A Lightweight Rail Defect Detection Method Based on Improved YOLOv5

Zongyao Wang [1, a], Zilong Lv [1], Jian Wang [1], Ronghui Bi [1, 2]

[1] School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026;Liaoning, China;

[2] China Communications Railway Operation Co., Ltd., Tianjin 300201 China

[a] wzy@dlmu.edu.cn

Abstract. With the rapid development of railway transportation infrastructure, tracks serve as a critical component of the railway system, and their safety is essential for reliable train operations. This study proposes an efficient, lightweight defect detection method based on the YOLOv5 model, leveraging deep learning to achieve real-time, high-accuracy identification of track defects. A standardized dataset comprising surface cracks and missing fasteners is constructed to facilitate model training. To address YOLOv5's limitations in small-object detection, the proposed approach incorporates shallow detection layers to enhance sensitivity to minor defects. Additionally, the integration of a Global Context (GC) attention mechanism improves the model's expressive and generalization capabilities. For computational efficiency, a C3 Master module is introduced, built upon the FasterNet lightweight architecture, significantly reducing parameters and accelerating inference speed. Comparative experiments with SSD, Faster R-CNN, and baseline YOLOv5 demonstrate the superiority of the proposed model, achieving a 95.3% mAP in track defect detection. The proposed lightweight solution enables precise, real-time defect detection, offering a novel approach to enhancing railway safety inspections.

Keywords: Rail defect detection; Deep learning; YOLOv5; Object detection.

1. Introduction

In the context of the ongoing process of economic globalisation, the railway has undergone significant development as an integral component of the transportation industry. According to statistical data, as of the beginning of January 2024, the total mileage of railway operation in China had surpassed 159,000 kilometres, with the mileage of high-speed railway operation accounting for approximately 45,000 kilometres, constituting 28.3% of the total mileage. The advent of high-speed railways and the concomitant freight requirements of heavy-haul trains have resulted in increased running pressure and impact load on the track, thereby augmenting the probability of track defects. Consequently, enhancing the track defect detection algorithm to improve detection accuracy and efficiency is imperative for enhancing the quality and efficiency of track flaw detection work, constructing a railway safety detection system, and ensuring the stability and safety of railway operation.

Most countries still rely on traditional detection methods, primarily manual visual inspection. While simple to implement, this approach suffers from low efficiency and inconsistent accuracy due to human fatigue and oversight. Recent technological advancements have facilitated the progressive implementation of non-destructive testing (NDT) techniques in rail defect detection, particularly ultrasonic and eddy current testing methods. Although these physical detection methods demonstrate superior precision and sensitivity, they present several practical limitations: elevated operational and maintenance costs, intricate system upkeep requirements, and relatively low inspection speeds. In addition, professional personnel are required to operate these methods. The advent of artificial intelligence has precipitated the widespread adoption of deep learning-based methodologies for the identification of track defects. The deep learning-based track defect detection method based on image processing has the characteristics of real-time performance and non-contact, and can be well applied in the field of track defect detection.

Recent research has demonstrated significant progress in applying deep learning techniques to rail defect detection. Shang [6] developed a neural network architecture based on the Inception-v3 framework, trained using a custom-built dataset. Building on this, Hao W. et al. [7] achieved enhanced positioning accuracy through modifications to the Mask R-CNN architecture, incorporating a multi-scale fusion module and implementing the complete intersection over union (CIOU) metric to address traditional intersection limitations.

Several studies have focused on multi-scale feature integration. Han Qiang et al. [8] proposed a hierarchical feature fusion network that combines image features from multiple receptive fields, significantly improving multi-scale defect recognition capabilities. Zhao Hongwei et al. [9] advanced this approach through their development of a dual-modal deep learning network, demonstrating that multi-source sensor data fusion substantially enhances detection robustness.

Innovative preprocessing techniques have also emerged. Luo Hui et al. [10] achieved notable improvements in image denoising and feature extraction by integrating Gabor filtering and HSV transformation into the Faster R-CNN framework. For specific applications, Hao R. et al. [11] designed a steel plate surface defect detection model employing deformable convolutions and a pyramid feature fusion network to generate high-resolution multi-scale feature maps.

The most precise results to date were reported by Wang et al. [12], who combined residual neural networks (ResNets) with fully convolutional networks (FCNs) to minimize feature-level transmission loss.

Most existing studies focus on optimizing model architecture by integrating multi-level feature fusion and attention mechanisms, demonstrating exceptional performance in improving defect detection accuracy. However, for mobile applications with limited computational resources and real-time requirements, further optimization is necessary. This paper proposes a lightweight track defect detection algorithm based on an improved YOLOv5 framework. The adapted model enhances YOLOv5's efficiency, achieving fast and accurate detection while maintaining a compact structure, thereby facilitating practical deployment."

2. Improved YOLOv5

YOLOv5, a lightweight variant of YOLOv4 proposed by Ultralytics LLC, achieves a balance between detection accuracy and speed while retaining its efficient architecture. This paper presents an enhanced YOLOv5 model tailored for track defect detection, with improvements in feature extraction, feature stacking/fusion, and network lightweighting. The proposed modifications integrate a GC (Global Context) attention mechanism and additional small object detection layers to boost detection accuracy, particularly for subtle defects. Furthermore, replacing the backbone's C3 module with FasterNet reduces model parameters without compromising performance. The resulting framework exhibits high precision, robustness, and real-time capability, addressing key challenges in track defect detection, including accuracy limitations and deployment constraints. Experimental validation confirms the model's effectiveness in industrial applications.

2.1 GC Attention Mechanism

As the YOLOv5 network layers deepen, the extracted information at the output becomes increasingly abstract, making it difficult to detect small objects in images. The present paper proposes a solution to this issue by integrating the GC attention mechanism and incorporating shallow detection layers.

The principle of the GC attention mechanism is primarily based on non-local global self-attention modelling, enabling the model to capture global contextual relationships (or features), thereby enhancing the model's feature extraction capability. Furthermore, the GC Block integrates the architecture of SENet, thereby substantially minimising computational expenditure and rendering it a nearly cost-free plug-and-play module.In summary, GCNet enhances performance in a range of computer vision tasks by introducing a global context attention mechanism, thereby

enabling the network to more effectively comprehend and utilise global image information. The fundamental principle underpinning this approach is the dynamic adjustment of attention weights, with the objective of focusing on the most pertinent information within the image.

2.2 Small Object Detection Layer

In order to enhance the mean precision of small object detection whilst circumventing the excessive loss of local detail features and small object information that is often associated with downsampling, this paper proposes a modification to the original YOLOv5 network structure. This modification involves the incorporation of a small object detection layer with a feature map scale of 160×160. This layer, following the process of deep feature transmission and fusion with shallow features, contains a greater quantity of contour and positional information regarding small objects. This facilitates their localization and recognition, whilst concomitantly reducing missed and false detection.

As demonstrated in Figure 1, the proposed paper incorporates a 160×160-scale small object detection layer (including supplementary fusion feature layers and an additional detection head) into the existing network architecture, thereby facilitating precise recognition of distant objects. Initially, the 80×80-scale feature layer from the sixth layer of the backbone network is stacked with the upsampled feature layer of the PANet structure. Subsequent to processing by the C3 module, CBS convolution, and upsampling, a deep semantic feature layer containing information regarding small objects is obtained. This feature layer is then stacked with the shallow positional feature layer from the fourth layer of the backbone network. This process enhances the expressive ability of the 160×160-scale fusion feature layer for small object semantic and positional information. In conclusion, the fusion feature layer is transmitted to the additional detection head for decoding and detection, subsequent to channel adjustment by the C3 module.

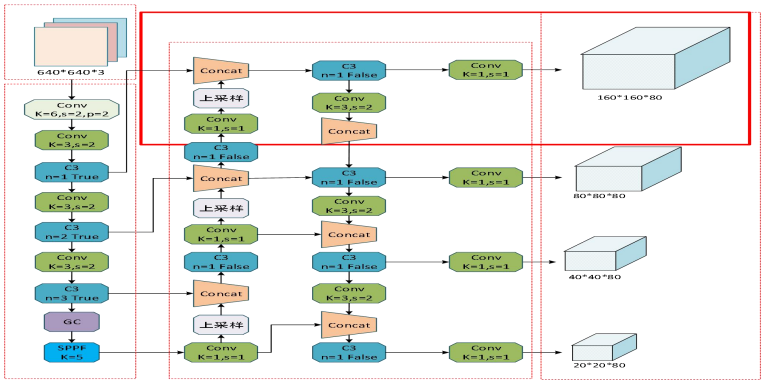

Fig.1 YOLOv5 network architecture diagram with added small object detection layer

2.3 Lightweight Network Model

At present, mainstream lightweight neural networks such as MobileNet, ShuffleNet, and GhostNet reduce network complexity through group convolution. However, fragmented computation patterns have been shown to be detrimental to recognition accuracy. In addition to convolutional neural networks, innovative models such as MobileViT and MobileFormer reduce computational complexity by integrating depthwise separable convolution (DWConv) and optimising attention mechanisms. Nevertheless, such hybrid architectures continue to exhibit memory access bottlenecks. The present paper puts forward a competitive alternative to PConv[13], which serves to reduce computational redundancy and the number of memory accesses.

The working principle of PConv is as follows: only a portion of the input feature map channels are used for feature extraction, while the remaining channels remain unchanged (i.e., from Cp to C channels). The number of channels used is Cp. It is hypothesised that the input and output feature maps will have an equal number of channels. The FLOPS of PConv can be expressed as follows:

$$FLOPS = h \times w \times k^2 \times c_p^2 \tag{2}$$

In the given context, the variables h and w denote the height and width of the feature map, respectively. The variable k is used to represent the size of the convolution kernel, while C_p is employed to denote the number of channels for conventional convolution. At this point in the process, PConv only requires h×w FLOPS, while also having a smaller memory access volume, as demonstrated in Equation (3):

$$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \tag{3}$$

PConv has the capacity to expeditiously execute a variety of visual tasks. The model is composed of four distinct stages, each preceded by an embedding layer or a merging layer, depending on the necessity for downsampling. The embedding layer refers to a 4×4 Conv with stride 4, whereas the merging layer involves a 2×2 Conv with stride 2. Each Faster Block is succeeded by two PWConv layers, and finally, a global pooling layer (Global Pool) and a fully connected layer (FC) are added.

3. Experiment

3.1 Dataset

Currently, large-scale public datasets for track defect detection are scarce. To address this gap, this study constructs a comprehensive dataset by aggregating images from two primary sources: (1) publicly available datasets and (2) original captures of railway track sections in open-access environments. The collected data, rigorously annotated in YOLO format, comprises 872 defect images categorized by standardized defect types (e.g., cracks, squats). These images are partitioned into training (80%), validation (10%), and testing (10%) sets to ensure robust model evaluation. This dataset not only facilitates the experiment but also serves as a benchmark for future research in railway maintenance.

Tab1 dataset type

Category	Number of Images	Number of Bounding Box Annotations
Spalling	364	1703
Crushing	303	529
Fracture	142	149
Missing Fasteners	284	380

3.2 Evaluation Metrics

The Mean Average Precision (mAP) is utilized as a metric to evaluate the performance of the algorithm. The mAP is a metric for the evaluation of object detection algorithms, with the capacity to assess both classification and localization performance. The following calculation formulas are to be used for the relevant computations:

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$mAP = \frac{1}{|Q|} \sum_{q=Q} AP(q) \tag{6}$$

In this study, P and R represent the precision and recall of the algorithm, respectively. TP denotes the number of true positive targets, FP is the number of false positive targets, FN is the number of false negative targets, AP is the average precision, and mAP is the mean average precision.

3.3 Experiment Result

The network model was designed based on the algorithm proposed in this study. As shown in Figure 2, the mean Average Precision (mAP) curve demonstrates that the model converges stably during training. After 200 iterations, the mAP on the training set reached a plateau, achieving a final value of 94.6%. This result validates both the rationality of the model architecture and its robust performance in track defect detection.

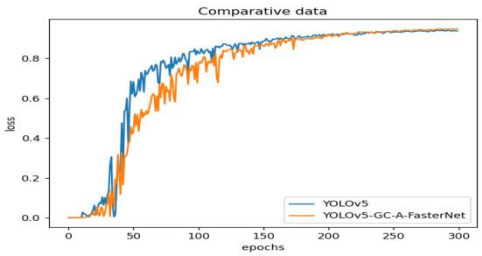

Fig.2 Average precision mean curve before and after improvement.

To further verify the superiority of the proposed algorithm, it was compared with commonly used object detection algorithms such as SSD and Faster R-CNN. The detection results of various algorithms are shown in Table 2.

Tab.2 Detection results of different algorithms

Network Model	Model Size (MB)	Parameters (Millions)	mAP@0.5 (%)
SSD	103	-	70.2
Faster-R-CNN	108	-	74.82
YOLOv5	14.5	7.02	93.8
YOLOv5-GC	14.6	7.09	94.3
YOLOv5-GC-A	15.6	7.23	95.3
YOLOv5-GC-A-FasterNet	11.7	5.78	94.6

As shown in Table 2, the final optimized algorithm performs well in track defect detection, achieving the best overall metrics. The mAP for several track defects reached 94.6%, an improvement of 0.8% over the original YOLOv5. The missed detection rate for dense defects was reduced, and the detection accuracy was significantly improved. In terms of algorithm speed, the integration of the FasterNet structure into the backbone network further reduced the model size and significantly decreased the number of parameters, outperforming other common object detection algorithms. In summary, the proposed algorithm exhibits high robustness, accuracy, and real-time performance, making it better suited for detection tasks. The detection results are mapped back to the original images, as shown in Figure 3.

(a)Missing Fasteners (b) Fracture (c) Spalling (d) Crushing

Fig.3 Track Defect Detection Results

4. Conclusion

This paper presents an improved YOLOv5-based algorithm for real-time and precise detection of four critical railway track defects: spalling, crushing, fractures, and missing fasteners. To enhance

feature representation, we integrate the Global Context (GC) attention mechanism into the YOLOv5 backbone, enabling the network to capture finer defect details and process them more accurately.

Given the limited variety of target defects, we further optimize the model by replacing the original backbone with FasterNet, a lightweight network designed for edge-device compatibility. This substitution significantly reduces computational overhead while maintaining detection accuracy, leveraging FasterNet advantages in latency reduction and throughput improvement. By fusing FasterNet with the C3 convolution module, we achieve a leaner model with fewer parameters and faster inference speeds.

Experimental results demonstrate that the proposed algorithm exhibits strong robustness, delivering high-precision defect detection across diverse environmental conditions. Compared to traditional methods, our approach addresses key limitations and provides a practical solution for automated track inspection tasks.

References

[1] Zhang Hui, Song Yanan, Wang Yaonan, et al. A Review of Non-destructive Testing and Evaluation Techniques for Rail Defects [J]. Journal of Instrumentation, 2019, 40(02): 11-25.

[2] Xu H, Cao J, Li G, et al. Discussing on How to Improve the Subway Rail Flaw Detection Quality [P]. Proceedings of the 2017 3rd International Forum on Energy, Environment Science and Materials (IFEESM 2017), 2018.

[3] Min Yongzhi, Yue Biao, Ma Hongfeng, et al. Surface Defect Detection of Steel Rails Based on Image Grayscale Gradient Features [J]. Journal of Instrumentation, 2018, 39(04): 220-229.

[4] Wang Baihui. Research and Application of Image Denoising and Edge Detection Algorithm Based on Wavelet Transform [D]. Southwest Jiaotong University, 2019.

[5] Tastimur C, Yetis H, Karakse M, et al. Rail Defect Detection and Classification with Real-Time Image Processing Technique [J]. 2019.

[6] Shang L, Yang Q, Wang J, et al. Detection of Rail Surface Defects Based on CNN Image Recognition and Classification [C]//2018 20th International Conference on Advanced Communication Technology (ICACT). 2018: 74-77.

[7] Hao W, Mengjiao L, Zhibo W. Rail Surface Defect Detection Based on Improved Mask R-CNN [J]. Computers and Electrical Engineering, 2022, 102.

[8] Han Qiang, Liu Junbo, Feng Qibo, et al. Steel Rail Surface Damage Detection Method Based on Multi-level Feature Fusion [J]. China Railway Science, 2021, 42(05): 41-49.

[9] Zhao Hongwei, Zheng Jiajun, Zhao Xinxin, et al. Surface Defect Detection Method for Steel Rails Based on Bimodal Deep Learning [J]. Computer Engineering and Applications, 2023, 59(07): 285-293.

[10] Luo Hui, Xu Guanglong. Steel Rail Surface Defect Detection Based on Image Enhancement and Deep Learning [J]. Journal of Railway Science and Engineering, 2021, 018(003): 623-629.

[11] Hao R, Lu B, Cheng Y, et al. A Steel Surface Defect Inspection Approach Towards Smart Industrial Monitoring [J]. Journal of Intelligent.

[12] Wang Yanshu, Yu Jianbo. Surface Defect Detection of Strip Steel Based on Adaptive Global Positioning Algorithm [J]. Journal of Automation, 2024, 50(08): 1550-1564. DOI: 10.16383/j.aas.c210467.

[13] Chen J , Kao S H , He H ,et al.Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).0[2024-11-26].DOI:10.1109/CVPR52729.2023.01157.