Received November 17, 2021, accepted January 25, 2022, date of publication February 1, 2022, date of current version February 11, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3148182

Detection of Locomotive Signal Lights and Pedestrians on Railway Tracks Using Improved YOLOv4

HUANMIN WANG[1,2,3], HUAYAN PEI[4], AND JIANBAI ZHANG[5]

[1]Mechatronics T&R Institute, Lanzhou Jiaotong University, Lanzhou 730070, China
[2]Engineering Technology Center for Informatization of Logistics & Transport Equipment, Lanzhou 730070, China
[3]Gansu Provincial Industry Technology Center of Logistics & Transport Equipment, Lanzhou 730070, China
[4]School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
[5]Lanzhou Railway Bureau Group Company Ltd., China Railway, Lanzhou 730000, China

Corresponding author: Huayan Pei (pei123com@126.com)

This work was supported by the Scientific Research Project of Universities, Gansu, China, under Grant 2018C-10.

ABSTRACT Intelligent detection of locomotive signal lights and pedestrians on railway tracks is of great significance to the safe operation of locomotives, especially under bad illumination conditions. Due to the highly complicated operation environment of railway locomotives, it is relatively difficult to apply deep neural network-based object detection methods in the recognition of locomotive signal lights and railway pedestrians. This work, for the first time, proposes a real-time detection method based on improved YOLOv4 to recognize locomotive signal lights and railway pedestrians. In our method, the Region-of-Interest is introduced into the original network of YOLOv4 to improve the detection precision of pedestrians on railway tracks. Importantly, we establish a dataset called detection of locomotive signal lights and railway pedestrians (DLSLRP), which is dedicated to the training, testing, and validation of related convolutional neural networks. We evaluated the proposed detector on DLSLRP dataset, the experimental results suggest that our method can detect locomotive signal lights and railway pedestrians with high speed and accuracy under different illumination conditions. The mAP reaches 93.52%, and the detection speed achieves 25 FPS.

INDEX TERMS Object detection, locomotive signal lights, pedestrians on railway tracks, improved YOLOv4, convolutional neural network.

I. INTRODUCTION

The goal of object detection is to identify the presence of an object instance from given categories (such as dogs, cats, humans, or cars) in a image or video. If present, the spatial location, and scale of every object instance are required to return via a bounding box [1], [2]. The traditional object detection methods mainly include optical flow [3], Hough transform methods [4], background subtraction [5], and frame difference [6], etc. Nevertheless, these algorithms have shortcomings such as limited application scenarios, narrow classification, and excessive hand-operated intervention. Moreover, they have disadvantages in robustness, which results in a deficiency in detection results and generalization ability.

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

With the vigorous development of deep learning, object detection algorithms based on deep learning have received considerable attention around the world, which generally have better performance compared to the conventional object recognition methods. Furthermore, deep learning-based object detection algorithms have a wider range of application scenarios, including intelligent transportation systems [7], medical object detection [8], military object detection [9], etc. To be specific, they can be used in pedestrians detection [10], face detection [11], [12], and skeleton detection [13], etc.

The current object detection methods based on deep learning can be primarily classified into two categories. One is the two-stage algorithms, which are mainly represented by the R-CNN series [14]–[16]. The other is the single-stage methods, the typical representatives of which are Single Shot MultiBox Detector (SSD) [17] and You Only Look Once (YOLO) series [18]–[21]. The primary difference between the two is that a great number of preselected boxes are required in the two-stage algorithms. As pioneers, AlexNet *et al.* [22] are the first to apply convolutional neural network (CNN) in image classification, who won the championship in the ILSVRC-2012 competition. The neural network that AlexNet *et al.* used in the ILSVRC-2012 consists of five convolutional layers. However, the performance of their network will experience degradation if a single convolutional layer is removed. In 2014, Girshick *et al.* combined region proposals with CNNs and presented a scalable object detection method called Regions with CNN features (R-CNN) [14], in which the selective search algorithm is employed to calculate feature similarity of adjacent regions. It is worth noting that in R-CNN, the last spatial pooling layer in image classification networks is converted into the Region-of-Interest (ROI) pooling layer in object detection networks. Nevertheless, R-CNN has to crop and scale the image area to the same size. To overcome this shortcoming, Girshick proposed a Fast Region-based Convolutional Network method (Fast R-CNN) [15] for object detection, which can be viewed as an update of R-CNN. Compared with the previous methods, Fast R-CNN not only improves the training and testing speed but also increases the detection accuracy. Whereas, the problem of region proposal computation is a bottleneck. To overcome this disadvantage, Faster R-CNN [16] introduces Region Proposal Network (RPN) into R-CNN, and merges RPN and Fast R-CNN into a single network by sharing their convolutional features. However, in the region-based object detection methods such as Fast/Faster R-CNN [15], [16], detectors utilize a costly per-region subnetwork hundreds of times, which indicates that if there are 2000 ROIs, the cost of running time is extremely high. To achieve the goal of acceleration, Dai *et al.* presented Region-based Fully Convolutional Networks (R-FCN) [23] for accurate and efficient object detection. R-FCN significantly reduces the amount of work required for each ROI by removing the fully-connected layers.

The single-stage algorithms are mainly represented by SSD [17] and YOLO series [18]–[21], which both use a single deep neural network to detect objects in images. Compared to algorithms requiring object proposal, SSD is simple since it completely removes proposal generation and feature resampling, and encapsulates all calculation in a single network [19]. The major advantages of SSD are fast detection speed and accurate boundary positioning effect. However, it is difficult for SSD to detect small-sized objects since it uses multi-level feature classification. YOLO [18] converts object detection into a regression problem, and adopts a single neural network to predict both bounding boxes and class probabilities for objects directly from full images in one evaluation. Thus, YOLO can be optimized end-to-end directly on detection effect. Whereas, it has limitations such as low recall rate, inaccurate location, and poor detection performance on small-sized objects. YOLOv2 [19] is an improved model of YOLO, it can run at varying sizes and offer a tradeoff between speed and accuracy by using a multi-scale training method. YOLOv3 [20] can be viewed as an incremental improvement of YOLO, which is a little bigger but more accurate and still fast. YOLOv4 [21] combines new features such as Cross-Stage-Partial-connections (CSP), Weighted-Residual-Connections (WRC), Self-adversarial-training (SAT), Cross mini-Batch Normalization (CmBN), and Mish-activation to achieve state-of-the-art results and to improve the accuracy of both classifier and detector. Convolutional neural network and YOLOv4 have been used in the recognition of a wide variety of objects. Chen et. al proposed an improved YOLOv4-based detector for the recognition of citrus in an orchard environment [24]. Chernov et. al presented a system for the detection of obstacles on railway on the basis of CNN [25]. Cheng and Zhang proposed a YOLOv4-based end-to-end detection method for the recognition of flowers [26]. Deng et. al presented a C-YOLOv4 network for the detection of iron material cracks and established an iron material polarization image crack dataset [27].

Over the past decade, various object detection algorithms based on deep learning have been proposed to recognize traffic lights for vehicles [28]–[31]. However, due to the extremely complex operation environment of railway locomotives, it is difficult to utilize these detection methods in the recognition of signal light for locomotives. Noteworthy, there is no publicly available dataset for the training and testing of related convolutional neural networks. While, locomotive signal lights play a crucial role in instructing locomotives entering and leaving train stations, controlling the operation speed of locomotives, and instructing the shunting operations, etc. Hence, real-time, rapid and accurate detection of signal lights is of great significance to the safe operation of locomotives. In the traditional operation scenarios of locomotives, the detection of signal lights primarily depends on the observation of drivers, who are prone to make misjudgments, especially under bad illumination conditions, and cause serious railway traffic accidents. Moreover, the conventional detection methods largely rely on the recognition of color features [32]–[34] and shape features [35]–[38] of locomotive signal lights, or the combination of the two. The former is drastically affected by illumination conditions, and is easy to make misjudgment in a complex background environment. Besides, it is difficult for the latter to detect small-sized signal lights, and does not have high levels of stability.

The main contributions of this work are summarized as following:

1) To realize detection of locomotive signal lights and pedestrians on railway tracks, this paper, for the first time, proposes an improved YOLOv4-based real-time detection method, which is able to recognize locomotive signal lights of different colors under bad illumination conditions and railway pedestrians.

2) To improve the detection accuracy of railway pedestrians, we incorporate the Region-of-Interest into the original network of YOLOv4.

3) We establish a comprehensive dataset called detection of locomotive signal lights and railway pedestrians (DLSLRP), which can be utilized in the training, testing,

and validation of deep neural network-based detection methods.

This article is organized as follows. The framework of our algorithm and the establishment of DLSLRP dataset are given in Section II. The experiments and discussions are presented in Section III. Conclusions are drawn in Section IV.

II. THE PROPOSED METHOD

As mentioned in the "Introduction", deep learning-based object detection methods can be classified into two-stage algorithms and single-stage methods. The typical representatives of the former are R-CNN series [14]–[16], it is necessary for which to generate region proposal, after that features of the image can be extracted, then the classification is predicted. Compared to the latter, the former have better performance on detection accuracy. Nevertheless, since the former have longer running time than the latter, they cannot meet the real-time application requirements of locomotive operation.

The single-stage algorithms are primarily represented by SSD [17] and YOLO series [18]–[21], both of which directly conduct convolutional operation on input images and output classification results, then the position of object in images are located. The former uses multi-scale feature layers for object detection, and combines the anchor mechanism of Faster R-CNN with the idea of YOLO grid. The latter utilizes multi-scale feature mapping to perform feature extraction and regression analysis for each position in the entire image. Since all feature data of SSD come from the last layer of feature pyramid, thus the number of its parameters is less than that of YOLOv4. Whereas, compared to SSD, YOLOv4 is better in terms of detecting small-sized objects. In other words, SSD cannot meet the detection requirements of locomotive operation.

Compared with other CNN-based object detection methods, YOLOv4 employs CSPDarknet53 as the backbone network, and introduces a feature pyramid network and the data augmentation algorithm. Importantly, the high-level features are merged with the low-level features by up-sampling in YOLOv4, which makes the feature extraction more comprehensive. Besides, in YOLOv4, the input image is firstly divided into $S \times S$ grids, each of which is responsible for detecting the category of a box, next the technique of non-maximum suppression is used, then an optimal classification result can be obtained. The detection speed of YOLOv4 reaches 45 FPS. Due to the above advantages, YOLOv4 is able to fulfill the real-time detection requirements of locomotive operation. Thus, we choose it as the basic model of our detector.

A. THE ARCHITECTURE

In this subsection, we mainly describe the network architecture of YOLOv4 that is employed in our method in detail. A modern object detector commonly contains two parts, a pre-trained backbone, and a head that is used to predict classes and bounding boxes of objects. The head part is usually classified into single-stage object detectors and two-stage object detectors. To be specific, an ordinary object detector generally consists of several parts [21]: Input, Backbones, Neck and Heads, as shown in Figure 1. As for single-stage object detectors, the head part uses Dense Prediction.

To address the problem of real-time detection of locomotive signal lights and pedestrians on railway tracks, and to obtain rapid and accurate detection results under bad illumination conditions, we utilize YOLOv4 as the basic model of our detection method, the architecture of which is presented in Figure 2. Specifically, as to the backbone network part, YOLOv4 adopts CSPDarknet53, which is better compared to CSPResNext50 in terms of detecting objects [21]. CSPDarknet53 is established on the basis of Darknet53 of YOLOv3, and learns from the experience of Cross Stage Partial Network (CSPNet) [39]. CSPNet includes five Cross-Stage-Partial-connections (CSP), which enhance the learning ability of CNN and maintain sufficient accuracy while being lightweightening. As to the neck part, YOLOv4 employs Spatial Pyramid Pooling (SPP) [40] and Path Aggregation Network (PANet) [41]. The former is an additional module of the neck part, and can greatly increase the receptive field, extract the most significant context features and hardly reduce the operation speed of network. The latter is a feature fusion module of the neck section, and is mainly used as the approach of parameter aggregation from distinct backbone levels for different detector levels.

B. THE ESTABLISHMENT OF DATASET

For the application of convolutional neural network, dataset plays an essential role in the training, testing, and validation of CNN. At present, there is no publicly available dataset can be used in the detection of locomotive signal lights and railway pedestrians. Hence, we establish a comprehensive dataset called detection of locomotive signal lights and railway pedestrians (DLSLRP). The Lanzhou Railway Bureau Group Co., Ltd of China Railway provides great help and support for the collection of sample images during the establishment of DLSLRP dataset. A large number of high-resolution images closely associated with locomotive signal lights and railway pedestrians are captured by cameras mounted in locomotives. A total number of 50k images with distinct backgrounds, under different illumination conditions are captured in real-time locomotive operation scenarios, among which 40k images are used in the training phase, and 10k images are utilized in the testing phase. The sample images of locomotive signal lights and railway pedestrians in DLSLRP dataset are illustrated in Figure 3 and Figure 4, respectively.

It is worth stressing that in DLSLRP dataset, the images are generally labeled much larger than the boundaries of real objects. It is mainly because of railway tracks make the recognition characteristic of locomotive signal lights more obvious, which can dramatically improve the detection accuracy. That is, to avoid detecting other unrelated signal lights, especially the traffic lights for vehicles, it is necessary to simultaneously recognize the railway tracks next to locomotive signal lights.

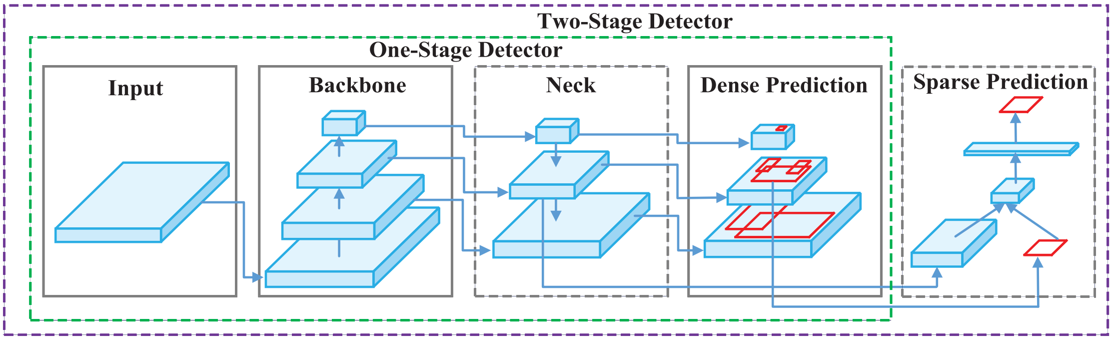


FIGURE 1. Object detector [21].

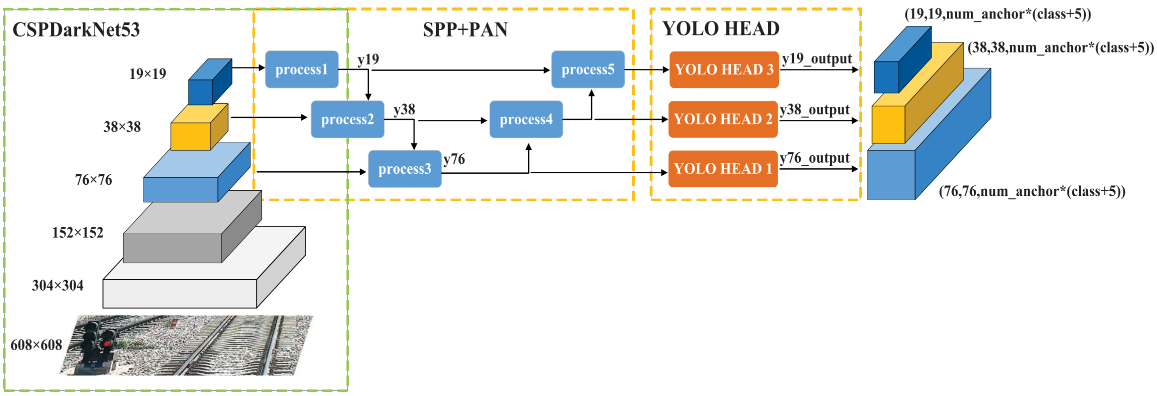


FIGURE 2. The architecture of YOLOv4.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

The detail list of hardware and software utilized in our work are as follows: NVIDIA GeForce GTX 2080 GPU, TensorFlow framework version 1.13, Windows 10 operation System, CUDA version 10, and CUDNN version 7.

A. DETECTION RESULTS OF LOCOMOTIVE SIGNAL LIGHTS

We verified the performance of our detector on DLSLRP dataset, the experimental results of locomotive signal lights detection are shown in Figure 5.

These results reveal that, irrespective of day or night, the proposed method is able to detect blue, red and white locomotive signal lights under bad illumination conditions. Most importantly, our detector only recognize the signal lights located at the left side of railway tracks, which satisfies the location requirements of signal lights detection for locomotives. In addition, we find that in the detection range of 1.5 kilometers, our method can detect and recognize locomotive signal lights with high accuracy at a fast speed, especially the small-sized locomotive signal lights.

B. DETECTION RESULTS OF PEDESTRIANS ON RAILWAY TRACKS

1) DETECTION RESULTS OF THE YOLOv4-BASED METHOD

During the operation of locomotives, pedestrians illegally crossing the railway tracks are one of the leading sources of casualties caused by railway traffic accidents. In some certain emergency cases, locomotives are forced to stop to prevent from hitting the pedestrians on railway tracks. Hence, rapid and accurate pre-detection of railway pedestrians plays a significant role in the safe operation of locomotives. Our detection method not only can recognize locomotive signal lights of different colors but is also able to detect pedestrians on railway tracks.

Figure 6 illustrates the detection results of railway pedestrians on DLSLRP dataset. Experimental results show that within the detection range of 1.5 kilometers, our method is able to recognize railway pedestrians with high detection speed and precision.

2) DETECTION RESULTS OF OUR METHOD

As shown in Figure 6, the pedestrians do not locate on railway tracks are also detected. Thus, to improve the detection accuracy of railway pedestrians, we integrate the Region-of-Interest (ROI) into the original network of YOLOv4, which greatly enhances the recognition ability of our detector. It is worth noting that we classify the area of railway tracks into hazardous and non-hazardous. Our method only detects pedestrians in the hazardous area by recognizing the ROI.

To further validate the detection effect of our method on railway pedestrians, we make a comparison between the YOLOv4-based method and our detector, the results are shown in Figure 7.

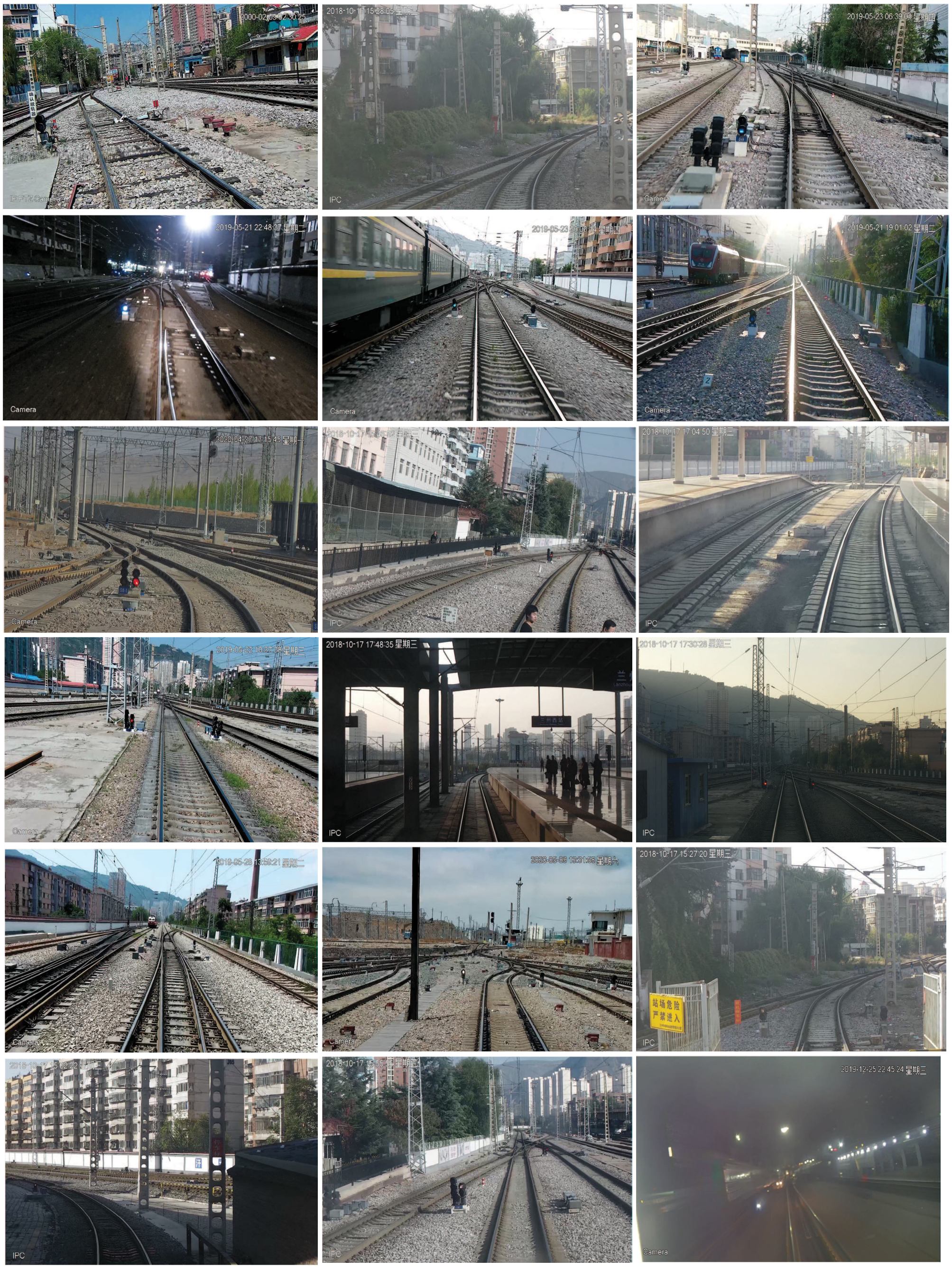


FIGURE 3. Sample images of locomotive signal lights in DLSLRP dataset.

As can be seen in Figure 7, the method based on YOLOv4 detects pedestrians in both hazardous and non-hazardous areas. In contrast, our method only recognizes the pedestrians in hazardous area. These findings suggest that, compared to YOLOv4-based detection method, our detector significantly improves the recognition precision of railway pedestrians, which reduces the false alarms for the pedestrians in non-hazardous area. That is to say, the proposed method can better meet the requirements of railway traffic safety for pedestrians detection.

C. PERFORMANCE EVALUATION AND DISCUSSIONS

1) THE mAP, PRECISION, AND RECALL OF OUR METHOD ON DISTINCT DETECTION CLASSES

To further evaluate the performance of our method, the mAP, precision, and recall of detection on blue signal lights, red signal lights, white signal lights, and railway pedestrians are calculated on DLSLRP dataset, which are illustrated in Figure 8-11, respectively.

As can be seen, mAP of the blue signal lights, red signal lights, white signal lights, and pedestrians reaches 98.89%,

FIGURE 4. Sample images of railway pedestrians in DLSLRP dataset.

TABLE 1. Comparison of the YOLOv3-based method and our detector.

Method	mAP/%	FPS	AP/%			
			Blue signal lights	Red signal lights	White signal lights	Pedestrians
YOLOv3-based method	86.93	20.0	94.36	97.12	97.38	58.85
Our method	93.52	25.0	98.89	100	98.72	76.47

100%, 98.72%, and 76.47%, respectively. The corresponding detection precision achieves 98.18%, 100%, 99.08%, and 96.04%, respectively. The recall reaches 94.74%, 97.24%, 94.74%, and 77.94%, respectively. The experimental results give elucidate that the proposed method has good performance on the detection of locomotive signal lights and pedestrians on railway tracks.

2) COMPARISON OF THE YOLOv3-BASED METHOD AND OUR DETECTOR

As mentioned in Section II, compared to the other CNN-based detection algorithms, YOLO series [18]–[21] can better meet the real-time detection requirements of locomotive operation. Thus, we further make a comparison between the YOLOv3-based method and our detector in the identical application scenario, the results are illustrated in Figure 12.

As shown in Figure 12, under bad illumination conditions, our detector is able to recognize both locomotive signal lights and pedestrians close to railway tracks. In contrast, in the same scenario, the YOLOv3-based method only detects the locomotive signal lights, and cannot recognize the pedestrians. Furthermore, the locomotive signal light located far away from the capturing camera, yet within the range of 1.5 kilometers, cannot be detected by the YOLOv3-based method. Whereas, our method is able to detect the far located signal light with an accuracy of 86%. The findings demonstrate that, compared with the YOLOv3-based method, our detector has better detection capacity.

The AP, mAP, and FPS of the YOLOv3-based method and our method are compared on DLSLRP dataset, which is shown in Table 1.

As illustrated in Table 1, compared to the YOLOv3-based detector, our method has greater AP and mAP on all detection categories, which indicate that it has better performance. Since the proposed detector is based on the improved YOLOv4, and has more complex network structure and parameters. Therefore, it has larger FPS than the YOLOv3-based method. Surprisingly, compared with the AP of locomotive signal lights, our method has lower AP in terms of detecting railway pedestrians. This is mainly because compared to the difference between railway pedestrians, locomotive signal lights of distinct colors have a small difference in appearance and shapes, which leads to the recognition of locomotive signal lights being comparatively easy. The mAP of our detector on all recognition classes achieves 93.52%, and the detection speed reaches 25 FPS, which indicate that it can meet the real-time detection requirements of locomotive operation.

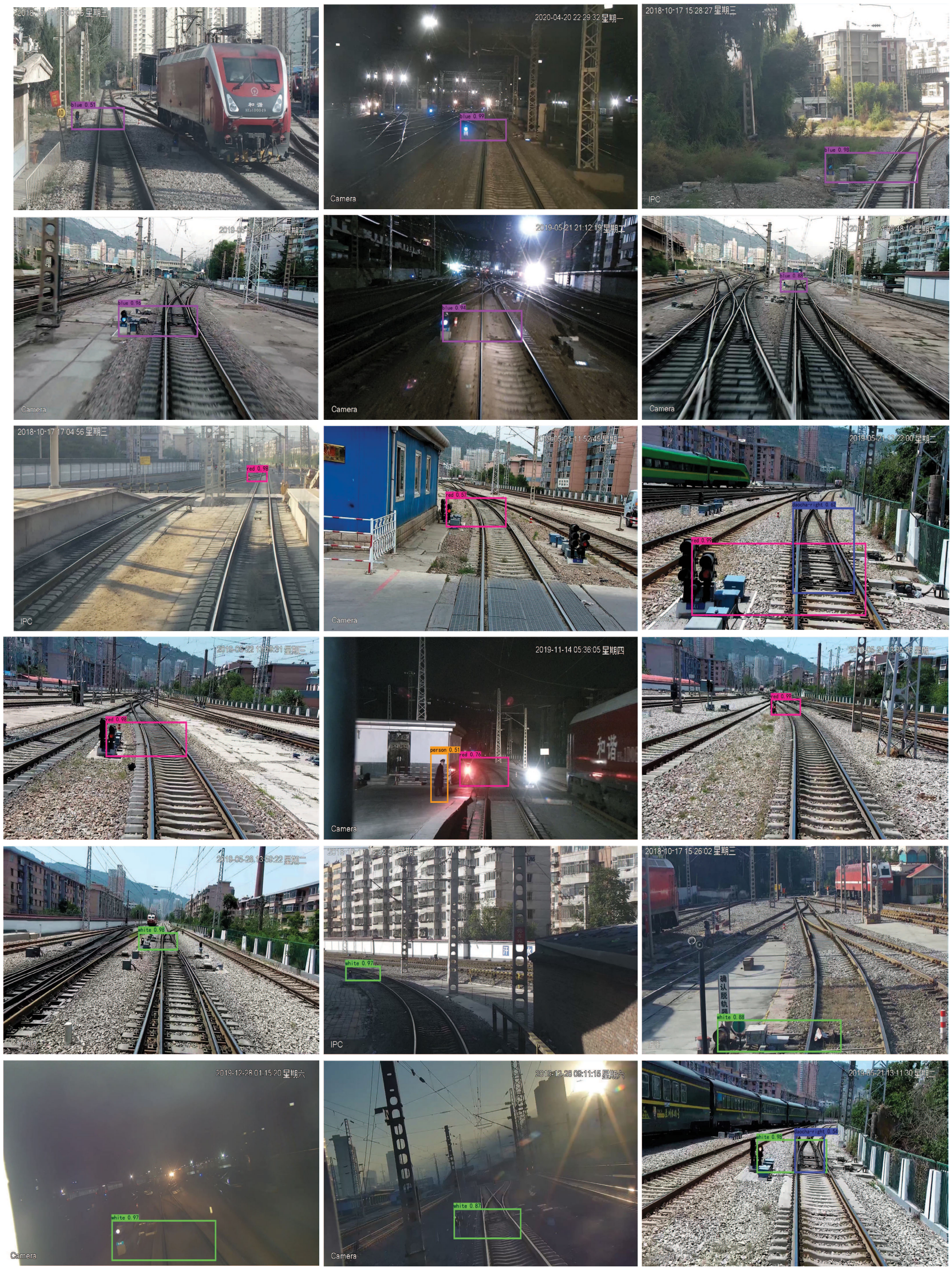


FIGURE 5. Detection results of locomotive signal lights of different colors under distinct illumination conditions.

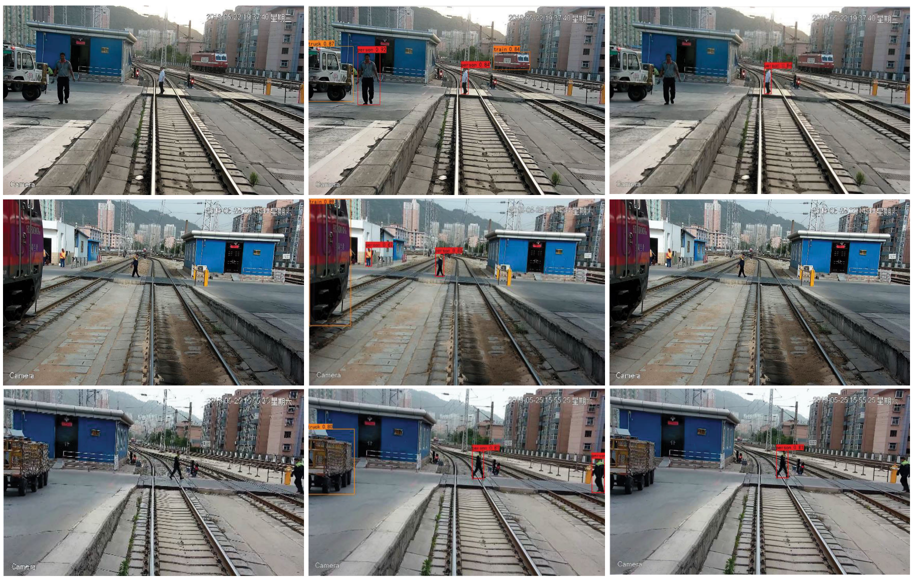
3) THE TOTAL LOSS OF OUR METHOD

Finally, the total loss of our detection method is shown in Figure 13. It takes approximately 5×10^4 steps to reach almost zero loss.

IV. CONCLUSION

Precise and speedy recognition of signal lights has a profound impact on the safe operation of locomotives. Attributed to the extremely complicated operation environment of locomotives, drivers are inclined to make misjudgments, which will cause serious railway traffic accidents. The conventional detection methods of locomotive signal lights have shortcomings such as being greatly influenced by illumination conditions, and difficult to detect small-sized signal lights. Thus, these methods can no longer meet the application requirements of locomotive safe operation. With the dynamic

FIGURE 6. Detection results of pedestrians on railway tracks.

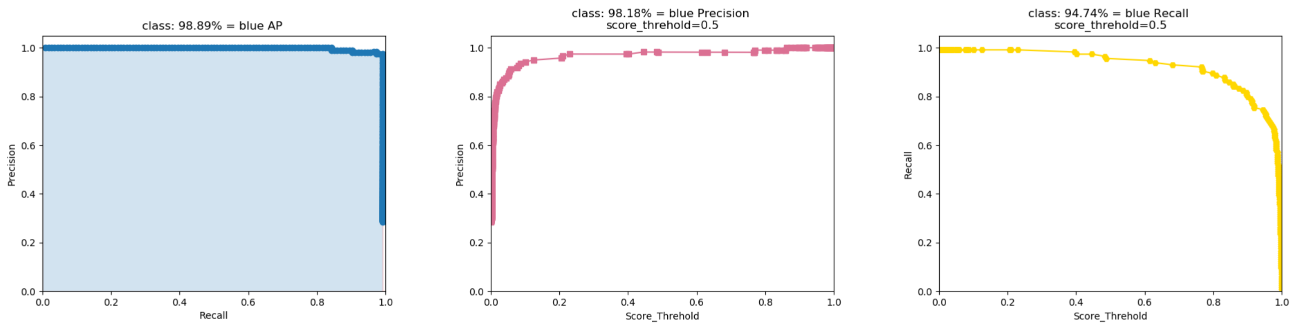


(a) The original image (b) YOLOv4-based method (c) Our method

FIGURE 7. Comparison of the YOLOv4-based method and our detector.

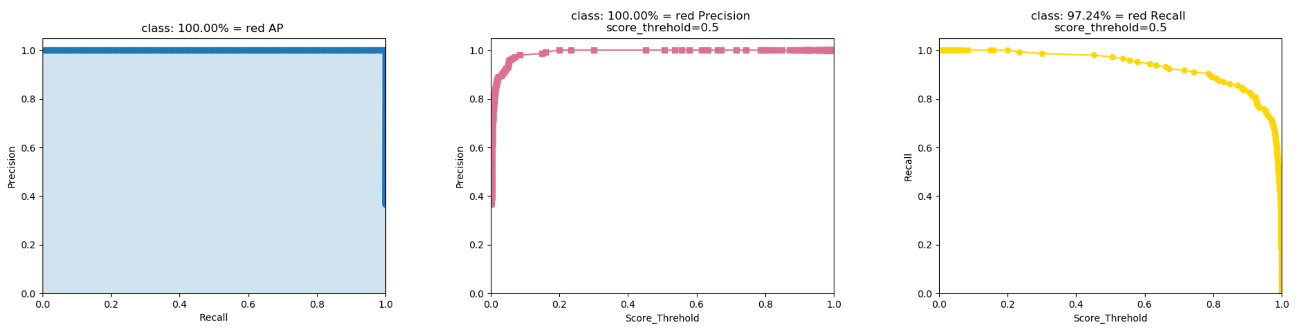
development of deep learning techniques, object detection algorithms based on deep learning have been extensively used in a wide range of fields. It is significant to apply CNN-based object detection methods in the recognition of locomotive signal lights and pedestrians on railway tracks.

This paper, for the first time, proposed a real-time detection method based on improved YOLOv4, which is able to recognize locomotive signal lights of different colors and railway pedestrians under bad illumination conditions. It is worth stressing that we established a comprehensive dataset called detection of locomotive signal lights and railway pedestrians (DLSLRP), which can be used in the training, validation, and testing of related convolutional neural networks. We evaluated our method on DLSLRP dataset, the experimental results



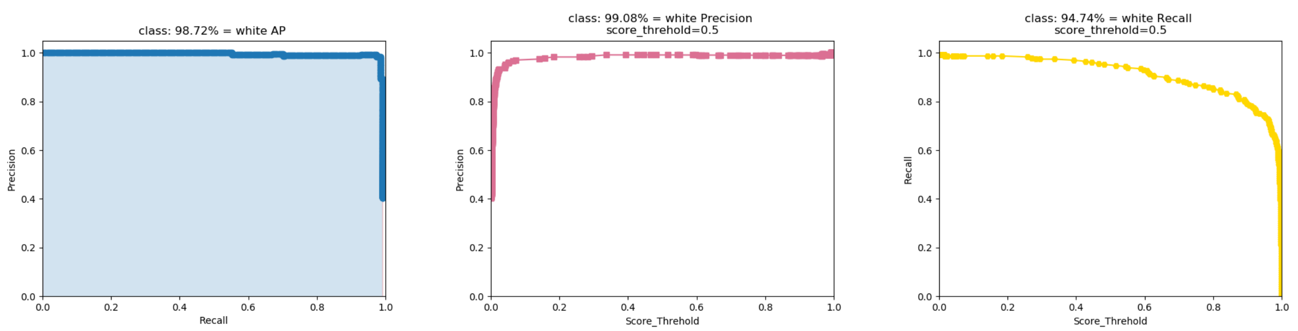
(a) mAP (b) Precision (c) Recall

FIGURE 8. **The mAP, precision, and recall of blue signal lights.**



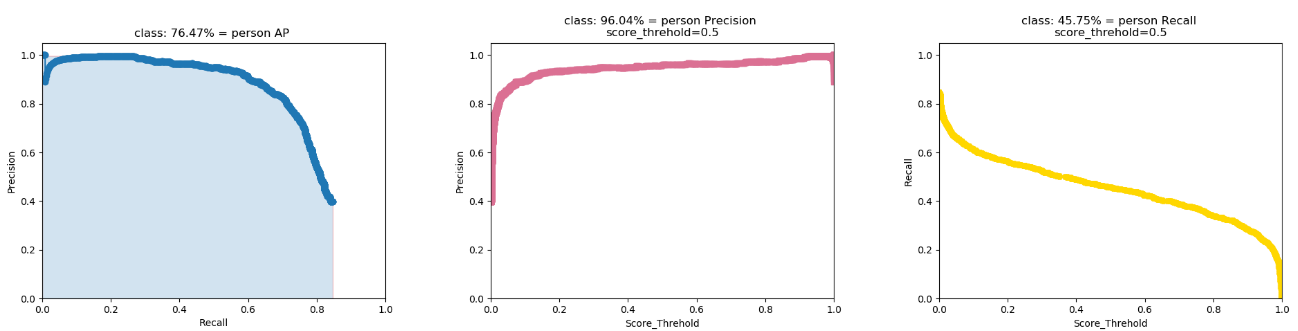
(a) mAP (b) Precision (c) Recall

FIGURE 9. **The mAP, precision, and recall of red signal lights.**



(a) mAP (b) Precision (c) Recall

FIGURE 10. **The mAP, precision, and recall of white signal lights.**



(a) mAP (b) Precision (c) Recall

FIGURE 11. **The mAP, precision, and recall of pedestrians.**

(a)YOLOv3-based method (b) Our method

FIGURE 12. Comparison of the YOLOv3-based method and our detector.

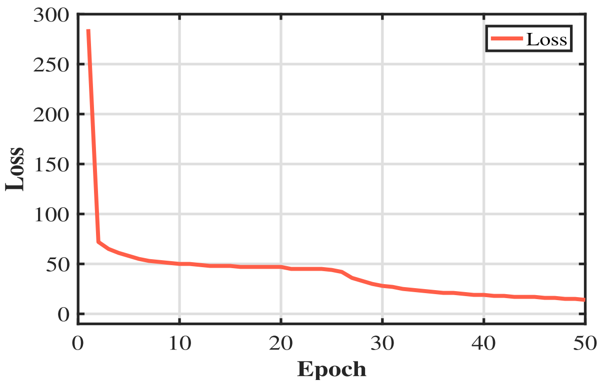


FIGURE 13. The total loss of our method.

reveal that the proposed detector can detect locomotive signal lights and railway pedestrians with high accuracy at a fast speed. The mAP achieves 93.52%, and the detection speed reaches 25 FPS. These findings indicate that our method meets the detection requirements of locomotive signal lights in real-time scenarios.

In the future, our focus is to further optimize the current model, such as improving its feature extraction network and feature expression ability, reducing the number of parameters, and further improving the recognition accuracy and real-time performance. Our work may provide a new perspective for the application of CNN-based object detection method in the recognition of locomotive signal lights and railway pedestrians.

ACKNOWLEDGMENT

The collection of sample images for this research was supported by Lanzhou Railway Bureau Group Company Ltd., China Railway. Their contributions are greatly appreciated.

REFERENCES

[1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, and S. Ma, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[3] L. Dougherty, J. C. Asmuth, A. S. Blom, L. Axel, and R. Kumar, "Validation of an optical flow method for tag displacement estimation," *IEEE Trans. Med. Imag.*, vol. 18, no. 4, pp. 359–363, Apr. 1999.

[4] D. Shi, L. Zheng, and J. Liu, "Advanced Hough transform using a multilayer fractional Fourier method," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1558–1566, Jun. 2010.

[5] M. Aronniemi, J. Sainio, and J. Lahtinen, "Chemical state quantification of iron and chromium oxides using XPS: The effect of the background subtraction method," *Surf. Sci.*, vol. 578, nos. 1–3, pp. 108–123, Mar. 2005.

[6] C. Li, "Dangerous posture monitoring for undersea diver based on frame difference method," *J. Coast. Res.*, vol. 103, no. 1, pp. 939–942, Jul. 2020.

[7] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[8] W. Zhu, W. Huang, Z. Lin, Y. Yang, S. Huang, and J. Zhou, "Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation," *Multimedia Tools Appl.*, vol. 75, no. 5, pp. 2815–2837, May 2015.

[9] X. Hua, X. Wang, D. Wang, J. Huang, and X. Hu, "Military object real-time detection technology combined with visual salience and psychology," *Electronics*, vol. 7, no. 10, p. 216, Sep. 2018.

[10] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrians detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.

[11] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 39–51, Jan. 1998.

[12] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[13] H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Trans. Med. Imag.*, vol. 15, no. 3, pp. 235–245, Jun. 1996.

[14] 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.

[15] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Montreal, QC, Canada, 2015, pp. 91–99.

[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[18] 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)*, Las Vegas, NV, USA, Sep. 2016, pp. 779–788.

[19] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

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

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

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[23] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, 2016, pp. 379–387.

[24] W. Chen, S. Lu, B. Liu, G. Li, and T. Qian, "Detecting citrus in orchard environment by using improved YOLOv4," *Sci. Program.*, vol. 2020, pp. 1–13, Nov. 2020.

[25] A. Chernov, M. Butakova, A. Guda, and P. Shevchuk, "Development of intelligent obstacle detection system on railway tracks for yard locomotives using CNN," in *Proc. Eur. Dependable Comput. Conf.*, Munich, Germany, 2020, pp. 33–43.

[26] Z. Cheng and F. Zhang, "Flower end-to-end detection based on YOLOv4 using a mobile device," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–9, Sep. 2020.

[27] H. Deng, J. Cheng, T. Liu, B. Cheng, and Z. Sun, "Research on iron surface crack detection algorithm based on improved YOLOv4 network," *J. Phys., Conf. Ser.*, vol. 1631, no. 1, Sep. 2020, Art. no. 012081.

[28] K. Behrendt, L. Novak, and R. Botros, "A deep learning approach to traffic lights: Detection, tracking, and classification," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 1370–1377.

[29] V. John, K. Yoneda, B. Qi, Z. Liu, and S. Mita, "Traffic light recognition in varying illumination using deep learning and saliency map," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2286–2291.

[30] V. John, K. Yoneda, Z. Liu, and S. Mita, "Saliency map generation by the convolutional neural network for real-time traffic light detection using template matching," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 159–173, Sep. 2015.

[31] J. Müller and K. Dietmayer, "Detecting traffic lights by single shot detection," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 266–273.

[32] S. D. Buluswar and B. A. Draper, "Color recognition in outdoor images," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 171–177.

[33] F. Gasparini and R. Schettini, "Color balancing of digital photos using simple image statistics," *Pattern Recognit.*, vol. 37, no. 6, pp. 1201–1217, Jun. 2004.

[34] M. Omachi and S. Omachi, "Traffic light detection with color and edge information," in *Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, Aug. 2009, pp. 284–287.

[35] M. Omachi and S. Omachi, "Detection of traffic light using structural information," in *Proc. IEEE 10th Int. Conf. Signal Process.*, Oct. 2010, pp. 809–812.

[36] F. Lindner, U. Kressel, and S. Kaelberer, "Robust recognition of traffic signals," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 49–53.

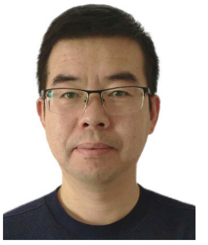
[37] R. de Charette and F. Nashashibi, "Traffic light recognition using image processing compared to learning processes," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 333–338.

[38] R. de Charette and F. Nashashibi, "Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2009, pp. 358–363.

[39] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[41] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8759–8768.



HUANMIN WANG received the M.S. degree in computer application from Lanzhou Jiaotong University, China, in 2008, where he is currently pursuing the Ph.D. degree in transportation Internet of Things and equipment technology. He is a Senior Lecturer with Lanzhou Jiaotong University. His research interests include the application of intelligent image recognition, the transportation internet of things equipment, and informatization theory. His awards and honors include the Second Prize for Progress in Science and Technology of Lanzhou, in 2014, and the Second Prize for Progress in Science and Technology of Gansu Province, in 2014.

HUAYAN PEI received the M.S. degree in computer application from Xidian University, China, in 2009, and the Ph.D. degree in intelligent transportation and information system engineering from Lanzhou Jiaotong University, China, in 2021. She is currently a Senior Lecturer with Lanzhou Jiaotong University. She has published works in several publications and venues, such as the *Physics Letters A*, *Europhysics Letters*, and *Chinese Physics B*. Her research interests include the application of object detection, game theory, complex systems, and cooperation dynamics.

JIANBAI ZHANG received the B.A. degree in locomotive engineering from Lanzhou Jiaotong University, China, in 1991. He is currently a Professor-Level Engineer with Lanzhou Railway Bureau Group Company Ltd., China Railway. His research interests include the locomotive overhaul technique, measurement and control technology, and locomotive application. His awards and honors include the Mao Yisheng Railway Engineer Award, in 2012, and the Second Prize of Gansu Mechanical Engineering Society, in 2020.