Dear Reviewer,

I would like to express my sincere gratitude for the time and effort you have dedicated to reviewing our manuscript. Your insightful comments and constructive critiques are immensely valuable to us and have provided a clearer direction for enhancing the quality of our work.

We have thoroughly considered each point raised in your review and have undertaken substantial revisions to address these concerns.

Once again, thank you for your invaluable contribution to the refinement of our research.

Best Regards,
Wenbo

Point-by-point response to reviewer (original comments by reviewer are in red color)

This suggests a deep learning approach for the identification of rail defects in ultrasonic B-scan images and compares the performance of YOLOv4, YOLOv3, and Faster R-CNN.

Comments 1. The work is interesting, however, the authors need to discuss in detail the problem being solved by this method, and what its position is in relation to the GCT-8C rail defect detector. Is it a complementary or an alternative?

Response1: Thank you for your constructive feedback and for acknowledging the interest level of our work. We appreciate the opportunity to clarify and expand upon the aspects you have highlighted, particularly concerning the problem addressed by our method and its relationship with the GCT-8C rail defect detector.

In our manuscript, we initially outline the specific problem our method aims to solve within the context of rail defect detection. However, upon reflection and guided by your valuable input, we recognize that a more detailed discussion would enhance the manuscript's clarity and contribution.

In the proposed framework, YOLOv8 is employed as an auxiliary solution, intended to supersede traditional manual discernment practices. This adoption of advanced model recognition serves as an imperative augmentation to the GCT-8C rail flaw detection apparatus in its damage detection and identification processes. The integration of YOLOv8 aims to bolster the accuracy and efficiency of flaw detection, thereby reducing reliance on manual intervention and enhancing overall inspection efficacy.

The discussion of our method and GCT-8C in latest manuscript is as follows:

[Page 24, Line 798, Section 6]

[In the context of rail defect detection, the methodology proposed in this paper serves as an efficacious complementary approach. It capitalizes on the data provided by the GCT-8C rail flaw detector, applying a series of processing techniques to this data. As a result, the approach achieves commendable outcomes in detecting rail damages. This method enhances the existing diagnostic capabilities by leveraging the strengths of the GCT-8C instrument's

data, which, when combined with the sophisticated data processing steps outlined in this study, significantly improves the accuracy and reliability of rail defect detection.]

Comments 2: Since GCT-8C rail defect detector is a major reference, the authors must include a discussion on it in the results section.

Response2: Thank you for your valuable feedback on our manuscript. Your point regarding the necessity of a more comprehensive discussion about the GCT-8C rail defect detector in the results section is well-received and appreciated.

In response to your suggestion, The GCT-8C rail flaw detector, utilized in this study, primarily functions to gather data on rail damages. Currently, the identification of defects detected by the GCT-8C largely relies on manual interpretation, leading to potential under-detection and misdiagnosis. Inherently, the GCT-8C device lacks the capability to autonomously recognize the type and location of rail defects. The algorithm proposed in this paper aims to automate the identification and localization of defects in data collected by the GCT-8C using deep learning techniques. This represents a significant enhancement to the functionalities of the GCT-8C rail flaw detector. However, it's important to note that the algorithm is not designed for direct comparison with the GCT-8C device itself, but rather serves as a complementary tool, augmenting the device's existing diagnostic capabilities.

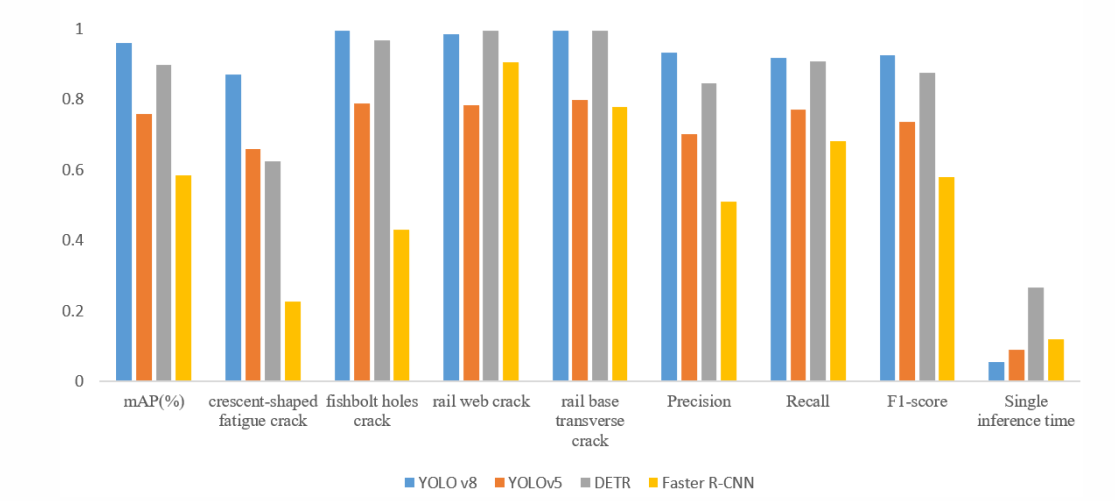
Comments 3: The data processing section is hard to understand, it is advised to include visual illustrations.

Response3: Thank you for your valuable feedback regarding the data processing section of our manuscript. We understand your concern about its clarity and agree that visual illustrations would significantly aid in comprehending this crucial part of our research.

Accordingly, we added new figures to showing the illustrate the model difference. We believe this will make the section more accessible and comprehensible to our readers. We appreciate your suggestion and are committed to improving the manuscript accordingly.

The figure of performance comparison of different models as follows:

[Page22, line 723, Section5.1]



Comments 4: To support the real-time claims it is important to present the computational time results.

Response4: Thank you for your insightful feedback regarding the substantiation of our real-time performance claims. We acknowledge the importance of presenting computational time results to validate these claims.

In Table 7, we have presented 'single inference time' as a comparative metric. The results indicate that YOLOv8 processes a single B-scan image with an impressive average speed of only 0.056 IPS. Such efficiency not only demonstrates the rapid processing capability of YOLOv8 but also highlights its potential as an advantageous supplement to the GCT-8C rail flaw detector.

Furthermore, in the conclusion section, the time requirements for different algorithms are also discussed, providing a comprehensive understanding of the time efficiency associated with each algorithm.

The Table 7 and in the latest manuscript are as follows:

[Page21, Line 722, in Section 5.1]

Model	IOU=0.5								
		AP(%)				Threshold=0.5			
	mAP (%)	crescent-shaped fatigue crack	fishbolt holes crack	rail web crack	rail base transverse crack	Precision	Recall	F1-score	Single inference time
YOLOv8	0.96	0.869	0.993	0.984	0.995	0.933	0.916	0.924	0.056
YOLOv5	0.757	0.659	0.789	0.783	0.799	0.702	0.771	0.735	0.089
DETR	0.896	0.625	0.968	0.995	0.995	0.845	0.908	0.875	0.267

Faster R-CNN	0.585	0.227	0.429	0.905	0.778	0.510	0.680	0.580	0.125

The discussion of time results in the latest manuscript are as follows:

[Page24, Line 817, in Section 6]

[Compared to Faster R-CNN model (mAP: 0.585, inference time: 0.125 IPS) with the problem of frequently missed and wrong de-tection, YOLO series can detect almost all categories of defects in B-scan images. DETR demonstrates commendable performance across multiple metrics, including mAP, precision, recall, and F1-score. Its overall performance closely approximates that of YOLOv8. However, in terms of single inference time, YOLOv8(0.056 IPS) possesses a definitive advantage, evidencing superior efficiency in processing speed compared to DETR (0.267 IPS). There comes out the first conclusion that the overall perfor-mance of YOLO series models is superior to that of Faster R-CNN on our datasets. Besides, The YOLOv8 model further improves ac-curacy (mAP:0.96 vs. 0.757 vs. 0.896) and inference time (0.056 IPS vs. 0.089 IPS vs. 0.067) over DETR and YOLOv5, and can accu-rately distinguish the interference items in B-scan images such as the image of bolt hole and its surrounding cracks.]

We appreciate your guidance in strengthening our manuscript.