

Doi:10.32604/cmc.2025.059389

ARTICLE





Deep Learning Based Online Defect Detection Method for Automotive Sealing Rings

Jian Ge¹, Qin Qin^{1,*}, Jinhua Jiang¹, Zhiwei Shen², Zimei Tu¹ and Yahui Zhang¹

¹School of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai, 201209, China ²School of Electrical Engineering and Telecommunications, UNSW, Sydney, NSW 2052, Australia

*Corresponding Author: Qin Qin. Email: qinqin@sspu.edu.cn

Received: 06 October 2024; Accepted: 24 February 2025; Published: 16 April 2025

ABSTRACT: Manufacturers must identify and classify various defects in automotive sealing rings to ensure product quality. Deep learning algorithms show promise in this field, but challenges remain, especially in detecting small-scale defects under harsh industrial conditions with multimodal data. This paper proposes an enhanced version of You Only Look Once (YOLO)v8 for improved defect detection in automotive sealing rings. We introduce the Multi-scale Adaptive Feature Extraction (MAFE) module, which integrates Deformable Convolutional Network (DCN) and Space-to-Depth (SPD) operations. This module effectively captures long-range dependencies, enhances spatial aggregation, and minimizes information loss of small objects during feature extraction. Furthermore, we introduce the Blur-Aware Wasserstein Distance (BAWD) loss function, which improves regression accuracy and detection capabilities for small object anchor boxes, particularly in scenarios involving defocus blur. Additionally, we have constructed a high-quality dataset of automotive sealing ring defects, providing a valuable resource for evaluating defect detection methods. Experimental results demonstrate our method's high performance, achieving 98.30% precision, 96.62% recall, and an inference speed of 20.3 ms.

KEYWORDS: Deep learning; automotive sealing ring; defect detection

1 Introduction

In the context of automotive intelligence, enhancing vehicle safety has become a widespread societal demand [1–3]. The Electronic Control Unit (ECU) is a key component in modern vehicles. Its malfunction can lead to significant safety risks [4]. Given the critical role of automotive sealing rings in protecting the ECU from external damage, even minor defects can lead to serious issues, such as electrical short-circuits, corrosion, and system failures [5]. Detecting such defects is essential for maintaining product quality and meeting vehicle safety standards [6,7].

Common defects in automotive sealing rings, such as blowout, fracture, excess material, and burr, are critical issues that can undermine the functionality of the ECU. However, defect detection faces unique challenges due to the small size of the sealing rings (1.3 cm \times 0.5 cm), compounded by harsh industrial conditions such as uneven lighting and defocus blur. Traditionally, defect detection in automotive sealing rings has relied heavily on manual inspection [8], limited by low accuracy and efficiency. To overcome these limitations, advanced automated detection technologies that integrate seamlessly into high-speed production lines are essential. The adoption of advanced technologies, such as deep learning-based systems,



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

offers a promising solution by accurately identifying defects, thereby reducing waste and improving overall productivity in automotive manufacturing [9,10].

Given the complexity and varied nature of defects in automotive sealing rings, selecting an appropriate deep learning detection algorithm is crucial. Deep learning detection methods are classified into one-stage and two-stage approaches [11]. While two-stage algorithms achieve high accuracy, their slower inference speed limits real-time applications [12,13]. In contrast, one-stage algorithms provide faster inference and are more suitable for real-time industrial scenarios, although they often struggle with small-object localization [14,15]. For real-time automotive sealing ring defect detection, the YOLO algorithm is preferable [16].

Building on YOLOv8's suitability for real-time detection, we propose an Enhanced YOLO. Our approach introduces the Multi-scale Adaptive Feature Extraction (MAFE) Module, which integrates deformable convolutional networks (DCN) and Space-to-Depth (SPD) to enhance feature extraction for small and irregular defects, minimizing spatial information loss. We also develop the Blur-Aware Wasserstein Distance (BAWD) loss function, which improves regression accuracy and detection capability for small object anchor boxes, particularly in scenarios with defocus blur.

We have constructed a high-quality dataset of automotive sealing ring defects. This dataset replicates real-world industrial conditions, providing a challenging benchmark for evaluating defect detection methods. Our experiments show significant improvements in defect detection using our Enhanced YOLOv8. Additionally, our model demonstrates robust performance on the COCO2017 dataset, validating its adaptability across diverse scenarios. The key contributions of this paper are outlined as follows:

(1) Introducing the MAFE module into YOLOv8. This module combines DCN for capturing long-range dependencies and adaptive spatial aggregation with SPD operations to reduce information loss during feature extraction, especially for small objects and low-resolution images.

(2) Developing the BAWD Loss to improve regression accuracy and detection capability for small object anchor boxes, especially in scenarios with defocus blur.

(3) Constructing a high-quality dataset of automotive sealing ring defects, providing a valuable resource for evaluating various defect detection methods. The dataset is available as an open-source resource at https://github.com/xika1234/automotive-seal-ring-dataset (accessed on 23 February 2025).

2 Related Work

2.1 Industrial Defect Detection Algorithm

Current mainstream approaches for defect detection include RCNN-based frameworks, Transformerbased models, and YOLO-based methods, each with distinct strengths and limitations. RCNN-based frameworks [17], achieve robust feature extraction and high accuracy, particularly for complex defects [18]. For instance, Faster R-CNN has been enhanced with deformable convolution for small weld defects [19], while Cascade R-CNN has been applied to multiscale defect localization in wheel hubs [20]. However, their two-stage nature limits real-time applicability. Transformer-based models, such as DETR [21], Deformable DETR [22], and Dynamic DETR [23], leverage attention mechanisms for global context modeling and have been applied to defect detection tasks like aero-engine blades and roller bearings [24,25]. Despite their potential, Transformers often face high computational costs, making them less practical for time-sensitive applications [26]. YOLO-based methods, such as YOLOv5 and YOLOv7 [27], balance speed and accuracy but struggle with small, subtle defects under harsh industrial conditions like motion blur. To address these limitations, methods such as Feature Pyramid Networks (FPN) and attention-based mechanisms have been explored. FPN enhances multi-scale feature representation by integrating spatial and semantic information from different feature levels [28,29], while attention mechanisms recalibrate feature importance to improve detection precision [30]. However, these approaches often rely on rigid structures or static attention designs, limiting their adaptability to irregular defect shapes and complex industrial scenarios.

To address these challenges, our Enhanced YOLOv8 introduces the MAFE module and the BAWD loss function. MAFE combines SPD operations and DCN. SPD expands the receptive field by rearranging spatial information, improving the extraction of long-range dependencies critical for small-object detection. Meanwhile, DCN introduces adaptive sampling to focus on irregular object shapes, addressing the limitations of conventional pooling and attention mechanisms. BAWD further strengthens the model by incorporating blur-aware adjustments into bounding box regression, mitigating uncertainties from defocus blur and enhancing robustness. Together, these innovations enable superior accuracy and efficiency in detecting small defects under challenging industrial conditions.

2.2 Automotive Sealing Rings Defect Detection

Defect detection in automotive sealing rings has been a critical area of research in quality control. Li et al. proposed a defect detection method based on the K-means clustering approach and particle swarm optimization [31]. Their approach involves preprocessing seal images, utilizing the SURF algorithm for feature point extraction, and performing image segmentation. However, their method relies on traditional machine learning techniques, which may lack the adaptability and efficiency required for diverse defect types and high-volume production environments. Scholz et al. both explored 3D-based approaches for defect detection [8]. They developed an automated optical inspection system based on sheet-of-light laser measurement, integrating a 3D data acquisition unit for rapid full-surface scanning of parts. Nevertheless, these approaches do not specifically address the challenges of real-time detection or the processing of high-volume production data.

In contrast, the enhanced YOLOv8 model proposed in this paper has been specifically optimized to address the challenges in industrial environments. This method is not only capable of adapting to diverse defect types but also possesses real-time detection capabilities, effectively handling high-frequency data input.

2.3 IoULoss

In object detection, Intersection over Union (IoU) serves as a key metric for evaluating anchor box matching accuracy, with various enhancements proposed to improve its effectiveness [32]. Variants like GIoU [33], EIoU [34], and SIoU [35] have been introduced to address specific limitations, such as handling non-overlapping bounding boxes or incorporating geometric considerations for better localization accuracy. While these advancements have improved object detection in many scenarios, they remain less effective for precise localization of small objects, often leading to positional offset errors that compromise performance in such cases [36].

To optimize the YOLOv8 model, we introduced BAWD loss, built on the foundation of NWD Loss [37,38]. NWD Loss models bounding boxes as Gaussian distributions, enabling a smooth and continuous distance metric. However, defocus blur in industrial images compromises this modeling by making object edges less distinct, leading to inaccuracies in estimating Gaussian parameters. BAWD loss mitigates this issue by incorporating blur awareness, dynamically adjusting covariance matrices based on blur levels. This refinement ensures more robust localization of small objects under blurred conditions, improving detection accuracy in challenging industrial environments.

3 Method

3.1 Automotive Sealing Ring Defect Detection Process

This paper presents a detection method composed of three parts, specifically designed for detecting surface defects in automotive sealing rings, as shown in Fig. 1.



Figure 1: Automotive sealing ring defect detection process

(1) **Input:** This module inputs the captured image into the automotive sealing ring defect detector without any specific size constraints.

(2) Automotive Sealing Ring Defect Detector: Initially, images are checked for their sizes, and those not meeting the 640×640 pixel requirement are compressed. This physically enlarges the defects and reduces the negative sample information. The Enhanced YOLO network is then trained using COCO dataset weights.

(3) **Output:** The defects are detected in the obtained image using Enhanced YOLO network and the detection results are obtained.

3.2 Enhanced YOLO

This section proposes an improved model Enhanced YOLO based on the network structure of YOLOv8. The specific structure is shown in Fig. 2. This study introduces two major improvements to YOLOv8. Firstly, the MAFE module is incorporated into the network to establish long-range dependencies during feature extraction, offering stronger spatial aggregation capabilities compared to standard convolutional modules. Additionally, the MAFE module minimizes parameters while mitigating inherent limitations of stride and pooling operations. This enhancement improves the focus on low-resolution images and boosts the detection of small object defects. Secondly, the BAWD loss function is introduced, which measures the similarity of the bounding box's Gaussian distribution, enabling more accurate regression of predicted boxes. This modification further strengthens small object detection and boosts overall precision.



Figure 2: Enhanced YOLO network structure

3.2.1 MAFE Module

The MAFE module combines SPD and DCN to enhance feature extraction specifically for small and irregular defects. SPD rearranges spatial information into channel dimensions, effectively expanding the receptive field and enhancing focus on fine-grained details. This is particularly advantageous for small defects such as burrs and fractures. Meanwhile, DCN introduces learnable spatial offsets, allowing convolutions to dynamically adapt to irregular defect shapes and deformations. This adaptation ensures critical spatial information is preserved, overcoming limitations of conventional pooling methods. These mechanisms collectively enable the model to capture long-range dependencies and provide robust feature representations under complex industrial conditions. The structure of the MAFE module is shown in Fig. 3.



Figure 3: MAFE module

The SPD operation rearranges the spatial information of the input feature map into the channel dimension. Assuming the input feature map is $X \in \mathbb{R}^{C \times H \times W}$, the SPD operation transforms it along the spatial dimensions, resulting in a new feature map $X' \in \mathbb{R}^{(C \times s^2) \times (H/s) \times (W/s)}$. The SPD can be expressed as:

$$X_{c',h',w'}^{'} = X_{c,h'\times s+i,w'\times s+j}$$
(1)

where $c' = c \times s^2 + i \times s + j$, i, j = 0, 1, ..., s - 1.

Specifically, first SPD converts the feature image $X(S, S, C_1)$ into intermediate feature image $X'\left(\frac{S}{scale}, \frac{S}{scale}, scale^2C_1\right)$. Then it is further converted into $X'\left(\frac{S}{scale}, \frac{S}{scale}, scale^2C_1\right) \rightarrow X''\left(\frac{S}{scale}, \frac{S}{scale}, C_2\right)$ by non-strided convolution. An illustration of the *scale* = 2 conversion process is given in Fig. 4.



Figure 4: The process of SPD-Conv operation

The SPD operation compresses spatial information into the channel dimension, expanding the receptive field, which is particularly effective for extracting long-range dependencies and features of small targets.

DCN introduces learnable spatial offsets, making the sampling locations of convolution kernels more flexible. The conventional convolution is expressed as:

$$y(p_c) = \sum_k w(k) \cdot x(p_0 + p_k)$$
⁽²⁾

where p_c represents the location of the current pixel, and p_k represents the sampling location of the standard convolution kernel.

With the introduction of learnable offsets, DCN enables adaptive adjustment of convolution kernel sampling locations. The mathematical expression is:

$$y(p_c) = \sum_k w(k) \cdot x(p_c + p_k + \Delta p_{gk})$$
(3)

By introducing grouped convolution, spatial feature aggregation across multiple groups is achieved. The corresponding formula is:

$$y(p_c) = \sum_k w_g \cdot m_{gk} \cdot x_g \left(p_c + p_k + \Delta p_{gk} \right)$$
(4)

where *G* denotes the total number of groups, $w_g \in \mathbb{R}^{C \times C'}$ denotes the weights of group *g*, indicating a position-independent projection. C' = C/G represents the channel dimension of each group. $m_{gk} \in \mathbb{R}$ is the normalization factor obtained by applying the Softmax function along the *k*-th dimension. $x_g \in \mathbb{R}^{C' \times H \times W}$ denotes the input feature map of group *g*, and $Deltap_{gk}$ represents the offset of the sampling point p_k within group *g*.

Finally, the feature map X', after the SPD operation, is fed into the DCN:

$$y(p_c) = \sum_k w(k) \cdot X'(p_c + p_k + \Delta p_{gk})$$
(5)

3.2.2 BAWD Loss

BAWD loss addresses the challenges posed by defocus blur in high-speed production environments by refining the covariance matrices in Gaussian modeling. Specifically, it adjusts the Gaussian parameters dynamically to reflect varying levels of blur, mitigating edge ambiguities that often compromise small-object localization. This approach ensures robust and precise regression by smoothing the Wasserstein distance calculations between predicted and ground-truth boxes. The effectiveness of BAWD in handling blurred scenarios has been validated through experiments, showing significant improvements over conventional losses like CIoU and NWD in detecting small and subtle defects.

A rotating bounding box $B = (x, y, w, h, \theta)$, is represented as a two-dimensional Gaussian $G(\mu, \Sigma)$. The Wasserstein distance between two Gaussians $G_p = N(\mu_p, \Sigma_p)$ and $G_g = N(\mu_g, \Sigma_g)$ is given by:

$$W_2^2(G_p, G_g) = \left\| \mu_p - \mu_g \right\|_2^2 + \left\| \Sigma_p^{1/2} - \Sigma_g^{1/2} \right\|_F^2$$
(6)

where $\|.\|_{F}$ is the Frobenius norm.

The Gaussian distributions G_p and G_g are defined using bounding boxes $A = (cx_p, cy_p, w_p, h_p)$ and $B = (cx_g, cy_g, w_g, h_g)$, Eq.(6) can be rewritten as:

$$W_2^2\left(G_p, G_g\right) = \left\| \left(\left[cx_p, cy_p, \frac{w_p}{2}, \frac{h_p}{2} \right]^{\mathrm{T}}, \left[cx_g, cy_g, \frac{w_g}{2}, \frac{h_g}{2} \right]^{\mathrm{T}} \right) \right\|_2^2$$
(7)

where cx_a and cy_a represent the center coordinates of the predicted box, w_a and h_a are box width and height; cx_b , cy_b , w_b , h_b corresponds to the ground truth box.

The NWD loss is as follows:

$$NWD(G_p, G_g) = \exp\left(-\frac{\sqrt{W_2^2(G_p, G_g)}}{C_s}\right)$$
(8)

Incorporating the blur-aware weight w(B), the BAWD loss is formulated as:

$$L_{NWD} = w(B) \times NWD(G_p, G_g)$$
⁽⁹⁾

Here, G_p represents the Gaussian model of the prediction, while G_g corresponds to the ground truth. The term C_s is a dataset-dependent constant.

4 Experiments

4.1 Dataset

The Enhanced YOLO object detection algorithm is validated on two datasets: 1) automotive sealing ring defect dataset constructed by us, and 2) the MS COCO2017 dataset. The performance on the automotive sealing ring dataset directly reflects the model's capability in the specific task of automotive sealing ring defect detection. Meanwhile, its performance on the COCO dataset demonstrates the model's generalization ability across diverse detection scenarios.

4.1.1 Automotive Sealing Ring Defect Dataset

This study manually constructed a dataset containing four common defects: blowout, fracture, excess material and burr. The defective automotive sealing rings are from a real-world project conducted by Shanghai Inspect Spring Automation Technology Company. This dataset was created to apply deep learning-based object detection models to automotive sealing ring defect detection. Purposefully, we set up harsh conditions in the dataset to ensure the images closely resemble actual industrial environments. This includes simulating situations such as uneven lighting, noise interference, and image blurring to enhance the robustness required of the algorithm. With such a dataset setup, this study aims to provide a challenging evaluation platform for testing and improving the performance of the sealing ring defect detection algorithm.

The automotive sealing ring defect dataset constructed by us consists of 4977 images, and the data acquisition equipment is shown in Fig. 5. The dataset includes four types of defects: blowout, fracture, excess material and burr. The quantity of each defect is shown in Fig. 6. To facilitate supervised learning, the dataset was split in an 8:1:1 ratio into a training set (4031 images), validation set (448 images), and test set (498 images).



Figure 5: The data acquisition system



Figure 6: Sealing ring defect data distribution

4.1.2 MS COCO2017 Dataset

The MS COCO2017 dataset, also called COCO2017, consists of 80 object categories, including humans, animals, vehicles, and household items. It is partitioned into 118,287 training images, 5000 validation images, and 40,670 test images, serving as a standard benchmark for evaluating object detection models and assessing their robustness.

4.2 Evaluation Metrics

The performance of the model will be evaluated by a variety of metrics such as Precision, Recall, mAP@0.5, mAP@0.5:0.95, Params and FLOPs.

Precision reflects the accuracy of the model in identifying positive samples, minimizing false positives. Recall represents the proportion of actual positives correctly predicted, with a high score indicating strong capability in detecting positive instances. The equations for Precision and Recall are as follows:

$$Precistion = \frac{TP}{FP + TP}$$
(10)

$$Recall = \frac{TP}{FN + TP}$$
(11)

TP indicates correctly classified positive samples, *FP* indicates misclassified positive samples, and *FN* indicates misclassified negative samples.

mAP@0.5 focuses on the average accuracy of the model's predictions concerning the actual objects (Ground Truth) with a bounding box overlap of 0.5. This evaluation emphasizes the model's performance at relatively high intersection and overlap ratio thresholds. In contrast, mAP@0.5:0.95, which covers multiple intersection and merging ratio thresholds ranging from 0.5 to 0.95, comprehensively assesses the model across different conditions. The AP and mAP formulas are given below:

$$AP = \int_0^1 p(r) dr \tag{12}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{13}$$

where p(r) represents the accuracy for a given recall and k represents the number of categories. Normally, this study uses approximation or interpolation to calculate the *AP* value.

Params denotes the model's trainable parameters, indicating the size of the model. FLOPs (Floating Point Operations) quantify the computational cost of model inference or training. Higher FLOPs indicate increased resource requirements. The equations of Params and FLOPs can be expressed as:

$$Params = (C_{in} \times K^2 + 1) \times C_{out}$$
⁽¹⁴⁾

$$FLOPs = 2 \times C_{in} \times C_{out} \times K^2 \times H \times W$$
⁽¹⁵⁾

where C_{in} and C_{out} represent the input and output feature map channels, respectively. *K* represents the convolution kernel size. *H* and *W* denote the output feature map's height and width, respectively.

4.3 Validation Results on the Automotive Sealing Ring Dataset

4.3.1 Cross-Validation

In order to avoid a single evaluation being affected by a specific data distribution, leading to unstable evaluation results, this study conducted five-fold cross-validation experiments on the automotive sealing ring dataset to more accurately assess the model's generalization ability. The results are shown in Table 1. The improved model proposed in this paper achieves 98.30%, 96.62%, 85.55% and 53.45% in precision, recall, mAP@0.5 and mAP@0.5:0.95. It is worth noting that the fluctuation range of each metric is minimal in different cross-validation experiments, which validates that the automotive sealing ring dataset we constructed has an excellent and balanced data distribution.

Order number	Precision	Recall	mAP@0.5	mAP@0.5:0.95
1	98.72%	96.95%	85.85%	54.64%
2	98.16%	96.10%	85.54%	52.82%
3	98.54%	97.32%	86.08%	53.76%
4	97.62%	96.04%	84.56%	51.67%
5	98.44%	96.68%	85.74%	54.36%
average	98.30%	96.62%	85.55%	53.45%

Table 1: Cross-validation results on automotive sealing ring dataset

4.3.2 Comparison with Other Methods

To assess the detection performance of the proposed model, we conducted a comparative analysis with other lightweight models on the automotive sealing ring dataset. The comparison results from the cross-validation experiment are shown in Table 2. While the Enhanced YOLO may not hold the edge in terms of inference speed and FLOPs, its performance in accuracy metrics stands out. Specifically, it achieves mAP@0.5 and mAP@0.5:0.95 scores of 85.73% and 54.36%, respectively, indicating superior performance levels. In addition, the parameters of proposed model are 2.8 M in total, lower than other existing methods. This makes it suitable for embedded devices and reduces hardware strain. This experiment offers a valuable benchmark for the industrial community when determining the optimal model for defect detection in automotive sealing rings.

Table 2: Results of different lightweight models on the automotive sealing ring dataset

Model	mAP@0.5	mAP@0.5:0.95	Inference time	FLOPs	Params
YOLOv4tiny	76.45%	45.62%	3.5 ms	6.9 G	5.8 M
YOLOv5s	79.36%	48.84%	4.2 ms	7.2 G	16.5 M
YOLOv6s	73.70%	43.60%	3.0 ms	4.7 G	11.4 M
YOLO7tiny	57.02%	27.36%	2.1 ms	6.2 G	5.8 M
YOLOv9-T	79.44%	49.23%	2.7 ms	7.7 G	2.0 M
YOLOv10-N	81.66%	50.87%	1.9 ms	6.7 G	2.3 M
Ours	85.74%	54.36%	20.3 ms	8.1 G	2.8 M

4.4 Validation Results on COCO2017 Dataset

To assess our model's robustness, we compared it with mainstream object detection models using the COCO dataset. Detailed numerical results are presented in the Table 3. As depicted in Table 3, although EfficientDet-D0, NanoDet, and MobileDet are comparable to our model in terms of the amount of parameter computation, the Enhanced YOLO outperforms in accuracy. Moreover, while SpineNet-49s achieves similar accuracy, and Deformable DETR and Efficient DETR perform slightly better than ours, it is noteworthy that our model retains a significant advantage in both parameter count and computational complexity. The outstanding performance on the COCO dataset demonstrates the robust generalization capability of the Enhanced YOLO across diverse scenarios.

Model	Backbone	mAP@0.5:0.95	FLOPs	Params
Efficientdet-D0	Efficientnet	34.3%	2.5 G	3.9 M
Retinanet	Resnet-50	37.0%	97 G	34 M
Nanodet	Darknet53	34.1%	2.97 G	2.5 M
DeformableDETR	Resnet-50	43.8%	173 G	40 M
Mobiledet	MobilenetV2	25.7%	1.37 G	6.4 M
SpineNet-49s	SpineNet-49s	39.5%	33.8 G	11.9 M
Efficient DETR	Resnet-50	44.2%	159 G	32 M
YOLOv8n	YOLOv8n	37.3%	8.7 G	3.2 M
Ours	Enhanced YOLO	40.8%	8.1 G	2.8 M

Table 3: Comparison results with state-of-the-art models on the COCO dataset

4.5 Ablation Experiments

To assess the effect of each enhancement, we performed a detailed ablation study. The visualization process of the model training is shown in Fig. 7. The specific results are shown in Table 4. For each enhancement strategy, respective control groups were set up to evaluate its impact on model performance. Specifically, "+DCN", "+SPD", "+BAWDLoss" and "Ours" represent the introduction of the DCN module, the introduction of the SPD module, adding the BAWD loss function, and our proposed model, respectively. As shown in Table 4, adding the DCN module, the SPD module, or changing the loss function each enhances performance metrics like Recall, mAP@0.5, and mAP@0.5:0.95. Moreover, introducing both modules consistently and substantially improved the model compared to the enhancements achieved by introducing a single module. The performance of the improvements with both modules also steadily improves compared to that of separate modules. Ultimately, when all three modules were simultaneously introduced, the model achieved optimal performance. In conclusion, the ablation study facilitated a precise understanding of the contribution of each enhancement. It also provided clarity on how individual and collective integrations of the strategies influenced the model's performance, guiding future optimizations and decisions. The detection results and the heatmaps of the original model and the improved model are shown in Figs. 8 and 9, respectively.



Figure 7: The model's performance variations with different improvements on the automotive sealing ring dataset. Different curves highlight changes in metrics like mAP and recall over epochs, evaluating the impact of each enhancement

Method	Recall	mAP@0.5	mAP@0.5:0.95	Inference time	Params
Baseline	91.62%	81.51%	50.03%	5.2 ms	3.1 M
+DCN	93.46%	83.48%	52.02%	17.7 ms	2.6 M
+SPD	92.58%	82.59%	51.36%	6.3 ms	3.3 M
+BAWDLoss	92.72%	82.37%	51.13%	5.8 ms	3.1 M
+MAFE	94.74%	84.64%	53.16%	18.8 ms	2.8 M
+DCN+BAWDLoss	95.49%	84.69%	52.14%	17.9 ms	2.6 M
+SPD+BAWDLoss	93.60%	83.95%	52.57%	8.1 ms	3.3 M
Ours	96.62%	85.74%	54.36%	20.3 ms	2.8 M

Table 4: Comparison of results with deconstructed models on the automotive sealing ring dataset



(b) Enhanced YOLO





Figure 9: The heatmaps of the original model and the improved model

5 Conclusion

This paper proposes an improved model based on the state-of-the-art model YOLOv8 in order to solve the task of automotive sealing ring defect detection. The introduction of the MAFE module enhances the model's ability to capture long-range dependencies and aggregate spatial features, significantly improving its robustness in detecting defects under complex industrial conditions. Additionally, the incorporation of the BAWD loss enhances the model's precision in identifying small objects by more accurately modeling the bounding box similarity, especially in challenging scenarios with blurry or uncertain object boundaries. These improvements enable the model proposed in this paper to obtain significant performance enhancement while maintaining the characteristics of lightweight design and high-speed inference. It is worth mentioning that this paper achieves advanced accuracy performance on a specially constructed automotive sealing ring dataset. This approach significantly reduces the leakage and false detection rates, providing an excellent example for the field of automotive sealing ring defect detection and other rubber industrial products.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Jian Ge; data collection: Jian Ge, Yahui Zhang; analysis and interpretation of results: Jian Ge, Zimei Tu, Jinhua Jiang; draft manuscript preparation: Jian Ge, Qin Qin, Zhiwei Shen. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- Muhammad K, Ullah A, Lloret J, Del Ser J, de Albuquerque VHC. Deep learning for safe autonomous driving: current challenges and future directions. IEEE Trans Intell Transp Syst. 2020;22(7):4316–36. doi:10.1109/TITS.2020. 3032227.
- 2. Wang Z, Zhan J, Duan C, Guan X, Lu P, Yang K. A review of vehicle detection techniques for intelligent vehicles. IEEE Trans Neural Netw Learn Syst. 2023;34(8):3811–31. doi:10.1109/TNNLS.2021.3128968.
- Chen S, Jin D, He H, Yang F, Yang J. Deep learning based online non-destructive defect detection for self piercing riveted joints in automotive body manufacturing. IEEE Trans Ind Inform. 2023;19(8):9134–44. doi:10.1109/TII. 2022.3226246.
- 4. Varol S, Odougherty P. A predictive analysis of electronic control unit system defects within automotive manufacturing. J Fail Anal Prev. 2022;22(3):918–25. doi:10.1007/s11668-022-01401-0.
- 5. Türköz M, Evcen DC. Investigation on the sealing performance of polymers at ultra high pressures. Int Polym Process. 2022;37(5):549–58. doi:10.1515/ipp-2022-4226.
- 6. Zhang Y, Wang Y, Jiang Z, Zheng L, Chen J, Lu J. Domain adaptation via transferable swin transformer for tire defect detection. Eng Appl Artif Intell. 2023;122(1):106109. doi:10.1016/j.engappai.2023.106109.
- 7. Parlak IE, Emel E. Deep learning-based detection of aluminum casting defects and their types. Eng Appl Artif Intell. 2023;118(2):105636. doi:10.1016/j.engappai.2022.105636.
- 8. Scholz O, Doerfler N, Seifert L, Zöller U. Fully-automatic surface inspection of O-ring seals. In: SAE technical paper; 2015.

- 9. Lu X, Jiang C, Ma Z, Li H, Liu Y. A simple and effective surface defect detection method of power line insulators for difficult small objects. Comput Mater Contin. 2024;79(1):373–90. doi:10.32604/cmc.2024.047469.
- 10. Kim G-I, Yoo H, Cho H-J, Chung K. Defect detection model using time series data augmentation and transformation. Comput Mater Contin. 2024;78(2):1713–30. doi:10.32604/cmc.2023.046324.
- 11. Jiang W, Li T, Zhang S, Chen W, Yang J. PCB defects target detection combining multi-scale and attention mechanism. Eng Appl Artif Intell. 2023;123(4968):106359. doi:10.1016/j.engappai.2023.106359.
- 12. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Analy Mach Intell. 2016;39(6):1137–49. doi:10.1109/TPAMI.2016.2577031.
- 13. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision; 2017; Venice, Italy. p. 2980–8.
- 14. Redmon J. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016.
- 15. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. Ssd: single shot multibox detector. In: Computer Vision-ECCV 2016: 14th European Conference; 2016 Oct 11–14; Amsterdam, The Netherlands: Springer.
- 16. Jiang P, Ergu D, Liu F, Cai Y, Ma B. A review of yolo algorithm developments. Procedia Comput Sci. 2022;199(11):1066-73. doi:10.1016/j.procs.2022.01.135.
- 17. Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018; Salt Lake City, UT, USA. p. 6154–62.
- Revathy G, Kalaivani R. Fabric defect detection and classification via deep learning-based improved mask RCNN. Signal Image Video Process. 2024;18(3):2183–93. doi:10.1007/s11760-023-02884-6.
- Chen C, Wang S, Huang S. An improved faster RCNN-based weld ultrasonic atlas defect detection method. Meas Control. 2023;56(3–4):832–43. doi:10.1177/00202940221092030.
- 20. Cheng S, Lu J, Yang M, Zhang S, Xu Y, Zhang D, et al. Wheel hub defect detection based on the DS-cascade RCNN. Measurement. 2023;206(1):112208. doi:10.1016/j.measurement.2022.112208.
- 21. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: European Conference on Computer Vision; 2020; Cham: Springer.
- 22. Zhu X, Su W, Lu L, Li B, Wang X, Dai J. Deformable DETR: deformable transformers for end-to-end object detection; 2020. arXiv:201004159. 2020..
- 23. Dai X, Chen Y, Yang J, Zhang P, Yuan L, Zhang L. Dynamic DETR: end-to-end object detection with dynamic attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021; Montreal, QC, Canada. p. 2968–77.
- 24. Zhou X, Ren Z, Zhang Y, Mi T, Zhou S, Jiang Z. A shunted-swin transformer for surface defect detection in roller bearings. Measurement. 2024;238(1):115283. doi:10.1016/j.measurement.2024.115283.
- 25. Sun X, Song K, Wen X, Wang Y, Yan Y. SDD-DETR: surface defect detection for no-service aero-engine blades with detection transformer. IEEE Trans Autom Sci Eng. 2024;22:6984–97. doi:10.1109/TASE.2024.3457829.
- 26. Zhang Q, Lai J, Zhu J, Xie X. Wavelet-guided promotion-suppression transformer for surface-defect detection. IEEE Trans Image Process. 2023;32:4517–28. doi:10.1109/TIP.2023.3293770.
- 27. Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023; Vancouver, BC, Canada. p. 7464–75.
- 28. Huang X, Zhu J, Huo Y. SSA-YOLO: an improved YOLO for hot-rolled strip steel surface defect detection. IEEE Trans Instrum Meas. 2024;73(10):5040017. doi:10.1109/TIM.2024.3488136.
- 29. Li H, Wang C, Liu Y. YOLO-FDD: efficient defect detection network of aircraft skin fastener. Signal Image Video Process. 2024;18(4):3197–211. doi:10.1007/s11760-023-02983-4.
- 30. Xiao D, Xie FT, Gao Y, Li ZN, Xie HF. A detection method of spangle defects on zinc-coated steel surfaces based on improved YOLO-v5. Int J Adv Manuf Technol. 2023;128(1–2):937–51. doi:10.1007/s00170-023-11963-4.
- 31. Li X, Zhu J, Shi H, Cong Z. Surface defect detection of seals based on k-means clustering algorithm and particle swarm optimization. Sci Program. 2021;2021(1):3965247–12. doi:10.1155/2021/3965247.

- 32. Yu J, Jiang Y, Wang Z, Cao Z, Huang T. UnitBox: an advanced object detection network. In: Proceedings of the 24th ACM International Conference on Multimedia; 2016. p. 516–520.
- Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized intersection over union: a metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 658–66.
- 34. Zhang Y-F, Ren W, Zhang Z, Jia Z, Wang L, Tan T. Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing. 2022;506(9):146–57. doi:10.1016/j.neucom.2022.07.042.
- 35. Gevorgyan Z. SIoU loss: more powerful learning for bounding box regression. arXiv:2205.12740. 2022.
- 36. Zhao J, Zhu H. CBPH-Net: a small object detector for behavior recognition in classroom scenarios. IEEE Trans Instrum Meas. 2023;72:2521112. doi:10.1109/TIM.2023.3296124.
- 37. Wang J, Xu C, Yang W, Yu L. A normalized gaussian wasserstein distance for tiny object detection. arXiv:2110.13389. 2021.
- 38. Li X, Wang W, Wu L, Chen S, Hu X, Li J, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection. Adv Neural Inf Process Syst. 2020;33:21002–12.