Structural Durability & Health Monitoring

Tech Science Press

Doi:10.32604/sdhm.2025.070589

ARTICLE

Optimized Industrial Surface Defect Detection Based on Improved YOLOv11

Hua-Qin Wu[1,2], **Hao Yan**[1,2], **Hong Zhang**[1,2,*], **Shun-Wu Xu**[1,2], **Feng-Yu Gao**[1,2] **and Zhao-Wen Chen**[1,2]

[1]Key Laboratory of Nondestructive Testing, Fujian Polytechnic Normal University, Fuqing, 350300, China
[2]School of Electronic and Mechanical Engineering, Fujian Polytechnic Normal University, Fuqing, 350300, China

*Corresponding Author: Hong Zhang. Email: zhangh@fpnu.edu.cn

Received: 19 July 2025; Accepted: 25 September 2025; Published: 31 December 2025

ABSTRACT: In industrial manufacturing, efficient surface defect detection is crucial for ensuring product quality and production safety. Traditional inspection methods are often slow, subjective, and prone to errors, while classical machine vision techniques struggle with complex backgrounds and small defects. To address these challenges, this study proposes an improved YOLOv11 model for detecting defects on hot-rolled steel strips using the NEU-DET dataset. Three key improvements are introduced in the proposed model. First, a lightweight Guided Attention Feature Module (GAFM) is incorporated to enhance multi-scale feature fusion, allowing the model to better capture and integrate semantic and spatial information across different layers, which improves its ability to detect defects of varying sizes. Second, an Aggregated Attention (AA) mechanism is employed to strengthen the representation of critical defect features while effectively suppressing irrelevant background information, particularly enhancing the detection of small, low-contrast, or complex defects. Third, Ghost Dynamic Convolution (GDC) is applied to reduce computational cost by generating low-cost ghost features and dynamically reweighting convolutional kernels, enabling faster inference without sacrificing feature quality or detection accuracy. Extensive experiments demonstrate that the proposed model achieves a mean Average Precision (mAP) of 87.2%, compared to 81.5% for the baseline, while lowering computational cost from 6.3 Giga Floating-point Operations Per Second (GFLOPs) to 5.1 GFLOPs. These results indicate that the improved YOLOv11 is both accurate and computationally efficient, making it suitable for real-time industrial surface defect detection and contributing to the development of practical, high-performance inspection systems.

KEYWORDS: YOLOv11; object detection; industrial surface defect; NEU-DET

1 Introduction

In industrial manufacturing, surface defect detection is crucial for ensuring product quality and improving efficiency. In hot-rolled steel production lines, surface defects such as cracks, inclusions, and rolled-in scales not only degrade the mechanical properties of steel products but also cause serious downstream processing issues, including equipment wear and product rejection [1]. Rapid and accurate defect detection is therefore essential for ensuring high-throughput production and meeting increasingly stringent quality standards [2].

From an operational perspective, steel manufacturers demand detection systems that can operate in real time to match high-speed production lines, maintain high accuracy to avoid costly false detections, and remain lightweight enough for deployment on resource-constrained industrial hardware [3]. These requirements highlight the necessity of developing efficient and robust algorithms for industrial defect detection [4].

Traditional manual inspection is slow and inconsistent, while machine vision–based automated methods greatly enhance detection speed and accuracy [5]. With the advancement of deep learning, object detection algorithms have shown strong performance in defect detection across steel, electronics, and automotive industries [6]. This study focuses on the NEU-DET dataset [2], which contains 1800 grayscale images of six common hot-rolled steel surface defects: Rolled-in Scale, Patches, Crazing, Pitted Surface, Inclusion, and Scratches. Optimizing deep learning algorithms for NEU-DET can significantly improve detection performance and advance intelligent industrial inspection.

Surface defect detection has evolved from traditional image processing to machine learning, and more recently, to deep learning. Early methods like edge detection and texture analysis worked in simple scenarios but lacked robustness for complex defects. Machine learning techniques such as Support Vector Machine (SVM) and Random Forest improved performance but relied heavily on handcrafted features [7]. Deep learning, especially Convolutional Neural Networks (CNNs), has transformed object detection, with models like Faster R-CNN [8], YOLO [9], and RetinaNet [10] widely used in industrial inspection. YOLO stands out for its end-to-end structure, high speed, and accuracy, making it ideal for real-time tasks. Two-stage methods (e.g., Faster R-CNN) offer higher accuracy but slower speed, while one-stage methods like YOLO are faster and more efficient [11,12]. YOLOv11, released in 2024, improves detection and efficiency through architectural and training upgrades [13]. In addition, several recent studies have demonstrated the potential of YOLOv11-based models in tasks such as surface defect detection [14], industrial visual inspection [15], and small-object detection [16].

However, challenges remain in detecting small defects and handling complex backgrounds in the NEU-DET dataset, highlighting the need for further optimization.

To address the aforementioned issues, this paper proposes three optimization strategies based on the YOLOv11 model and validates their effectiveness through experiments:

- Guided Attention Feature Module (GAFM) is integrated to replace the traditional structure, enhancing multi-scale feature fusion through attention-guided mechanisms. It improves robustness in complex backgrounds by enabling more effective semantic and spatial information exchange across scales.
- Aggregated Attention Mechanism combines global and local attention to highlight key defect regions and suppress background noise, enhancing feature representation, especially for small and low-contrast defects.
- Ghost Dynamic Convolution uses lightweight dynamic kernels to reduce computation while maintaining rich feature extraction, boosting inference efficiency without sacrificing accuracy.

2 Methodology and Improvements

2.1 Overview of the YOLOv11 Architecture

As a new-generation improvement in the YOLO series, YOLOv11 introduces multiple optimizations in its architecture to enhance detection accuracy, robustness, and inference efficiency. As shown in Fig. 1, the overall framework retains the three-part structure of the YOLO series: the backbone, the neck, and the detection head. However, several innovative design strategies have been incorporated into each module to improve feature representation and computational efficiency.

YOLOv11 is an recently released model family released by Ultralytics, with its architecture introduced in a recent arXiv preprint [13] and implemented in their official open-source codebase [17]. Despite its recent release, YOLOv11 has already been widely adopted in both academia and industry, serving as a practical baseline for numerous computer vision tasks. In our experiments, we use the lightweight YOLOv11n variant following the Ultralytics default training configuration.

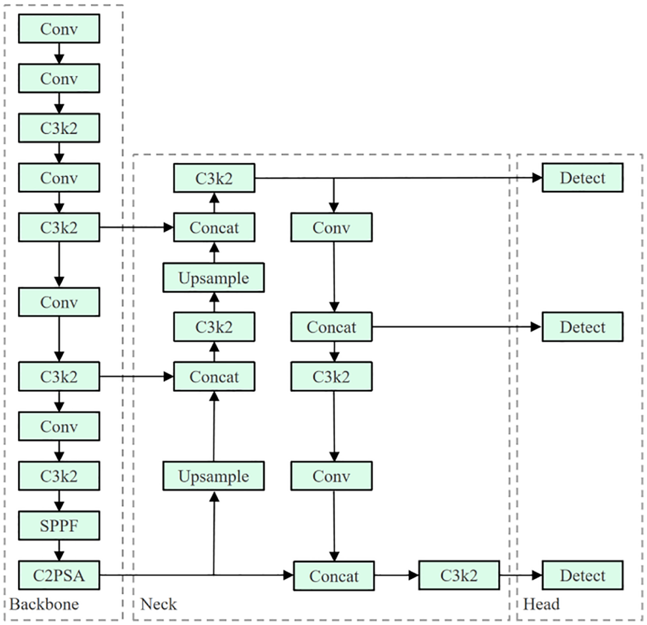

Figure 1: Structure of YOLOv11

2.2 *Detailed Improvement Strategies*

As shown in Fig. 2, to enhance the detection accuracy, multi-scale adaptability, and computational efficiency of YOLOv11 in industrial surface defect detection, this study introduces three targeted improvements: GAFM, Aggregated Attention, and Ghost Dynamic Convolution. These modules are designed to optimize feature fusion across different scales, enhance the model's focus on critical defect regions, and reduce redundant computations while maintaining high-quality feature representation. The following sections provide detailed descriptions of each enhancement

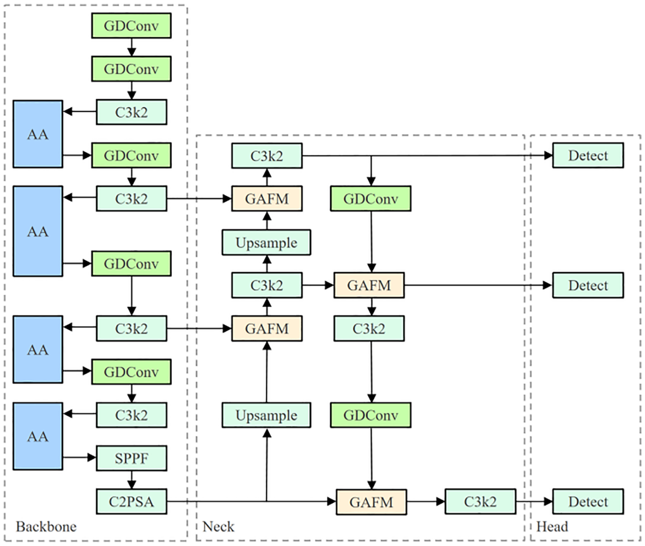

Figure 2: Improved YOLOv11 structure

2.2.1 Guided Attention Feature

The Guided Attention Feature Module (GAFM) is introduced to enhance the multi-scale feature fusion of YOLOv11 by enabling bidirectional guidance between low-level and high-level representations under a unified channel attention mechanism. Given a low-level feature map $F_l \in \mathrm{R}^{C_l \times H \times W}$ and a high-level feature map $F_\mathrm{h} \in \mathrm{R}^{C_h \times H \times W}$, the two are concatenated along the channel dimension to form a joint representation $F_c = \mathrm{Concat}(F_l, F_\mathrm{h})$.

A squeeze-and-excitation (SE) [18] operation is then applied to F_c. Specifically, global average pooling followed by two fully connected layers with non-linear activations produces a channel-wise weight vector, which is partitioned according to the channel dimensions of F_l and F_h to recalibrate each branch as $\tilde{F}_l = F_l \odot s_l$ and $\tilde{F}_\mathrm{h} = F_\mathrm{h} \odot s_\mathrm{h}$. To strengthen the complementary relationship across scales, a cross-residual fusion strategy is adopted, in which the recalibrated shallow feature is combined with the original high-level feature, while the recalibrated high-level feature is combined with the original low-level feature. The final fused representation is then obtained by concatenating the two residual outputs, expressed as $F_{fu} = \mathrm{Concat}(\widetilde{F_l} + F_\mathrm{h}, \tilde{F}_\mathrm{h} + F_l)$.

This design preserves fine-grained spatial information from shallow layers while simultaneously reinforcing the semantic consistency of deeper layers, thereby enhancing the model's ability to focus on defect-sensitive regions with limited computational overhead. Within the overall architecture, GAFM is integrated at critical fusion points, namely between the backbone C3_k2 block and the upsampling path, as well as within the neck after C3_k2 and following the Ghost Dynamic Convolution (GDConv) operation, as illustrated in Fig. 3.

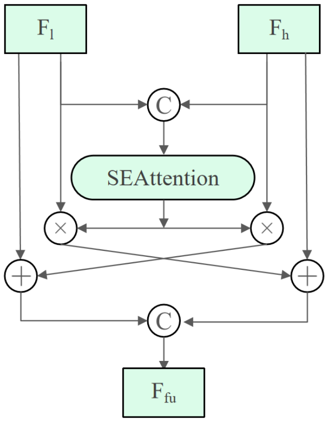

Figure 3: GAFM structure

2.2.2 Aggregated Attention

To strengthen the feature representation of YOLOv11 in complex and multi-scale defect detection scenarios, this study integrates the Aggregated Attention (AA) module [19]. As illustrated in Fig. 4, the AA structure follows a "preprocessing–attention aggregation–restoration" design. The input feature map $\mathrm{X} \in \mathbb{R}^{\mathrm{B} \times \mathrm{C} \times \mathrm{H} \times \mathrm{W}}$ is first processed to align dimensions, then forwarded to the Aggregated Attention Layer, and finally restored to produce the output $\mathrm{Y} \in \mathbb{R}^{\mathrm{B} \times \mathrm{C} \times \mathrm{H} \times \mathrm{W}}$.

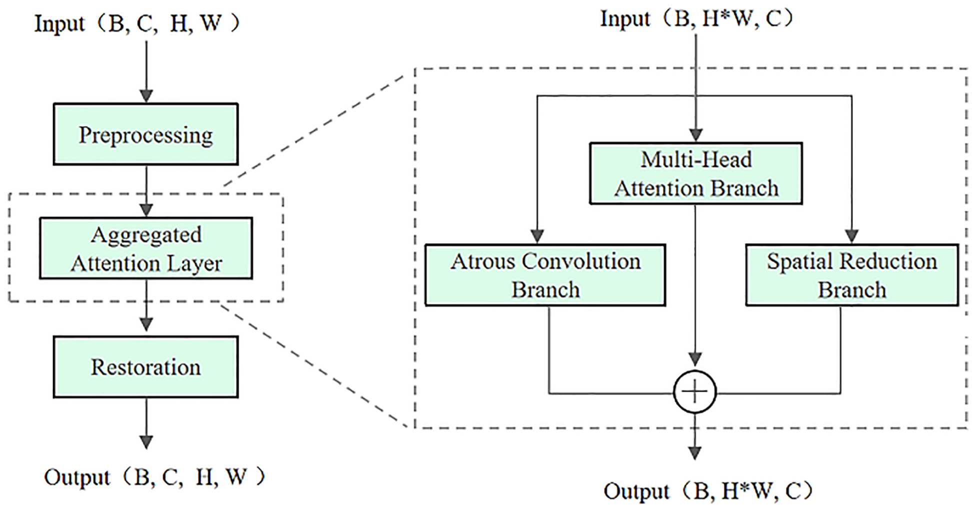

Figure 4: Aggregated attention structure

The Aggregated Attention Layer is composed of three complementary branches.

- Multi-head attention branch. The input is reshaped to $\mathbb{R}^{B\times N\times C}$ with $N = H \times W$, and the attention is computed as:

$$\mathrm{Attention}(Q, K, V) = \mathrm{Softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{1}$$

where Q, K, V are linear projections of the input. This branch models long-range dependencies and captures global contextual cues.

- Atrous convolution branch. To enhance local detail representation, atrous (dilated) convolution is applied:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k] \cdot w[k] \tag{2}$$

where r is the dilation rate. This allows an enlarged receptive field without increasing parameters.

- Spatial reduction branch. The Key/Value matrices are downsampled by a factor s, reducing computational complexity from $O(N^2)$ to $O(N \cdot \frac{N}{s})$:

$$\tilde{K}, \tilde{V} \in \mathbb{R}^{B\times\backslash\mathrm{tfracNs}\times C}. \tag{3}$$

This improves efficiency for high-resolution features while preserving discriminative information.

Finally, the outputs of the three branches are fused and reshaped back to $\mathbb{R}^{B\times C\times H\times W}$, forming the enhanced representation:

$$Y = \mathrm{Fusion}(Y_{attn}, Y_{atrous}, Y_{sr}). \tag{4}$$

By combining global attention, local receptive fields, and efficient spatial reduction, the AA module provides stronger adaptability to industrial surface defect detection, where defects vary in scale, morphology, and background complexity.

2.2.3 Ghost Dynamic Convolution

In this study, all standard convolution layers in the YOLOv11 backbone are replaced with the Ghost Dynamic Convolution (GDC) module to achieve a more efficient yet expressive feature extraction process. As shown in Fig. 5, GDC integrates the lightweight design of GhostNet [20] with the adaptive capability of dynamic convolution.

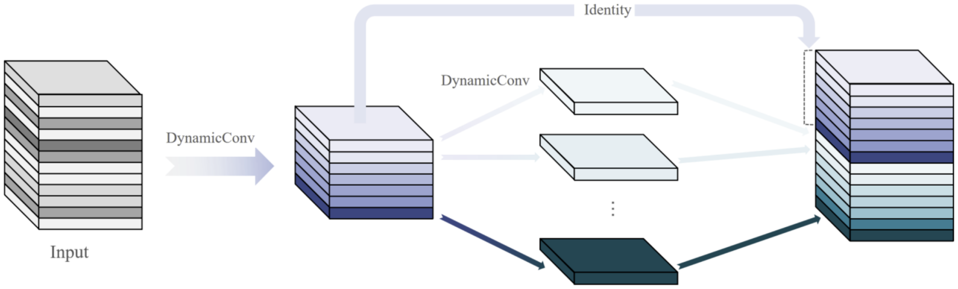

Figure 5: Ghost dynamic convolution structure

The Ghost module first reduces redundancy in convolutional feature maps by generating a small set of intrinsic feature maps through standard convolution, and then employs inexpensive linear transformations to produce additional "ghost" features. For an input tensor $X \in \mathbb{R}^{c\times h\times w}$, a conventional convolution operation is expressed as

$$Y = X * f + b, \tag{5}$$

where $f \in \mathbb{R}^{c\times k\times k\times n}$ represents the convolutional filters. In contrast, the Ghost module derives intrinsic features Y', followed by cheap transformations Φ to generate supplementary ghost features:

$$y_{i,j} = \Phi_{i,j}\left(y_i'\right), i = 1, \ldots, m, j = 1, \ldots, s \tag{6}$$

This results in the complete feature set

$$Y = [y_{11}, y_{12}, \ldots, y_{ms}], \tag{7}$$

with significantly fewer parameters and floating point operations (FLOPs), providing a theoretical compression ratio and speed-up of approximately s–fold [20].

On top of this efficient mechanism, dynamic convolution adaptively reweights multiple convolution kernels according to the input, allowing the model to capture richer, context-dependent representations. By embedding dynamic kernel selection into the ghost feature generation process, GDC achieves a favorable trade-off between efficiency and adaptability.

3 Dataset and Experimental Setup

3.1 Overview of the NEU-DET Dataset

The NEU-DET (Northeastern University Surface Defect Database) is a publicly available dataset released by Northeastern University, widely adopted in research on surface defect detection in industrial scenarios. It consists of 1800 grayscale images with a resolution of 200 × 200 pixels, covering six typical types of defects: crazing, inclusion, patches, pitted_surface, rolled-in_scale, and scratches, with 300 images per class. All samples were captured using industrial cameras and manually annotated. An example of the dataset is shown in Fig. 6.

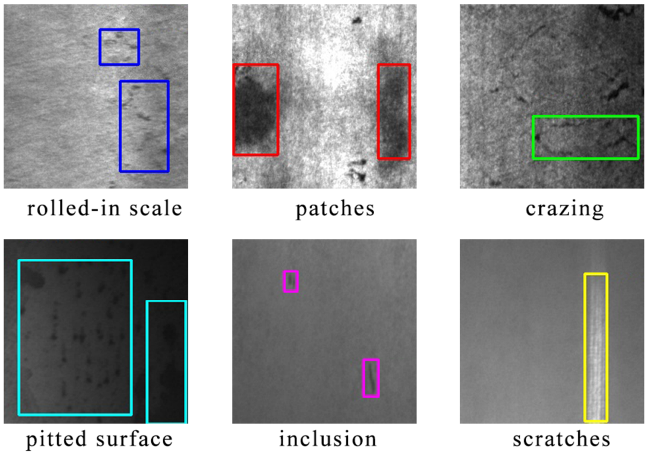

Figure 6: Example images from the NEU-DET dataset (defect regions highlighted)

To enhance the robustness and generalization capability of the model, data augmentation was applied to the original data set. The augmentation techniques included adding Gaussian noise, random rotations, brightness adjustments, and horizontal flipping. These basic transformations expanded the dataset size to three times its original scale, resulting in a total of 5400 images. The augmentation process ensured a balanced distribution of defect types and preserved the semantic integrity of the original images, providing a richer and more stable training foundation for the model.

3.2 Experimental Environment and Hardware Configuration

The experimental setup was based on a Windows 10 operating system, equipped with 64 GB of RAM, an NVIDIA RTX 4090D GPU, and an Intel Core i7-14700K CPU. All model training and evaluations were conducted using Python 3.9, PyTorch 2.3.1, and CUDA 12.1. The NEU-DET dataset was used throughout the experiments to train and optimize the proposed defect detection models. The primary hyperparameter settings used during training are summarized in Table 1.

Table 1: Training hyperparameters

Parameter	Value
Learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Optimizer	SGD

(Continued)

Table 1 (continued)

Parameter	Value
Batch size	16
Epochs	300

3.3 *Evaluation Metrics*

To comprehensively evaluate the performance of the proposed detection models and assess their suitability for real-world industrial defect detection, the following metrics were adopted: Precision (P), Recall (R), mean Average Precision (mAP), number of Parameters (Params), and computational complexity (GFLOPs).

- Precision quantifies the proportion of true positive predictions among all defect predictions, as in Eq. (8). A high precision indicates strong ability to distinguish defects from background, reducing false alarms.
- Recall measures the proportion of actual defect instances that are correctly identified, as expressed in Eq. (9). High recall reflects the model's effectiveness in detecting as many defects as possible, reducing the risk of missed detections.
- mAP is a widely used metric in object detection that summarizes the model's performance across all defect categories. It is calculated as the average area under the Precision-Recall (P-R) curve for each class, as in Eqs. (10) and (11).
- To assess model complexity and deployment feasibility, we also report the number of parameters (Params) and floating-point operations (FLOPs). These metrics provide insight into the model's computational requirements, which are critical for optimizing inference speed and ensuring compatibility with resource-constrained industrial applications.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{AP} = \int_0^1 P(R)\mathrm{d}R \tag{10}$$

$$\text{mAP} = \frac{1}{C}\sum_{i=1}^{C} \text{AP}_i \tag{11}$$

where:

- *TP* (True Positives): Correctly predicted defective samples.
- *FP* (False Positives): Normal samples incorrectly predicted as defects.
- *FN* (False Negatives): Defects not detected by the model.
- *C*: Number of defect classes in the dataset.

4 Experimental Results and Analysis

4.1 Comparison before and after Improvement

To evaluate the effectiveness of the proposed enhancements, a series of experiments was conducted on the NEU-DET dataset to compare the performance of the original YOLOv11 model and the improved version. The results are summarized in Table 2.

Table 2: Comparison of YOLOv11 before and after improvement

Model	mAP@0.5	Params	GFLOPs
Original YOLOv11n	0.815	2.58 M	6.3
Improved YOLOv11n	0.872	2.46 M	5.1
Improvement	↑0.057	↓0.12	↓1.2

In terms of detection accuracy, the improved YOLOv11 model achieved a mAP@0.5 of 87.2%, compared to 81.5% for the baseline. Precision improved from 77.8% to 84.7%, and recall increased from 73.1% to 83.0%. These results demonstrate that the integration of GAFM, Aggregated Attention, and Ghost Dynamic Convolution significantly enhances the model's ability to detect small and multi-scale defects under complex backgrounds.

In terms of model complexity, the number of parameters was reduced from 2.58 to 2.46 M, and the computational cost (GFLOPs) decreased from 6.3 to 5.1. This indicates that the proposed modifications not only improve detection accuracy but also increase inference efficiency, making the model more suitable for real-time industrial applications.

4.2 Ablation Study

To evaluate the individual contributions of each proposed module, we conducted ablation experiments based on the YOLOv11n baseline by progressively integrating GAFM, Aggregated Attention (AA), and Ghost Dynamic Convolution (GDC). The results are summarized in Table 3.

Table 3: Ablation results of proposed improvements

Method	mAP@0.5	Precision	Recall	Params	GFLOPs
Baseline	0.815	0.731	0.778	2.58 M	6.3
+GAFM	0.852	0.834	0.813	2.68 M	6.6
+AA	0.846	0.817	0.826	2.76 M	6.9
+GDC	0.816	0.743	0.769	1.78 M	3.7
+GAFM + AA	0.870	0.844	0.812	3.04 M	7.4
+GAFM + AA + GDC	0.872	0.830	0.847	2.46 M	5.1

Introducing GAFM alone significantly boosts the mAP@0.5 from 81.5% to 85.2%, demonstrating improved multi-scale feature fusion with only a minor increase in computational cost. The AA module further enhances the model's attention to key defect regions, especially improving recall.

Compared with GAFM and AA, GDC provides a relatively smaller accuracy improvement when used in isolation (from 81.5% to 81.6%). However, its primary advantage lies in substantially reducing model complexity, cutting parameters from 2.58 to 1.78 M and FLOPs from 6.3 to 3.7 G, thus achieving nearly

41% model compression and 41% computational savings. This indicates that GDC can serve as an efficient replacement for standard convolutions, maintaining accuracy while improving efficiency.

When combined with GAFM and AA, the complementary nature of the modules becomes evident. GAFM enriches multi-scale feature representations, AA enhances spatial attention, while GDC ensures the compactness of the backbone. For instance, combining GAFM and AA achieves 87.0% mAP@0.5, while integrating all three modules further boosts mAP@0.5 to 87.2%, with 83.0% precision and 84.7% recall, all while reducing parameters to 2.46 M and GFLOPs to 5.1. This confirms that the proposed improvements strike a favorable balance between accuracy and efficiency, with GDC playing a key role in enabling lightweight deployment without compromising detection performance.

4.3 Comparative Experiments

To further validate the effectiveness of the proposed improvements, the enhanced YOLOv11n model was compared with several mainstream YOLO variants, as well as the two-stage Faster R-CNN detector. The evaluation was conducted on the NEU-DET dataset, and the results are shown in Table 4.

Table 4: Comparison with YOLO series on NEU-DET dataset

Model	mAP@0.5	Precision	Recall	Params	GFLOPs
Faster R-CNN	0.759	0.745	0.712	41.75 M	135
YOLOv3	0.781	0.677	0.791	103.6 M	283
YOLOv5n	0.776	0.739	0.727	2.51 M	7.2
YOLOv6n	0.799	0.755	0.743	4.24 M	11.9
YOLOv8n	0.803	0.752	0.750	3.01 M	8.2
YOLOv11n	0.815	0.731	0.778	2.58 M	6.3
Ours	0.872	0.830	0.847	2.46 M	5.1

In terms of accuracy, the improved YOLOv11n achieved the highest performance, reaching a mAP@0.5 of 87.2%, which is 6.9 percentage points higher than YOLOv8n and 5.7 percentage points higher than the original YOLOv11n. Precision and Recall also showed noticeable improvements, demonstrating enhanced feature extraction and small-defect detection capabilities.

Compared with the two-stage Faster R-CNN, which achieved 75.9% mAP@0.5 with 41.75 M parameters and 135 GFLOPs, our model not only outperformed it in accuracy (+11.3 percentage points) but also required significantly fewer computational resources. This highlights the advantage of our lightweight design in balancing accuracy and efficiency.

Regarding computational efficiency, the improved model maintains a lightweight structure with only 2.46 M parameters and 5.1 GFLOPs, significantly outperforming the large-scale YOLOv3 model in terms of inference cost while delivering better accuracy.

In comparison, Li et al. [21] proposed a YOLOv5m-based method with MRAM and MAEH modules, achieving 82.7% mAP on NEU-DET but requiring over 20 M parameters, which indicates a higher computational burden despite good accuracy.

To provide a more comprehensive evaluation, Precision-Recall (PR) curves were also analyzed. Figs. 7 and 8 show the PR curves of the baseline YOLOv11n and the improved YOLOv11n on the NEU-DET dataset, respectively. The improved model demonstrates larger areas under the curves across most defect categories, indicating more stable detection performance and better robustness against small and low-contrast defects.

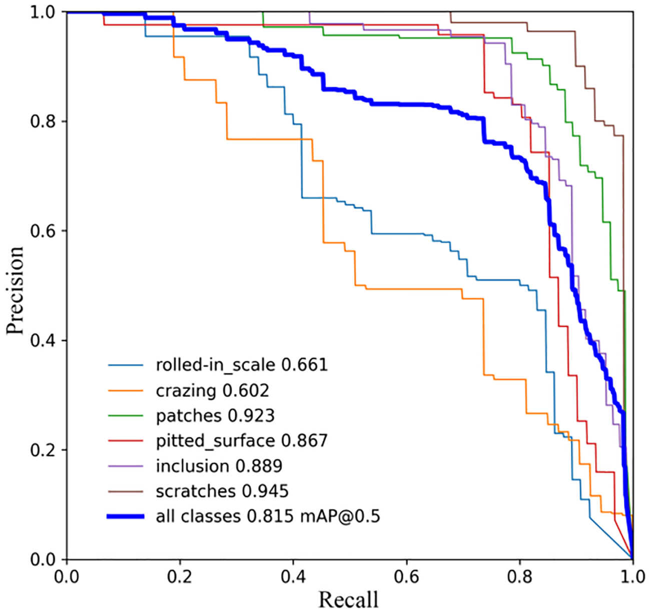

Figure 7: YOLOv11n P-R curve

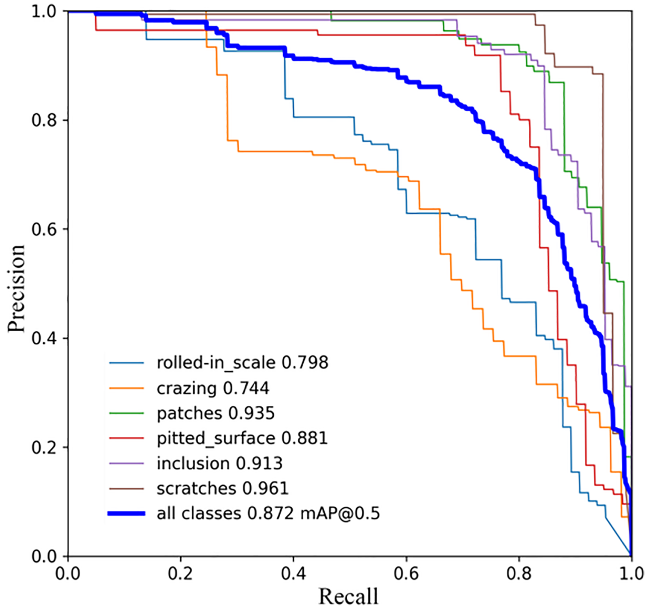

Figure 8: Improved YOLOv11n P-R curve

Overall, the proposed enhancements strike a good balance between detection accuracy and model efficiency, making the method well-suited for real-time industrial defect detection applications.

4.4 Visualized Detection Results

To more comprehensively illustrate the effectiveness of the proposed method, Fig. 9 provides a visual comparison between the baseline YOLOv11n and the improved YOLOv11n on representative samples from the NEU-DET dataset. The comparison covers six common defect types: Crack (Cr), Inclusion (In), Patches (Pa), Pitted Surface (Ps), Rolled-in Scale (Rs), and Scratches (Sc). In each group, the upper row shows the detection results of the improved YOLOv11n model, while the lower row shows the baseline YOLOv11n results.

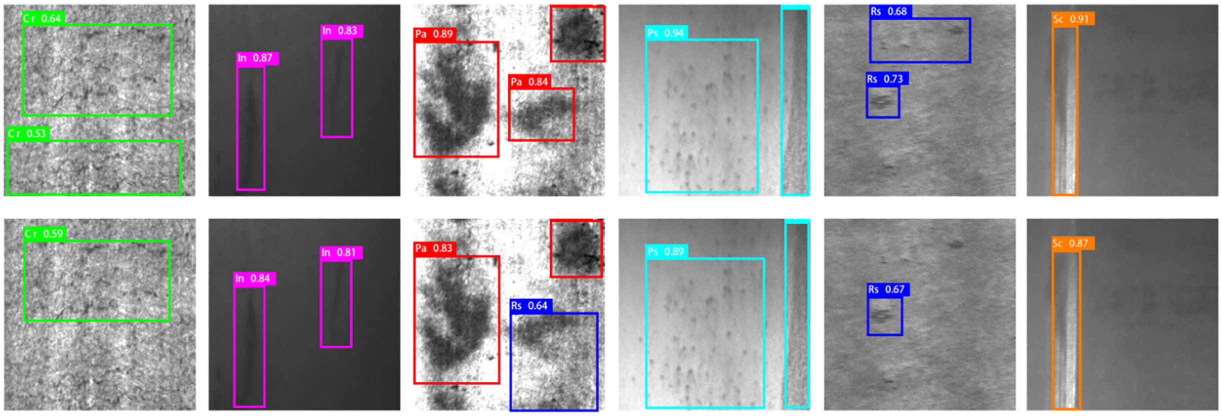

Figure 9: Visual comparison between the improved YOLOv11n (upper row) and the baseline YOLOv11n (lower row) on NEU-DET

In the first group (Cr), the baseline YOLOv11n fails to detect a crack defect, while the improved model successfully localizes it with a higher confidence score. In the fifth group (Rs), the baseline model misses one rolled-in scale defect, whereas the improved model correctly detects it. Additionally, in the third group (Pa), the baseline model incorrectly classifies a patch defect as rolled-in scale, while the improved model avoids this misclassification and produces the correct label.

Overall, these qualitative results demonstrate that the improved YOLOv11n effectively reduces missed detections and false classifications compared with the baseline model, especially for small or low-contrast defects, thereby enhancing the robustness of industrial defect detection.

4.5 Generalization Experiments

To further validate the generalization capability of the proposed method, we also conducted experiments on the GC10-DET dataset [22]. GC10-DET is a real-world industrial surface defect dataset containing 3570 grayscale images with ten categories of defects, such as punch, weld line, and crescent gap. Compared with NEU-DET, it covers a broader variety of defect patterns and is therefore well suited for evaluating model robustness in practical scenarios. The experimental results are summarized in Table 5.

Table 5: Comparison with mainstream detectors on GC10-DET

Model	mAP@0.5	Precision	Recall	Params	GFLOPs
Faster R-CNN	0.649	0.566	0.611	41.75 M	135
YOLOv3	0.632	0.600	0.615	103.6 M	283
YOLOv5n	0.683	0.705	0.610	1.77 M	4.2
YOLOv6n	0.662	0.680	0.625	4.24 M	11.9
YOLOv8n	0.690	0.692	0.638	3.01 M	8.1
YOLOv11n	0.712	0.720	0.650	2.58 M	6.3
Ours	0.737	0.745	0.665	2.46 M	5.3

The improved YOLOv11n (Ours) achieved the best overall performance with a mAP@0.5 of 73.7%, Precision of 74.5%, and Recall of 66.5%, surpassing the baseline YOLOv11n (71.2% mAP) by +2.5 percentage points while maintaining a lightweight architecture of only 2.46 M parameters and 5.3 GFLOPs. Compared with YOLOv8n (69.0% mAP) and YOLOv5n (68.3% mAP), the proposed model consistently demonstrated higher accuracy, confirming the benefits of the introduced optimization modules.

In terms of efficiency, our method not only significantly reduces computational overhead compared to large-scale detectors such as YOLOv3 (283 GFLOPs) and Faster R-CNN (135 GFLOPs) but also achieves superior detection accuracy. This balance between performance and efficiency is especially valuable in real-time industrial inspection scenarios, where computational resources are often limited.

Overall, the results on GC10-DET provide further evidence that the proposed improvements generalize effectively beyond a single dataset. By consistently outperforming both one-stage and two-stage detectors, the enhanced YOLOv11n proves to be a robust and practical solution for industrial surface defect detection.

5 Conclusion and Future Work

5.1 Summary of Findings

This paper addresses the task of surface defect detection in industrial hot-rolled steel strips by proposing three enhancements to the YOLOv11n model: integrating GAFM for improved multi-scale feature fusion, incorporating Aggregated Attention to strengthen critical feature representation, and employing Ghost Dynamic Convolution to reduce computational cost. Experimental results on the NEU-DET dataset demonstrate that the improved model achieves superior performance in terms of mAP@0.5, precision, and recall, compared to both the original YOLOv11n and other mainstream lightweight YOLO models. Moreover, the proposed approach maintains a lower parameter count and computational load, achieving a good balance between detection accuracy and inference efficiency. These characteristics make it suitable for real-time defect detection in resource-constrained industrial environments.

5.2 Limitations and Future Directions

Despite the promising results achieved in this study, several limitations remain. First, all experiments were conducted solely on the NEU-DET dataset, lacking validation across multiple domains and real-world industrial scenarios. Second, the current model still faces challenges in detecting extremely small defects and handling noisy backgrounds. Future research will focus on the following directions: (1) incorporating more diverse and realistic industrial datasets to improve generalization and robustness; (2) exploring more efficient attention mechanisms and model pruning techniques to further enhance inference speed and deployment efficiency; and (3) integrating semi-supervised or unsupervised learning strategies to reduce the reliance on manually annotated training data.

Acknowledgement: Not applicable.

Funding Statement: This research was supported in part by the National Natural Science Foundation of China (Grant No. 62071123), in part by the Natural Science Foundation of Fujian Province (Grant Nos. 2024J01971, 2022J05202), and in part by the Young and Middle-Aged Teacher Education Research Project of Fujian Province (Grant No. JAT210370).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Hong Zhang; Methodology, Hong Zhang, Hua-Qin Wu, Hao Yan; Software, Hua-Qin Wu, Hao Yan; Validation, Shun-Wu Xu, Zhao-Wen Chen; Formal Analysis, Hua-Qin Wu, Feng-Yu Gao; Investigation, Hua-Qin Wu, Hao Yan; Resources, Hong Zhang; Data Curation, Hao Yan, Feng-Yu Gao, Shun-Wu Xu; Writing—Original Draft Preparation, Hua-Qin Wu, Zhao-Wen Chen; Writing—Review and Editing, Hao Yan, Shun-Wu Xu; Visualization, Hua-Qin Wu, Feng-Yu Gao; Supervision, Hong Zhang; Project Administration, Hong Zhang; Funding Acquisition, Hong Zhang, Hua-Qin Wu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in the NEU-DET and GC10-DET datasets. The NEU-DET dataset can be accessed from He et al. [2] (doi:10.1109/TIM.2019.2890706), and the GC10-DET dataset can be accessed from Lv et al. [22] (doi:10.3390/s20061562).

Ethics Approval: Not applicable.

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

References

1. Bai D, Li G, Jiang D, Yun J, Tao B, Jiang G, et al. Surface defect detection methods for industrial products with imbalanced samples: a review of progress in the 2020s. Eng Appl Artif Intell. 2024;130(1):107697. doi:10.1016/j.engappai.2023.107697.
2. He Y, Song K, Meng Q, Yan Y. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Trans Instrum Meas. 2019;69(4):1493–504. doi:10.1109/TIM.2019.2890706.
3. Tang B, Chen L, Sun W, Lin Z. Review of surface defect detection of steel products based on machine vision. IET Image Process. 2023;17(2):303–22. doi:10.1049/ipr2.12647.
4. Leite D, Andrade E, Rativa D, Maciel AM. Fault detection and diagnosis in industry 4.0: a review on challenges and opportunities. Sensors. 2024;25(1):60. doi:10.3390/s25010060.
5. Chen Y, Ding Y, Zhao F, Zhang E, Wu Z, Shao L. Surface defect detection methods for industrial products: a review. Appl Sci. 2021;11(16):7657. doi:10.3390/app11167657.
6. Saberironaghi A, Ren J, El-Gindy M. Defect detection methods for industrial products using deep learning techniques: a review. Algorithms. 2023;16(2):95. doi:10.3390/a16020095.
7. Abdullah DM, Abdulazeez AM. Machine learning applications based on SVM classification: a review. Qubahan Acad J. 2021;1(2):81–90. doi:10.48161/qaj.v1n2a50.
8. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28(6):91–9. doi:10.1109/tpami.2016.2577031.
9. Redmon J, Divvala S, Girshick R, Farhadi A. You only Look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun 27–30; Las Vegas, NV, USA. Piscataway, NJ, USA: IEEE; 2016. p. 779–88. doi:10.1109/CVPR.2016.91.
10. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision; 2017 Oct 22–29; Venice, Italy. Piscataway, NJ, USA: IEEE; 2017. p. 2980–8. doi:10.1109/ICCV.2017.324.
11. Zhang H, Cloutier RS. Review on one-stage object detection based on deep learning. EAI Endorsed Trans E-Learn. 2022;7(23):e7. doi:10.4108/eai.7-12-2022.173383.
12. Du L, Zhang R, Wang X. Overview of two-stage object detection algorithms. J Phys Conf Ser. 2020;1544(1):012033. doi:10.1088/1742-6596/1544/1/012033.
13. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements. arXiv:2410.17725. 2024.
14. Tian Z, Yang F, Yang L, Wu Y, Chen J, Qian P. An optimized YOLOv11 framework for the efficient multi-category defect detection of concrete surface. Sensors. 2025;25(5):1291. doi:10.3390/s25051291.
15. Shi M. Research on robot autonomous inspection visual perception system based on improved YOLOv11 and SLAM. In: Proceedings of the IEEE 7th International Conference on Communications, Information System and Computer Engineering; 2025 Jun 15–17; Wuhan, China. Piscataway. NJ, USA: IEEE; 2025. p. 850–4. doi:10.1109/CISCE.2025.00018.
16. Gong X, Yu J, Zhang H, Dong X. AED-YOLO11: a small object detection model based on YOLO11. Digit Signal Process. 2025;166:105411. doi:10.1016/j.dsp.2025.105411.
17. Ultralytics. YOLO11 Documentation and Code. [cited 2025 Aug 22]. Available from: https://docs.ultralytics.com/models/yolo11/.

18. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–22; Salt Lake City, UT, USA. Piscataway, NJ, USA: IEEE; 2018. p. 7132–41. doi:10.1109/CVPR.2018.00745.
19. Shi D. Transnext: robust foveal visual perception for vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024 Jun 16–20; Seattle, WA, USA. Piscataway, NJ, USA: IEEE; 2024. p. 17773–83. doi:10.1109/CVPR46437.2024.01747.
20. Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C. GhostNet: more features from cheap operations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020 Jun 14–19; Seattle, WA, USA. Piscataway, NJ, USA: IEEE; 2020. p. 1580–9. doi:10.1109/CVPR42600.2020.00164.
21. Li X, Xu C, Li J, Zhou X, Li Y. Multi-scale sensing and multi-dimensional feature enhancement for surface defect detection of hot-rolled steel strip. Nondestruct Test Eval. 2024:1–24. doi:10.1080/10589759.2024.2101234.
22. Lv X, Duan F, Jiang JJ, Fu X, Gan L. Deep metallic surface defect detection: the new benchmark and detection network. Sensors. 2020;20(6):1562. doi:10.3390/s20061562.