c) Image of the improved YOLOv5s detection

FIGURE 13. Comparision between original YOLOv5s and improved YOLOv5s.

model's ability to extract complex features such as image details. As a result, the values of the evaluation indicators increased by 0.5 per cent, 0.7 per cent, 1.8 per cent and 1.2 per cent, respectively. Changing the loss function to Focal EIoU further improved the quality of the sample dataset, resulting in an increase in the evaluation metrics by 0.4%, 1.3%, 1.2%, and 3.2%, respectively. The trend analysis of the data curves before and after the improvement shows that the accuracy rate of the improved model gradually stabilised and remained unchanged at the 50th training epoch. From the recall and mAP@0.5 evaluation metrics, the algorithm proposed in this paper can be smoothed faster. Therefore, the accuracy improvement of this paper's algorithm in detecting and recognising targets in complex scenes shows that the improvement of CBAM attention mechanism, BiFPN structure and MobileNet network reduced-parameter structure is very effective.

B. COMPARISON OF TEST RESULTS

Figure 13 shows a comparison of some of the detected images during the experiment. From the observation in Figure 13(b), it can be seen that the original YOLOv5 suffers from a serious leakage problem in the multi-target situation, and the accuracy and number of detections are poorly effective for small targets as well as complex background environment interference. Comparing Figures 13(b) and 13(c), it can be seen that the original YOLOv5 model had two missed targets, whereas the improved YOLOv5 not only detected the missed dangling string targets well, but also the accuracy of each target was greatly improved. The experimental results show that the improved algorithmic model has stronger anti-interference and robustness.

C. COMPARISON WITH RELATED METHODS

To further verify the superiority of the detection efficiency and classification accuracy of the algorithm in this paper, the algorithm in this paper is compared with the mainstream target detection algorithms SSD, YOLOv4, YOLOv4-tiny, YOLOv5s, YOLOX-S, IN-YOLO, LFF-YOLO and AED-YOLOv5 models in the current stage of experiments [24], [25], [26], [34]. The same dataset is used in the experiment, and the same experimental parameters, hardware conditions

TABLE 4. Comparision of detection performance of different algorithms.

Model	Precision	Recall	Map@0.5	Map@0.5:0.95	Params	FPS
SSD	97.9	89.5	94.6	81.4	23.7×10^6	62
YOLOv4	96.8	87.9	92.6	77.2	63.9×10^6	24
YOLOv4-tiny	97.1	88.3	92.5	78.4	8.9×10^6	165
YOLOv5s	97.2	88.6	93.7	77.5	7.03×10^6	209
IN-YOLO[34]	97.5	88.9	94.2	79.7	48.7×10^6	36
YOLOX-S	97.7	89.2	94.8	82.5	9.01×10^6	163
LFF-YOLO[26]	98.8	92.3	97.9	86.4	6.05×10^6	243
AED-YOLOv5[27]	98.2	91.0	96.9	84.8	5.04×10^6	295
Ours	98.5	91.5	97.3	85.3	5.06×10^6	290

and software environment are set. The experimental results are shown in Table 4.

Table 4 illustrates that the algorithm proposed in this study has significantly fewer parameters, only 5.06×10^6, as compared to the SSD and YOLOv4 algorithms, which have 23.7×10^6 and 63.9×10^6 parameters, respectively. It is noteworthy that, despite having only half the number of parameters, the proposed algorithm outperforms the YOLOX-S algorithm by enhancing evaluation indexes by 0.8 percent, 2.3 percent, 2.5 percent, and 2.8 percent for different parameters. Compared with the current lightweight YOLOv4-tiny algorithm, the model detection accuracy is improved by 1.4% with a reduction in the number of parameters. In order to further illustrate that the model proposed in this paper is better than other models for detection, it is again compared with the LFF-YOLO and AED-YOLOv5 algorithms, respectively. The comparison of experimental data reveals that although the accuracy of our model is only 0.3% lower than that of the LFF-YOLO model, our model has one-i-fifth fewer parameters. Although the number of parameters of the AED-YOLOv5 model is almost the same as in this paper, the evaluation indexes of the algorithm in this paper are higher than it by 0.3%, 0.5%, 0.4%, 0.5% respectively. Therefore, combining the accuracy and the number of parameters, the algorithm in this paper is more suitable to be deployed for the task of identifying defects in contact network hanging chord lines.

VI. CONCLUSION

1) Aiming at the current problems of long cycle time, low efficiency and insufficient robustness and generalisation ability of traditional image processing in manual inspection, this proposes a model with YOLOv5 target algorithm model as the basic framework and MobileNetv3 lightweight module fused into the model according to the characteristics of the over-hanging chord faults, which effectively reduces the parameters of the model and improves the speed of computation while guaranteeing the accuracy of the positioning. Also, by introducing the BiFPN feature pyramid structure in the neck, adding the CBAM attention mechanism, and changing the loss function to Focal EIoU at the prediction end, the ability of extracting feature maps is improved, and the problem of imbalance between positive and negative samples is solved. Finally, the superiority of the improved algorithm is verified on this dataset, and the values of the evaluation indices are all improved, which can meet the field requirements of detection accuracy and detection speed.

2) To demonstrate the progress and superiority of the algorithm, and in comparison with current mainstream algorithmic models, it is shown that the algorithm has some value in practical applications. However, there are still some shortcomings in this model, and the next step is to further optimise the model's algorithm to improve recognition accuracy, while further reducing memory consumption and recognition time, so that it can be better applied to more mobile devices.

REFERENCES

[1] H. Liang, C. Zuo, and W. Wei, "Detection and evaluation method of transmission line defects based on deep learning," *IEEE Access*, vol. 8, pp. 38448–38458, 2020.

[2] Z. Wang, X. Liu, H. Peng, L. Zheng, J. Gao, and Y. Bao, "Railway insulator detection based on adaptive cascaded convolutional neural network," *IEEE Access*, vol. 9, pp. 115676–115686, 2021.

[3] Z. Li, Y. Zhang, H. Wu, S. Suzuki, A. Namiki, and W. Wang, "Design and application of a UAV autonomous inspection system for high-voltage power transmission lines," *Remote Sens.*, vol. 15, no. 3, p. 865, Feb. 2023.

[4] 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.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[6] 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.

[7] S. Han, F. Yang, H. Jiang, G. Yang, N. Zhang, and D. Wang, "A smart thermography camera and application in the diagnosis of electrical equipment," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[8] X. Zhang, Y. Zhang, J. Liu, C. Zhang, X. Xue, H. Zhang, and W. Zhang, "InsuDet: A fault detection method for insulators of overhead transmission lines using convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[9] X. Lei and Z. Sui, "Intelligent fault detection of high voltage line based on the faster R-CNN," *Measurement*, vol. 138, pp. 379–385, May 2019.

[10] W. Wanguo, W. Zhenli, L. Bin, Y. Yuechen, and S. Xiaobin, "Typical defect detection technology of transmission line based on deep learning," in *Proc. CAC*, Hangzhou, China, Nov. 2019, pp. 1185–1189.

[11] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, "YOLO-FIRI: Improved YOLOv5 for infrared image object detection," *IEEE Access*, vol. 9, pp. 141861–141875, 2021, doi: 10.1109/ACCESS.2021.3120870.

[12] Y. Fan, Y. Li, Y. Shi, and S. Wang, "Application of YOLOv5 neural network based on improved attention mechanism in recognition of Thangka image defects," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 1, pp. 245–265, 2022.

[13] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, Apr. 2020, pp. 12993–13000.

[14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[16] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.

[17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[18] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768, doi: 10.1109/CVPR.2018.00913.

[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. 11534–11542.

[20] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.

[21] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Comput. Vis.*, Sep. 2018, pp. 3–19.

[22] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022.

[23] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 2778–2788.

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

[25] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[26] X. Qian, X. Wang, S. Yang, and J. Lei, "LFF-YOLO: A YOLO algorithm with lightweight feature fusion network for multi-scale defect detection," *IEEE Access*, vol. 10, pp. 130339–130349, 2022.

[27] S. Xu, Q. Feng, J. Fei, G. Zhao, X. Liu, H. Li, C. Lu, and Q. Yang, "A locating approach for small-sized components of railway catenary based on improved YOLO with asymmetrically effective decoupled head," *IEEE Access*, vol. 11, pp. 34870–34879, 2023.

[28] Y. H. Shao, D. Zhang, and H. Y. Chu, "A review of YOLO object detection based on deep learning," *J. Electron. Inf. Technol.*, vol. 44, no. 10, pp. 3697–3708, Oct. 2022, doi: 10.11999/JEIT210790.

[29] Q. Wang, M. Cheng, S. Huang, Z. Cai, J. Zhang, and H. Yuan, "A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed solanum rostratum dunal seedlings," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107194, doi: 10.1016/j.compag.2022.107194.

[30] J. Zhong, Z. Liu, H. Wang, W. Liu, C. Yang, Z. Han, and A. Núñez, "A looseness detection method for railway catenary fasteners based on reinforcement learning refined localization," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: 10.1109/TIM.2021.3086913.

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

[32] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 131–135.

[33] R. Chen, Y. Lin, and T. Jin, "High-speed railway pantograph-catenary anomaly detection method based on depth vision neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2022.3188042.

[34] D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "IN-YOLO: Real-time detection of outdoor high voltage insulators using UAV imaging," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1599–1601, Jun. 2020.

SHUO ZHANG was born in Handan, Hebei, China, in 1996. He received the B.S. degree from Langfang Normal University, China, in 2019. He is currently pursuing the M.S. degree in electrical engineering with Shijiazhuang Tiedao University. His current research interests include deep learning algorithms and fault diagnosis for electrical equipment.

YUJIAN CHANG was born in Taiyuan, Shanxi, China, in 1972. She received the B.E. degree from the Department of Automation, East China Institute of Metallurgy, China, in 1994, the M.E. degree from the University of Science and Technology Beijing, China, in 2005, and the Ph.D. degree from Shijiazhuang Tiedao University, China, in 2021. She is currently a Professor with the School of Electrical and Electronic Engineering, Shijiazhuang Tiedao University. Her research interests include microgrid traction power failure detection and image defect detection.

SHUOHE WANG was born in Shijiazhuang, Hebei, China, in 1968. He received the B.E. degree in industrial electrical automation from the Hebei University of Engineering, China, in 1990, the M.E. degree in signal and information processing from Xidian University, China, in 1999, and the Ph.D. degree from Tianjin University, China, in 2009. He is currently a Professor with the School of Electrical and Electronic Engineering, Shijiazhuang Tiedao University. His research interests include the key technologies of traction power supply, power quality analysis and control, digital signal processing and pattern recognition, and renewable energy.

YUESONG LI was born in Shijiazhuang, Hebei, China, in 1998. He received the B.S. degree from Hebei Agricultural University, China, in 2021. He is currently pursuing the M.S. degree in electrical engineering with Shijiazhuang Tiedao University. His current research interests include railway catenary system analysis and image defect recognition.

TANGQI GU was born in Shijiazhuang, Hebei, China, in 1999. He received the B.S. degree from the Hebei University of Architecture, China, in 2022. He is currently pursuing the M.S. degree in electrical engineering with Shijiazhuang Tiedao University. His current research interests include railway catenary system analysis and image defect recognition.