

Article

Not peer-reviewed version

Improved YOLOv10 Lightweight Bearing Surface Defect Detection Algorithm

[Lei Ding](#)^{*}, Yuhua Tong, Yunhao Cui, Zhijie Sun

Posted Date: 14 January 2025

doi: 10.20944/preprints202501.1035.v1

Keywords: defect detection; sampling module; YOLOv10; attention mechanism



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Improved YOLOv10 Lightweight Bearing Surface Defect Detection Algorithm

Lei Ding ^{1,*}, Yuhua Tong ¹, Yunhao Cui ², Zhijie Sun ¹

¹ Department of intelligent manufacturing, Shandong Vocational College of Information Technology, Weifang 261000, China

² School of Mechatronics Engineering, Henan University of Science and Technology, Luoyang 471000, China

* Correspondence: 15965368562@163.com

Abstract: Aiming at the problems of insufficient lightweight degree and incompatibility between accuracy and model lightweight of existing rolling bearing surface defect detection algorithms, based on the study of previous related algorithms, the YOLOv10 model algorithm was improved and the optimized model algorithm YOLO-ARE was obtained. The undersampling module ADown is innovatively introduced in Backbone and Head, which enables the model to capture the features of bearing defect images at a higher level, while reducing the amount of computation and optimizing the accuracy of target detection. In the backbone network, the fusion C2f-RVB-EMA module is inserted to optimize the model to extract the image defect features of the rolling bearing, which further improves the lightweight level of the model. The experiment was carried out on the self-built rolling bearing defect data set. Compared with the original YOLOv10n model, GFLOPs of the improved model decreased by 9.2%, Parameters decreased by 13.6%, model size decreased by 8.6%, accuracy P and mAP50 increased by 0.6% and 0.3% respectively. The detection speed is 158.1f/s, and the model can be arranged on the edge end to meet the requirements of industrial batch detection of bearing appearance defects.

Keywords: defect detection; sampling module; YOLOv10; attention mechanism

0. Introduction

In the process of industrial automation production, the rotation of the robot, belt running, and other rotary movements, are inseparable from the rolling bearing, its performance directly affects the stability of the equipment, and thus affects the transfer, positioning, and assembly of products. In the rolling bearing production process, by the production process, assembly process, and other factors, rolling bearing surface pits, scratches, and other defects, affect the normal use of bearings, which in turn affects the industrial automation of the product, resulting in a reduction in the yield of the product and other issues. Therefore, in the process of automatic production of rolling bearings, it is necessary to carry out high-precision defect detection on the bearing surface.

In recent years, an increasing number of experts and scholars have begun to apply machine learning and deep learning technologies in the detection of surface defects on rolling bearings. Presently, the primary algorithms utilized for rolling bearing defect detection encompass R-CNN [1], Fast R-CNN [2], Faster R-CNN [3], Mask R-CNN [4], and the YOLO series [5–9]. Gao Liming et al.[10] introduced an enhanced Faster R-CNN algorithm for detecting surface defects on insulated bearings. They employed the K-means++ algorithm for clustering the defect dataset, utilized ROI Align to minimize localization errors, and incorporated an online hard example mining strategy. Although the accuracy of the improved model increased by 4.8%, it still remained relatively low at 91.2%, with room for improvement in terms of lightweight performance. Yuan Tianle et al.[11] proposed automatic extraction and preprocessing of the detection area, along with improvements using the multi-head self-attention mechanism module from the Transformer to enhance the

precision and recall rate of bearing surface defect detection and the model's anti-interference capability. Despite the accuracy of the improved model increasing by 1.5%, the recall rate by 7.3%, and the average precision by 7.9%, the mAP@0.5 was only 86.1%, indicating a relatively low precision and insufficient lightweight performance. Gu Yunpeng et al.[12] developed a U-Net-based model for detecting appearance defects on bearing rollers. They used a decoder with fused partial convolution to improve computational efficiency, introduced an efficient channel attention mechanism to enhance the capture and understanding of important image features, and designed a Focal Loss function to mitigate the impact of sample imbalance. However, the model's accuracy was relatively low, with precision and recall rates of only 91.58% and 91.74%, respectively. Yao Jingli et al.[13] enhanced attention to defect features by adding SimAM attention, embedded the C2f_SCConv module in the neck network to reduce computational load, and utilized Wise-IoU to accelerate network convergence. The mAP50 increased by 2.9%, but the computational cost (GFLOPs) was 11.8, higher than that of algorithms such as YOLOv5s and EfficientDet. Peng Yanfei et al.[14] optimized the EfficientViT-B0 module and restructured YOLOv5s, introducing a dynamic head based on a multi-head self-attention mechanism. The improved model achieved reductions in parameters, computational cost, and complexity, yet the mAP@0.5 was only 93.8%, indicating a relatively low precision.

Although the research of the above scholars has made certain progress, the degree of lightweight is insufficient, and the precision and lightweight are not coordinated enough. This study aims to reduce the complexity of the algorithm, and achieve the coordination of algorithm lightweight and detection accuracy by improve YOLOv10 algorithm.

1. YOLOv10 Target Detection Algorithm

YOLO series algorithms are widely used in the field of defect detection because of their efficient real-time detection performance. There are six types of YOLOv10 algorithm models: v10-N, v10-S, v10-M, v10-B, v10-L and v10-X[15]. Considering that the algorithm model needs to be arranged at the edge for detection, the original model in this study uses YOLOv10n, and the network structure is shown in Figure 1.

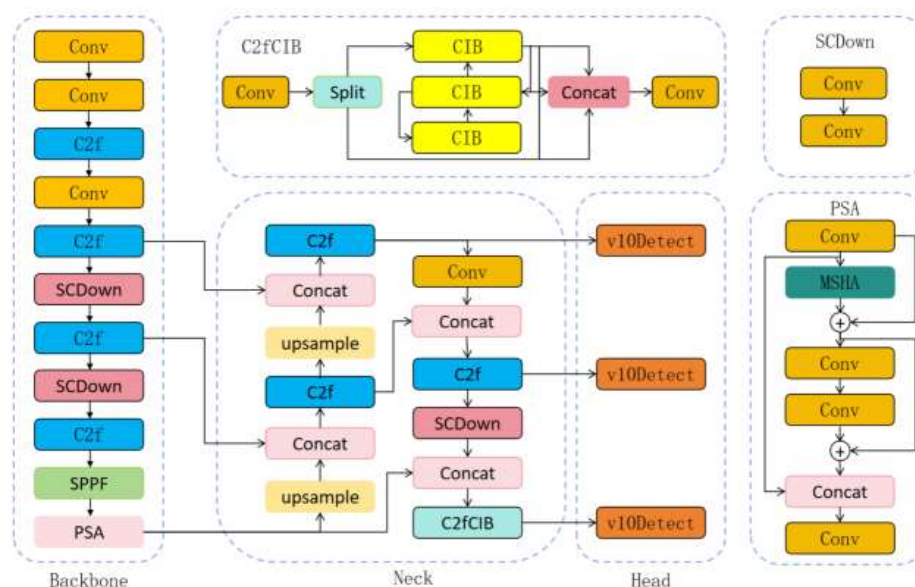


Figure 1. Structure diagram of YOLOv10 model.

YOLOv10 is an advanced object detection algorithm that achieves real-time detection through a single neural network inference, featuring high accuracy and speed. It is suitable for fine-grained object detection and provides a new solution for the surface defect detection of rolling bearings. However, YOLOv10 still faces certain challenges when dealing with the complex textures and tiny

defects on the surface of rolling bearings. Therefore, targeted improvements to YOLOv10 are necessary to enhance its applicability.

2. Improved YOLOv10 Algorithm

2.1. Integrate Subsampling Module ADown

ADown is a convolution block for downsampling operation in target detection tasks, which can reduce the spatial dimension of the rolling bearing defect feature map, help the model to capture the bearing defect feature in the image at a higher level, and reduce the amount of computation. The ADown network structure is shown in Figure 2.

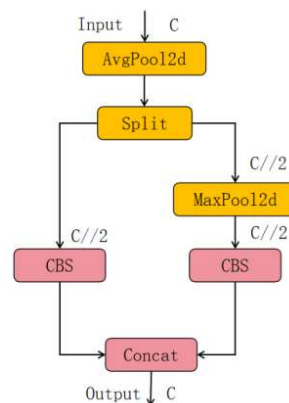


Figure 2. ADown network structure diagram.

The ADown module helps enhance the performance of the model. Firstly, it has a prominent advantage in lightweight design. By effectively reducing the number of parameters, it successfully lowers the model's complexity, enabling the model to maintain an efficient operation state in resource-constrained application scenarios, greatly improving the operational efficiency and facilitating actual deployment. Secondly, when processing feature maps, it can minimize the spatial resolution while preserving key image information to the greatest extent through its ingenious design, ensuring detection accuracy. Thirdly, it has excellent learnability. In the face of diverse data scenarios, it can adapt itself based on data characteristics and dynamically optimize its performance to meet the requirements of different tasks. Fourthly, according to relevant research data, after introducing the ADown module, the model not only becomes more compact in size but also achieves a significant improvement in target detection accuracy, achieving a dual optimization of model miniaturization and high precision. Fifthly, the module has extremely high flexibility and can be seamlessly integrated into the backbone network and detection head of YOLOv10, equipped with rich configuration options, allowing researchers to flexibly combine them according to specific improvement needs and explore diverse optimization paths. Additionally, it has good compatibility and can be used in conjunction with other cutting-edge technologies such as the HWD module to further enhance model performance [16].

ADown module uses convolution layer to extract useful information in feature graph, reduces the spatial dimension of feature graph by adjusting the stride of convolution layer, and reduces the complexity of model by optimizing the number of parameters in convolution layer. The ADown module provides an efficient downsampling solution for real-time object detection models through lightweight design and flexibility.

2.2. Integration of C2f, RepViT Block, and EMA Attention

C2f-rvb-ema module combines the advantages of C2f, RepViT Block and EMA, improves performance through efficient re-parameterization and feature extraction, realizes lightweight, and improves accuracy to some extent [17].

The C2f module incorporates multi-scale features and utilizes lightweight ELAN attention mechanisms to mitigate gradient disappearance and efficiently extract key features while maintaining a lightweight structure. Consisting of multiple layers with deep convolution and efficient feature fusion mechanisms, the original C2f focused on capturing rich feature representations while balancing computational costs.

As the core component of the RepViT model, the RepViT Block re-arranges the 3x3 depthwise separable convolution of the MobileNetV3 module and integrates it into a unified branch. Its network structure is shown in Figure 3. The RepViT Block introduced in the C2f_RVB module further enhances the functionality of C2f by adding re-parameterized convolution [17]. This enables the module to adapt to different computing environments by more effectively fusing features during the inference process.

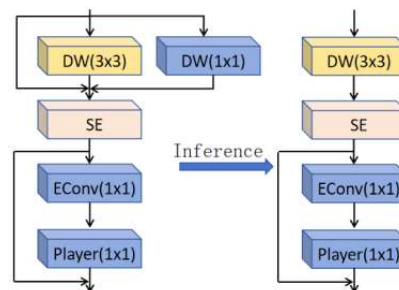


Figure 3. RepViT Block network structure diagram.

The EMA attention mechanism effectively reduces the influence of noise or abnormal Q values by performing an exponential moving average on historical data, thereby enhancing the model's robustness. Its network structure is shown in Figure 5. The C2f-RVB introduces the EMA attention mechanism to smooth feature weights and mitigate noise from low-level features in the detection task[17]. This mechanism enables the network to more accurately focus on important features while reducing the interference caused by background noise, thereby improving detection robustness. Its network structure is illustrated in Figure 4.

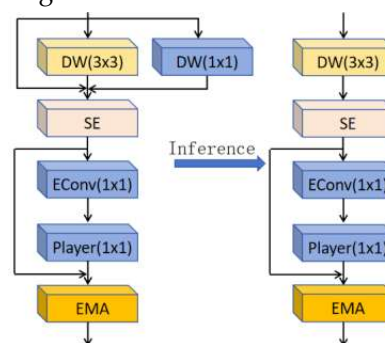


Figure 4. RepViT-EMA Block network structure.

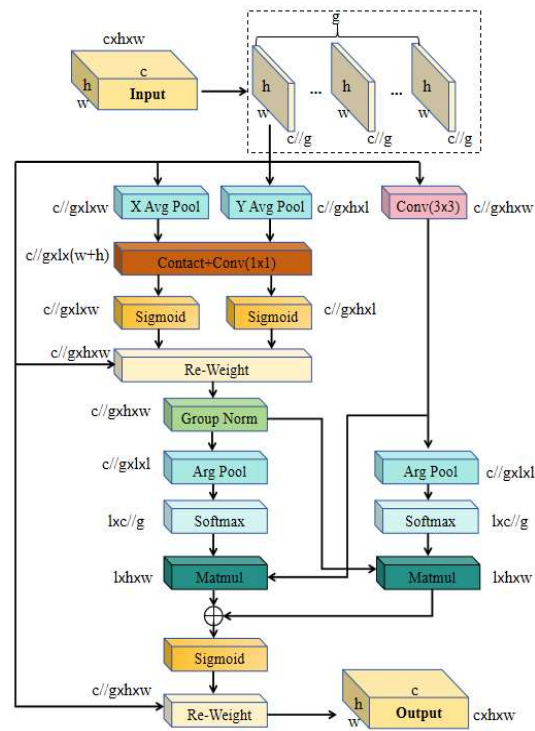


Figure 5. Network structure of EMA module.

2.3. Improve the Structure of YOLOv10 Model

To achieve a lightweight rolling bearing surface defect detection algorithm that is in harmony with detection accuracy, this experiment designed a lightweight detection algorithm named YOLOv10-ARE based on YOLOv10n, as shown in Figure 6. By adding the downsampling module ADown in the backbone and neck to replace the Conv module in the original model, the structure of the model was optimized, enabling it to capture defect features of rolling bearings at a higher level, reducing the computational load of the model, and improving its detection accuracy. In the backbone, the 8-layer C2f module was replaced with the C2f-RVB-EMA module, further achieving the lightweight of the model and optimizing its detection performance.

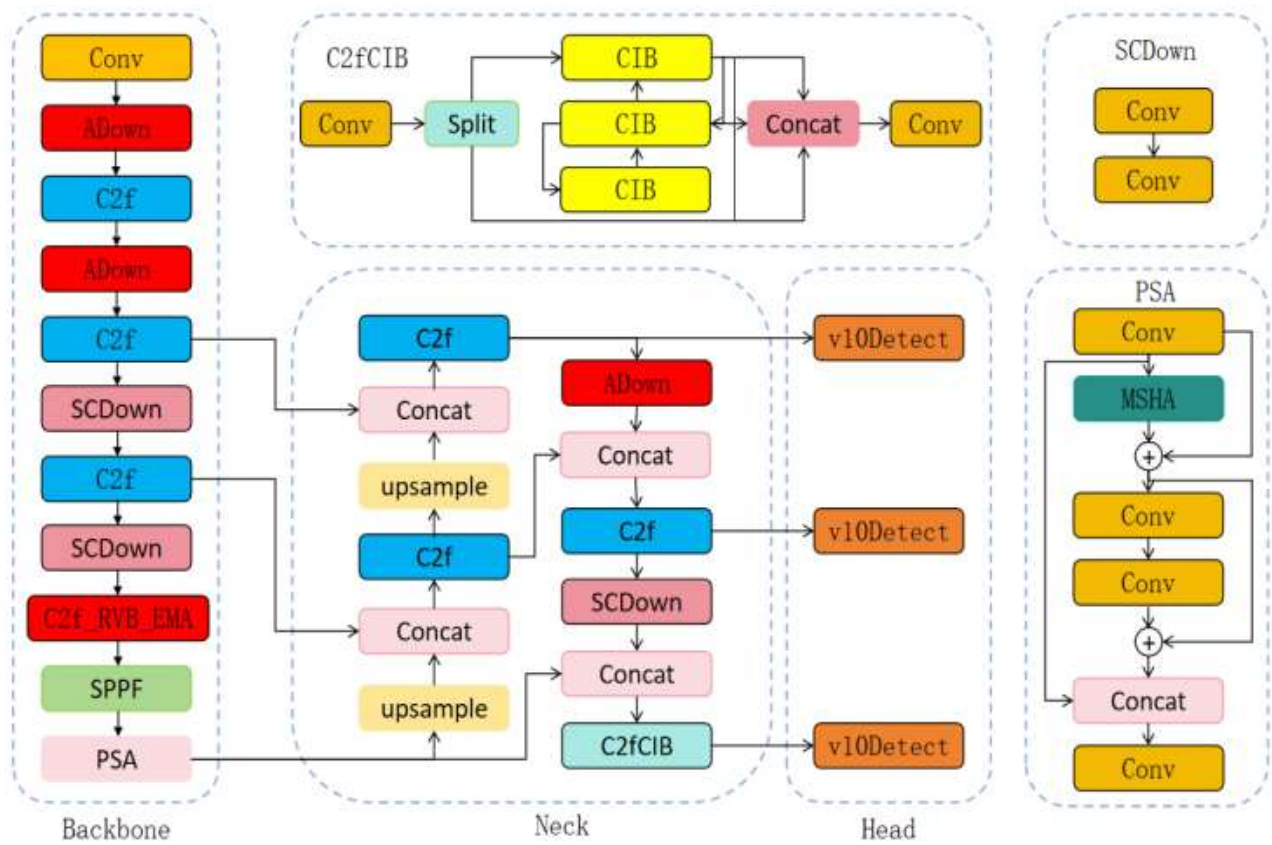


Figure 6. YOLOv10-ARE model architecture

3. Analysis of Experimental Results

3.1. Experimental Environment and Parameter Setting

This experiment uses Ali Cloud server, GPU is NVIDIA A10, memory is 30G. Python 3.10.14, PyTorch2.3.0, Cuda12.1. The default parameters of the model were used to train the rolling bearing defect images. The image size was 640×640, the initial learning rate was 0.01, the number of training rounds was 300, and batch=32.

3.2. Production of Bearing Defect Data Set

Defective rolling bearings from the industrial production process were collected, and additional defects were artificially introduced. Surface defect images of the bearings were captured and categorized into three types:



scratches, abrasions, and grooves. This dataset comprises a total of 3,984 defect images. To augment the dataset, certain defect images underwent transformations such as flipping and blurring. Representative defect images are illustrated in Figure 7. For comprehensive model evaluation, the dataset was divided into training, validation, and test sets in a ratio of 7:1:2.

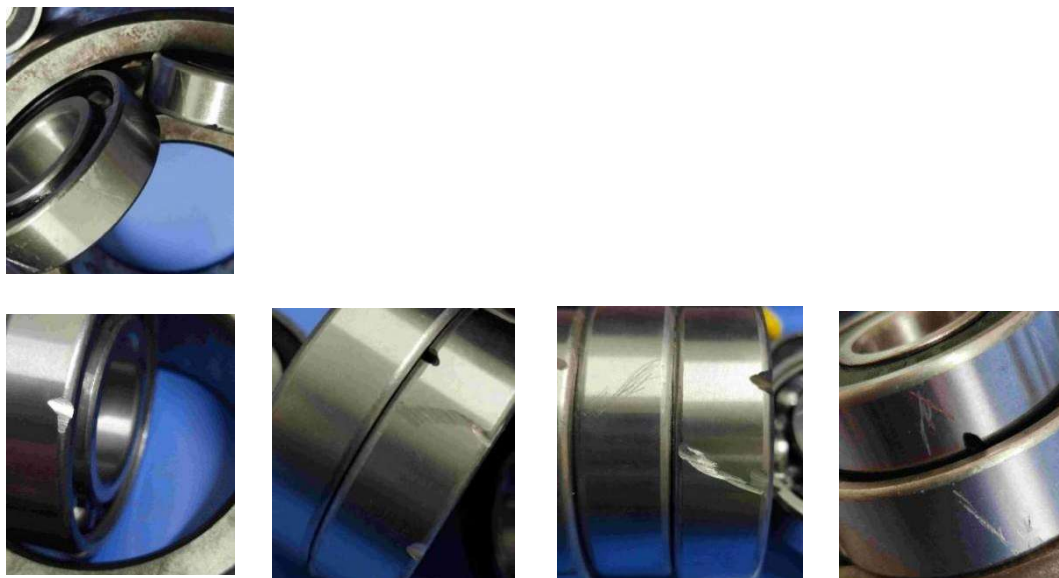


Figure 7. Pictures of part of the bearing defect data set.

3.3. Evaluation Index

In this experiment, the number of model parameters, model size and model operation complexity (GFLOPs) were used to evaluate the lightweight of the improved model. Accuracy rate P, recall rate R, mAP50, MAP50-95 were used to evaluate the accuracy of the improved model. The

detection rate of the improved model was evaluated using the frame rate FPS. The calculation formula of each index is as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \tag{3}$$

3.4. Comparative Analysis of Experimental Results

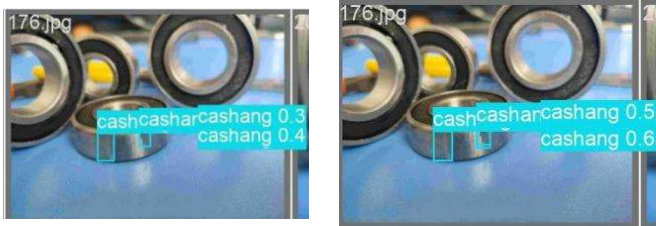
In order to evaluate the influence of ADown module and C2f-RVB-EMA module on the original algorithm and the effectiveness of YOLO-ARE, and to provide a valuable reference for improving the performance of the rolling bearing flaw detection algorithm, this study designed an ablation experiment on a self-made data set. The results are shown in Table 1.

Table 1. Results of ablation experiment.

Serial number	YOLOv10n	ADown	C2f-RVB-EMA	P	R	mAP50	mAP50-95	Parameter	GFLOPs	Model size
1	√	×	×	0.938	0.919	0.959	0.624	2.2M	6.5G	5.8M
2	√	√	×	0.938	0.921	0.961	0.624	2.1M	6.1G	5.7M
3	√	√	√	0.944	0.92	0.962	0.623	1.9M	5.9G	5.3M

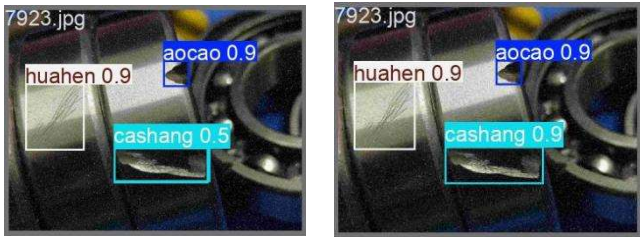
This experiment was conducted to optimize and enhance YOLOv10n. By comparing the first and second groups of experiments in the table, it is evident that after incorporating the ADown module into YOLOv10n, the recall (R) increased by 0.3%, mean average precision at 50% overlap (mAP50) improved by 0.2%, while precision (P) and mean average precision from 50% to 95% overlap (mAP50-95) remained constant. Additionally, the number of parameters decreased by 0.1 million (M), GFLOPs reduced by 0.4 billion (G), and the model size shrank by 0.1M. These results demonstrate that the addition of the ADown module not only enhanced the lightweight nature of the model but also improved its detection accuracy and ability to extract surface defect features of bearings. By comparing the second and third groups of experiments in the table, after integrating the C2f-RVB-EMA module into YOLOv10n+ADown, precision (P) increased by 0.6%, mAP50 improved by 0.1%, while recall (R) and mAP50-95 decreased slightly by 0.1%. The number of parameters decreased by 0.2M, GFLOPs reduced by 0.2G, and the model size decreased by 0.4M. These findings indicate that the integration of the C2f-RVB-EMA module further achieved model lightweighting without significantly compromising the model's accuracy.

Compared with the model before and after improvement, P increased from 0.938 to 0.944, an increase of 0.6%. R increased by 0.1% from 0.919 to 0.92. mAP50 increased from 0.959 to 0.962, an increase of 0.3%. mAP50-95 increased from 0.624 to 0.623, decreased by 0.1%, and the model detection accuracy was improved to a certain extent. The size of the model decreased from 5.8M to 5.3M, decreased by 8.6%, the number of parameters decreased from 2.2M to 1.9M, decreased by 13.6%, and the calculation volume decreased from 6.5G to 5.9G, decreased by 9.2%. While the model is lightweight, the model detection accuracy is improved. The detection effect before and after model optimization is shown in Figure 8.



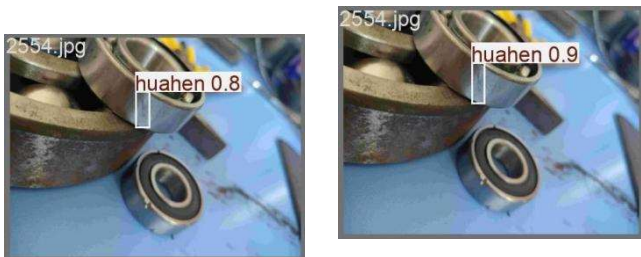
(a) Original model
detection effect 1

(b) Detection effect
after optimization 1



(c) Original model
detection effect 2

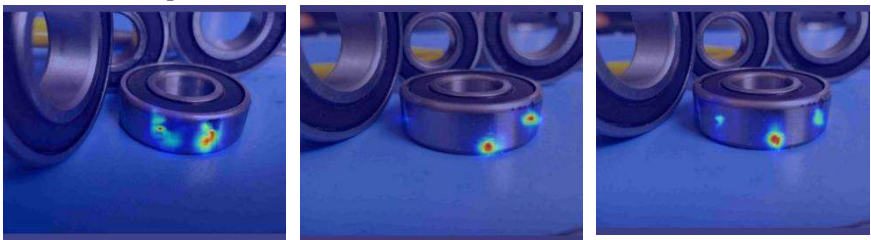
(d) Detection effect
after optimization 2



(e) Original model
detection effect 3

(f) Detection effect
after optimization 3

Figure 8 Comparison of model recognition effect before and after improvementIn order to display the detection performance of model YOLO-ARE and visualize the detection effect of rolling bearing characteristic defects, the thermal map before and after model optimization is shown in Figure 9. It can be



seen from the figure that YOLO-ARE focuses more on the extraction of fine defect features of rolling bearings, and the highlighted area in the figure has a higher coincidence degree with the defect part of the rolling bearings in the original figure, indicating that the algorithm model proposed in this experiment focuses more on the defect features of rolling bearings and can better suppress background noise.

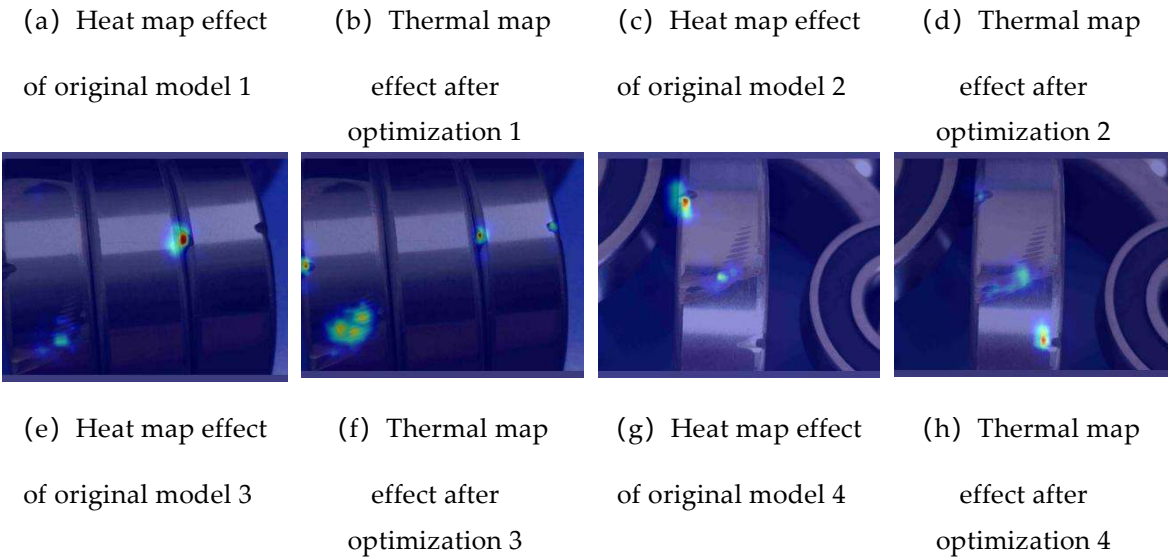


Figure 9. Comparison of model heat maps before and after improvement.

The comparison between the original model and the improved model YOLOv10-ARE in terms of performance indicators P, R, mAP50 and MAP50-95 is shown in Figure 10. The orange curve represents the original model, and the blue curve represents the improved model YOLOv10-ARE. The comparison figure shows that under the same number of training rounds, the convergence speed, P, R, mAP50 and MAP50-95 of the YOLOv10-ARE model are slightly better than the original model, indicating that the overall performance of the improved model YOLOv10-ARE is better than the original model.

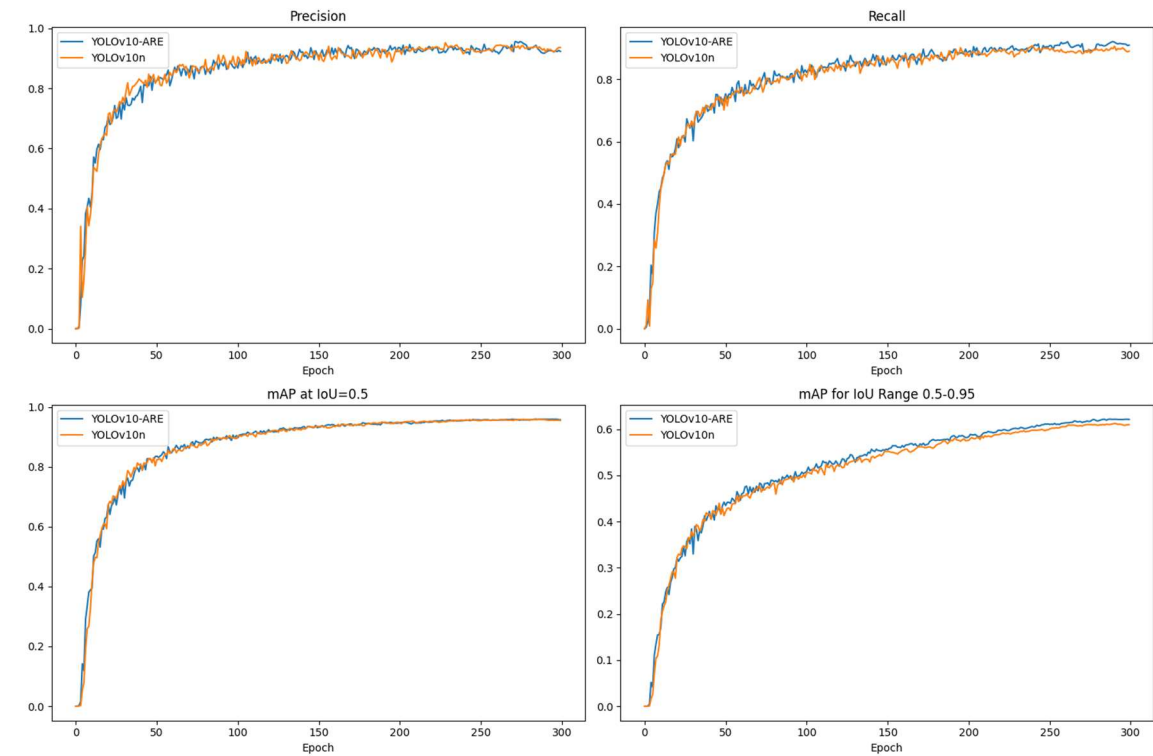


Figure 10. Comparison of model performance before and after improvement.

3.5. Comparison Experiment with Other Algorithms

In order to verify the performance of the model YOLOv10-ARE in this experiment in detecting surface defects of rolling bearings, under the same conditions, the model is compared with several common mainstream algorithms YOLOv7-tiny, YOLOv8n and YOLOv10n respectively, and the experimental results are shown in Table 2.

Table 2. Performance comparison experiments of different algorithms.

Model	P	R	mAP50	mAP50-95	FPS / (f/s)	GFLOPs/G	Parameter/M	Model size/MB
YOLOv7-tiny	0.926	0.914	0.951	0.572	98	13.0	5.7	12.3
YOLOv8n	0.952	0.924	0.964	0.649	114.2	8.1	2.9	6.3
YOLOv10n	0.938	0.919	0.959	0.624	176.2	6.5	2.2	5.8
Model of this paper	0.944	0.92	0.962	0.623	158.1	5.9	1.9	5.3

By comparing the model parameters in Table 2, it can be found that the experimental model YOLO-ARE is basically consistent with YOLOv10n in terms of recall rate R and mAP50-95, which is inferior to YOLOv8n. The accuracy of P and mAP50 is better than that of YOLOv10n and YOLOv7-tiny, but slightly lower than that of YOLOv8n. However, GFLOPs of this model reach 5.9G, model parameters 1.9M and model size 5.3MB, which is significantly superior to other models in terms of

lightweight. Considering the existing requirements of industrial field testing, YOLOv10-ARE model can be placed on the edge end to meet the requirements of rolling bearing testing.

4. Conclusion

Aiming at the problems of insufficient lightweight and incompatibility between accuracy and model complexity of previous rolling bearing surface defect detection algorithms, this study proposed a rolling bearing surface defect detection algorithm based on improved YOLOv10. The innovative introduction of the undersampling module ADown in the Backbone and neck Head is conducive to the model to capture bearing defect image features, reduce the amount of computation, and optimize the accuracy of target detection. The fusion C2f-RVB-EMA module was inserted into the backbone network to optimize the model's focus on the surface defect features of the rolling bearing, further reduce the complexity of the model and improve the lightweight level of the model. The experimental results show that the model YOLO-ARE performs better than other models in parameters number, model size and operational complexity, and the detection accuracy has been improved slightly, although it is slightly worse than YOLOv8n, it achieves the coordination of high detection accuracy and lightweight, and meets the needs of industrial detection. The next step is to improve the accuracy and detection efficiency under the premise of ensuring the lightweight of the model.

Fund Project: Henan Science and Technology Research and Development Program Joint Fund (232103810038); Key Research Project of Colleges and Universities of Henan Province (24A460009).

References:

1. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Colum- bus: IEEE, 2014: 580–587.
2. Girshick R. Fast R-CNN [J]. 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
3. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards realtime object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
4. He K, Gkioxari G, Dollar P, et al. Mask R-CNN [C]// 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2961–2969.
5. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779–788.
6. Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263–7271.
7. Redmon J, Farhadi A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08). <https://arxiv.org/pdf/1804.02767.pdf>.
8. Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection [EB/OL]. (2020-04-23). <https://arxiv.org/pdf/2004.10934 v1.pdf>.
9. Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. (2021-08-06). <https://arxiv.org/pdf/2107.08430.pdf>.
10. Gao Liming, Jia Shuhai, Zhang Guolong, et al. Surface Defect Detection Method of insulated Bearing Based on Improved Faster R-CNN [J]. Journal of Bearings, 2023,(4):1-8.
11. Yuan Tian-Le, YUAN Julong, ZHU Yong-Jian, et al. Surface Defect Detection Algorithm of thrust Ball Bearing Based on Improved YOLOv5 [J]. 2022,56(12):2349-2357.
12. Gu Yunpeng, Ma Chao, Zang Shaofei, et al. U-Net segmentation model of bearing roller appearance defects based on partial convolution and ECA mechanism [J/OL]. Bearings. <https://link.cnki.net/urlid/41.1148.th.20231116.1132.002>
13. Yao Jingli, Cheng Guang, Wan Fei, et al. Improve YOLOv8 lightweight bearing defect detection algorithm [J/OL]. Computer engineering and application. <https://link.cnki.net/urlid/11.2127.TP.20240815.1528.014>.

14. Peng Yanfei, LI Dongxue, Chen Xitao. Improve YOLOv5s lightweight bearing defect detection method [J/OL]. Mechanical science and technology. <https://doi.org/10.13433/j.cnki.1003-8728.20230342>.
15. Ao Wang, Hui Chen, Lihao Liu, et al. YOLOv10: Real-Time End-to-End Object Detection[J], arXiv preprint
16. Chien-Yao Wang, I-Hau Yeh², Hong-Yuan Mark Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information[J], arXiv:2402.13616, 2024. <https://doi.org/10.48550/arXiv.2402.13616>.
17. Daliang Ouyang, Su He, Guozhong Zhang, et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning[J], arXiv:2305.13563, 2023. <https://arxiv.org/abs/2305.13563>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.