# PGA-YOLO: Toward Efficient and Robust Feature Interaction for Defect Detection in Complex Industrial Scenes

Jin Zhang[1], Zhiwen Wu[1], Cheng Sun[1,*], Bin Hu[1], Qi Cao[1], and Tie Wang[2]
[1]Changsha University of Science & Technology, Changsha, China
[2]Hunan Sanyue Shuwei Technology Co., Ltd., Changsha, China
Email: {mail_zhangjin, sun_ching}@163.com, {1580413032, caoqi}@qq.com,
hubin@csust.edu.cn, wangtie@i3yoo.com

***Abstract*—Unlike conventional object detection tasks, industrial metal surface defect detection is challenged by severe background interference and extreme scale imbalance among defect instances. To address these challenges, we propose PGA-YOLO, a detection framework co-designed from the perspectives of feature interaction and detection architecture. Specifically, a refined feature interaction mechanism is introduced to enhance the discriminability of deep features while effectively suppressing background-induced confusion. Meanwhile, efficient convolutional operators together with an adaptive detection head are incorporated to improve small-object localization accuracy while maintaining a lightweight model profile. This synergistic design effectively alleviates the trade-off between detection precision and computational efficiency. Extensive experiments demonstrate that PGA-YOLO achieves a superior balance between accuracy and efficiency. With only 5.78M parameters, the proposed method attains 81.43% mAP@50 and 50.67% mAP@50:95 on the NEU-DET dataset. Furthermore, PGA-YOLO consistently outperforms state-of-the-art methods in terms of generalization capability and real-time performance on the GC10-DET and AL10-DET datasets.**

***Index Terms*—Surface defect detection; Feature interaction; Small object localization; Lightweight model.**

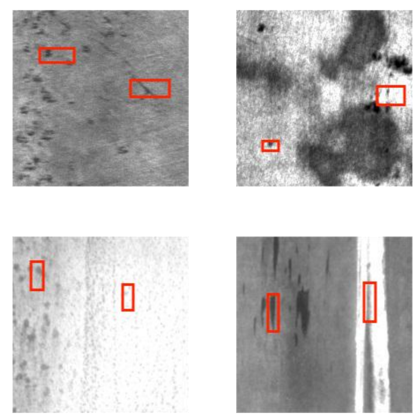

Fig. 1. Metal surface defect images are characterized by severe suppression from complex background interference, extreme scale heterogeneity, and high inter-class similarity among certain defect types.

## I. Introduction

In contemporary industrial manufacturing, the surface integrity of metallic components serves as a fundamental determinant of product performance and structural reliability, where undetected flaws may precipitate catastrophic material failures [1]. Despite the emergence of deep learning as a transformative paradigm for defect identification, practical deployment remains fraught with persistent obstacles—most notably, severe environmental noise, extreme cross-scale variance among defect instances, and the subtle inter-class similarities that blur decision boundaries (as illustrated in Fig. 1).

This work was supported by the National Key R&D Program of China (No. 2024YFB2605500), the Natural Science Foundation of Hunan Province, China (Nos. 2025JJ60410, 2025JJ60421), the Science and Technology Innovation Program of Hunan Province (No. 2025RC1050), the Open Research Fund of Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation (Nos. B202401, B202405), the Ministry of Education (MOE) Industry-University Cooperation Collaborative Education Project (No. 2505063327), and the Humanity and Social Science Youth Foundation of the MOE (No. 25YJCZH079).

Corresponding author: Cheng Sun (email: sun_ching@163.com).

To surmount these hurdles, the research community has explored diverse architectural innovations. From the perspective of feature enhancement and multi-scale representation, Zhang [2] and Ning [3] leveraged BiFPN- and ResNet-FPN-based architectures to bolster the perception of defects across diverse spatial scales, while Liu [4] expanded the effective receptive field through parallel atrous convolutions and multi-scale pooling strategies. In the realm of attention-based refinement, Fang [5] demonstrated that integrating attention modules can effectively calibrate models to concentrate on salient defect regions. Furthermore, to satisfy the stringent constraints of lightweight deployment, Ren [6] and Yu [7] introduced LFEBNet and depthwise separable convolutions to substantially curtail computational overhead. However, despite these advancements, existing frameworks often exhibit a sub-optimal synergy between sophisticated feature interaction and structural efficiency, frequently failing to maintain high discriminative power under extremely limited resource budgets.

While the aforementioned methodologies have advanced the field to varying degrees, three pivotal challenges persist in

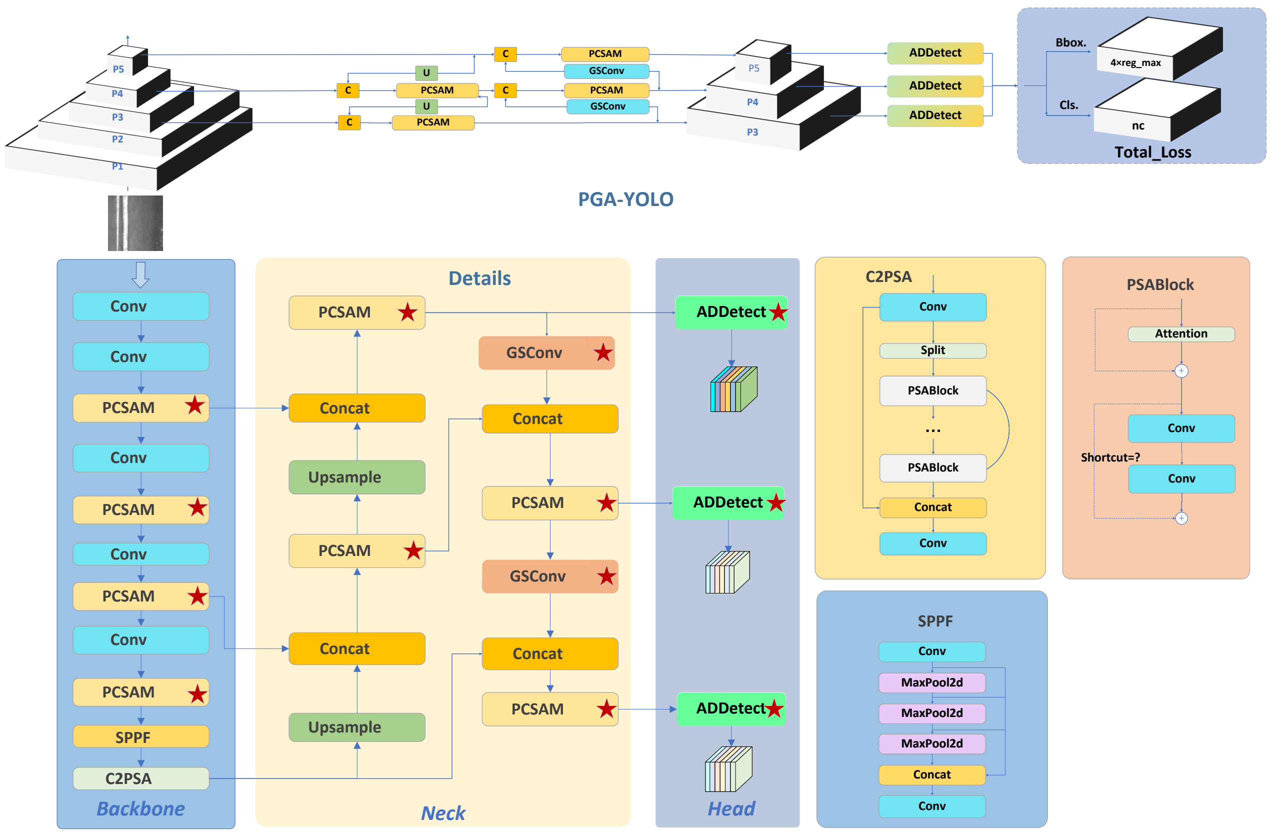


Fig. 2. Overall architecture of the PGA-YOLO network. The red stars indicate the improved modules.

reconciling detection sensitivity with architectural efficiency. Specifically, current attention paradigms often fail to achieve optimal coordination between spatial and channel dimensions, rendering fine-grained defect features vulnerable to occlusion by complex background noise. Furthermore, high-performance enhancement strategies are frequently accompanied by prohibitive computational overhead, which severely impedes their deployment on resource-constrained industrial embedded devices. Compounding these issues, conventional fixed label assignment strategies [8] struggle to accommodate the drastic scale variations inherent in defect instances, frequently leading to localization drift and increased missed detection rates.

To address the aforementioned challenges in metal surface defect detection, this paper presents the following contributions:

- Development of the PGA-YOLO framework: We construct a novel detector integrating the Parallel Coordinated Spatial-Channel Attention Module (PCSAM), GSConv [9], and Adaptive Dual-Head Collaborative Assignment (AD). Through deep integration of these modules, the model effectively mitigates complex industrial interference while maintaining an extremely low parameter count.
- Innovative design of core components: We design the PCSAM to suppress background noise via parallel spatial-channel interaction. GSConv is incorporated to optimize computational efficiency, balancing precision and lightweight requirements. Furthermore, the AD strategy is proposed to precisely rectify the localization drift of microscopic targets through a differentiated label assignment mechanism.
- Extensive validation and generalization: Experiments demonstrate that PGA-YOLO achieves superior detection accuracy and real-time performance across the NEU-DET, GC10-DET, and AL10-DET datasets, providing a robust lightweight solution for industrial-grade deployment in resource-constrained environments.

## II. THE PROPOSED METHOD

Building upon YOLOv11 [10], we develop PGA-YOLO, a synergistic framework optimized for high-precision, lightweight industrial defect detection. The architecture strengthens feature discriminability by replacing standard C3k2 units with self-developed PCSAM, while streamlining feature fusion through the integration of GSConv in the neck to minimize computational overhead. Furthermore, an Adaptive Dual-head (AD) design is employed to refine the localization and classification of defects across drastic scales. The overall topology is illustrated in Fig. 2, with the underlying mechanisms elaborated in the following sections.

### A. Parallel Coordinated Spatial-Channel Attention Module (PCSAM)

To address the feature obscuration caused by specular reflection and texture interference on metal surfaces, and inspired by CBAM [11], we propose the PCSAM, as illustrated in Fig. 3. By leveraging parallel joint modeling to achieve deep feature interaction, this module effectively suppresses industrial background noise while significantly enhancing the feature capture precision for microscopic targets.

The input feature map $F \in \mathbb{R}^{H\times W\times C}$ is processed through parallel branches to generate the channel descriptor $D_c$ and the spatial descriptor $D_s$(as illustrated in Fig. 3). Specifically, global information is extracted from $F$ via both maximum and average pooling. Subsequently, the correlations are modeled through shared convolutional layers, followed by an element-wise summation to fuse the statistical features(see Eqs. 1–3). The symbol $|$ denotes the selection for $D_c$ (left) and $D_s$ (right), the terms on the left belong to $D_c \in \mathbb{R}^{1\times 1\times C}$, while those on the right belong to $D_s \in \mathbb{R}^{H\times W\times 1}$.

$$\mathbf{F}_{c,\text{max}} = \text{MaxPool}(\mathbf{F}) \in \mathbb{R}^{1\times 1\times C} \mid \mathbb{R}^{H\times W\times 1} \tag{1}$$

$$\mathbf{F}_{c,\text{avg}} = \text{AvgPool}(\mathbf{F}) \in \mathbb{R}^{1\times 1\times C} \mid \mathbb{R}^{H\times W\times 1} \tag{2}$$

$$\begin{aligned}\mathbf{D}_c \mid \mathbf{D}_s &= \text{CONV}(\mathbf{F}_{c|s,\text{max}}) + \text{CONV}(\mathbf{F}_{c|s,\text{avg}}) \\ &\quad \in \mathbb{R}^{1\times 1\times C} \mid \mathbb{R}^{H\times W\times 1}\end{aligned} \tag{3}$$

Subsequently, $D_c$ is flattened to $1 \times C$ and $D_s$ to $HW \times 1$ via reshaping, which are then concatenated to form the joint descriptor $D_{joint}$ (see Eqs.4). A joint convolutional layer (CONV) is utilized to learn the deep interaction between the two, followed by a split operation to obtain the intermediate representations $\mathbf{M}'_c$ and $\mathbf{M}'_s$. These intermediate states are then used to derive the final channel attention map $\mathbf{M}_c \in \mathbb{R}^{1\times 1\times C}$ and spatial attention map $\mathbf{M}_s \in \mathbb{R}^{H\times W\times 1}$(see Eqs.5). Finally, $\mathbf{M}_c$ performs a channel-wise broadcast multiplication with the input $F$, followed by a spatial-wise broadcast multiplication between $\mathbf{M}_s$ and the intermediate result to produce the refined feature map(see Eqs.6).

$$\begin{aligned}\mathbf{D}_{\text{joint}} &= \text{Concatenate}(\text{Flatten}(\mathbf{D}_c), \text{Flatten}(\mathbf{D}_s)) \\ &\quad \in \mathbb{R}^{(HW+C)\times 1}\end{aligned} \tag{4}$$

$$\mathbf{M}_{c|s} = \sigma\left(\mathbf{M}'_{c|s}\right) \in \mathbb{R}^{1\times 1\times C} \mid \mathbb{R}^{H\times W\times 1} \tag{5}$$

$$\mathbf{F}_{\text{out}} = \mathbf{M}_c \otimes \mathbf{F} \otimes \mathbf{M}_s \tag{6}$$

### B. GSConv Module

GSConv [9] adopts a parallel structural design that extracts features through dual branches. The SC branch utilizes standard convolutions to preserve spatial details and channel dependencies, while the DSC branch employs depthwise separable convolutions to expand the receptive field with minimal overhead. After processing, GSConv concatenates the features from both branches along the channel dimension (see Fig. 4). To break the feature barriers between branches, a Shuffle mechanism is introduced to achieve deep interaction and fusion of channel information. This design enables the lightweight features of the DSC to complement the dense features of the SC. Consequently, it maintains high sensitivity to small targets such as micro-cracks while significantly reducing the computational cost for real-time industrial detection.

### C. Adaptive Dual-Head Collaborative Assignment (AD)

To address the missed detection issues caused by uniform assignment strategies in traditional structures, the Adaptive Dual-Head Collaborative Assignment (AD) is designed (see Fig. 5). The AD introduces an Adaptive Assigner (AA), which dynamically generates two sets of customized labels for the lead and auxiliary heads based on the Ground Truth (GT), thereby strengthening the capture capability for microscopic defects. Specifically, the AA employs a high-quality positive sample selection mechanism for the lead head to ensure it learns the most discriminative features. For the auxiliary head, a soft-label assignment mechanism is used to enable learning from features of suspected positive samples (see Eqs. 7–8). Here, $b_i^*$ and $c_i^*$ represent the optimal GT box and its category, $\alpha_{ij}$ denotes the soft-label weight, and $[T_l, T_m]$ indicates the range of Intersection over Union (IoU) values.

$$\mathbf{T}_l(j) = \begin{cases} (\mathbf{b}_i^*, \mathbf{c}_i^*, 1), & \text{if } \max_i \text{IoU}(\mathbf{a}_j, \mathbf{b}_i) > \tau_h \\ & \text{and } j \in \text{TopK}(s_j) \\ (0,0,0), & \text{otherwise} \end{cases} \tag{7}$$

$$\mathbf{T}_a(j) = \begin{cases} (\mathbf{b}_i^*, \mathbf{c}_i^*, 1), & \text{if IoU}(\mathbf{a}_j, \mathbf{b}_i) > \tau_m \\ (\mathbf{b}_i^*, \mathbf{c}_i^*, \alpha_{ij}), & \text{if } \tau_l < \text{IoU}(\mathbf{a}_j, \mathbf{b}_i) \le \tau_m \\ (0,0,0), & \text{otherwise} \end{cases} \tag{8}$$

The AA adopts distinct loss functions for the lead and auxiliary heads (see Eqs. 11–12). To ensure collaborative learning and prevent feature representation conflicts, we introduce a prediction consistency loss $\mathcal{L}_{consist}$ and a complementary guidance loss $\mathcal{L}_{complement}$, which are weighted and summed into the final total loss function (see Eqs. 9–13). In these formulations, $P_l$ and $P_a$ denote the category predictions of the lead and auxiliary heads, while $B_l$ and $B_a$ represent their respective bounding box predictions. The distance between the feature representations $f_l^{(j)}$ and $f_a^{(j)}$ of the two detection heads in the soft-label positive set $\mathcal{S}$ is not less than the threshold $\gamma$. Furthermore, $\mathcal{P} = \{j|T_l(j) > 0 \text{ or } T_a(j) > 0\}$ represents the positive sample set, and $\mathcal{S} = \{j|0 < T_a(j) < 1\}$ is the soft-label positive sample set.

$$\mathcal{L}_{lead} = \mathcal{L}_{cls}(\mathbf{P}_l, \mathbf{T}_l) + \mathcal{L}_{reg}(\mathbf{B}_l, \mathbf{T}_l) \tag{9}$$

$$\mathcal{L}_{aux} = \mathcal{L}_{cls}(\mathbf{P}_a, \mathbf{T}_a) + \mathcal{L}_{reg}(\mathbf{B}_a, \mathbf{T}_a) \tag{10}$$

$$\mathcal{L}_{cons} = \tfrac{1}{N_{pos}} \textstyle\sum_{j\in\mathcal{P}} \text{JS}\big(\sigma(\mathbf{p}_l^{(j)}), \sigma(\mathbf{p}_a^{(j)})\big) \tag{11}$$

$$\mathcal{L}_{comp} = \tfrac{1}{N_{soft}} \textstyle\sum_{j\in\mathcal{S}} \max(0, \gamma - \|\mathbf{f}_l^{(j)} - \mathbf{f}_a^{(j)}\|_2) \tag{12}$$

$$\mathcal{L}_{total} = \mathcal{L}_{lead} + \lambda_1\mathcal{L}_{aux} + \lambda_2\mathcal{L}_{cons} + \lambda_3\mathcal{L}_{comp} \tag{13}$$

## III. EXPERIMENTS

Experiments are conducted on the NEU-DET industrial dataset, which contains 1,800 grayscale images (640×640 resolution) across six defect categories, including Crazing (Cr) and Patches (Pa). To evaluate the generalization capability, the GC10-DET and AL10-DET datasets are also introduced.

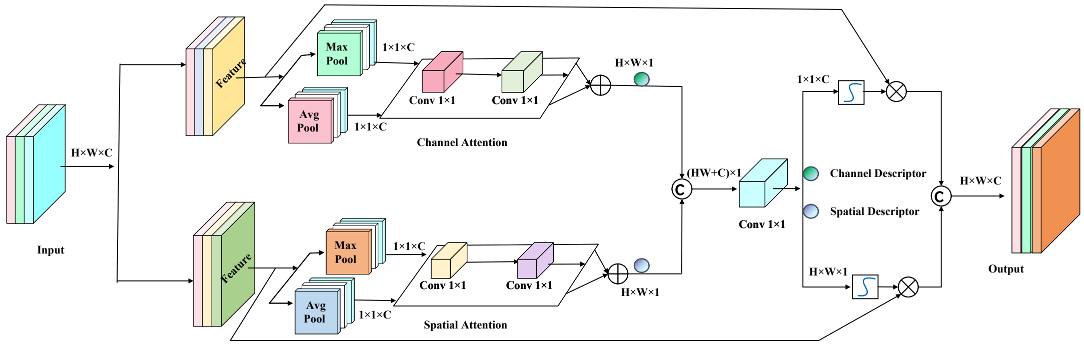


Fig. 3. Detailed architecture of the PCSAM module. Unlike conventional methods that model channel and spatial attention in isolation, this structure integrates channel-wise dependencies into the spatial domain to achieve joint modeling and deep feature interaction.

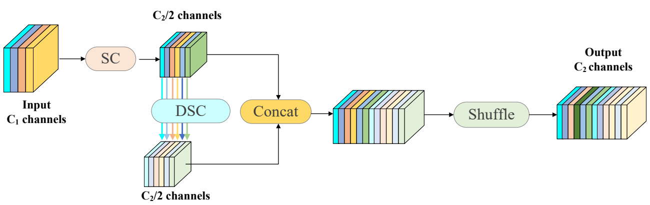


Fig. 4. Structural diagram of the GSConv module

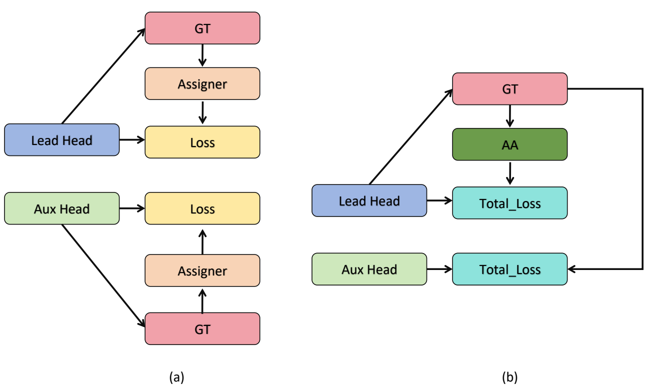


Fig. 5. (a) Conventional independent label assigner; (b) Proposed Adaptive Assigner (AA) with modified loss functions.

The development environment consists of Ubuntu 22.04 and PyTorch 2.3.1, with hardware acceleration provided by an NVIDIA A30 GPU. For training, the SGD optimizer is employed with an initial learning rate of 0.01 and a batch size of 32 for 300 epochs. Evaluation metrics include Precision (P), Recall (R), mAP50, mAP50:95, Parameters (Params), and FLOPs. The experimental evaluation is organized into five subsections (A-E). Section D (Generalization Experiments) is conducted on various datasets (GC10-DET and AL10-DET) to assess the model's adaptability. All other experiments, including ablation studies and state-of-the-art (SOTA) comparisons, are performed on the NEU-DET dataset.

## A. *Ablation experiment*

The effectiveness of each proposed module is quantified in Table I. Starting from the YOLOv11s baseline, the integration of PCSAM yields a 1.55% improvement in mAP50, primarily attributed to its enhanced capability in capturing discriminative defect features. By incorporating GSConv, the model significantly reduces the computational overhead by 2.43M parameters while maintaining competitive accuracy. Finally, the addition of AD elevates the mAP50:95 to 50.45%, effectively addressing the challenges of false positives and missed detections for microscopic defects. These results validate the synergistic effect of the proposed components in balancing detection precision and model efficiency.

TABLE I
ABLATION EXPERIMENT.BOLD INDICATES BEST.

| PCSAM | GSConv | AD | R/% | P/% | mAP50/% | mAP50:95/% | Params/M | Flops/G |
|---|---|---|---|---|---|---|---|---|
| - | - | - | 79.32 | 81.12 | 79.24 | 45.32 | 8.63 | 18.45 |
| ✓ | - | - | 83.42 | 82.39 | 80.79 | 49.84 | 9.81 | 19.34 |
| - | ✓ | - | 81.65 | 80.87 | 79.62 | 48.76 | 6.43 | 17.12 |
| - | - | ✓ | 84.36 | 79.91 | 80.67 | 50.45 | 8.79 | 19.54 |
| ✓ | ✓ | - | 83.84 | 81.35 | 80.14 | 49.12 | 7.55 | 17.92 |
| - | ✓ | ✓ | 84.51 | 80.52 | 81.02 | 50.12 | 6.59 | 18.21 |
| ✓ | - | ✓ | 84.36 | 82.88 | 81.31 | 50.55 | 9.97 | 20.43 |
| ✓ | ✓ | ✓ | **84.45** | **83.53** | **81.43** | **50.67** | **5.78** | **16.28** |

## B. *Comparison with State-of-the-Art Methods.*

The results in Table II demonstrate that PGA-YOLO achieves the optimal balance between detection performance and lightweight design. Compared with the YOLOv11s baseline, its mAP50 and mAP50:95 are improved by 2.19% and 5.35%, respectively, while the Params and FLOPs are significantly reduced by 2.85M and 2.17G. Furthermore, PGA-YOLO outperforms recent SOTA models, such as YOLOv10s and YOLOv12s, in terms of both accuracy and parameter efficiency. These results verify its superior efficiency for practical industrial deployment.

## C. *Comparative experiments of improved modules*

As shown in Table III, PCSAM achieves superior performance in key metrics, with Recall (R) and Precision (P)

TABLE II
COMPARISON OF THE PROPOSED PGA-YOLO WITH STATE-OF-THE-ART MODELS.BOLD INDICATES BEST.

| Model | R/% | P/% | mAP50/% | mAP50:95/% | Params/M | Flops/G |
|---|---|---|---|---|---|---|
| Faster R-CNN | 66.31 | 63.4 | 65.98 | 35.1 | 41.33 | 120.99 |
| YOLOv5s | 75.92 | 67.1 | 69.25 | 41.7 | 7.08 | 26.52 |
| YOLOv7 | 66.97 | 71.7 | 74.01 | 38.8 | 37.41 | 105.78 |
| YOLOv7-tiny | 77.12 | 72.45 | 63.82 | 34.87 | 6.03 | 13.22 |
| YOLOXs | 74.68 | 75.48 | 76.33 | 38.86 | 8.94 | 26.77 |
| YOLOv8s | 76.32 | 74.65 | 74.42 | 48.14 | 11.14 | 28.67 |
| YOLOv10s | 78.91 | 78.57 | 73.42 | 47.16 | 8.07 | 24.58 |
| YOLOv11s | 79.32 | 81.12 | 79.24 | 45.32 | 8.63 | 18.45 |
| YOLOv12s | 77.56 | 79.87 | 78.45 | 46.89 | 9.26 | 21.53 |
| MSI-YOLO [12] | 80.64 | 81.76 | 77.98 | 47.56 | 9.86 | 25.34 |
| Ref [13] | 81.46 | 82.41 | 80.14 | 49.46 | 9.82 | 25.46 |
| Ref [14] | 82.76 | 79.34 | 78.75 | 48.96 | 11.48 | 29.87 |
| **PGA-YOLO** | **84.45** (↑5.13) | **83.53** (↑2.41) | **81.43** (↑2.19) | **50.67** (↑5.35) | **5.78** (↓2.85) | **16.28** (↓2.17) |

TABLE III
EXPERIMENTAL RESULTS OF THE PCSAM MODULE.BOLD INDICATES BEST.

| Modules | R/% | P/% | mAP@0.5/% | Params/M | Flops/G |
|---|---|---|---|---|---|
| ECA [15] + GSConv + AD | 79.32 | 80.32 | 78.49 | 10.12 | 28.86 |
| CBAM [11] + GSConv + AD | 83.15 | 82.59 | 80.78 | 7.84 | 44.63 |
| CA [16] + GSConv + AD | 74.38 | 79.78 | 77.54 | 9.45 | 28.88 |
| GAM [17] + GSConv + AD | 77.81 | 81.85 | 80.49 | 21.23 | 32.45 |
| SimAM [18]+ GSConv + AD | 82.32 | 81.98 | 80.85 | 18.48 | 29.22 |
| **PCSAM + GSConv + AD** | **84.45** | **83.53** | **81.43** | **5.78** | **16.28** |

reaching 84.45% and 83.53%, respectively. This demonstrates that by leveraging parallel joint modeling for deep feature interaction, PCSAM can more effectively suppress background interference and enhance the capture capability for microscopic defects. Table IV indicates that GSConv significantly optimizes model efficiency; it achieves a mAP50 of 81.43% while maintaining the lowest parameter count (5.78M) and computational cost (16.28G). Experimental results prove that this module, by suppressing feature redundancy, serves as the core component for balancing the lightweight design and high performance of PGA-YOLO. Furthermore, Table V illustrates that AD yields promising results across all critical metrics, with mAP50:95 reaching 50.67%. This underscores AD's ability to refine the perception accuracy for tiny defects through an optimized detection architecture.

TABLE IV
EXPERIMENTAL RESULTS OF THE GSCONV MODULE.BOLD INDICATES BEST.

| Module | R/% | P/% | mAP@0.5/% | Params/M | Flops/G |
|---|---|---|---|---|---|
| PCSAM + PConv [19] + AD | 76.32 | 65.7 | 76.84 | 6.48 | 20.41 |
| PCSAM + SCConv [20] + AD | 74.43 | 67.5 | 80.86 | 6.87 | 18.32 |
| PCSAM + AKConv [21] + AD | 71.7 | 66.3 | 79.91 | 8.23 | 26.41 |
| PCSAM + SAConv [22]+ AD | 72.3 | 66.6 | 77.72 | 7.10 | 19.85 |
| **PCSAM + GSConv + AD** | **84.45** | **83.53** | **81.43** | **5.78** | **16.28** |

### D. Generalization experiments

The results in Table VI indicate that PGA-YOLO maintains highly consistent detection performance and lightweight advantages across various industrial datasets. Specifically, its mAP@0.5 remains stable between 78% and 82%, showing minimal fluctuations compared to the baseline dataset. This demonstrates that the model possesses exceptional anti-interference capabilities and cross-domain generalization stability, enabling it to robustly adapt to diverse and complex inspection requirements on metal production lines.

TABLE V
EXPERIMENTAL RESULTS OF THE AD HEAD.BOLD INDICATES BEST.

| Head | R/% | P/% | mAP50:95/% | Params/M | Flops/G |
|---|---|---|---|---|---|
| PCSAM + GSConv + LDH [23] | 83.41 | 82.12 | 46.34 | 6.78 | 23.31 |
| PCSAM + GSConv + EMAH [24] | 84.12 | 79.23 | 45.66 | 8.43 | 31.78 |
| PCSAM + GSConv + MLLAH [25] | 79.95 | 81.37 | 48.74 | 7.46 | 27.34 |
| PCSAM + GSConv + LWH [26] | 82.42 | 83.14 | 49.72 | 11.23 | 34.25 |
| **PCSAM + GSConv + AD** | **84.45** | **83.53** | **50.67** | **5.78** | **16.28** |

TABLE VI
EXPERIMENTAL RESULTS ON DIFFERENT DATASETS

| Dataset | R/% | P/% | mAP50/% | mAP50:95/% | Params/M | Flops/G |
|---|---|---|---|---|---|---|
| NEU-DET | 84.45 | 83.53 | 81.43 | 50.67 | 5.78 | 16.28 |
| GC10-DET | 83.79 | 84.56 | 80.27 | 51.13 | 5.78 | 16.28 |
| AL10-DET | 84.67 | 79.24 | 78.78 | 49.97 | 5.78 | 16.28 |

### E. Visualization Analysis

To intuitively compare performance and quantitatively evaluate the algorithms' capability to focus on defect regions, visualization comparisons and Heatmaps (Grad-CAM) between PGA-YOLO and YOLOv11s on the NEU-DET dataset are conducted.

Fig. 6 illustrates that PGA-YOLO significantly enhances feature extraction, effectively compensating for the missed detections in the baseline model. For extremely small-scale targets such as Inclusions (In) and Patches (Pa) in Figs. 6(a) and (b), PGA-YOLO exhibits higher precision; meanwhile, Fig. 6(e) confirms its superior anti-interference ability in complex backgrounds, markedly improving localization reliability in industrial environments.

The results in Fig. 7 reveal that the red regions in PGA-YOLO's heatmaps are deeper and more closely aligned with the actual defect contours, proving its ability to precisely lock onto critical discriminative features. For low-contrast micro-targets like Scratches (Sc) and Crazing (Cr), PGA-YOLO's response intensity significantly outperforms the baseline. This validates that the AD detection head provides superior supervisory signals, effectively enhancing the capture capability for subtle defects in complex environments.

## IV. CONCLUSION

This paper presents PGA-YOLO, a lightweight algorithm optimized for metal surface defect detection. By incorporating the Parallel Channel-Spatial Attention Module (PCSAM), our approach enables deep parallel interaction between features, effectively suppressing interference from complex industrial

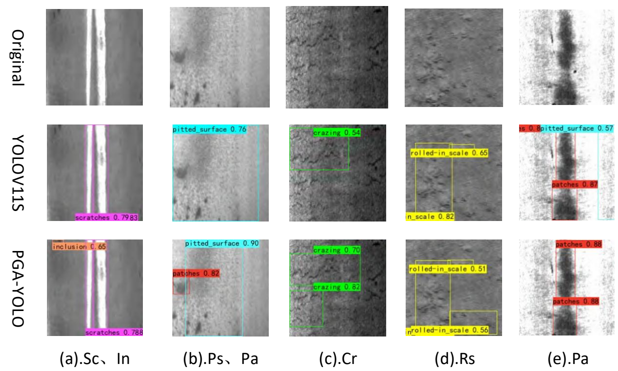


Fig. 6. Comparative visualization of detection performance between the baseline YOLOv11s and the proposed PGA-YOLO. The results demonstrate that PGA-YOLO effectively mitigates missed detections and false negatives in complex industrial scenarios.

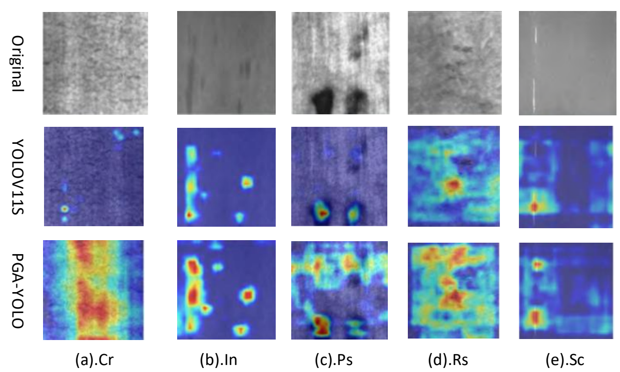


Fig. 7. Heatmap comparison between the baseline and PGA-YOLO. Our model shows more precise feature localization and higher response intensity for subtle defects. This demonstrates that the PCSAM and AD significantly improve feature discriminability and supervision for low-contrast, multi-scale targets.

backgrounds. Furthermore, the integration of an Adaptive Dual-head (AD) structure and efficient convolutional operators significantly improves the localization of microscopic defects. Extensive evaluations on the NEU-DET, GC10-DET, and AL10-DET datasets demonstrate that PGA-YOLO achieves superior detection accuracy and generalization with minimal computational overhead, making it highly suitable for real-time industrial applications.

## References


[1] R. Mordia and A. K. Verma, "Visual techniques for defects detection in steel products: A comparative study," *Eng. Fail. Anal.*, vol. 134, p. 106047, March 2022.

[2] Y. Zhang, W. Wang, Z. Li, et al., "Development of a cross-scale weighted feature fusion network for hot-rolled steel surface defect detection," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105628, Jan. 2023.

[3] T. Ning, W. Wu, and J. Zhang, "Small object detection based on YOLOv8 in UAV perspective," *Pattern Anal. Appl.*, vol. 27, no. 3, p. 103, Aug. 2024.

[4] R. Liu, M. Huang, Z. Gao, et al., "MSC-DNet: An efficient detector with multi-scale context for defect detection on strip steel surface," *Measurement*, vol. 209, p. 112467, Feb. 2023.

[5] J. Fang, X. Tan, and Y. Wang, "ACRM: Attention cascade R-CNN with Mix-NMS for metallic surface defect detection," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, 2021, pp. 423–430.

[6] X. Ren, Y. Bai, G. Liu, et al., "YOLO-Lite: An efficient lightweight network for SAR ship detection," *Remote Sens.*, vol. 15, no. 15, p. 3771, Aug. 2023.

[7] T. Yu, X. Luo, Q. Li, et al., "CRGF-YOLO: An optimized multi-scale feature fusion model based on YOLOv5 for detection of steel surface defects," *Int. J. Comput. Intell. Syst.*, vol. 17, no. 1, p. 154, Dec. 2024.

[8] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 7464–7475.

[9] H. Li, J. Li, H. Wei, et al., "Slim-neck by GSConv: A lightweight-design for real-time detector architectures," *J. Real-Time Image Process.*, vol. 21, no. 3, p. 62, June 2024.

[10] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, Oct. 2024.

[11] S. Woo, J. Park, J. Y. Lee, et al., "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[12] W. Xu, Y. Zhang, X. Jiang, et al., "An efficient steel defect detection model based on multi-scale information extraction," *Robot. Intell. Autom.*, vol. 44, no. 6, pp. 817–829, Nov. 2024.

[13] L. J. Liu, Y. Zhang, and H. R. Karimi, "Resilient machine learning for steel surface defect detection based on lightweight convolution," *Int. J. Adv. Manuf. Technol.*, vol. 134, no. 9, pp. 4639–4650, Oct. 2024.

[14] C. You and H. Kong, "Improved steel surface defect detection algorithm based on YOLOv8," *IEEE Access*, vol. 12, pp. 99570–99577, July 2024.

[15] Q. Wang, B. Wu, P. Zhu, et al., "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11534–11542.

[16] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 13713–13722.

[17] Y. Liu, Z. Shao, and N. Hoffmann, "Global attention mechanism: Retain information to enhance channel-spatial interactions," *arXiv preprint arXiv:2112.05561*, 2021.

[18] L. Yang, R. Y. Zhang, L. Li, et al., "Simam: A simple, parameter-free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 11863–11874.

[19] J. Chen, S. Kao, H. He, et al., "Run, don't walk: chasing higher FLOPS for faster neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 12021–12031.

[20] J. Sun, Y. Peng, C. Chen, et al., "ESC-YOLO: optimizing Apple fruit recognition with efficient Spatial and channel features in YOLOX," *J. Real-Time Image Process.*, vol. 21, no. 5, p. 162, Oct. 2024.

[21] X. M. Jin, X. Y. Liang, and P. F. Deng, "Lightweight daylily grading and detection model based on improved YOLOv10," *Smart Agric.*, vol. 6, no. 5, pp. 108–118, Oct. 2024.

[22] S. Qiao, L. C. Chen, and A. Yuille, "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 10213–10224.

[23] X. Dong, J. Yuan, and J. Dai, "Study on Lightweight Bridge Crack Detection Algorithm Based on YOLO11," *Sensors*, vol. 25, no. 11, p. 3276, May 2025.

[24] Z. Ye, X. Peng, D. Li, et al., "YOLOv11-MSE: A Multi-Scale Dilated Attention-Enhanced Lightweight Network for Efficient Real-Time Underwater Target Detection," *J. Mar. Sci. Eng.*, vol. 13, no. 10, p. 1843, Oct. 2025.

[25] T. Ye, S. Huang, W. Qin, et al., "YOLO-FIX: Improved YOLOv11 with Attention and Multi-Scale Feature Fusion for Detecting Glue Line Defects on Mobile Phone Frames," *Electronics*, vol. 14, no. 5, p. 927, March 2025.

[26] J. Huang, K. Wang, Y. Hou, et al., "LW-YOLO11: a lightweight arbitrary-oriented ship detection method based on improved YOLO11," *Sensors*, vol. 25, no. 1, p. 65, Dec. 2024.