IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 February 2023, accepted 29 March 2023, date of publication 5 April 2023, date of current version 12 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264441

RESEARCH ARTICLE

A Locating Approach for Small-Sized Components of Railway Catenary Based on Improved YOLO With Asymmetrically Effective Decoupled Head

SHUAI XU, QIANG FENG, JIYOU FEI, GENG ZHAO, XIAODONG LIU, HUA LI, CHANG LU, AND QI YANG

College of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University, Dalian 116028, China

Corresponding author: Jiyou Fei (fjy@djtu.edu.cn)

This work was supported by the Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science & Technology University, under Grant OMTIKF2021001.

ABSTRACT In the routine inspection process of railway catenary systems, the primary task is to find out the locations of various components accurately. The complex composition of the components in the catenary system and their large dimensional differences make the inspection of small components considerably difficult. Aiming at the problem of the difficulty in locating small components, a new locating method, named asymmetrically effective decoupled head-you only look once (AED-YOLO), for locating small components of the catenary has been proposed in this study. In this method, firstly, a small object detection layer has been added to improve the detection accuracy of the small-sized components such as fastener nuts and bracing wire. Secondly, to reduce missed and false detection errors of small components, the improved bidirectional feature pyramid network with high-order spatial interactions and recursive gated convolution has been used to fuse the features of different scales to further enhance the ability to detect small objects. Finally, an asymmetrically effective decoupled head has been proposed using different decoupled branches to decouple the classification and localization processes, thus further reducing the error in small-sized object classification and location. Experiments performed on the railway catenary dataset collected on-site show that the proposed localization method can efficiently improve detection accuracy, achieving a mean average precision of 93.5%. Thus, compared to the other methods, the proposed method can accurately locate small-sized components.

INDEX TERMS Railways catenary, component detection, YOLO, asymmetrically effective decoupled head.

I. INTRODUCTION

In a high-speed railway transportation system, the catenary system plays an important role as the power transmission equipment. As shown in Fig. 1, the composition of the railway catenary, erected along the railway line, contains a total of more than 40 types of components of various sizes. Under long-term working load conditions, the catenary inevitably suffers from defects and abnormalities in the components. For high-speed trains, the quality status of the catenary seriously affects the safety of the trains. Therefore, routine maintenance and quality inspection of the catenary is particularly important [1], [2].

Even in the current technological background, the routine maintenance of the catenary is still dominated by manual visual inspection, in which the determination of whether defects or abnormalities have occurred in the catenary is done by visual inspection by the maintenance personnel [3]. However, it cannot be ignored that owing to the presence of a large number of different-sized components in the catenary system, relying on manpower alone cannot meet the requirement of timeliness of railway maintenance [4], [5], [6]. In addition, as shown in Fig. 2, due to the complex composition of the catenary, the size and shape of the small and large components are quite different. It is very easy to

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

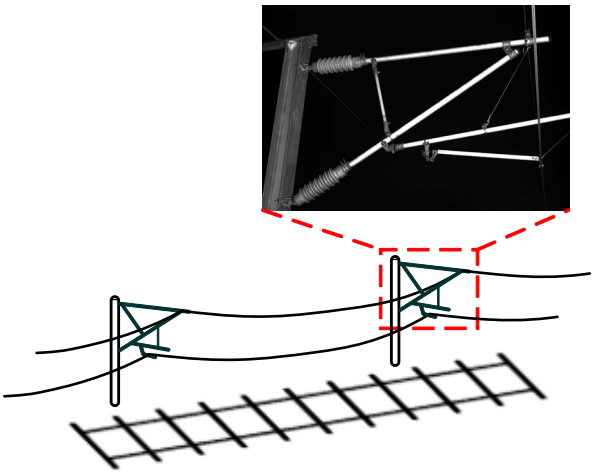


FIGURE 1. Schematic showing the location of the catenary in a railway line.

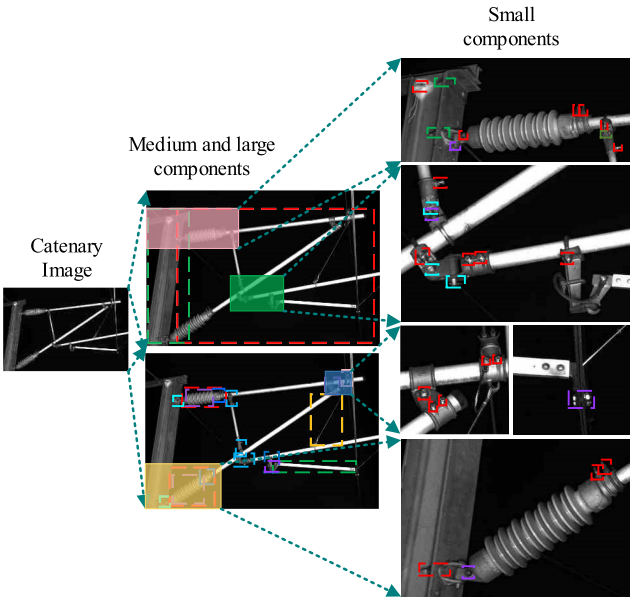


FIGURE 2. Structural of the catenary.

produce visual fatigue and a large subjective impact by manual inspection alone, which is unable to meet the requirement of maintenance of small components [7], [8].

The most critical step in railway catenary maintenance is accurately locating the catenary components, based on which defects and abnormalities can be analyzed [5]. With the development of computer vision technology, various visual inspection methods based on deep learning have been applied in the industry. In this paper, we summarize the existing research literature on catenary component location detection as well as the advantages and disadvantages of the existing methods in their practical applications.

Zhong, et al. [1] proposed a forced learning refined localization-based catenary fasteners defect detection method, which utilized the faster region-based convolutional neural network (Faster RCNN) [9] and residual network 101 (ResNet101) [10] to perform coarse localization of the catenary to obtain a coarse localized feature map. Then, the obtained feature map was enhanced at a fine level by reinforcement learning [11]. Experimental results showed that the proposed reinforcement learning-based fastener defect and looseness detection method was effective. This has important implications for the progress of catenary health condition detection methods. A CNN-based fastener detection method, named PVANET++, was proposed by Zhong, et al. [4], in which the input images were localized by PVANET++, and the Hough transform was used to locate the pins. Subsequently, a new anchoring mechanism was applied to generate multiple defect candidate frames. Finally, multiple hidden layer features and candidate frames were combined to obtain the final defect location. The results obtained by applying their method on an actual railway catenary dataset showed that the proposed PVANET++ has advanced localization detection capability. Tan, et al. [12] proposed a mask region-based convolutional neural network (Mask RCNN) [13] and multi-feature clustering-based catenary detection method. The proposed method used gradient, texture, and grey feature fusion to detect and locate defects such as broken objects, dirt, and foreign objects in insulators with high accuracy. Tan, et al. [14] used Faster RCNN as the base network to locate and identify catenary pendants. This method incorporated fine locating algorithms, edge fitting algorithms, and bending zoom algorithms to achieve sub-pixel-level defect detection, which could be used to efficiently detect defects such as micro deformation, loosening, and so on. A pantograph-body anomaly detection method for high-speed railways was proposed by Chen, et al. [15], which used a deep vision network to locate and detect pantographs in railway catenary systems. In addition, this method combined the image vision feature extraction algorithm to detect the state of the connection point between the pantograph and the body of the car. A fault diagnosis method was proposed for the current-carrying ring components [16], which used an improved RetinaNet [17] algorithm. The proposed method can improve the efficiency and safety of the railway transportation system and avoid unnecessary accidents.

The above analysis shows that most of the current defect detection and fault diagnosis methods for catenary equipment belong to the analysis of a single component. However, research on the detection and location techniques for more than 40 components of the entire catenary system, especially small components, has not been done fully. At present, the dimensions of the different types of components vary considerably. Thus, it is difficult to detect components of different sizes.

To solve the difficult problem of detecting small targets and the problem of missed detections, a method has been proposed in this study for accurately locating small-sized components in the catenary. The main contributions of this study can be summarized as follows:

1. Based on you only look once (YOLO) network, an enhanced method, named asymmetrically effective decoupled head-YOLO (AED-YOLO), has been proposed for locating small-sized components in a high-speed railway catenary system.
2. Combined with HorNet [18], an improved version of the bidirectional feature pyramid network has been proposed for bidirectional feature fusion to improve the detection performance of small-sized objects, thus improving the detection accuracy for catenary components.
3. An asymmetrically effective decoupled head (AED-Head) has been proposed to provide different decoupled methods for different task properties of classification and localization using an asymmetrical decoupled structure. It can match the spatial feature requirements of different tasks and thus further improve the classification and localization accuracy.

The remainder of this paper has been organized as follows. A few related works on object detection based on deep learning have been introduced in Section II. The catenary full-component locating detector, named AED-YOLO, has been described in Section III. In Section IV, experiments performed using the proposed method and the results thus obtained have been presented to illustrate the effectiveness of the proposed method. Finally, conclusions from this study have been given in Section V.

II. RELATED WORK

A. YOLOv5 NETWORK

The YOLO series of object detection networks are used in a wide range of industries and agriculture [19], [20], [21], [22]. Version V5 combines most of the steps from V1 to V4 for optimal combination. The architecture of the YOLOv5 network can be split into three parts: backbone, neck, and head. In the backbone network, cross stage partial dark network (CSPDarkNet) is used for image feature extraction and image downsampling. In the neck network, the path aggregation network is used for feature fusion at different scales. Image feature extraction and semantic information capture for large, medium, and small targets are achieved by fusing feature maps of 19 px (pixels) × 19 px, 38 px × 38 px, and 76 px × 76 px sizes. Finally, a simple full convolutional prediction head is used in the prediction head section to generate localized candidate frames for different-sized targets, resulting in a final prediction output.

After the success of YOLO, some improvements based on the YOLO architecture, such as YOLOX [23], TPH-YOLO [24], and other methods, have been done which will not be described in this paper.

B. FEATURE PYRAMID NETWORK

Shallow feature maps maintain rich shallow features of images, such as their morphology, size, and other features, whereas deep feature maps preserve semantic information, such as contextual information and location relationships [25], [26]. Better detection performance can be achieved by fusing the feature maps of the different phases. The first feature pyramid network (FPN) [27], whose structure is shown in Fig. 3(a), was proposed for the first time to build a deep-to-shallow feature fusion path, thus enriching the semantic information in the shallow feature maps and obtaining different degrees of accuracy improvement in several application scenarios. Based on the FPN, Liu, et al. [28] proposed a secondary fusion model named path aggregation network (PANet), which adds a bottom-up secondary fusion path, to enhance the effectiveness of feature fusion and thus improve the detection accuracy. The structure of the PANet is shown in Fig. 3(b). Although PANet adds a secondary fusion path, it is a simple structured bidirectional fusion method, which still has certain limitations. To solve this problem, Tan, et al. [29] further proposed a complex bidirectional fusion network named Bi-FPN based on the original FPN, thus establishing a bottom-up and top-down bidirectional fusion path. This increases the semantic information in the shallow feature map while further enriching the morphological information of the deeper features. Its structure is shown in Fig. 3(c).

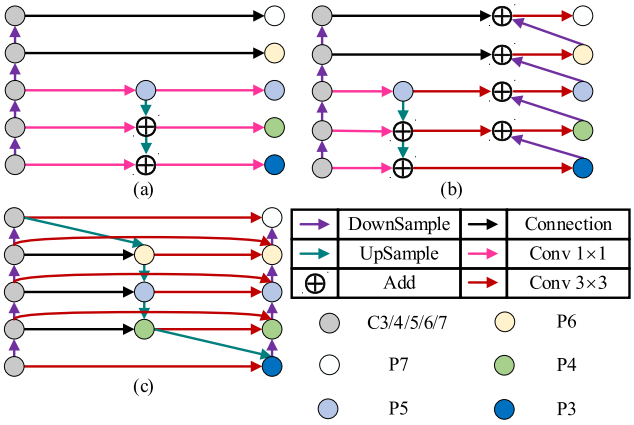


FIGURE 3. Structural of the FPN (a), PANet (b) and BiFPN (c).

C. SPPCSPC MODULE

The spatial pooling pyramid cross-stage partial convolution (SPPCSPC) module has been proposed for feature fusion at different scales, thus improving the ability of the network to perceive the image features [30]. However, in the SPPCSPC module, the introduction of convolution adds a large number of parameters. In this study, the SPPCSPC structure was introduced in YOLOv5 for feature fusion but with the difference that the group convolution was introduced in the SPPCSPC structure for parametric number compression. The comparison of the original SPPCSPC structure and the SPPCSPC structure with group convolution for parametric number compression is shown in Table 1. The architecture of the SPPCSPC module is shown in Fig. 4.

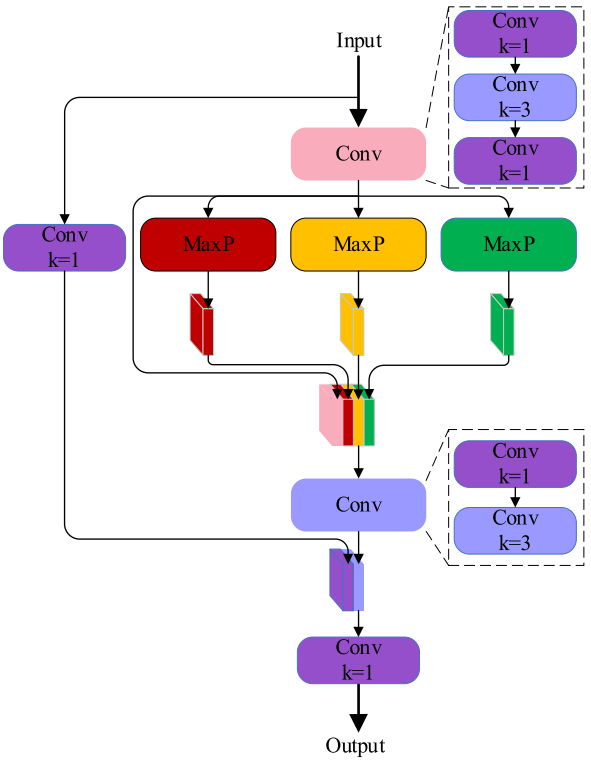


FIGURE 4. The architecture of the SPPCSPC module.

TABLE 1. Parameters of SPPCSPC and SPPCSPC-Group (based on YOLOv5L).

Methods	Parameters/M	GFLOPs
SPPF (YOLOv5l)	46.3	108.6
YOLOv5l+SPPCSPC	72.0	129.2
YOLOv5l+SPPCSPC-Group	50.8	112.2

It can be seen from Table 1 that the improved SPPCSPC module with group convolution reduces the parameters by ∼21.2 M and reduces the computation by ∼17 GFLOPs.

III. METHOD

A. OVERVIEW

In this work, YOLOv5 has been used as the baseline on which a bidirectional feature pyramid network, based on the HorNet plug-in, has been extended. In addition, an AED head to decouple the information for detecting each component of the catenary has been added to the network. The overall architecture of the network is shown in Fig. 5. Further, the feature map upsampling is done using the bilinear interpolation method.

B. BI-FPN WITH HORBLOCK

Based on the original FPN, HorBlock has been added for feature extraction to obtain higher detection accuracy with a small increase in parameters and computational cost. The structure of the proposed improved Bi-FPN with HorBlock, named Hor-Bi-FPN, is shown in Fig. 6, and the structure of HorBlock is shown in Fig. 7. This block compresses the number of parameters while ensuring good feature extraction capabilities. The output feature map for the neck section comes directly from HorBlock and is fed directly into the parsing head used for localization and classification tasks.

In HorBlock, the main process that is performed is the high-order gate convolution. The gated convolution and recursive design enable efficient, scalable, and translational higher-order spatial interactions to improve the utilization of effective information. The calculation process can be described as follows:

$$x = LayerNorm(Input) \tag{1}$$

$$y^{C/4}, x^{7C/4} = Split[Proj^{2C}_{1\times1}(x)] \tag{2}$$

$$x_1^{C/4}, x_2^{C/2}, x_3^{C} = Split[Dwconv^{2C}_{1\times1}(x^{7C/4})] \tag{3}$$

$$y^{C/2} = Proj^{C/2}_{1\times1}[x_1^{C/4} \times y^{C/4}] \tag{4}$$

$$y^{C} = Proj^{C}_{1\times1}(x_2^{C/2} \times y^{C/2}) \tag{5}$$

$$y^{C} = Proj^{C}_{1\times1}(x_3^{C} \times y^{C}) \tag{6}$$

where *Input* means input features, the *LayerNorm* is layer normalization, $x^{C/4}$, $y^{C/4}$ is output feature with channel *C/4*. *Split* is the channel split, and $Proj^{C}_{1\times1}$ represents the convolution with a kernel size of 1×1 and an output channel number of C.

The addition of a large-scale feature map size of 152 px × 152 px for small-sized components detection improves the ability of the network to sense them, thus avoiding the problem of small object loss due to the sampling process at high depth, consequently improving the detection rate of the network for catenary components.

C. ASYMMETRICALLY EFFICIENT DECOUPLED HEAD

In YOLOX, a form of predictive branch decoupled detection head has been proposed to solve the problem of convergence and the inability of YOLOv3 to decouple the detection head. In the predictive branching decoupled head, the category and position information are decoupled independently, thus improving its detection performance. However, it cannot be ignored that the performance of the network is potentially affected when the coupled detection head is decoupled. In addition, in the proposed decoupled head, the decoupled category and location information employ the same decoupled approach, which not only increases the number of parameters but also affects the effective acquisition of location information by the detection head, and thus affects the detection performance. Based on the above analysis, AED-Head has been proposed in the catenary detection method. Its structure is shown in Fig. 8.

In the classification branch, the input is concatenated together after convolution with the kernel size of 1×1. The output of a convolution is inputted into the feature enhancement module (FEM). The output of the FEM is summed with the convolved feature map and concatenated together with the input again. Lastly, the final classification result is obtained after convolution with a kernel size of 1×1.

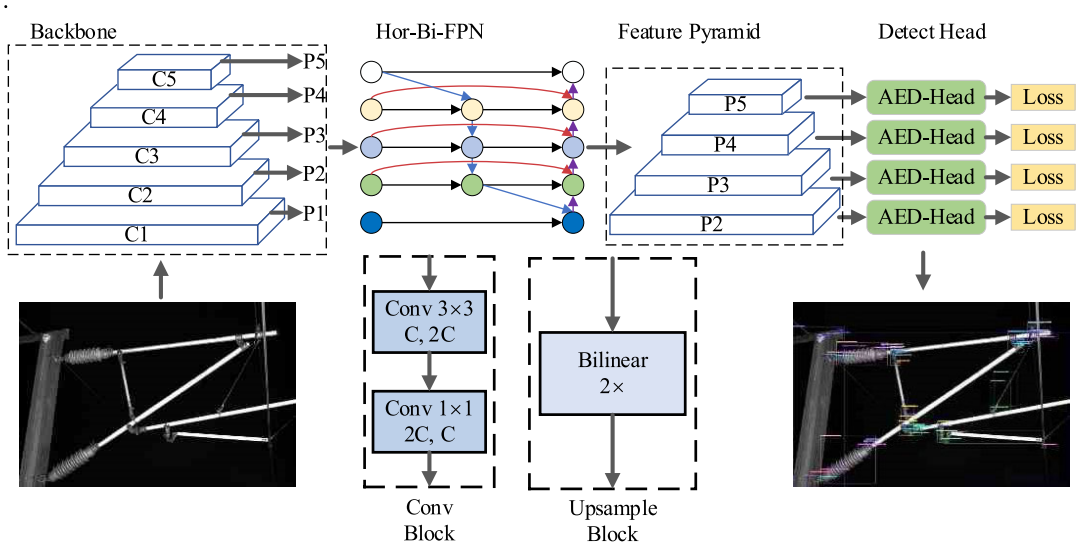


FIGURE 5. The architecture of the proposed AED-YOLO method.

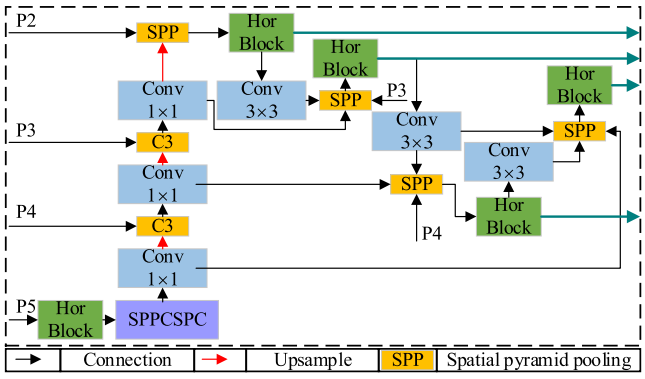


FIGURE 6. The architecture of the proposed Hor-Bi-FPN.

The classification branch calculation process can be described as:

$$x_1 = Conv^{C}_{1\times1}(Input) \tag{7}$$

$$Cls = Conv^{C}_{1\times1}\left\{Cat\left[\begin{array}{l} Input, \\ (FEM\ (Cat\ (Input, x_1)) \oplus x_1)\end{array}\right]\right\} \tag{8}$$

Where Input means input features, and *Cls* is the classification prediction results.

In FEM, the calculation process can be described as follows:

$$Output = Conv^{C}_{1\times1}\ (Input) \oplus Conv^{C}_{3\times3,6}\ (Input) \tag{9}$$

where $Conv^{C}_{3\times3,6}$ represents the convolution with a kernel size of 1×1, the dilated ratio is 6, and the output channel number is C. $\oplus$ is the add operation and *Cat* represents concatenate operation.

In the location branch, the output of a convolution is given as input to the FEM. The output of the FEM is summed using the convolved feature map. Finally, the result is obtained after convolution with a kernel size of 1×1. The calculation process can be described as follows:

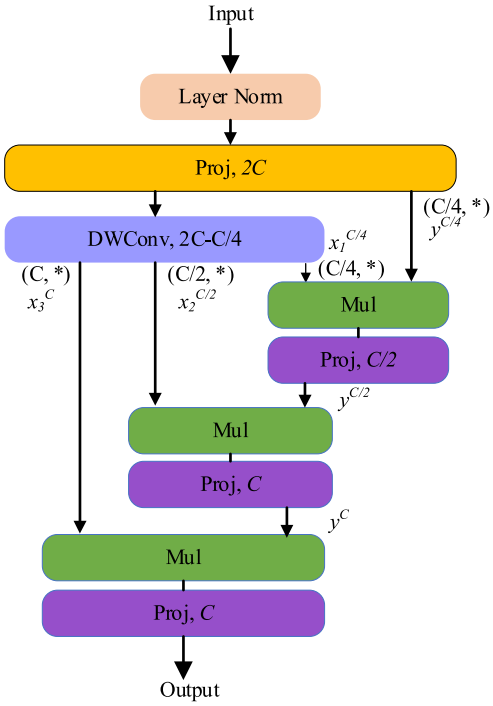


FIGURE 7. The architecture of the HorBlock.

$$Reg = Conv^{4}_{1\times1}\left\{Cat\left[\begin{array}{l} Input, \\ (FEM\ (x_1) \oplus x_1)\end{array}\right]\right\} \tag{10}$$

$$IoU = Conv^{1}_{1\times1}\left\{Cat\left[\begin{array}{l} input, \\ (FEM\ (x_1) \oplus x_1)\end{array}\right]\right\} \tag{11}$$

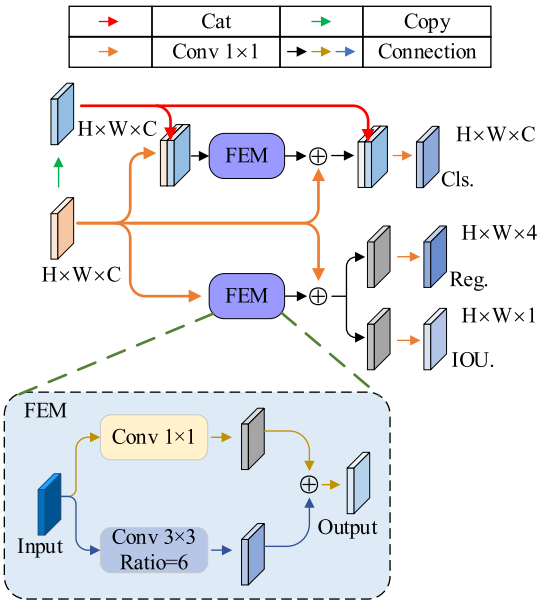


FIGURE 8. The architecture of the proposed AED-Head.

TABLE 2. Parameters of AED-Head and decoupled head (based on YOLOv5L).

Methods	Parameters/M	GFLOPs
SPPF (YOLOv5l)	46.3	108.6
YOLOv5l+Decoupled Head	53.8	149.6
YOLOv5l+AED-Head	50.4	123.9

Where Input means input features, *Reg* is the regression prediction results. *IoU* means the location accuracy.

This improvement not only speeds up the convergence of the network but also reduces the number of parameters. The parameters of the AED-Head and the decoupled head are shown in Table 2.

It can be seen from Table 2 that the proposed AED-Head reduces the parameters by 3.4 M and the computation by 25.7 GFLOPs.

The asymmetrical structure of the decoupled head improves the accuracy of the classification of different-sized components and thus ensures the accuracy of locating them. The decoupled head of the size of 152 px ×152 px improves locating the small components such as fastener nuts.

IV. EXPERIMENTS AND RESULTS

A. DATASET

Data from 300km/h High-Speed Rail China High-Speed Rail 400km test data. The catenary images are acquired from the high-speed railway catenary inspection vehicle and manually labeled. The dataset for the 40 catenary components is shown in Table 3.

TABLE 3. The distribution of the dataset.

Type	Train Dataset	Val Dataset	All
Numbers	4703	523	5226

TABLE 4. Parameter setting.

Parameters	Value
Image size	768
Batch size	4
Epochs	300
Initial learning rate	0.01
Momentum	0.937

TABLE 5. Results were obtained by applying the different methods to the catenary dataset.

Methods	mAP/%
YOLOv3	91.7
YOLOv5	91.2
TPH-YOLO	86.7
SSD	76.2
Faster RCNN	52.1
Proposed method	**93.5**

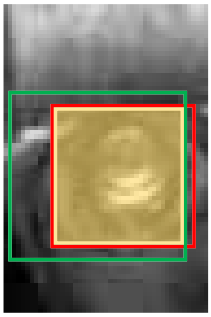


FIGURE 9. Diagram of IoU.

B. IMPLEMENTS DETAILS

1) PARAMETER SETTINGS

The initialization parameters of the backbone network were transformed from the pre-trained YOLOv5l network. Then, the initialization weights of the other convolution layers were initialized using the "Kaiming" initialization method. The proposed AED-YOLO model was trained using the mini-batch stochastic gradient descent method with a batch size of 4. Other parameters were set as shown in Table 4.

2) COMPUTATION PLATFORM

The proposed method was implemented in Python and PyTorch and was run on Ubuntu 18.04 and a computing platform with GeForce RTX Titan Xp.

TABLE 6. Results of the different methods for some small and medium-sized components.

Components	YOLOv3/%	YOLOv5/%	TPH-YOLO/%	SSD/%	Faster RCNN/%	Proposed method/%
Screw_pin	80.0	80.1	75.5	47.9	80.9	**82.9**
Insulator_flow_porcelain_hook	66.3	82.8	39.8	83.3	50.0	**99.5**
Bracing_wire	47.4	49.7	32.2	34.9	33.9	**53.6**
Screw_cotter_nut	97.9	97.3	97.4	46.1	0.00	**98.3**
Screw_cotter_nonut	92.6	92.3	97.3	86.2	40.8	**97.8**
Screw_pin_fix	76.6	72.8	76.0	65.9	21.8	**86.3**
Screw_nut_wire	65.8	59.8	60.9	14.4	0.00	**79.6**
Insulator_rod_composite_connector	96.0	98.6	96.7	85.6	18.2	**99.3**

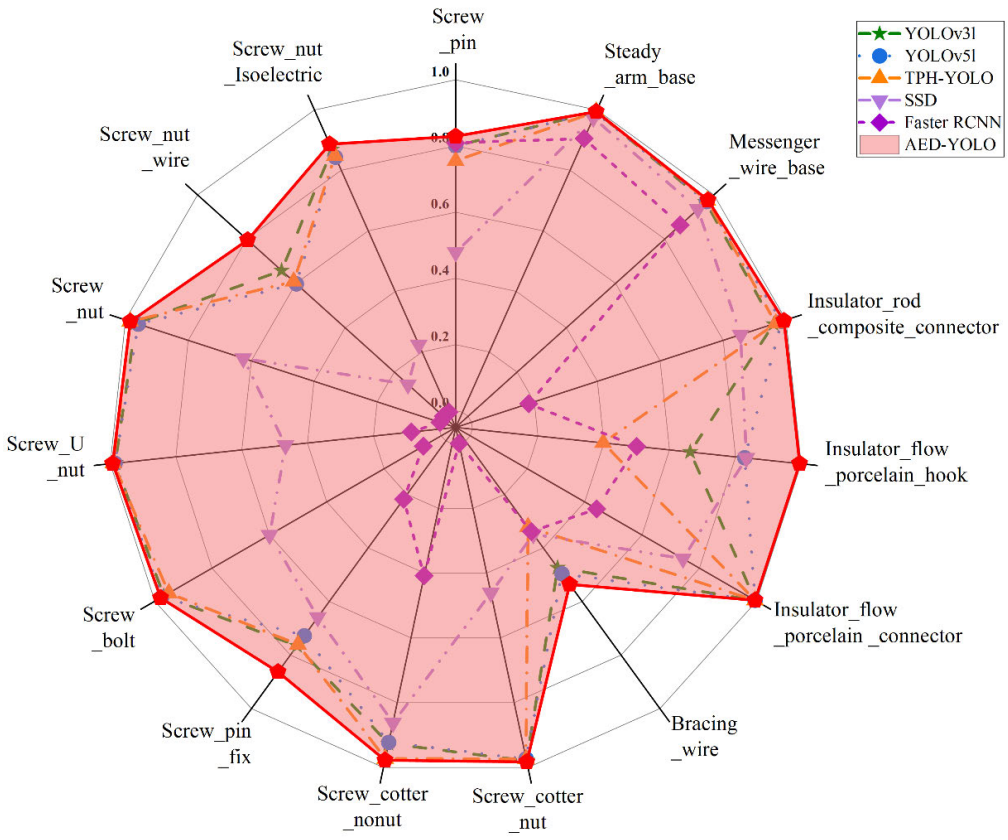


FIGURE 10. The average detection accuracy for some small components.

C. EVALUATION METRIC

In the field of object detection, the intersection over union (IoU) ratio and confidence are the most used evaluation metrics. The IoU metric is used to evaluate the degree of coincidence between two regions and is defined as:

$$IOU = \frac{I_{area}}{U_{area}} \tag{12}$$

where I_{area} is the intersection area, and U_{area} is the union area. As shown in Fig. 9, the red bounding box is the real bounding box containing the detection object, the green bounding box is the prediction box output by the algorithm, and the overlapping area of the two is the yellow bounding box area. Then the IoU of the target recognition is expressed as follows:

$$IOU = \frac{B_1 \cap B_2}{B_1 \cup B_2} \tag{13}$$

where B_1 is the area of the red box and B_2 is the area of the green box.

In addition to IoU, the following four evaluation metrics were used for evaluating the test results of detection: accuracy (Acc), precision (P), recall (R), and average precision (AP).

$$P = \frac{TP}{TP + FP} \times 100\% \tag{14}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{15}$$

$$AP = \frac{P + R}{2} \times 100\% \tag{16}$$

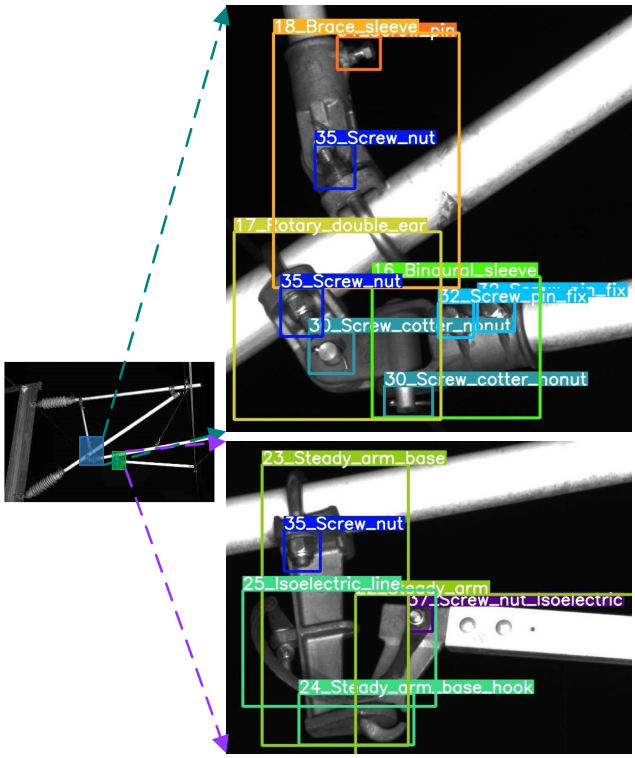


FIGURE 11. Locating effect of some small components.

where *TP* is the predicted defect number of IoU > 0.5, *FP* is the predicted defect number of IoU $\leqslant 0.5$, and *FN* indicates the missed inspection number. The mAP is the average APs of all categories.

D. RESULTS

In the experimental part, classical methods such as YOLO3, SSD, Faster RCNN, and TPH-YOLO were selected for comparing their performance with that of the proposed method, and the results thus obtained are shown in Table 5.

The data in Table 5 leads to the conclusion that the proposed AED-YOLO method achieves the highest average accuracy of 93.5% when applied to the catenary dataset of 40 components. This exceeds the results of YOLOv3 by 1.8% and YOLOv5 by 2.3%. Compared to the other methods, the mean average precision is also greatly improved. The detection results for some of the small-sized components are shown in Fig. 10. It can be seen from the figure that compared to the other methods, the proposed method achieves high mean precision for 14 small-sized components. In particular, a significant improvement in the detection precision for screw_nut_isoeletric, screw_pin_fix, and insulator_flow_porcelain_hook is observed. The results of the inspection of some of the key components are shown in Fig. 11. As can be seen from the figure, the proposed method exhibits good detection performance for small-sized components in critical areas.

TABLE 7. Results of the ablation study.

Methods	mAP/%
YOLOv5l	91.2
YOLOv5l+Hor-Bi-FPN	92.3
YOLOv5l+Hor-Bi-FPN+AED-Head	**93.5**

TABLE 8. Parameter comparison.

Methods	Parameters/M	GFLOPs
SPPF (YOLOv5l)	46.3	108.6
YOLOv5l+Decoupled Head	53.8	149.6
YOLOv5l+AED-Head	**50.4**	**123.9**

In catenary systems, a large proportion of the components are small, and the inspection of small components is a top priority. Table 6 gives the inspection results for eight small-sized components. It can be seen from the results that the detection accuracy of the proposed AED-YOLO method is high, compared to the other methods, for the eight small-sized components, indicating that the proposed method exhibits better detection and positioning capability for small-sized components.

E. ABLATION STUDY

In the ablative experimental test, YOLOv5l was selected as the baseline network and combined with other modifications to perform a series of ablation experiments. The results of the experiments are given in Table 7.

It can be seen from the table that compared to the baseline, the mAP of the YOLOv5l + Hor-Bi-FPN method is improved by 1.1%, and thus this method will be more accurate in locating catenary components. The detection accuracy can be further improved by 1.2% by adding an AED-Head. Compared to the original YOLOv5, it can be seen from Table 6 that the proposed method can improve the detection accuracy by more than 10% when detecting Insulator_flow_porcelain_hook, Screw_pin_fix, and Screw_nut_wire. The accuracy of several other components is also observed to be improved by varying degrees. This shows that the method proposed in this work can efficiently improve the positioning accuracy of small- and medium-sized components of the catenary system.

F. DISCUSSION

Although the method proposed in this work is effective in improving the positioning accuracy of small-sized components in contact network systems, it also introduces a large number of parameters. Table 8 shows the changes in the number of parameters and the number of calculations before and after the improvements.

As can be seen from the table, the method proposed in this work increases the number of parameters by 19.6 M and the number of calculations by 90.5 GFLOPs, compared to the original YOLOv5. As a result, it is slower in providing results.

V. CONCLUSION

In this work, a precision location network for catenary components has been proposed. In conjunction with HorNet, an improved version of the bidirectional feature pyramid network has been proposed for bidirectional feature fusion to increase the perceptibility of the small-sized components. Secondly, an AED head has been introduced to provide different decoupled methods for performing the tasks of classification and localization using an asymmetrical decoupled structure. Experiments demonstrate that the proposed approach is advanced compared to the existing approaches in terms of detecting 40 types of components located in a high-speed railway catenary system and exhibits an mAP of 93.5%. In particular, for locating small-sized components, the detection precision is observed to be improved by varying degrees; thus, providing superior locating detection capability for small-sized components.

In future research, we plan to further improve the detection performance and optimize the framework to accelerate the proposed approach to meet real-time and high-accuracy requirements. In addition, integrating more efficient and less parametric detection methods will also be an aim of our next research work.

REFERENCES

[1] J. Zhong, "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.

[2] Y. Chen, B. Song, Y. Zeng, X. Du, and M. Guizani, "A deep learning-based approach for fault diagnosis of current-carrying ring in catenary system," *Neural Comput. Appl.*, pp. 1–13, Jul. 2021, doi: 10.1007/s00521-021-06280-4.

[3] C. Cheng, J. Wang, H. Chen, Z. Chen, H. Luo, and P. Xie, "A review of intelligent fault diagnosis for high-speed trains: Qualitative approaches," *Entropy*, vol. 23, no. 1, p. 1, Dec. 2020, doi: 10.3390/e23010001.

[4] J. Zhong, Z. Liu, Z. Han, Y. Han, and W. Zhang, "A CNN-based defect inspection method for catenary split pins in high-speed railway," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2849–2860, Aug. 2019, doi: 10.1109/tim.2018.2871353.

[5] Y. Han, Z. Liu, Y. Lyu, K. Liu, C. Li, and W. Zhang, "Deep learning-based visual ensemble method for high-speed railway catenary clevis fracture detection," *Neurocomputing*, vol. 396, pp. 556–568, Apr. 2019, doi: 10.1016/j.neucom.2018.10.107.

[6] D. Cheng, Q. Wang, F. Wu, X. Wang, Y. Niu, and Z. Ye, "Research on intelligent detection of state of catenary puller bolt based on deep learning," *J. China Railway Soc.*, vol. 43, no. 11, pp. 52–60, 2021.

[7] B. Li and F. Cui, "Inception module and deep residual shrinkage network-based arc fault detection method for pantograph-catenary systems," *J. Power Electron.*, vol. 22, no. 6, pp. 991–1000, Jun. 2022, doi: 10.1007/s43236-022-00415-z.

[8] Y. Han, Z. Liu, X. Geng, and J. P. Zhong, "Fracture detection of ear pieces in catenary support devices of high-speed railway based on HOG features and two-dimensional Gabor transform," *J. China Railway Soc.*, vol. 39, no. 2, pp. 52–57, 2017.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 770–778.

[11] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++ better real-time instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 1108–1121, Feb. 2022, doi: 10.1109/TPAMI.2020.3014297.

[12] P. Tan, X.-F. Li, J. Ding, Z.-S. Cui, J.-E. Ma, Y.-L. Sun, B.-Q. Huang, and Y.-T. Fang, "Mask R-CNN and multifeature clustering model for catenary insulator recognition and defect detection," *J. Zhejiang Univ.-Sci. A*, vol. 23, no. 9, pp. 745–756, Sep. 2022, doi: 10.1631/jzus.A2100494.

[13] K. He, G. Georgia, D. Piotr, and G. Ross, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2017, doi: 10.1109/TPAMI.2018.2844175.

[14] P. Tan, X. Li, Z. Wu, J. Ding, J. Ma, Y. Chen, Y. Fang, and Y. Ning, "Multialgorithm fusion image processing for high speed railway dropper failure-defect detection," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 7, pp. 4466–4478, Jul. 2021, doi: 10.1109/tsmc.2019.2938684.

[15] 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.

[16] Y. Chen, B. Song, Y. Zeng, X. Du, and M. Guizani, "Fault diagnosis based on deep learning for current-carrying ring of catenary system in sustainable railway transportation," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106907, doi: 10.1016/j.asoc.2020.106907.

[17] X. Cheng and J. Yu, "RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/tim.2020.3040485.

[18] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S.-N. Lim, and J. Lu, "HorNet: Efficient high-order spatial interactions with recursive gated convolutions," 2022, *arXiv:2207.14284*.

[19] 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.

[20] 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.

[21] X. Zhou, H. Fang, X. Fei, R. Shi, and J. Zhang, "Edge-aware multi-level interactive network for salient object detection of strip steel surface defects," *IEEE Access*, vol. 9, pp. 149465–149476, 2021, doi: 10.1109/access.2021.3124814.

[22] W. Wu and Q. Li, "Machine vision inspection of electrical connectors based on improved YOLO v3," *IEEE Access*, vol. 8, pp. 166184–166196, 2020, doi: 10.1109/access.2020.3022405.

[23] 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, doi: 10.1109/CVPR.2016.91.

[24] 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, doi: 10.1109/ICCVW54120.2021.00312.

[25] Y. Zhang, J. H. Han, Y. W. Kwon, and Y. S. Moon, "A new architecture of feature pyramid network for object detection," in *Proc. IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2020, pp. 1224–1228, doi: 10.1109/ICCC51575.2020.9345302.

[26] G. Yang, Z. Wang, and S. Zhuang, "PFF-FPN: A parallel feature fusion module based on FPN in pedestrian detection," in *Proc. Int. Conf. Comput. Eng. Artif. Intell. (ICCEAI)*, Aug. 2021, pp. 377–381, doi: 10.1109/ICCEAI52939.2021.00075.

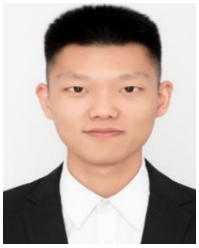
[27] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2117–2125.

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

[29] 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. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.

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

SHUAI XU received the B.Sc. degree from Dalian Jiaotong University, China, in 2005, and the M.S. degree in traffic information engineering and control from the School of Electrical Engineering, Dalian Jiaotong University, in 2010, where he is currently pursuing the Ph.D. degree with the College of Mechanical Engineering. His research interests include the intelligent detection of catenary for high-speed railways and deep learning.



QIANG FENG received the B.S. degree from Qingdao Agricultural University. He is currently pursuing the master's degree with the College of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University, China. His research interests include computer vision, intelligent measurement, and image processing.

JIYOU FEI received the B.Sc. degree from The Second Artillery Engineering College, in 1987, the M.Sc. degree from the Shaanxi University of Science and Technology, in 1998, and the Ph.D. degree from Xi'an Jiaotong University, in 2003. He is currently a Professor with Dalian Jiaotong University. His research interests include image processing, computer vision, and non-contact precision measurement technology.

GENG ZHAO is currently pursuing the Ph.D. degree with the College of Mechanical Engineering, Dalian Jiaotong University, China. His research interests include technical transformation and intelligent measurement.

XIAODONG LIU received the B.Eng. degree, the M.E. degree in mechatronic engineering, and the Ph.D. degree in mechanical manufacturing and automation from Dalian Jiaotong University, in 2005, 2008, and 2013, respectively. She is currently a Senior Laboratory Master with the College of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University. Her research interests include intelligent measurement and image processing.

HUA LI received the B.S. degree from the Nanjing University of Science and Technology, in 1996, and the M.S. and Ph.D. degrees from Pukyong National University, South Korea, in 2005 and 2008, respectively. She is currently a Professor with Dalian Jiaotong University. Her research interests include computer vision, deep learning, and non-contact precision measurement technology.

CHANG LU received the B.Sc. and M.S. degrees in mechanical engineering from Dalian Jiaotong University, in 2013 and 2015, respectively. She is currently a Lecturer with the Army Artillery and Air Defense Academy Sergeant School of PLA. Her research interest includes high-speed precision machining.

QI YANG received the B.E. degree in information security from the Software College, Northeastern University. He is currently pursuing the degree in electronic information technology with the College of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University. His research interests include machine learning and computer vision.