RESEARCH ON BEARING SURFACE DEFECT DETECTION BASED ON IMPROVING YOLOV5S

Changli ZHA*[1],Shining TANG[2],Jianglin CHENG[3],Tao CHENG[4]

Deep learning has advanced intelligent defect detection in bearing manufacturing. This research introduces an improved algorithm to address YOLOv5s network limitations in identifying minor defects in micro bearings. The original backbone network is replaced by the innovative PoolFormer module, while the lightweight CARAFE upsampler substitutes the initial feature sampler, reducing computational complexity and enhancing token feature aggregation. Additionally, a specialized detection layer is integrated into the feature fusion network of the YOLOv5s model to improve small target identification. Modifications to the loss function include the incorporation of WIoU loss, confidence scaling loss, and classification Bce loss. Experimental results demonstrate that the enhanced YOLOv5s network model achieves a 93.7% accuracy rate on a unique bearing defect dataset. In platform-based detection, the updated YOLOv5s model surpasses the original by 18% in defect detection accuracy, showcasing significant improvements in accuracy and reliability.

Keywords: Deep learning; YOLOv5s; Bearing defect; Machine vision; Small object detection

1. Introduction

The identification of surface defects on bearings has been a long-standing area of study. Besides manual examination, common methods include eddy current

[1]* Prof., Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, China, e-mail: zhachangli@126.com (corresponding author)

[2] Ms., Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, China, e-mail: 1054422023@qq.com

[3] Ms., Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, China, e-mail: 645122900@qq.com

[4] Ms., Electronic Engineering and Intelligent Manufacturing, Anqing Normal University, China, e-mail: 2392778076@qq.com

[1] and magnetic flux leakage inspections [2]. The primary challenge is detecting defects efficiently and effectively while minimizing manual labor and inspection expenses. In response, researchers have explored traditional image processing techniques like scale-invariant feature transformation [3], oriented gradient histograms [4], and local binary patterns [5]. However, these methods face drawbacks in terms of subjectivity and cost, particularly when identifying small targets and irregular defects. The advent of convolutional neural networks (CNN) has shifted the focus from traditional feature extraction algorithms to neural network technology for target recognition. Currently, the leading deep learning algorithms for object recognition include the two-stage framework based on the region-CNN (R-CNN) [6-8] series and the single-stage algorithm, which treats target detection as a regression issue. The R-CNN framework uses a selective search algorithm to generate proposals, which are then input into a CNN model for feature extraction. A support vector machine uses the extracted features for object prediction and classification. Although this framework offers high object detection accuracy, it operates slowly. In contrast, single-stage algorithms such as the YOLO [9] series and SSD [10] directly determine the category and location of targets from input images. These methods offer rapid processing speed but have lower target detection accuracy.

The utilization of advanced deep learning models for target detection has revolutionized the field of object surface defect identification. Chen Qi [11] used local topography and image subtraction to extract defect characteristics on the inner surface of bearings, employing a support vector machine (SVM) classifier for defect categorization. Kunakornvong et al. [12] detected air bearing defects by analyzing variations in brightness. They used a co-occurrence matrix to mitigate the impact of brightness alterations on air bearing images, defining characteristic parameters based on four distinguishing features and utilizing the Euclidean distance method to determine defect identification thresholds. Deng et al. [13] employed fitting techniques and circular scanning to delineate the detection area of bearings, enhancing image quality through contrast enhancement and low-pass filtering methods. Feng et al. [14] proposed a defect detection network based on a priori models to detect pit defects at different scales on the bearing surface. Experimental results show that the accuracy of this method for detecting pit defects of different sizes reaches 99.3%, which surpasses other existing methods by 2 to 5%. Gao et al. [15] improved Faster R-CNN method for image detection of surface defects. The outcomes showcased an enhancement in the detection accuracy of

surface defects in insulating bearings to 91.2%, which is 4.8% higher than the original Faster R-CNN method.

Advancements have been made in detecting defects on bearing surfaces, yet limitations persist in key areas such as accuracy and speed for identifying small targets. This research project employs the YOLOv5s algorithm as the base for detecting bearing surface defects, substituting the original backbone network with the PoolFormer module. Additionally, the lightweight CARAFE upsampler replaces the original sampler for feature maps, reducing computational complexity and improving feature aggregation. Moreover, a layer for detecting small targets is integrated into the feature fusion network based on YOLOv5s to enhance small target identification. The implementation of a new loss function, which includes WIoU for localization, confidence scaling, and Bce for classification, further boosts detection capabilities. The study concludes with an experimental setup to validate the theoretical findings.

2. YOLOv5

The YOLO series has revolutionized object detection with its innovative algorithm that combines a region of interest (ROI) module and a detection stage to increase detection speed. In 2020, the release of YOLOv5 [16], the fifth generation, marked a significant advancement in deep learning algorithms for target detection. Building on the success of YOLOv4, YOLOv5 further improves detection accuracy. Despite its compact size of just 28 MB, YOLOv5's weight files are highly effective, making it an ideal starting model. The impressive performance of YOLOv5 is evident in its results on the COCO dataset. The YOLOv5 model comprises four main elements: Input, Backbone, Neck, and Head.

3. Algorithm improvement

3.1 Algorithm framework

This study introduces a suggested framework for identifying surface defects on bearings employing a Transformer model structure and merging features at multiple scales. The procedural framework includes the following steps: (1) Capture images of flawed bearings using a camera, then filter, label, augment, and segment the images to create the dataset. (2) Enhance the YOLOv5s network by replacing the convolutional and C3 modules of the foundational network with the Poolformer module. Add a layer for detecting small targets in the YOLOv5s feature

fusion network to enable multi-scale feature fusion. (3) Train the model using the training dataset and evaluate its performance with the test set images from the dataset.

3.2 Poolformer

The PoolFormer [17] approach, developed by Yan Shuicheng's research team, validates the effectiveness of the Transformer model's structure. Within this design, the token mixer merges information across tokens, while the channel MLP consists of two MLP layers and non-linear activation functions. PoolFormer replaces the Transformer's attention module with a non-parametric spatial average pooling layer known as Pooling. Unlike the attention module, Pooling lacks trainable parameters, reducing computational complexity and improving the aggregation of neighboring token attributes. Additionally, the model features four stages with embedding dimensions of 64, 128, 320, and 512. Assuming a total of L blocks in the PoolFormer model, stages 1, 2, 3, and 4 contain L/6, L/6, L/2, and L/6 blocks, respectively.

3.3 Upsampling operators: CARAFE

Upsampling is used in modern convolutional network designs for object detection, instance segmentation, and scene resolution enhancement. Wang et al. introduced a CARAFE[18] upsampling operator that implements content-aware feature rearrangement. CARAFE predicts the recombined kernel and reorganizes features within a specified neighborhood based on the content at each location. This operator achieves significant improvements with minimal additional parameters and computational workload. CARAFE surpasses traditional upsampling operators like interpolation or deconvolution by using adaptive and optimized recombined kernels at different positions.

3.4 Adding the small size detection module

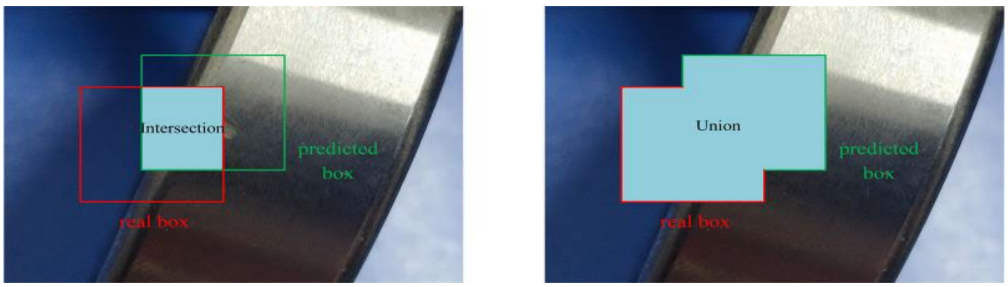
The initial YOLOv5 model uses feature maps of different sizes to detect targets of varying scales within an image. The shallow network feature map provides detailed location information, while the deep network feature map offers more semantic details. Larger targets are identified using an 80×80 feature map, medium-sized targets with a 40×40 feature map, and smaller targets with a 20×20 feature map. In images depicting bearing defects, most defects are considered small targets, representing only a small portion of the overall image. The original network might overlook positional details from the shallow network when detecting these

small targets, potentially impacting the detection process. To address this issue, an additional small-scale detection layer has been incorporated into the YOLOv5 network structure, increasing the number of detection layers from 3 to 4. This addition introduces minimal parameters, reducing the risk of losing important feature information and enhancing the model's ability to detect small targets.

3.5 Loss Function Improvements

1) Optimize the regression box loss function

The IoU loss is calculated by determining the Intersection over Union (IoU) between the real and predicted boxes. Fig. 1 illustrates the plot of IoU loss.



(a) Intersection (b) Union

Fig. 1. The plot of IoU loss.

Box C is the smallest box that can enclose both the real box A and the predicted box B. Equation 1 defines the IoU loss.

$$L_{IoU} = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

In YOLOv5s, Complete IoU (CIoU) is used to calculate the localization loss, and the calculation formula is shown in Equation 2.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b^A, b^B\right)}{c^2} + \alpha v \tag{2}$$

Among them: b^A and b^B represent the centroid points of the prediction box and the real box respectively. The Euclidean distance between the two points is denoted as ρ. The diagonal length of the minimum circumscribed rectangle of the prediction box and the real box is represented by c. Parameter α indicates the balance between these distances, while parameter v evaluates the aspect ratio consistency between the predicted frame and the target frame.

The CIoU algorithm considers overlapping areas, center point distances, and aspect ratio consistencies in boundary regression to enhance the convergence accuracy of the detection frame and stabilize the regression process. However, the

variable v in Equation 3 may not accurately reflect the height-width difference and confidence, limiting the model's ability to optimize similarity. CIoU relies on high-quality examples in the training data to improve bounding box loss fitting through a monotonic focusing mechanism. Yet, blindly reinforcing boundaries can reduce detection performance if the target detection dataset includes low-quality examples. To address this issue, the article suggests using WIoU (Wise-IoU) [19] as the border regression loss function, calculated as follows.

$$L_{WIOU} = \mathrm{R}_{WIoU} L_{IoU} \tag{3}$$

$$\mathrm{R}_{WIoU} = \exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{\left(W_g^2 + H_g^2\right)^*}\right) \tag{4}$$

Among them: R_{WIoU} is the distance metric; W_g and H_g are the width and height of the minimum bounding box; x_{gt} and y_{gt} are the center points of the real box. $\mathrm{R}_{WIoU} \in [1,e)$, significantly amplify the L_{IOU} of the ordinary mass anchor box. $L_{IoU} \in [0,1]$, significantly reduce the R_{WIoU} of the high-quality anchor box and its attention to the distance of the center point when the anchor box coincides well with the target box.

Based on the distance metric, WioU develops a distance attention mechanism that reduces the geometric metric's impact on the model training without significantly hindering the overlapping of the anchor frame and target frame.

2) Adding Varifocal loss confidence loss function

To address the issue of dense targets and uneven positive and negative samples in the dataset, this article introduces the varifocal loss function. The varifocal loss function enhances the confidence loss function by ensuring equal treatment of positive and negative samples and mitigating category imbalance during training. This is defined as shown in Equation 5.

$$VFL(p,q) = \begin{cases} -q(q\log(p) + (1-q)\log(1-p)), & q > 0 \\ -ap^{\gamma}\log(1-p), & q = 0 \end{cases} \tag{5}$$

In the context of the equation, p represents the forecasted classification score, while q represents the desired score. For positive instances, q corresponds to the Intersection over Union (IoU) value, comparing the actual frame to the predicted frame, and for negative instances, q indicates a target score of 0. The varifocal loss formula incorporates a parameter that regulates scaling by reducing the loss associated with negative instances while maintaining the loss for positive instances. This approach enhances the overall training effectiveness and model performance.

3) The improved loss function

This paper introduces a novel loss function that integrates the WIoU loss function for localization, the varifocal loss function for confidence, and the BCE (Binary Cross Entropy) loss function for classification. The improved loss function addresses specific categories and the overall loss function in detail,

$$Loss_{IoU} = \exp\left(\frac{\left(x-x_{gt}\right)^2+\left(y-y_{gt}\right)^2}{\left(W_g^2+H_g^2\right)^*}\right)L_{IoU} \tag{6}$$

$$Loss_{obj} = \begin{cases} -q(q\log(p)+(1-q)\log(1-p)), & q>0 \\ -ap^{\gamma}\log(1-p), & q=0 \end{cases} \tag{7}$$

$$Loss_{\text{cls}} = -\sum_{i=1}^{N}\left(t_i\log\left(p_i\right)+\left(1-t_i\right)\log\left(1-p_i\right)\right) \tag{8}$$

$$Loss = \frac{1}{N}\left(\lambda_{\text{IoU}}Loss_{\text{IoU}}+\lambda_{\text{obj}}Loss_{\text{obj}}+\lambda_{\text{cls}}Loss_{\text{cls}}\right) \tag{9}$$

Among the variables listed, N represents the number of samples determined by the batch size. The positioning loss function, reliability loss function, classification loss function, and overall batch size loss function are denoted by $Loss_{\text{IOU}}$, $Loss_{\text{obj}}$, $Loss_{\text{cls}}$, and $Loss$, respectively, during the training process. Each step employs a global loss function for back propagation and weight adjustments.

3.6 Improved YOLOv5s Network Structure

The YOLOv5s algorithm network exhibited reduced detection accuracy on the selected dataset, with instances of false positives and missed detections. To address this issue. Initially, the convolution module and C3 module of the backbone network are replaced with the Poolformer module. The original sampler for feature map sampling is also swapped with the lightweight CARAFE upsampler. These modifications aim to reduce computational complexity and improve the approximation of token feature combinations. A 160×160 small target detection layer is incorporated into the feature fusion network to improve the model's detection performance on small target samples. Subsequently, a new loss function is introduced, combining the WIoU loss function for positioning, the varifocal loss function for confidence, and the Bce loss function for classification. The enhanced YOLOv5s network model is depicted in Fig. 2.

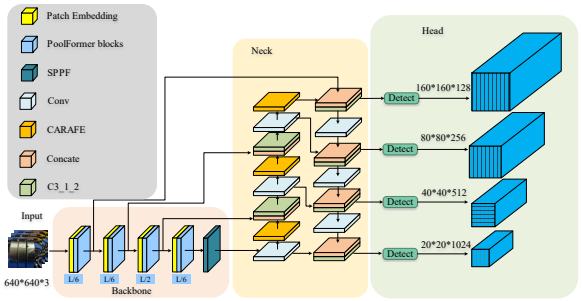


Fig. 2. Improved YOLOv5s Network Structure.

4 Experiment

The study described in this manuscript employed the PyTorch 1.7.1 framework for deep learning. The hardware setup included an Intel i5-12400F CPU running at 2.50 GHz and an NVIDIA GeForce RTX 3060 GPU with 12GB of VRAM.

4.1 Dataset

Data collection was conducted in real-time at the workbench, categorizing surface imperfections into bruises, scratches, and grooves. A total of 600 images were collected for each type of imperfection, resulting in a dataset of 1,800 images. Some images were randomly rotated and scaled, increasing the dataset size to 5,800 images. Flawed regions were annotated using Labelme software to label rectangular boxes of various dimensions. Fig. 3 showcases the three defect classifications.

(a) (b) (c) (d)

Fig. 3. Types of defects. (a) All three defects, (b) Grooves, (c) Bruised and (d) Scratched.

4.2 Model training

The input image size was to 640 × 640 pixels during model training. An IOU threshold of 0.5 was used. The network model underwent 300 training iterations with a batch size of 8, using the SGD optimizer and a linear decay learning rate scheduling strategy. The initial learning rate was set to 0.01 and gradually decreased to 0.0001. The momentum parameter was set to 0.937, and the weight decay coefficient to 0.0005. All models were trained with these specified parameters.

4.3 Evaluation index

To objectively assess the proposed detection method, this study utilizes evaluation criteria consistent with MS COCO. The model accuracy evaluation criteria include precision (P), recall (R), average precision (AP), and mean average precision (mAP). Detection speed is assessed using frames per second (FPS) as the performance index. To compare the computational burden of different networks, computational time complexity (FLOPs) is selected for method differentiation.

4.4 Loss function curve

Fig. 4 illustrates the training loss function curve for the target detection algorithm. It reveals that the loss values of the YOLOv5s_improved model consistently remain lower than those of the YOLOv5s model. These findings demonstrate that the YOLOv5s_improved model proposed in this study converges more rapidly during training. Moreover, the loss curve for the YOLOv5s_improved model shows smoother fluctuations and reduced numerical deviations, indicating that the proposed model exhibits improved stability during the training process.

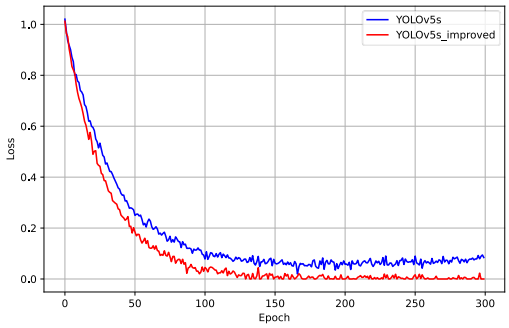


Fig. 4. Comparison of YOLOv5s and improved YOLOv5s loss function cu

rves.

5 Results

5.1 Ablation experiment

1) WioU

Experiments were conducted to determine the most effective model for dataset detection, comparing the three models. Fig. 5 presents the PR curve analysis of the YOLOv5s_WIoU model using different WIoU versions. The analysis shows that the WIoUv1 loss function, incorporating the attention mechanism, yielded the best detection results. Consequently, WIoUv1 was chosen as the final model for the experiment.

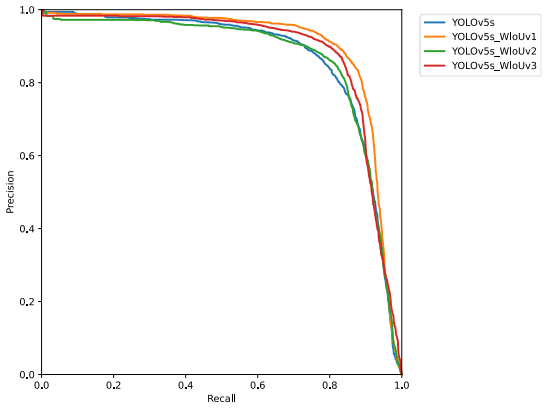


Fig. 5. Comparison PR curve of detection performance under different WIoU models.

2) Loss weight optimization

To detect bearing defects effectively, this study aimed to reduce background interference and enhance the recognition of target regions by adjusting the weights of the loss function components. Initially, equal importance was assigned to localization, objectness, and classification tasks, with weights set to $Loss_{\mathrm{IOU}}=1$, $Loss_{\mathrm{obj}}=1$, and $Loss_{\mathrm{cls}}=1$. Systematic adjustments were then made to explore different task priorities. Increasing the IoU loss weight to 1.5 while reducing the objectness and classification weights to 0.75 improved bounding box localization precision. Similarly, increasing the objectness loss weight to 1.5 enhanced the model's ability to identify target regions, and increasing the classification loss weight to 1.5 improved the recognition of defect categories. To balance localization precision and target presence detection, a configuration with $Loss_{\mathrm{IOU}}=1.1$,

$Loss_{\text{obj}} = 1.1$, and $Loss_{\text{cls}} = 0.8$ was selected as achieving optimal overall performance. The experimental results are shown in Table 1.

Table 1

Performance Comparison Under Different Loss Weight Configurations

λ_{IoU}	λ_{obj}	λ_{cls}	mAP50 (%)	mAP50:95 (%)
1	1	1	87.5	43.3
1.5	0.75	0.75	88.3	45.3
0.75	1.5	0.75	88.1	44.9
0.75	0.75	1.5	87.0	42.8
1.1	1.1	0.8	88.7	45.8

Table 1 presents the impact of different loss weight configurations on bearing defect detection performance. Based on the comparison of experimental results, the weight configuration $Loss_{\text{IOU}} = 1.1$, $Loss_{\text{obj}} = 1.1$, and $Loss_{\text{cls}} = 0.8$ was ultimately determined as the final parameter setting. Under this configuration, mAP50 reached 88.7% and mAP50:95 reached 45.8%, achieving the best detection performance. Therefore, this configuration is adopted as the final parameter indicator in this study.

3) Upsampling operators: CARAFE

The precision and speed of the upsampling technique CARAFE can be affected by the hyperparameters $k_{encoder}$ k_{up} . To find the right balance between precision and speed, this study experiments on these two parameters are conducted. The outcomes of these experiments are highlighted in Table 2.

Table 2

Comparison of model checking performance of CARAFE with different parameters

$k_{encoder}$	k_{up}	mAP50 (%)	mAP50:95 (%)	GFLOPs(G)
1	3	89.3	44.0	15.8
1	5	89.4	44.2	15.8
3	3	89.4	44.6	15.9
3	5	90.9	46.8	16.0
3	7	90.2	46	16.3
5	5	90.5	45.5	16.5
5	7	91.2	47.9	17.1
7	7	90.8	47.3	17.3

Table 2 shows that enhancing $k_{encoder}$ and k_{up} can improve the model's performance. However, boosting one parameter alone does not significantly impact the model's performance. The findings suggest that higher values for both parameters improve performance, but increasing the parameters also increases the model's computational complexity. To achieve a balance between precision and efficiency, $k_{encoder}=3$ and $k_{up}=5$ were chosen as the optimal parameter values for the experiment.

4) The detection effect of the improved model

The loss weight configuration in this experiment is $Loss_{\mathrm{IOU}}=1.1$, $Loss_{\mathrm{obj}}=1.1$ and $Loss_{\mathrm{cls}}=0.8$. Abilative experiments were conducted to assess the impact of six integrated modules in the YOLOv5s network model. The results from the experiments are outlined in Table 3. The incorporation of the small target detection layer significantly enhances accuracy, leading to a 4.8% increase in mAP. While WIoUV1 and Varifocal loss modules improve detection accuracy, they also introduce additional parameters that marginally affect detection speed. In contrast, the CARAFE upsampling operator and Poolformer module are streamlined models that enhance network speed. Combining these modules results in a 6.2% increase in mAP, demonstrating enhanced detection accuracy. Despite a slight decrease in detection speed, the model remains viable for real-time bearing defect detection.

Table 3

Ablation experiment

Number	WIoU	vfloss	CARAFE	New detection layer	Poolformer	mAP50 (%)	mAP50:95 (%)	FPS
1						88.7	45.8	114
2	√					90.6	46.4	108
3		√				89.9	45.6	110
4			√			91.7	47.8	118
5				√		92.8	51.5	94
6					√	91.6	48.8	112
7	√	√				91.3	47.9	98
8	√	√	√			92.1	51.2	112
9	√	√	√	√		92.8	51.8	93
10	√	√	√	√	√	94.4.	57.6	104

5.2 Comparative experiment

The effectiveness of the proposed algorithm in detecting bearing defects was thoroughly assessed through comparisons with various other algorithms, including YOLOv5s, YOLOv3, YOLOv3-tiny, SSD, and Faster R-CNN. The results of these experiments are presented in Table 4. The proposed approach outperforms SSD and YOLOv3 by 21.1% and 10.2% in terms of mAP, respectively. It also surpasses Faster R-CNN by 8.4%, indicating superior performance over single-stage and two-stage detection models. Although there is a decrease in detection speed compared to the original YOLOv5s model, the improved model still shows potential for real-time detection. Additionally, detection accuracy increased from 87.5% to 93.7%, significantly enhancing precision in detecting bearing defects and meeting industry standards for production.

Table 4

Comparison of detection performance of different models

Method	mAP50 (%)	mAP50:95 (%)	FPS
SSD300	72.6	37.1	77
Faster R-CNN	85.3	42.4	56
YOLOv3	83.5	41.3	83
YOLOv3-tiny	69.6	35.9	125
YOLOv5s	87.5	43.3	111
YOLOv5s_improved	94.4	57.6	104

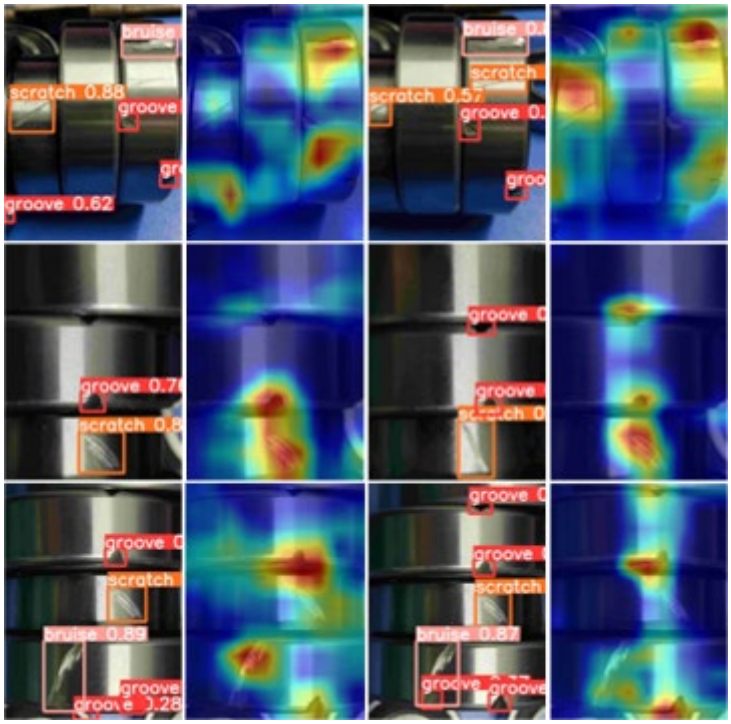
5.3 Bearing defect detection results

To evaluate the effectiveness of the improved YOLOv5s algorithm, surface defect detection was performed on 50 small bearings with grooves, scratches, and abrasions using a visual inspection platform. The results were then compared with those obtained using the original YOLOv5s algorithm, as presented in Fig 6 and Table 5.

Table 5

Comparison of defect detection accuracy

Method	False (EA)	Missed (EA)	Accuracy rate (%)
Yolov5s	11	5	68
YOLOv5s_improved	3	4	86



(a) (b) (c) (d)

Fig. 6. Bearing defect detection effect diagram and Grad-CAM. (a) using the original YOLOv5s network model, (b) Grad-CAM, (c) improved YOLOv5s network model and (d) Grad-CAM

The research revealed that the initial YOLOv5s network structure had limitations in precisely identifying minor defects on bearing surfaces. However, after enhancements were made, there was a notable 18% improvement in accuracy. Results for defect identification with the basic model are shown in Fig. 6(a) and 6(b), while the outcomes with the upgraded model are displayed in Fig. 6(c) and 6(d). Fig. 6(a) illustrates instances where the original model failed to identify certain defects, while Fig. 6(b) utilizes Grad-CAM to pinpoint areas requiring attention. In contrast, the improved model effectively identified either deeper or less conspicuous defects, as shown in Fig. 6(c) and 6(d). These results validate the efficacy of the proposed approach in enhancing defect detection accuracy.

6. Conclusions

The YOLOv5s model demonstrates poor detection accuracy on the specific dataset. To address this issue, an upgraded algorithm based on the YOLOv5s design is proposed in this study. The Backbone network is replaced by the Poolformer

module, which incorporates the CARAFE upsampler instead of the original sampler for sampling feature maps. These modifications aim to decrease computational complexity and improve the aggregation of token features. A multi-scale detection strategy is employed to enhance the detection of small surface defects in real-world scenarios. A specialized 160x160 layer for small targets is integrated to improve the extraction of small targets and features. Utilizing shallow layers for feature fusion enhances the model's ability to extract spatial information. Furthermore, the loss function is fine-tuned to boost network convergence and regression performance during training. Results from experiments indicate that the advanced YOLOv5s model achieved a detection accuracy of 93.7% on the custom surface defect dataset and in platform-based detection, the upgraded YOLOv5s showed an 18% improvement in accuracy compared to the original YOLOv5s model.

Acknowledgement

This work was financially supported by the Quality Engineering Project for Education in the New Era (Graduate Education) of Anhui Province of China (Grant No. 2023qygzz029 and 2023cxcysj139).

REFERENCES

[1]. *G-M, Javier, J Gómez-Gil, and E Vázquez-Sánchez*,"Non-destructive Techniques Based On Eddy Current Testing", in Sensors., vol. 11, no. 3, 2011, pp. 2525-2565.

[2]. *Y.Shi, C.Zhang, R.Li, and G.Jia,"*Theory And Application Of Magnetic Flux Leakage Pipeline Detection", in Sensors., vol. 15, no. 12, 2015, pp. 31036-31055.

[3]. *D.G.Lowe*,"Object Recognition From Local Scale-invariant Features", in Proceedings of the seventh IEEE international conference on computer vision., vol. 2, 1999, pp. 1150-1157.

[4]. *N.Dalal, B.Triggs,*"Histograms Of Oriented Gradients For Human Detection", in 2005 IEEE computer society conference on computer vision and pattern recognition., vol. 1, Jun. 2005, pp. 886-893.

[5]. *Z.Guo, L.Zhang and D.Zhang*,"A Completed Modeling Of Local Binary Pattern Operator For Texture Classification", in IEEE transactions on image processing., vol. 19, no. 6, 2010, pp. 1657-1663.

[6]. *R.Girshick, J.Donahue and T.Darrell*,"Rich Feature Hierarchies For Accurate Object Detection And Semantic Segmentation", in Proceedings of the IEEE conference on computer vision and pattern recognition., 2014, pp. 580-587.

[7]. *K.He, X.Zhang, S.Ren and J.Sun*,"Spatial Pyramid Pooling In Deep Convolutional Networks For Visual Recognition", in IEEE transactions on pattern analysis and machine intelligence., vol. 37, no. 9, 2015, pp. 1904-1916.

[8]. S.*Ren, K.He and R.Girshick*,"Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks", in IEEE Transactions on Pattern Analysis & Machine Intelligence., vol. 39, no. 6, 2017, pp. 1137-1149.

[9]. *J.Redmon, S.Divvala, R.Girshick and A.Farhadi,*"You Only Look Once: Unified, Real-Time Object Detection", in Proceedings of the IEEE conference on computer vision and pattern recognition., 2016, pp. 779-788.

[10]. *W.Liu, D.Anguelov, D.Erhan, C.Szegedy, S.Reed, C.Y.Fu, A.C.Berg*,"Ssd: Single Shot Multibox Detector", in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing., Oct. 2016, pp. 21-37.

[11*]. Chen Qi*. Research on automatic detection system of inner surface defects of sliding bearing based on machine vision [D]. Zhenjiang: Jiangsu University, 2017.

[12*]. P.Kunakornvong, C.Tangkongkiet and P.Sooraksa*,"Defect Detection On Air Bearing Surface With Gray Level Co-occurrence Matrix", in The 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering., Mar. 2014, pp. 1-4.

[13]. *S.Deng, W.Cai, Q.Xu and B.Liang,*"Defect Detection Of Bearing Surfaces Based On Machine Vision Technique", in 2010 International Conference on Computer Application and System Modeling., vol. 4, Oct. 2010, pp. V4−548−V4−554.

[14]. *H.Feng, J.Zhuang, K.Chen, K.Song, J.Xiao and S.Ye*,"The Prior Model-Guided Network For Bearing Surface Defect Detection", in Electronics., vol. 12, no. 5, 2023, pp. 1142.

[15]. *Gao Liming, Jia Shuhai, Zhang Guolong et al.* Surface Defect detection method of Insulating Bearing Based on Improved Faster R-CNN [J]. Bearing,2023(04):1-8.

[16]. YOLOv5. Available online: https://github.com/ultralytics/yolov5 (accessed on 9 June 2020).

[17]. *W.Yu, M.Luo, P.Zhou, C.Pi, Y.Zhou, X.Wang, and S.Yan,"*Metaformer Is Actually What You Need For Vision 2022", in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition., 2022, pp. 10819-10829.

[18*]. J.Wang, L.Chen, R.Xu, Z.Liu, C.C.Loy and D.LIN,"*Carafe: Content-aware Reassembly Of Features", in Proceedings of the IEEE/CVF International Conference on Computer Vision., Nov. 2019, pp. 3007–3016.

[19]. *Z.Tong, Y.Chen, Z.Xu and R.Yu*,"Wise-IoU: Bounding Box Regression Loss With Dynamic Focusing Mechanism", in arXiv preprint arXiv., Feb. 2023, pp. 2301-10051.

[20]. *Z.ZHU, D.Liang, S.Zhang, X.Huang, B.Li and S.Hu Z*,"Traffic-sign Detection and Classification in the Wild", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., 2016, pp. 2110-2118.