

RESEARCH ON GEAR CRACK DETECTION METHOD BASED ON IMPROVED YOLOV11N

ZHIYI WAN¹, JIALIN CUI^{2,3,*}, BIN LIAN², ZHE-MING LU⁴, YANGMING ZHENG⁴
AND HAO LUO⁴

¹School of Information Science and Engineering
Zhejiang Sci-Tech University
No. 928, Second Avenue, Xiasha Higher Education Zone, Hangzhou 310018, P. R. China
2023220705031@mails.zstu.edu.cn

²School of Information Science and Engineering
NingboTech University
No. 1, South Qianhu Road, Ningbo 315100, P. R. China
lianbin_1@nbt.edu.cn

*Corresponding author: cuijialin@nbt.edu.cn

³Huanjiang Laboratory
Zhuji 311826, P. R. China

⁴School of Aeronautics and Astronautics
Zhejiang University
No. 38, Zheda Road, Hangzhou 310027, P. R. China
{zheminglu; zymsun2002; luohao}@zju.edu.cn

Received July 2025; revised November 2025

ABSTRACT. *In order to improve the accuracy of automatic gear crack detection, this paper proposes a method based on deep learning network. Using the improved method based on YOLOv11n, the Inner-MPDIoU (Inner-Intersection over Union with Minimum Points Distance) loss function is introduced first, which effectively improves the detection accuracy of small targets. Then, the overall performance of crack detection was further improved by the LQEHead (Localization Quality Estimation Head). Considering the diversity of gear crack morphology and size, the Cross Stage Partial Multi-Scale Feature Aggregation module is used to enhance the detection ability and maintain the computational efficiency. In addition, the Wavelet Feature Upsampling technology is combined to enhance the detail retention ability in the feature extraction process, which successfully improves the recognition rate of fine cracks. The findings indicate that the precision and mAP@0.5 of the improved YOLOv11n model have increased by 3.8 percentage points and 3 percentage points, respectively, compared to the baseline YOLOv11n model. Experimental results demonstrate that the improved algorithm offers a reliable and efficient solution for gear crack detection, significantly boosting accuracy while maintaining detection speed. These improvements highlight its strong practical value and promising application potential.*

Keywords: Gear crack detection, YOLOv11n, Multi-scale feature fusion, Loss function, Wavelet feature enhancement, Localization quality estimation

1. Introduction. Metal workpieces are key components in mechanical transmission systems, which are widely used in automobiles, aerospace, rail transit, industrial machine tools, and other mechanical equipment. Their operating status directly affects the safety and reliability of these mechanical systems [1]. Workpiece cracks are one of the common types of faults, mostly caused by material wear, manufacturing process defects, long-term

overload operation, and other factors. If not detected in time, the crack may continue to expand, causing the transmission system to be paralyzed, and even causing serious consequences such as production safety accidents. Therefore, it is important to study an efficient and accurate method for detecting cracks in workpieces.

In the field of crack detection, traditional detection methods provide important technical support for equipment fault diagnosis through years of practice accumulation. Technicians can directly observe whether there are cracks in the workpiece through the naked eye or with the help of tools such as magnifying glasses, which are easy to operate and low cost, but the visual inspection effect is highly dependent on the subjective factors of the inspector, which can easily cause missed detection. Penetrant testing [2] uses the capillary action of the liquid to coat the surface of the workpiece with the permeate to penetrate into the cracks, and then apply the developer to showing the cracks, but this method has high requirements for the cleanliness of the workpiece surface, many interference factors, and the detection process is cumbersome, and the detection efficiency is low. Ultrasonic testing [3] determines the existence and location of cracks based on the echo signal of ultrasonic waves by emitting ultrasonic waves into the workpiece. However, this method is significantly affected by complex geometries and material properties, making it challenging to guarantee the accuracy of the detection results. Eddy current testing [4] uses the principle of electromagnetic induction to detect cracks by analyzing the eddy current fraction changes of the detection coil, but this method is only suitable for conductive materials and is easily disturbed by the conductivity, permeability, and other factors of metal materials.

In recent years, deep learning methods have achieved remarkable results in the field of image detection, among which the YOLO (You Only Look Once) algorithm has been widely used in object detection tasks due to its high efficiency and accuracy. Dehaerne et al. [5] significantly improved the mAP for semiconductor line spacing defect detection by adjusting the hyperparameters of YOLOv7 and using vertical image flipping to augment the training data. They also effectively integrated predictions from different models through the Weighted Box Fusion strategy, thereby enhancing the overall detection performance. Based on the improved YOLOv8 network structure, Long and Jia [6] proposed three key improvements: optimizing multi-scale feature extraction with the Diverse Branch Block module, enhancing defect detection accuracy via the Multi-Scale Channel attention mechanism, and reducing parameters and computational load by optimizing the GDetect module. Experiments verified that this method significantly improved the accuracy and efficiency of steel surface defect detection. Zhang et al. [7] proposed a new model named GDM-YOLO based on the YOLOv8s network. This model incorporates three modules: the SPD Ghost Convolution downsampling module, the C2f-Dilated-Reparam-Block module, and the Multiscale Feature Enhancement Block module. While maintaining computational efficiency, it significantly improves the accuracy of steel surface defect detection and meets the real-time detection requirements of industrial production. Lu et al. [8] proposed a small target defect detection method for photovoltaic power plants, which integrates the dual-path mechanism of Token Mixer and Channel Mixer as the backbone network, optimizes the neck network with a bidirectional weighted feature pyramid network, and incorporates the triple attention mechanism and deformable convolution in the detection head, which effectively enhances detection accuracy and real-time performance. Shi et al. [9] proposed an improved DFP-YOLO network model, which reduces parameters/complexity via C3_D and C3_F modules, improves detection accuracy with SIoU loss, accelerates convergence using transfer learning, and achieves model lightweight through channel pruning. Jiang and Zhao [10] improved YOLOv7 by designing the ECBAM attention module, the SFPN shallow-weighted feature fusion network, and adopting the SIoU

loss function, aiming to address the characteristic of extreme aspect ratios of wood crack defects, thereby enhancing the detection capability of wood crack defects. Zhao et al. [11] proposed a detection model integrating multi-scale features. By designing the efficient inverted residual block module, fusing CSP efficient dilated reparam aggregation network, embedding the dual attention module attention mechanism, and building the lightweight detection head, they realized the requirement of real-time detection of crack defects in downhole scenarios. Chaowalittawin et al. [12] proposed an adaptive convolutional neural network ensemble model combined with multi-light channels for duck egg crack detection. After image preprocessing, they evaluated the model using Adaptive MobileNetV2, which confirmed that the model exhibited excellent detection performance under different combinations of lighting and preprocessing methods. Nigam et al. [13] proposed a noise-robust crack detection algorithm that combines the discrete wavelet transform, windowed variance, and coefficient averaging, which can effectively locate cracks. Zhang et al. [14] proposed an optimized model based on YOLOv8. By introducing the SimAM attention mechanism into the backbone network, adopting the lightweight C3Ghost module, and replacing the standard neck with a bidirectional multi-scale feature fusion structure, the model enhances the detection effect of fine cracks while ensuring real-time performance.

The application of deep learning in crack detection is also expanding, and models can accurately identify the location of cracks in large-scale objects, such as railway tracks [15], building concrete [16], and roads [17], through training on a large number of sample data. However, in the field of crack detection of small objects, due to the small scale of cracks in small objects, the lack of obvious features, and the lack of sufficient annotation data, the training and optimization of deep learning models face many challenges, and there are few related studies and applications, and more exploration and breakthroughs are urgently needed.

In this paper, the automatic crack detection of workpieces based on the YOLOv11n algorithm is performed for small workpieces, with gears as an example, and efficient and accurate gear crack defect detection is achieved. Aiming at the small size of gear cracks, the Inner-MPDIoU (Inner-Intersection over Union with Minimum Points Distance) loss function was introduced to improve the accuracy of small target detection. The improvement of the LQE (Localization Quality Estimation) detection head further enhances the overall performance of crack detection. Due to the existence of multiple morphologies and sizes of gear cracks, the CSP-PMSFA (Cross-Stage Partial-Partial Multi-Scale Feature Aggregation) module is used to improve multi-scale target detection performance, achieving a good balance between computational efficiency and feature expression ability. WFU (Wavelet Feature Upsampling) technology can enhance the ability to retain details during feature extraction, thereby improving the identification of fine cracks.

The first part of this article introduces the current research status of small target defects. Section 2 provides a detailed description of the models used and the research methods employed. Section 3 mainly introduces the experimental data and processing procedures. Section 4 verifies the performance advantages of the improved model through comparative experiments and ablation experiments. The conclusions drawn from our research are provided in Section 5.

2. Improved Methods Based on YOLOv11n.

2.1. Fundamentals of YOLOv11n algorithm. As an advanced single-stage object detection algorithm, YOLOv11 transforms the object detection problem into a regression problem, allowing it to directly predict the bounding box coordinates, category, and confidence level of the target in the image [18]. However, when dealing with subtle and

complex targets, such as gear cracks, the original YOLOv11n algorithm still has issues, including insufficient detection accuracy, which needs to be improved and optimized. In order to solve these problems, the original model structure was adjusted, the loss function was improved by using Inner-MPDIoU, the LQEHead was designed to enhance the performance of the detection head, and the CSP-PMSFA module and WFU technology were integrated to optimize the feature extraction and fusion process and the improved YOLOv11n model is shown in Figure 1.

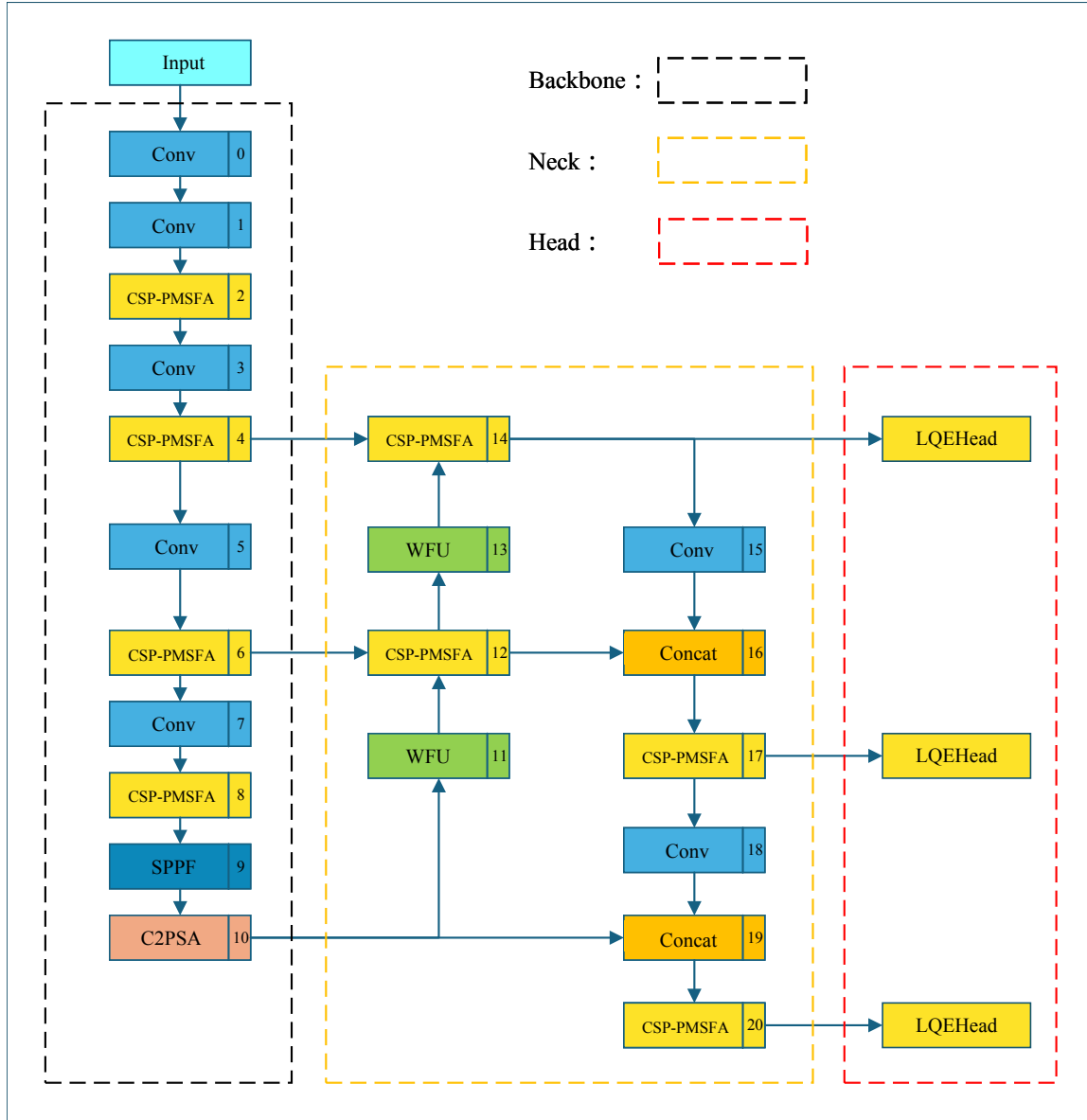


FIGURE 1. Architecture diagram of the improved YOLOv11n

2.2. Inner-MPDIoU loss function. In the object detection task, the calculation of the loss function is the key link. YOLOv11n uses Complete Intersection over Union (CIoU) and Distribution Focal Loss (DFL) as the regression loss of the bounding box, and comprehensively considers the core factors such as the spacing between the center point, the difference in aspect ratio and the overlapping area between the prediction box and the bounding box based on the IoU framework. However, there are limitations in dealing with slender cracks, and it is difficult to accurately describe the similarity between the

predicted bounding box and the actual bounding box. To tackle this challenge, Inner-IoU is integrated with MPDIoU to formulate Inner-MPDIoU. By calculating IoU through an auxiliary bounding box and focusing on the core region of the bounding box, this method effectively measures the overlapping area, resolves the bounding box overlap problem, and thus improves target localization accuracy and model robustness.

2.2.1. Inner-IoU. The IoU loss demonstrates insufficient effectiveness in handling complex geometric shapes or non-overlapping scenarios, and it fails to adaptively match diverse detectors and detection tasks, thereby leading to compromised generalization capability. In order to compensate for the weak generalization and slow convergence of the IoU loss in different detection tasks, the scale factor ratio is introduced to control the scale size of the auxiliary bounding box to calculate the loss, so as to speed up the regression process of the bounding box [19]. Figure 2 illustrates the structural and positional relationships of the bounding boxes in Inner-IoU. The center points of the Ground-Truth (GT) box and its inner GT box are denoted as $b(x_c, y_c)$, while the center points of the Anchor box and its inner Anchor box are $b^{gt}(x_c^{gt}, y_c^{gt})$. Here, h^{gt} and w^{gt} represent the height and width of the GT box, respectively, and h and w denote the height and width of the Anchor box.

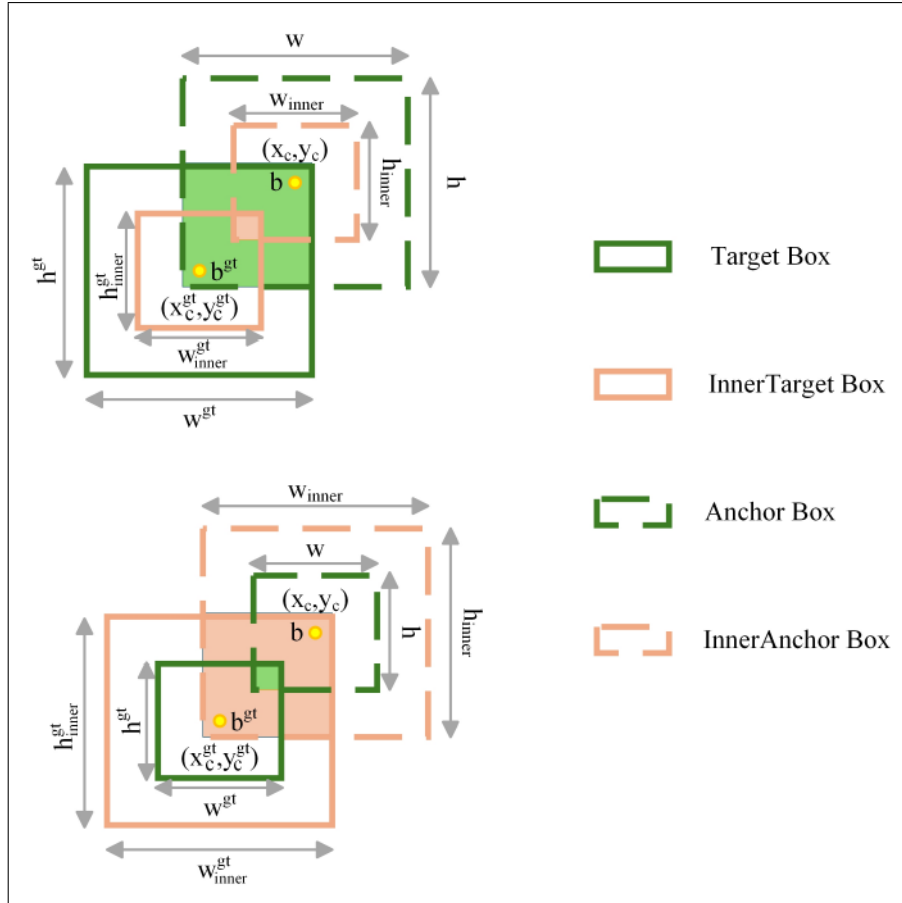


FIGURE 2. Diagram of structural and positional relationships of Inner-IoU bounding boxes

The calculation process of Inner-IoU is shown in Equations (1)-(5). It uses the scale factor ratio to control the size of the auxiliary box: for samples with high IoU, a small auxiliary box is used to calculate the loss, whose effective regression range is smaller than that of IoU loss but has a larger absolute gradient value, which can accelerate the convergence of high IoU samples; for samples with low IoU, a large auxiliary box is

adopted to calculate the loss, which expands the effective regression range and enhances the regression effect of low IoU samples.

$$b_l = x_c - \frac{w \times \text{ratio}}{2}, \quad b_r = x_c + \frac{w \times \text{ratio}}{2} \quad (1)$$

$$b_t = y_c - \frac{h \times \text{ratio}}{2}, \quad b_b = y_c + \frac{h \times \text{ratio}}{2} \quad (2)$$

$$\text{inter} = (\min(b_r^{\text{gt}}, b_r) - \max(b_l^{\text{gt}}, b_l)) \times (\min(b_b^{\text{gt}}, b_b) - \max(b_t^{\text{gt}}, b_t)) \quad (3)$$

$$\text{union} = (w^{\text{gt}} \times h^{\text{gt}}) \times (\text{ratio})^2 + (w \times h) \times (\text{ratio})^2 - \text{inter} \quad (4)$$

$$\text{IoU}^{\text{inner}} = \frac{\text{inter}}{\text{union}} \quad (5)$$

When dealing with slender cracks, Inner-IoU can target samples with different IoU conditions in this way, allowing for more targeted processing and achieving more refined evaluation. This enables more accurate measurement of the overlap between the predicted box and the ground truth box, laying a foundation for the precise detection of slender cracks.

2.2.2. MPDIoU. For multi-scale targets, MPDIoU directly minimizes the distance between the upper-left and lower-right corners of the predicted and ground-truth bounding boxes, which effectively addresses the bounding box overlap problem and simplifies the calculation process [20]. The formula is given in Equation (6), where $(x_1^{\text{gt}}, y_1^{\text{gt}})$ and $(x_2^{\text{gt}}, y_2^{\text{gt}})$ denote the coordinates of the upper-left and lower-right corners of the ground-truth bounding box, respectively; $(x_1^{\text{pred}}, y_1^{\text{pred}})$ and $(x_2^{\text{pred}}, y_2^{\text{pred}})$ represent the coordinates of the upper-left and lower-right corners of the predicted bounding box, respectively; and w and h correspond to the width and height of the input image, respectively.

$$\text{MPDIoU} = \text{IoU} - \frac{(x_1^{\text{pred}} - x_1^{\text{gt}})^2 + (y_1^{\text{pred}} - y_1^{\text{gt}})^2}{w^2 + h^2} - \frac{(x_2^{\text{pred}} - x_2^{\text{gt}})^2 + (y_2^{\text{pred}} - y_2^{\text{gt}})^2}{w^2 + h^2} \quad (6)$$

For a target with a special shape, such as a slender crack, the accuracy of its bounding box's position and shape is crucial. By focusing on the key point distances of the bounding box, MPDIoU makes the vertices of the predicted box closer to those of the ground truth box. This enables more accurate overall localization of the slender crack, resolving the issue that traditional methods struggle to accurately describe the similarity of key point distances between the predicted box and the ground truth box when dealing with slender cracks.

In summary, this paper introduces the Inner-MPDIoU loss function, which uses Inner-IoU to replace the IoU calculation in MPDIoU, and its formula is shown in Equation (7). Inner-MPDIoU combines the advantages of Inner-IoU and MPDIoU: it can leverage Inner-IoU's advantage of dynamically adjusting the auxiliary box scale according to the IoU conditions of different samples to refine the evaluation of overlapping regions, and at the same time, it can utilize MPDIoU's characteristic of directly minimizing key point distances to accurately align the vertices of bounding boxes. This effectively solves the localization problem of slender cracks and improves the detection accuracy and convergence speed of the model.

$$L_{\text{Inner-MPDIoU}} = L_{\text{MPDIoU}} + \text{IoU} - \text{IoU}_{\text{inner}} \quad (7)$$

2.3. LQE detector head. In YOLOv11n, the model solely predicts object bounding boxes and category information, while lacking an accurate assessment of localization quality. Gear cracks are characterized by large target scale differences, complex background

textures, and heavy interference from low-quality boxes. Relying solely on classification scores to select positive samples will lead to the introduction of low-quality samples, which damage the consistency of feature learning and ultimately result in an increased false detection rate. To address this limitation, this study integrates LQEHead into YOLOv11n. Specifically, an independent branch is employed to predict the localization quality score of detection boxes, which is then fused with the classification score to enhance the model's capability to screen high-quality positive samples and improve the overall quality of detection boxes. The core concept of LQE is to independently estimate the localization quality of detection boxes at each feature layer and combine it with the classification score [21].

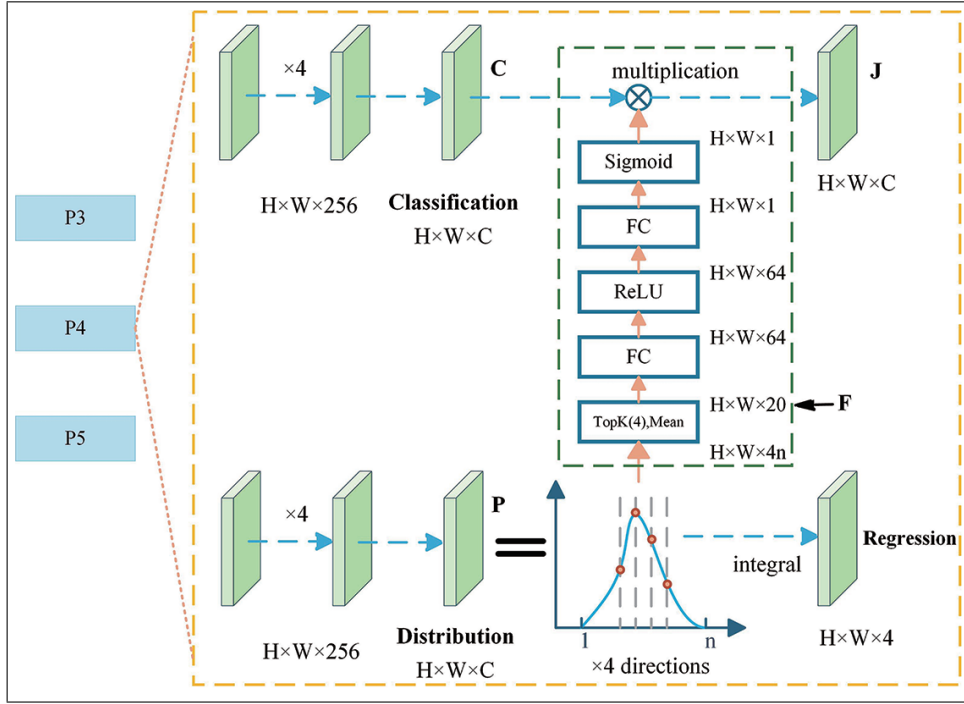


FIGURE 3. Network architecture diagram of LQEHead

2.3.1. Multi-scale feature adaptation. The integration of LQEHead further optimizes the multi-scale feature fusion mechanism in the detection head of the YOLOv11n model. LQEHead structures are added independently at each feature layer, with each detection head processing feature maps of different resolutions to ensure the accuracy of localization quality estimation at multiple scales. High-resolution feature maps are processed using smaller convolution kernels to extract detailed localization information about small targets with greater precision. For low-resolution feature maps, larger convolution kernels are employed to enhance the overall localization quality assessment of large targets. Through multi-scale independent processing, LQEHead achieves an accurate estimation of the localization quality of gear cracks of different sizes. This approach resolves the contradiction present in traditional single-scale detection heads, which tend to provide evaluations that are too coarse for small targets and too fine for large targets. As a result, it offers a reliable multi-scale quality basis for subsequent score fusion.

2.3.2. Score fusion mechanism. LQEHead uses features from both the classification and regression branches to construct a comprehensive detection score, as defined in Equation (8), where C denotes the feature output from the classification branch, and I represents the IoU feature information, which serves as an estimate of the IoU quality.

$$J = C \times I, \quad C = [C_1, C_2, \dots, C_m], \quad C_i \in \{0, 1\}, \quad I \in [0, 1] \quad (8)$$

Conventional YOLO models depend on classification scores for positive sample selection, which may introduce low-quality bounding boxes. The integration of LQEHead enables the classification branch to jointly consider both classification scores and predicted bounding box quality scores. The comprehensive score serves as a key metric for sorting positive samples, with priority given to samples with high classification and localization quality, thereby improving the consistency of sample quality during the training phase. In gear crack detection, employing a composite score instead of a single classification score for sorting and screening significantly reduces false positives caused by inflated classification scores, thereby enhancing the reliability of detection results.

In summary, addressing the large target scale difference of gear cracks, LQEHead establishes a hierarchical quality evaluation system through a multi-scale feature adaptation mechanism. This hierarchical evaluation logic enables targeted and accurate measurement of the localization quality of cracks of different scales. LQEHead fuses the semantic features output by the classification branch and the IoU quality features associated with the regression branch. This fusion method can effectively avoid quality misjudgment caused by background texture interference and insufficient semantic information of small targets in a single classification score, ultimately reducing the false detection rate and improving the detection accuracy of the detection model in the gear crack detection task.

2.4. CSP-PMSFA module. The CSP-PMSFA module, an innovative module proposed in this paper, which combines the ideas of partial convolution and multi-scale feature fusion, is designed to efficiently extract and integrate feature information across diverse scales. In the Backbone, the CSP-PMSFA module is adopted to replace the C3k2 module of YOLOv11, and in the Neck, the CSP-PMSFA module is used to replace the Concat and C3k2 modules. The corresponding network architecture diagram is shown in Figure 4.

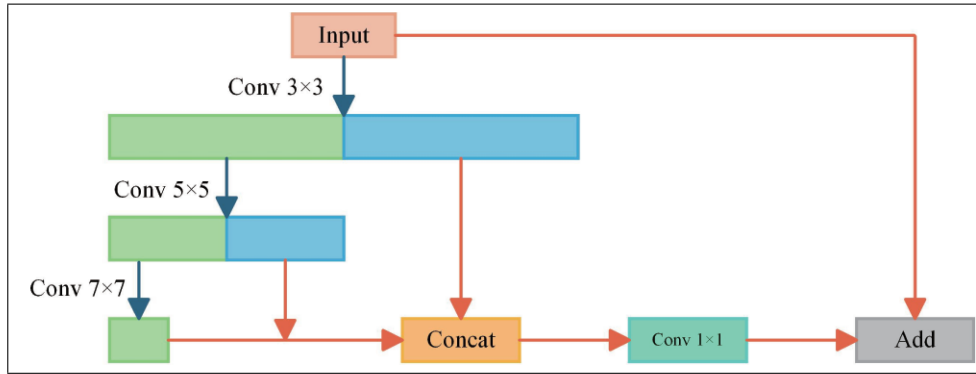


FIGURE 4. CSP-PMSFA network architecture diagram

2.4.1. Partial multi-scale feature extraction. Unlike standard convolution, which processes all input channels, partial convolution performs multi-scale feature extraction on a subset of channels, avoiding exhaustive calculations across all channels and thereby reducing redundant operations. Combining the concept of the CSP [22] structure, the input feature map is divided into two parts in the channel dimension. One part is propagated backward through the convolution layer, whereas the other is directly merged with the output of the convolution module through the cross-stage hierarchical structure. In gear crack detection, cracks of varying sizes necessitate feature representations at different

scales. This partial multi-scale feature extraction method improves computational efficiency significantly while ensuring feature extraction efficacy, thereby effectively reducing the complexity of model calculations.

2.4.2. Enhanced feature fusion. After partial multi-scale feature extraction, 1×1 convolution is employed to adjust channel dimensions and fuse information from feature maps of different scales, generating a more compact feature representation. Meanwhile, residual connections are utilized to integrate the original input features into the fused features. This fusion method effectively retains the original information while introducing new multi-scale information, enhancing the model's capacity to express gear crack features and enabling it to better identify gear cracks of different shapes and sizes.

In summary, addressing the multi-scale and morphologically heterogeneous characteristics of gear cracks, the CSP-PMSFA module abandons the redundant full-channel computation mode of standard convolutional neural networks. It employs a channel splitting strategy that combines partial convolution with a CSP structure to extract multi-scale features only on key channels, thereby reducing the model's computational complexity while ensuring effective feature capture. Meanwhile, it builds an enhanced feature fusion mechanism: considering that crack features are susceptible to background interference and small-scale crack features are prone to loss, it realizes channel adaptation and information compression of multi-scale features through convolution, and introduces residual connections to retain the underlying information of original input features, thereby strengthening the model's ability to express features of cracks with different morphologies. Through an innovative design approach that combines partial computation with multi-scale fusion, the CSP-PMSFA module achieves a good balance between computational efficiency and feature representation capabilities.

2.5. Wavelet feature upsampling. YOLOv11n employs traditional nearest neighbor interpolation methods for upsampling. These methods cannot effectively separate and fuse feature components of varying frequencies. This limitation results in blurred high-frequency details or distorted low-frequency structures. As a result, the model's accuracy in multi-scale object detection is affected. WFU conducts upsampling through wavelet transform-based multiscale decomposition and dynamic frequency fusion. This approach effectively enhances detail retention and multiscale representation in the feature extraction process [23]. The wavelet feature upsampling procedure is depicted in Figure 5. Initially, the input feature F_S is decomposed into high-frequency components D_{RR} and low-frequency components A_{LL} through wavelet transformation. Enhancement and optimization of the high-frequency component D_{RR} are performed using residual blocks. The residual structure can alleviate the gradient vanishing problem in deep networks. It prevents high-frequency information from being interfered with by noise during transmission and makes high-frequency details more prominent. For the low-frequency component A_{LL} , it is fused with the feature map of the next stage. The structural integrity of the low-frequency component is supplemented by leveraging the semantic information of subsequent features. This results in enhanced low-frequency information. The optimized high-frequency component and the fused low-frequency component are jointly input into the inverse wavelet transform module. This module uses specific wavelet basis functions to re-integrate the two types of components into a complete feature map F'_S . This not only preserves the clarity of high-frequency details but also maintains the accuracy of the low-frequency structure. Ultimately, this achieves feature upsampling enhancement.

In the gear crack detection scenario, the wavelet transform can decompose an image into sub-bands of different frequencies to extract detailed information. Based on this,

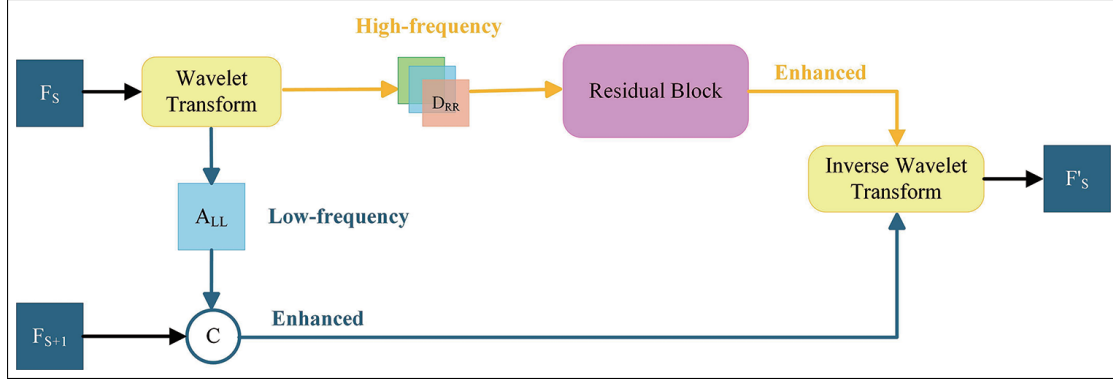


FIGURE 5. Wavelet feature upsampling architecture

WFU first accurately separates the high-frequency and low-frequency components of features through wavelet transform, then performs targeted optimization on information of different frequencies: it enhances high-frequency components to highlight subtle details. It fuses low-frequency components to maintain structural integrity. This processing method is better suited to the needs of multi-scale target detection and can significantly improve the detection accuracy of small-sized targets. Combined with the characteristics of complex gear surface textures and subtle crack features, WFU enhances the detailed information in features, making crack features stand out from complex textures. This helps the model capture the presence of cracks more easily and further improves the ability to detect gear cracks.

3. Experimental Data and Processing.

3.1. Evaluation metrics. This paper employs common target detection evaluation metrics, such as Precision (P), Recall (R), and mean Average Precision (mAP), to evaluate model performance. P is defined as the proportion of true positive samples among all samples predicted as positive, measuring the accuracy of positive prediction. R represents the proportion of actual positive samples correctly identified by the model among all actual positive samples, measuring the model's capability to capture positive cases. The evaluation method for mAP is as follows: first, calculate the Average Precision (AP) of each category at different recall rates, and then take the average of all categories to comprehensively evaluate the overall performance of the object detection model in multi-category tasks. The calculation process is detailed in Equations (9) to (12), where TP and FP denote the number of targets correctly detected and incorrectly detected by the model. FN represents the number of targets missed by the model.

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \int_0^1 P(R) dR = \sum_{k=0}^n P(k) \Delta R(k) \quad (11)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (12)$$

3.2. Experimental environment setup. The fundamental configuration information for this experiment is as follows: The experimental environment consists of the Ubuntu 22.04 operating system, equipped with a GeForce RTX 3080 GPU and a 12-core Intel (R) Xeon (R) Silver 4214R CPU. The model is developed and trained based on the PyTorch 2.3.0 framework, leveraging CUDA 12.1 and Python 3.12 runtime environment.

The training parameters are set as follows: The model undergoes 100 training epochs with a batch size of 8. Optimization is performed using the SGD optimizer, initialized with a learning rate of 0.01, a momentum coefficient of 0.937, and a weight decay coefficient of 0.0005.

3.3. Dataset construction. To verify the performance of the improved YOLOv11n model in gear crack detection, an image acquisition hardware platform consisting of a camera, a lens, a light source, and a bracket was constructed in a laboratory environment. This paper employed an industrial CCD sensor camera with higher sensitivity, lower noise, and faster response speed, as well as a ring light source that can provide uniform and high-intensity illumination. A substantial number of images were extracted from video clips recorded by the camera. After removing blurry and incomplete images, images of workpieces containing crack defects were selected as image data for the dataset.

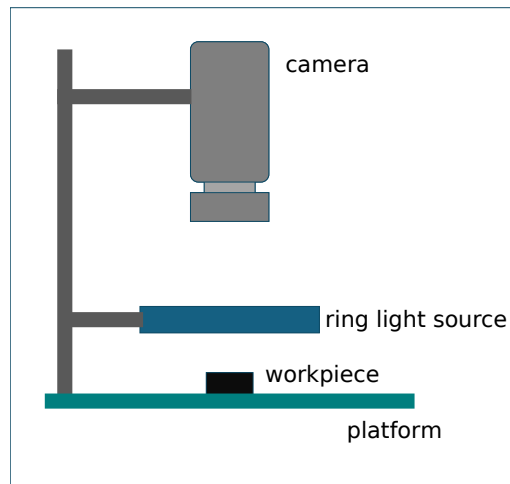


FIGURE 6. Hardware composition of dataset collection

Using the image annotation software LabelImg, the collected gear images were annotated according to the annotation format of the YOLO dataset. They were then divided into training, validation, and test sets in a ratio of 8 : 1 : 1 to establish a gear crack defect dataset. Figure 7 shows four images of cracks of different scales from the gear crack dataset created in this paper.

Figure 8(a) illustrates the distribution of gear crack bounding boxes, where the horizontal and vertical axes represent the width and height of the bounding boxes, respectively. The color intensity corresponds to the proportion of the bounding boxes, such that darker shades indicate a higher proportion. Figure 8(b) shows the distribution of gear crack bounding boxes across the entire dataset. The horizontal and vertical axes represent the normalized horizontal and vertical coordinates of the bounding boxes relative to the entire image, respectively. Darker colors indicate denser distributions. As can be seen from Figure 8(a), there are significant differences in the size of the gear crack bounding boxes, with multiple crack shapes present. Small targets account for a relatively large proportion of the dataset, so the model needs to learn to distinguish the features of targets of different

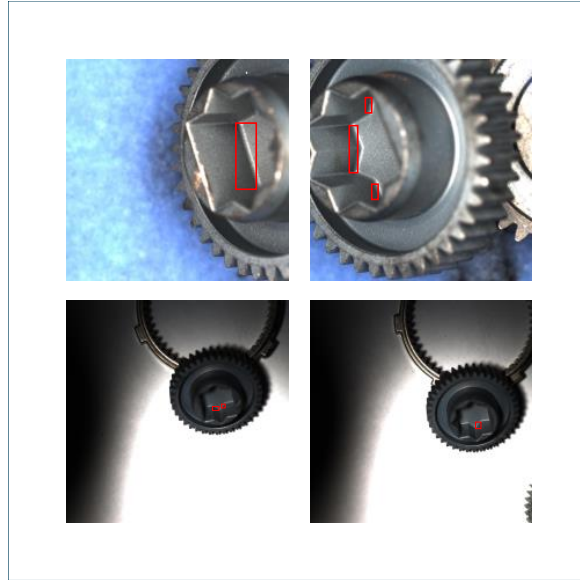


FIGURE 7. Images of gear crack dataset

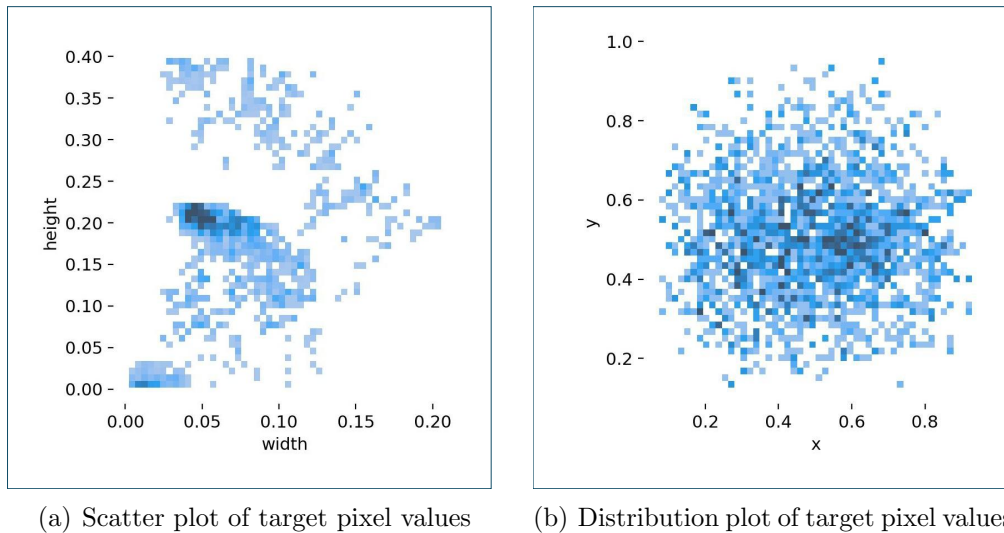


FIGURE 8. Scatter plots of targets in the gear crack dataset

sizes. As can be seen from Figure 8(b), the gear crack bounding boxes are not uniformly distributed in the image, but tend to appear in the middle of the image.

4. Experimental Results and Analysis.

4.1. Comparative experiments. The YOLO algorithm is a typical representative of single-stage detection network architectures. With its efficient detection speed and excellent real-time performance, it has significant advantages in industrial applications [24]. To assess the performance of the improved YOLOv11n-based algorithm proposed in this paper, comparative experiments were conducted against several mainstream object detection algorithms. During the experiment implementation, variables including experimental equipment, training hyperparameters, and iteration counts were strictly standardized to ensure consistency and comparability. Through systematic analysis of the experimental

TABLE 1. Comparison table of performance metrics for YOLO series models

Model	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv5n	90.3	92.6	92.1	67.5
YOLOv8n	93.2	90.3	91.5	68.7
YOLOv10n	90.5	85.7	89.6	65.0
YOLOv11n	90.6	92.0	92.5	69.2

results, comprehensive consideration of detection speed, accuracy, expandability, and robustness, YOLO11n was ultimately selected as the basic detection network.

4.2. Ablation experiments. The experimental results of the improved model are shown in Table 2. The synergistic integration of the Inner-MPDIoU loss function, the detection head LQEHead, the CSP-PMSFA module, and the WFU technique effectively improved the performance of the YOLOv11n model in the task of gear crack defect detection.

TABLE 2. Comparison table of performance metrics for improved YOLOv11n model

Group	YOLOv11n	Inner-MPDIoU	LQEHead	CSP-PMSFA	WFU	P/%	R/%	mAP@0.5/%
1	✓					90.6	92.0	92.5
2	✓	✓				93.8	91.5	93.5
3	✓	✓	✓			91.9	88.6	94.5
4	✓	✓	✓	✓		92.1	90.9	95.0
5	✓	✓	✓	✓	✓	94.4	88.6	95.5

Experimental data indicate that the introduction of the Inner-MPDIoU module leads to increased model precision and mAP@0.5, alongside a slight decrease in recall. This demonstrates that Inner-MPDIoU plays a significant role in optimizing bounding box regression accuracy and enhancing detection precision. When the LQEHead module is further incorporated, the model's mAP@0.5 improves, while precision and recall decline. This phenomenon suggests that while the LQEHead module focuses on boosting small-target detection accuracy, it may cause specific interference with the feature extraction of some medium and large targets. With the introduction of the CSP-PMSFA module, the model's precision, recall, and mAP@0.5 all increase. This shows that CSP-PMSFA effectively enhances the model's adaptability to targets of different scales through multi-scale feature fusion and channel attention optimization. When the WFU module is added based on Experiment Group 4, the model's precision and mAP@0.5 are both improved, and the recall remains consistent with that of Experiment Group 3. This result verifies that the WFU module further enhances detection precision by dynamically adjusting the weight of feature channels. In production lines, detection results are usually directly used for automated rejection; excessively high false positives will increase re-inspection costs and production delays. Therefore, this study holds that on the premise that the recall rate is within an acceptable range, prioritizing the improvement of precision and mAP@0.5 has a higher value for practical applications.

By synthesizing the results of the five experimental groups, it can be concluded that the synergistic effect of the proposed improved modules can significantly enhance the target detection performance of the YOLOv11n model: The mAP@0.5 value of the improved YOLOv11n model is an improvement of 3 percentage points over YOLOv11n. The precision of the improved YOLOv11n model is an improvement of 3.8 percentage points over YOLOv11n. The improved YOLOv11n model has certain advantages in detection

accuracy while ensuring detection speed. It enables more precise detection of gear cracks, effectively reducing both missed detection and false detection.

4.3. Comparative analysis of test image effects. In order to more intuitively demonstrate the detection performance of the improved model, the same gear crack defect images were selected to conduct a contrast experiment between the YOLOv11n model and the improved model. The results are shown in Figure 9, where Figures 9(a) and 9(b) show the detection results of the YOLOv11n model and the improved model, respectively. A comparison shows that the baseline YOLOv11n model suffers from false negatives and false positives, resulting in poor detection accuracy. The improved model effectively reduces the false negative rate and significantly reduces false positives, while also providing more accurate defect localization. In summary, the improved model enhances the accuracy of gear crack defect detection, significantly reducing false negatives and false positives, thereby demonstrating a notable improvement in detection performance compared to the baseline model.

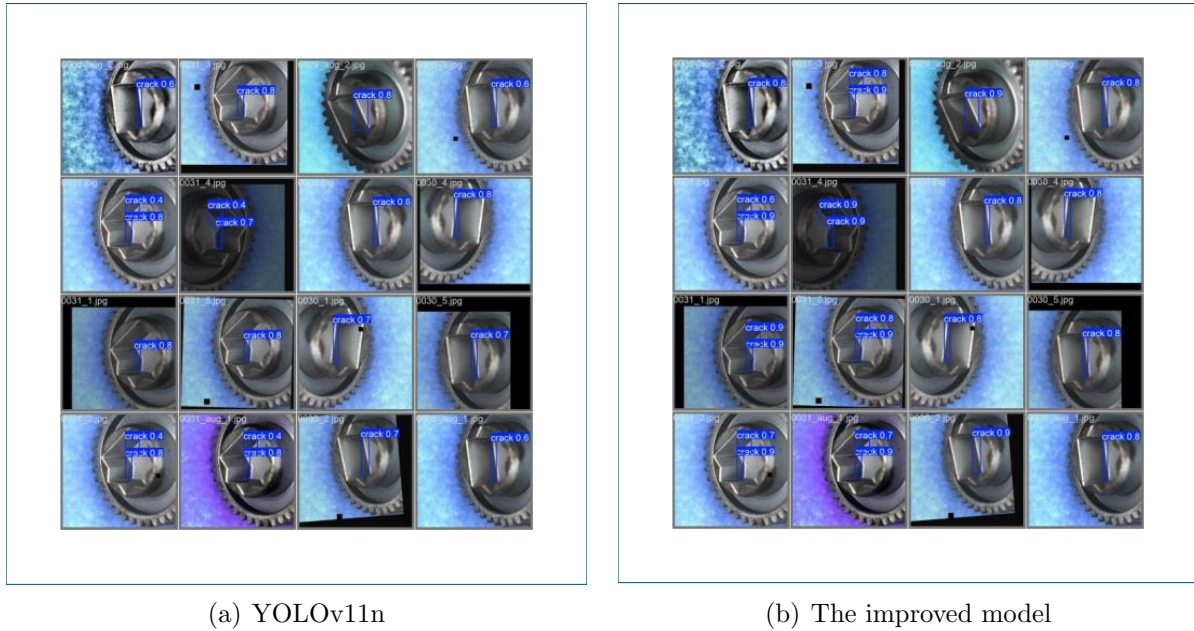


FIGURE 9. Comparison of detection performance between YOLOv11n and the improved model

5. Conclusion. This study focuses on the challenge of automated detection of cracks in small workpieces. Aiming at the requirements of gear crack detection, it proposes a detection method based on the improved YOLOv11n model. To address the limitations of traditional bounding box loss functions in modeling elongated cracks, we integrate the Inner-MPDIoU loss function into the training framework. This hybrid loss function combines the Inner-IOU bounding box regression loss, which leverages auxiliary bounding boxes, with the MPDIoU similarity metric, which employs minimum point distance, to enhance geometric feature constraint. To mitigate issues caused by complex textures on workpiece surfaces, the LQEHead detection head is introduced. It performs quality scoring for each detection box based on the peak value of the probability distribution of boundaries, significantly improving the model's localization accuracy and anti-interference capability. Considering the diversity of crack sizes, the CSP-PMSFA module is employed. While reducing computational redundancy, it effectively integrates multi-scale features

across different levels, thereby enhancing the model's computational efficiency and expressive power. To tackle the common problems of noise interference and loss of crack details in workpiece images, the WFU technique is adopted to replace traditional upsampling layers. This resolves issues of aliasing and detail loss during conventional upsampling, thereby improving the model's ability to extract and fuse features.

Future research will focus on lightweight model structure design, exploring methods such as multimodal data fusion and adaptive learning strategies to optimize model performance further. This aims to meet the higher requirements for real-time and accurate detection of gear cracks in industrial production, promoting the widespread application of this technology in practical engineering.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grant 61972350, the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY23F020013, and the Ningbo Science and Technology Plan Project under Grants 2023Z040 and 2024Z261.

REFERENCES

- [1] M. Xiao, W. Wang, X. Shen, Y. Zhu, P. Bartos and Y. Yiliyasi, Research on defect detection method of powder metallurgy gear based on machine vision, *Machine Vision and Applications*, vol.32, pp.1-13, 2021.
- [2] Y. Li, F. Liu, B. Du and W. Ren, Application of nondestructive testing technology in pressure vessel and pressure pipeline, *China High-Tech*, no.2, pp.140-142, 2024.
- [3] W. Liu, P. Hu, J. Xiao, Q. Yin, X. Zhou, H. Li and S. Shen, High precision detection of artificial defects in additively manufactured Ti6Al4V alloy via laser ultrasonic testing, *Journal of Materials Research and Technology*, vol.30, pp.8740-8748, 2024.
- [4] W. Matsunaga, X. Xu, K. Mizukami, Y. Mizutani and A. Todoroki, Identification of crack orientation in electrically conductive materials using eddy current testing powered by rotational directive filter, *NDT and E International*, vol.150, 103273, 2025.
- [5] E. Dehaerne, B. Dey, S. Halder and S. De Gendt, Optimizing YOLOv7 for semiconductor defect detection, *Metrology, Inspection, and Process Control XXXVII*, vol.12496, no.1, pp.635-642, 2023.
- [6] Y. Long and Y. Jia, Steel surface defect detection method based on improved YOLOv8, *Academic Journal of Science and Technology*, vol.8, no.2, pp.55-56, 2023.
- [7] T. Zhang, H. Pang and C. Jiang, GDM-YOLO: A model for steel surface defect detection based on YOLOv8s, *IEEE Access*, vol.12, pp.148817-148825, 2024.
- [8] Y. Lu, C. Du, X. Li, S. Liang, Q. Zhang and Z. Zhao, A high-precision defect detection approach based on BiFDRRep-YOLOv8n for small target defects in photovoltaic modules, *Energies*, vol.18, no.9, 2299, 2025.
- [9] T. Shi, Y. Ding, K. Zhu and Y. Su, DFP-YOLO: A lightweight machine tool workpiece defect detection algorithm based on computer vision, *The Visual Computer*, vol.41, no.7, pp.1-13, 2024.
- [10] X. Jiang and X. Zhao, Improved YOLOv7 algorithm for wood surface defect detection, *Computer Engineering and Applications*, vol.60, no.7, pp.175-182, 2024.
- [11] B. Zhao, R. Cheng and X. Jia, YOLOv8 crack defect detection algorithm based on multi-scale features, *Computer Engineering and Applications*, vol.60, no.22, pp.261-270, 2024.
- [12] V. Chaowalittawin, W. Krungseanmuang, P. Sathaporn and B. Purahong, Duck egg crack detection using an adaptive CNN ensemble with multi-light channels and image processing, *Applied Sciences*, vol.15, no.14, 7960, 2025.
- [13] R. Nigam, S. K. Singh and R. Kumar, Development of a noise-robust algorithm for crack detection in beams: Numerical and experimental analysis, *Journal of Vibration Engineering & Technologies*, vol.13, no.6, 422, 2025.
- [14] J. Zhang, Z. V. Beliaeva and Y. Huang, Accuracy-efficiency trade-off: Optimizing YOLOv8 for structural crack detection, *Sensors*, vol.25, no.13, 3873, 2025.
- [15] G. Xu, S. Zhang and T. Bai, Crack detection method of CRTSII track slab based on Faster R-CNN improvement, *China Railway Science*, vol.44, no.1, pp.106-113, 2023.
- [16] W. Ding, K. Yu and J. Shu, Method for detecting cracks in concrete structures based on deep learning and UAV, *China Civil Engineering Journal*, vol.54, no.S1, pp.1-12, 2021.

- [17] S. Zhou, D. Yang, Z. Zhang, J. Zhang, F. Qu, P. Punetha, W. Li and N. Li, Enhancing autonomous pavement crack detection: Optimizing YOLOv5s algorithm with advanced deep learning techniques, *Measurement*, vol.240, 115603, 2025.
- [18] R. Khanam and M. Hussain, YOLOv11: An overview of the key architectural enhancements, *arXiv Preprint*, arXiv: 2410.17725, 2024.
- [19] H. Zhang, C. Xu and S. Zhang, Inner-IoU: More effective intersection over union loss with auxiliary bounding box, *arXiv preprint*, arXiv: 2311.02877, 2023.
- [20] M. Siliang and X. Yong, MPDIoU: A loss for efficient and accurate bounding box regression, *arXiv Preprint*, arXiv: 2307.07662, 2023.
- [21] X. Li, W. Wang, X. Hu, J. Li, J. Tang and J. Yang, Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.11632-11641, 2021.
- [22] C. Wang, H. Liao, Y. Wu, P. Chen, J. Hsieh and I. Yeh, CSPNet: A new backbone that can enhance learning capability of CNN, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp.390-391, 2020.
- [23] W. Li, H. Guo, X. Liu, K. Liang, J. Hu, Z. Ma and J. Guo, Efficient face super-resolution via wavelet-based feature enhancement network, *Proc. of the 32nd ACM International Conference on Multimedia*, pp.4515-4523, 2024.
- [24] S. Wang, L. Zhao, M. Shao and R. Ge, Insulator defect detection method for transmission line based on improved YOLOv5s, *Electronic Measurement Technology*, vol.45, no.21, pp.181-188, DOI: 10.19651/j.cnki.emt.2210118, 2022.

Author Biography



Zhiyi Wan received the B.E. degree in Engineering from Zhejiang Sci-Tech University, China, in 2023. She is currently pursuing the M.E. degree in Control Engineering at Zhejiang Sci-Tech University, China. Her current research interests include computer vision related applications.



Jialin Cui received the M.S. degree in Automation from Zhejiang University, China, in 2005. He is currently an Associate Professor at NingboTech University. His current research interests include machine vision, artificial intelligence and application of artificial intelligence technology in information security.



Bin Lian received the M.S. degree in Cryptography from Southwest Jiaotong University, China, in 2005 and the Ph.D. degree in Cryptography from Shanghai Jiao Tong University, China, in 2015. He is with the NingboTech University, China. His research interests include cryptography, cryptographic protocol, and technology of network security.



Zhe-Ming Lu received the B.S. and M.S. degrees in Electrical Engineering and the Ph.D. degree in Measurement Technology and Instrumentation from the Harbin Institute of Technology (HIT), China, in 1995, 1997, and 2001, respectively. He became a Lecturer with HIT in 1999. Since 2003, he has been a Professor with the Department of Automatic Test and Control, HIT. He is currently a Full Professor with the School of Aeronautics and Astronautics, Zhejiang University, China. In the areas of multimedia signal processing and information hiding, he has published more than 350 papers, seven monographs in Chinese, two monographs in English and three book chapters in English. His current research interests include multimedia signal processing, information security, and complex networks.



Yangming Zheng received his Ph.D. degree from Xi'an Electronic Science and Technology University, China, in 2005. Following his doctoral studies, he embarked on a postdoctoral research journey at Zhejiang University, China, from 2005 to 2008, subsequently earning an appointment as an Associate Professor at the same institution. Currently, he holds a full-time faculty position at Zhejiang University. His prevailing research interests encompass machine intelligence, machine vision, and unmanned factories, wherein he continues to explore and innovate within these technological and industrial domains. His dedication to research and development in these areas aims to further the advancements and applications of intelligent machinery and autonomous systems in various industrial settings.



Hao Luo received the B.S., M.S., and Ph.D. degrees from Harbin Institute of Technology, China, in 2002, 2004, and 2008, respectively. He is currently an Associate Professor with the School of Aeronautics and Astronautics, Zhejiang University, China. From 2014 to 2016, he was a Visiting Scholar with the School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Australia. His current research interests include computer vision, embedded artificial intelligence, and signal processing.