Available Online: 10 March 2026 DOI: 10.54254/2977-3903/2026.32185

YOLOv12-enhanced: multi-scale attention and edge information fusion for industrial valve nozzle detection

Bo Liu, Jian Zhang*

School of Mechanical Engineering, Tongji University, Shanghai, China

*Corresponding Author. Email: jianzh@tongji.edu.cn

Abstract. Accurate valve nozzle detection is an important component of industrial visual inspection systems; however, structural complexity, scale variation, illumination fluctuation, and partial occlusion remain challenging factors that affect detection stability. This study presents YOLOv12-Enhanced, a refined single-stage detection framework developed for industrial valve nozzle scenarios. The proposed approach incorporates three architectural modifications: a RepViT backbone to enhance hierarchical feature representation through structural re-parameterization and global–local modeling, a Spatial Pyramid Pooling Fast (SPPF) module combined with C2PSA attention to strengthen multi-scale contextual feature extraction, and a Global Edge Information Fusion (GEIF) module to integrate shallow edge information with deep semantic features for improved boundary alignment. Experimental evaluation on the Pascal Visual Object Classes (VOC) dataset shows that the proposed model achieves 71.0% mAP50 and 54.4% mAP50–95 under identical training conditions, exceeding the baseline YOLOv12n. Ablation experiments further demonstrate that each module contributes incremental performance gains. Evaluation on a self-constructed valve nozzle dataset consisting of 500 real industrial images indicates stable detection behavior under varying illumination and partial occlusion conditions. The experimental findings suggest that the proposed structural refinements provide a balanced enhancement in feature representation and localization precision while maintaining comparable computational complexity.

Keywords: YOLOv12-enhanced, valve nozzle detection, multi-scale attention, edge information fusion, industrial inspection

1. Introduction

In recent years, the rapid development of deep learning and the availability of large-scale annotated datasets have significantly advanced object detection technologies. These developments have enabled wide-ranging applications in autonomous driving, intelligent surveillance, and industrial inspection systems. In industrial environments, vision-based detection plays a crucial role in component positioning, structural verification, and automated assembly. However, practical deployment scenarios often involve complex backgrounds, scale variation, illumination fluctuation, and partial occlusion, which can degrade detection stability and localization precision.

Traditional feature-based approaches, such as edge detection methods [1] and the Hough Transform [2], rely heavily on handcrafted features and are generally sensitive to noise and illumination changes. With the emergence of convolutional neural networks, deep learning–based detectors have become the dominant solution. Two-stage detectors represented by the Region-based Convolutional Neural Network Features (R-CNN) family [3] typically achieve high detection accuracy but incur considerable computational overhead. In contrast, single-stage detectors such as YOLO [4] and Single Shot MultiBox Detector (SSD) [5] provide a more favorable balance between accuracy and efficiency, making them more suitable for real-time industrial applications.

The YOLO series has undergone continuous architectural refinement to improve representation capability and computational efficiency. YOLOX [6] introduced a decoupled detection head to mitigate classification–regression conflicts. YOLOv6 adopted RepOptimizer and re-parameterization strategies to reduce training–inference discrepancies [7, 8]. YOLOv7 further optimized feature aggregation structures to enhance parameter utilization efficiency [9]. More recent developments have incorporated attention mechanisms to strengthen contextual modeling ability. For example, RepViT [10] demonstrated that structural re-parameterization combined with efficient backbone design can achieve improved accuracy–latency trade-offs. The C2PSA module introduced in YOLOv11 integrates channel and spatial attention for multi-scale feature enhancement [11], while SPPF-based architectures have been shown to expand receptive fields and improve contextual representation in detection tasks [12].

Despite these advancements, two challenges remain significant in industrial-oriented component detection tasks such as valve nozzle inspection. First, valve nozzle assemblies typically consist of multiple substructures (e.g., nozzle body, handle, protective cap) with distinct scale characteristics, requiring stable multi-scale hierarchical feature modeling. Second, industrial environments frequently introduce partial occlusion, structural interference, and uneven illumination, which can weaken boundary contrast and affect bounding box regression accuracy. Conventional convolution-based architectures may not sufficiently integrate shallow structural cues with deep semantic representations under such conditions.

To address these issues, this study proposes YOLOv12-Enhanced, a refined detection framework tailored for industrial valve nozzle detection. The proposed method integrates a RepViT backbone to improve hierarchical feature extraction, incorporates SPPF combined with C2PSA attention to strengthen multi-scale contextual modeling, and introduces a Global Edge Information Fusion (GEIF) module to enhance boundary-aware localization. Through comprehensive evaluation on both a public benchmark dataset and a self-constructed industrial dataset, the effectiveness and robustness of the proposed structural refinements are systematically analyzed.

2. Fundamental YOLOv12 framework

The YOLO series has been widely recognized for achieving a favorable balance between detection accuracy and inference efficiency, making it suitable for real-time object detection tasks [13, 14]. YOLOv12 further enhances this balance by introducing attention-oriented architectural refinements while maintaining the lightweight characteristics of earlier versions.

The backbone of YOLOv12 adopts the C3K2 module, which combines feature fusion mechanisms with multi-scale convolutional kernels to expand the receptive field and improve contextual representation. In addition, the framework incorporates structural optimization strategies such as Area Attention and R-ELAN (Residual Efficient Layer Aggregation Network) to enhance feature aggregation efficiency and stabilize training behavior. The Area Attention mechanism reduces computational redundancy by partitioning feature

maps while preserving a relatively large receptive field. The structure of C2f and R-ELAN are shown in Figure 1.

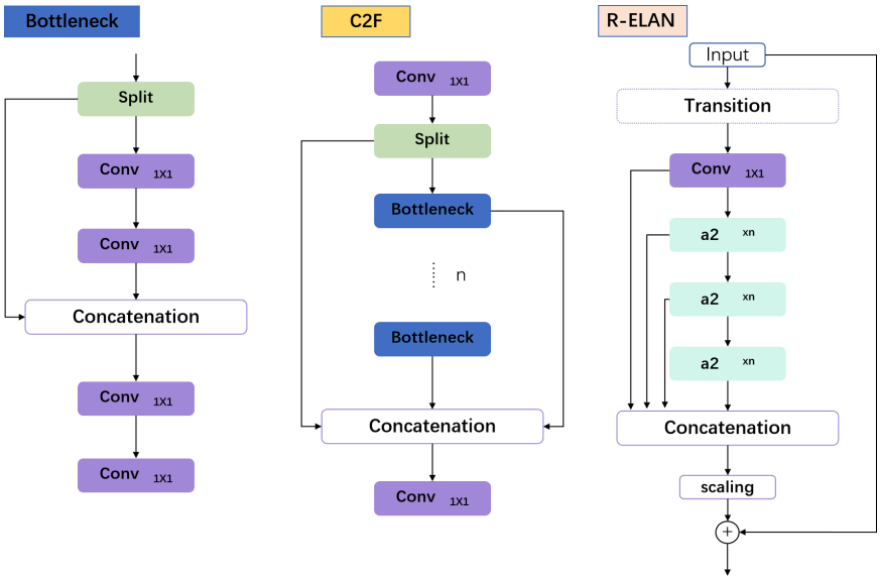

Figure 1. Structure of C2f and R-ELAN

Although these architectural improvements enhance contextual modeling and parameter utilization, ablation studies reported for lightweight variants (e.g., YOLOv12-N) indicate that certain residual aggregation strategies may contribute limited performance gains under constrained model capacity. Therefore, rather than modifying the entire structural aggregation mechanism, this study focuses on backbone refinement and feature enhancement strategies tailored for industrial valve nozzle detection. The objective is to improve hierarchical representation, multi-scale modeling capability, and boundary localization precision while preserving computational efficiency.

3. Enhanced YOLOv12 framework

3.1. Optimization of the backbone network based on RepViT

Revisiting Mobile CNN From ViT Perspective (RepViT) is inspired by RepVGG and aims to improve model performance while maintaining lightweight characteristics. By incorporating efficient design principles from Vision Transformers, such as structural re-parameterization and channel mixer separation, RepViT enhances representation capability compared with conventional lightweight CNNs.

As illustrated in Figure 2, RepViT adopts a four-stage hierarchical architecture with progressively reduced spatial resolution. To address latency issues in early-stage feature extraction, RepViT introduces an Early Convolution strategy composed of two 3×3 convolutions with stride 2, which improves feature expressiveness while maintaining efficiency.

During training, RepViT employs a multi-branch structure to enhance representation capacity. In the inference stage, structural re-parameterization converts it into an equivalent single-branch form, thereby reducing computational overhead. By combining the local feature extraction ability of CNNs with improved global contextual modeling, RepViT strengthens hierarchical feature representation, providing a more expressive backbone for object detection tasks.

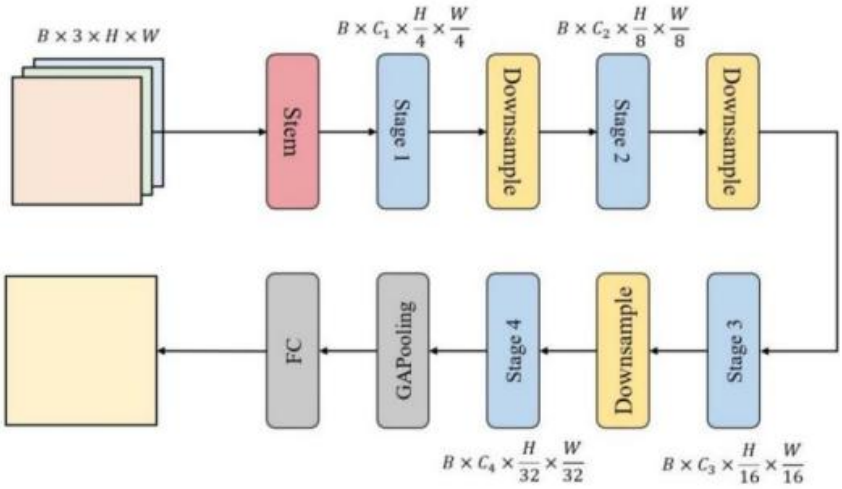

Figure 2. Structure of RepViT

3.2. Spatial Pyramid Pooling-Fast (SPPF) and attention module C2PSA

To compensate for the potential reduction in feature extraction capacity after replacing the original backbone with RepViT, this study integrates the SPPF module and the C2PSA attention mechanism into the enhanced framework.

Spatial Pyramid Pooling-Fast (SPPF) is an efficient multi-scale pooling structure derived from SPP [15] and its structure is illustrated in Figure 3. By sequentially combining convolution layers with multi-scale max pooling operations, SPPF enlarges the effective receptive field and aggregates contextual information across different spatial scales with minimal computational overhead. This design enhances the model's ability to detect objects with varying sizes.

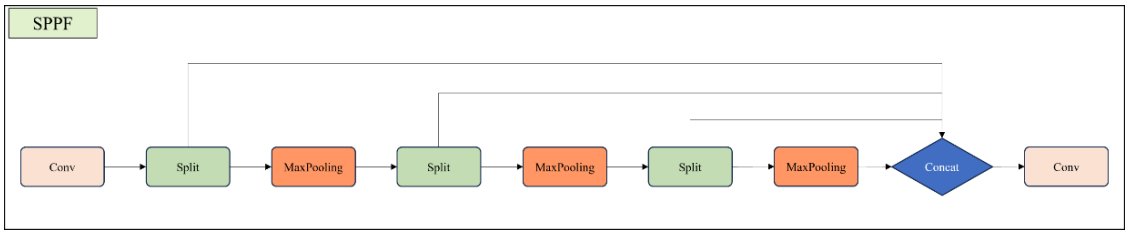

Figure 3. Structure of spatial pyramid pooling

The C2PSA module combines the Cross Stage Partial (CSP) [16] structure with the Pyramid Squeeze Attention (PSA) attention mechanism [17], thereby enhancing multi-scale feature extraction capability [18]. It is primarily used for object detection in complex scenarios, particularly excelling in handling multi-scale objects. The structure of C2PSA and its submodules is illustrated in Figure 4.

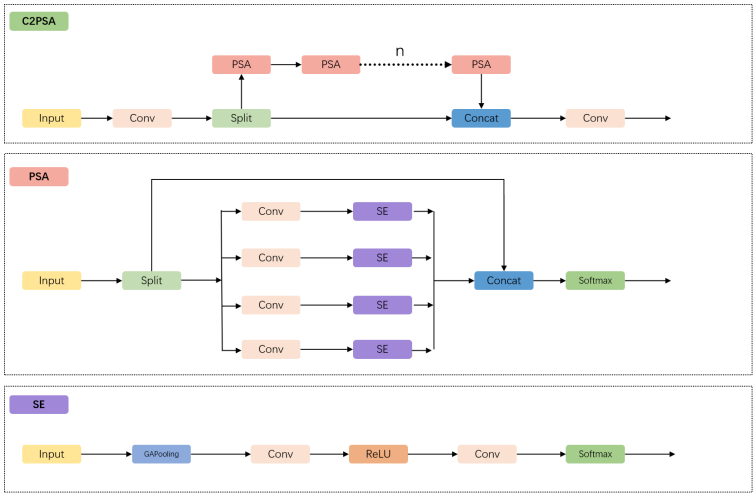

Figure 4. Structure of C2PSA

The C2PSA module inherits the segmented feature processing concept of CSP, where features are split into two parts after a 1×1 convolution—one part is directly transmitted, while the other undergoes processing by the PSA attention module. The two feature parts are then concatenated and passed through another 1×1 convolution to restore the original channel count.

The PSA module is an efficient attention mechanism designed to enhance the performance of convolutional neural networks in multi-scale feature extraction, it extracts multi-scale spatial information by incorporating convolution kernels of varying sizes, while integrating the Squeeze-and-Excitation (SE) module to weight feature channels, thereby strengthening the network's attention focus on targets of different scales. The SE module, one of the submodules of PSA, is an attention mechanism module first proposed by Jie Hu et al. in 2018 [19]. It extracts features through three steps: squeeze, excitation, and reweighting, not only integrating the global statistical information of feature maps in the spatial domain but also significantly improving computational efficiency and robustness by reducing model parameters.

By incorporating SPPF and C2PSA, the backbone output is enriched with stronger multi-scale contextual representation and refined feature weighting. This combination improves discriminative capability across objects with scale variation and structural complexity.

3.3. Global Edge Information Fusion module

Accurate boundary alignment is essential for improving bounding box regression in object detection tasks. To enhance localization precision, this study proposes a Global Edge Information Fusion (GEIF) module that integrates shallow edge cues with multi-scale semantic features.

The GEIF module consists of three components: a Sobel-based edge extraction module [20], a Global Edge (GE) module, and a Convolutional Edge Fusion (CEF) module. The Sobel module extracts directional edge features from shallow layers, providing structural information with reduced noise sensitivity. The GE module organizes edge representations in a hierarchical manner, enabling effective propagation of structural cues. The CEF module performs cross-channel feature fusion by projecting edge features into the same feature space as convolutional representations, followed by adaptive weighting and convolutional refinement. The structures of its submodules are illustrated in Figure 5.

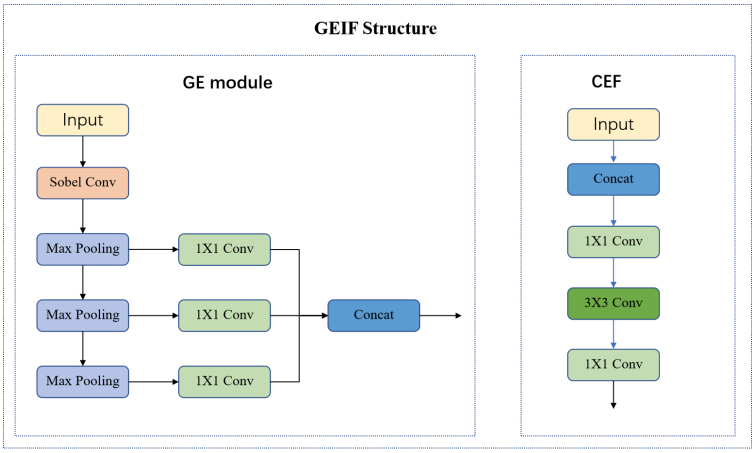

Figure 5. Structure of GEIF

Through this design, shallow edge information is systematically incorporated into deeper feature representations, improving boundary sensitivity and alignment between predicted bounding boxes and object contours. This integration enhances localization stability, particularly in scenarios involving partial occlusion or weak boundary contrast.

3.4. Enhanced YOLOv12 network structure

The enhanced network architecture, as illustrated in Figure 6, comprises three main components:

(1) An optimized backbone structure on the left.

(2) A neck module featuring a feature pyramid for multi-scale feature fusion.

(3) A detection head module.

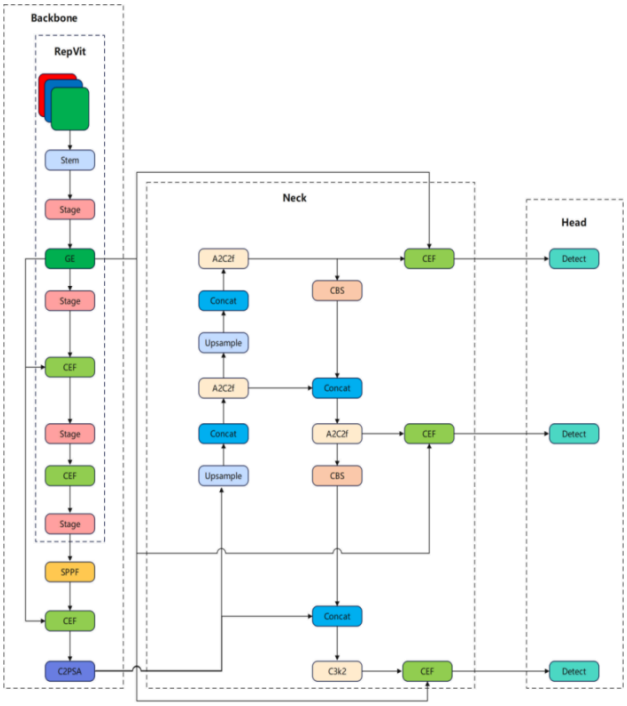

Figure 6. Structure of enhanced YOLOv12

4. Experimental section

4.1. Datasets and environment

Pascal VOC. Following standard practice, the Pascal VOC dataset was employed to evaluate general detection performance. A total of 5,200 images were used, with a 9:1 train-to-validation split. The dataset covers multiple object categories with relatively balanced distribution, making it suitable for multi-class detection tasks.

To verify the engineering applicability of the proposed method in industrial oriented object detection tasks, a self-constructed valve nozzle dataset was established in this study. The dataset contains 500 images collected from real industrial environments, covering various illumination conditions and complex backgrounds. All samples were manually annotated using the LabelMe tool, with bounding boxes as the annotation format. Three object categories were defined in the dataset: valve_nozzle, valve_handle, and valve_cap, corresponding to the nozzle body, operating handle, and protective cap, respectively. These categories exhibit variations in scale and pose, posing certain challenges to the model's multi-scale representation and target discrimination capabilities. The dataset was divided into training and validation sets according to a fixed ratio and used for model training and performance evaluation.

The experimental environment and configurations employed in this study are detailed in Table 1 and Table 2.

Table 1. CONFIGURATION

network training hyper-parameters	Project	Image resolution	Epochs	Optimizer	queue	Learning rate	batch size	momentum coefficient	data mosaic augmentation
	Parameter	640X640	300	SGD	4	0.01	16	0.937	10

Table 2. Experimental environment

	content	parameter		content	parameter
hardware	CPU	Intel(R) Xeon(R) Silver 4210R 2.40GHz	software	Operating system	Windows10
	GPU	NVIDIA RTX A4000		Python	3.10.5
	RAM	32G		Pytorch	2.2.2
	VRAM	16G		CUDA	12.6

4.2. Methods for evaluating algorithm performance

In this experiment, we used the following metrics to evaluate the performance of the model: accuracy, precision, recall, mAP, PR curve, etc. Here, TP represents True Positive, TN represents True Negative, FP represents False Positive, FN represents False Negative. Besides, AP represents Average Precision, $p(r)$ is the precision-recall curve, r represents recall. C is the total number of classes.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = TPTP + FN \tag{3}$$

$$AP = \int_0^1 Precision(R)\, dR \tag{4}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{5}$$

Under different IOU threshold conditions, mAP will have different forms. Generally, in object detection, mAP50 and mAP50-95 are the most common indicators. The area under the Precision-Recall (PR) curve quantifies the model's ability to identify positive samples across varying classification thresholds. When the PR curve exhibits a more pronounced convex shape toward the upper right corner, the integrated area—represented by the Average Precision (AP) value—increases accordingly.

In multi-class object detection tasks, the mean Average Precision (mAP), calculated as the arithmetic mean of AP values across all categories, serves as an indicator of the model's cross-category generalization capability. Therefore, an increase in the area enclosed by the PR curve corresponds to a simultaneous improvement in both AP and mAP, directly validating the performance enhancement of the model in object detection tasks.

4.3. Ablation experiments and result analysis

To evaluate the contribution of each proposed module, ablation experiments were conducted on the Pascal VOC dataset. The results are summarized in Table 3.

Table 3. Experimental results

module	precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLOv12n (Baseline)	71.1	62.3	67.8	50.5
Baseline+RepViT	73.4	60.7	67.9	51.7
Baseline+RepViT+SPPF+C2PSA	75.4	62.4	70.7	54.1
Baseline+RepViT+SPPF+C2PSA+GEIF	76.8	62.7	71.0	54.4

Effect of RepViT Backbone: Replacing the baseline backbone with RepViT improves hierarchical feature representation. As a result, mAP50-95 increases from 50.5% to 51.7%, indicating more discriminative feature extraction. This confirms that combining CNN-based local perception with Transformer-based global modeling enhances the network's representation capability.

Effect of Adding SPPF and C2PSA: When the SPPF and C2PSA modules are further integrated, the model achieves substantial improvements, with mAP50 rising to 70.7% and mAP50-95 to 54.1%. These results demonstrate that multi-scale pooling and channel–spatial attention significantly strengthen contextual representation, leading to more accurate recognition across objects of varying scales.

Effect of Adding GEIF Module: Finally, incorporating the GEIF module refines boundary localization. The final model achieves the highest performance, with precision improving from 71.1% (baseline) to 76.8%, mAP50 to 71.0%, and mAP50-95 to 54.4%. This confirms that integrating shallow edge cues with deep semantic features effectively enhances boundary sensitivity and contributes to higher overall detection accuracy. Notably, FLOPs decrease slightly compared to the previous stage, indicating that GEIF not only enhances boundary sensitivity but also maintains computational efficiency.

Overall, the ablation results clearly validate the effectiveness of each proposed module. RepViT improves feature representation, SPPF and C2PSA enhance multi-scale discriminability, and GEIF further refines boundary localization. Together, these components enable YOLOv12-Enhanced to achieve consistent and significant gains in detection accuracy.

4.4. Comparative experiment

To evaluate the performance of the proposed enhanced model against current mainstream object detection frameworks, a series of comparative experiments were conducted under identical hardware, software environments, and datasets. The selected models for training and validation include YOLOv5, YOLOv8n, YOLOv10n, YOLOv12n and the enhanced model presented in this study. Table 4 presents the comparison results (mAP50 and mAP50-95) for five sets of comparative experiments carried out on the PASCAL VOC datasets.

Table 4. Comparison table of object detection models

Models	mAP50 (%)	mAP50-95 (%)	GFLOPs
YOLOv5n	66.2	47.3	4.5
YOLOv8n	67.1	49.6	8.9
YOLOv10n	66.9	50.5	7.4
YOLOv12n	67.8	50.5	6.7
Enhanced	71.0	54.4	7.1

As illustrated in Table 4, the proposed YOLOv12-Enhanced model achieves the best overall detection performance among all compared lightweight detectors. Specifically, it attains an mAP50 of 71.0% and an mAP50–95 of 54.4%, which are the highest values across all evaluated models. Compared with YOLOv12n, the enhanced version improves mAP50 by 3.2 percentage points (from 67.8% to 71.0%) and mAP50–95 by 3.9 percentage points (from 50.5% to 54.4%), demonstrating clear gains in both classification confidence and localization accuracy under stricter IoU thresholds.

When compared with other mainstream lightweight models, including YOLOv5n, YOLOv8n, and YOLOv10n, the proposed model consistently outperforms them in both evaluation metrics, indicating stronger cross-scale representation and boundary regression capability. In terms of computational complexity, the enhanced model requires 7.1 GFLOPs, which is only slightly higher than YOLOv12n (6.7 GFLOPs) and remains lower than YOLOv8n (8.9 GFLOPs). This modest increase in computation leads to a measurable improvement in detection accuracy

Overall, the results confirm that the proposed architectural refinements effectively enhance detection precision while maintaining lightweight characteristics suitable for practical deployment.

As shown in Figure 7, the PR curves of the improved model outperform those of YOLOv12n across most categories. In particular, the improved model maintains relatively stable precision in the medium-to-high recall range, indicating that it achieves higher recall while effectively suppressing false positives. From an overall perspective, the improved model presents a larger envelope area under the PR curve, demonstrating superior average detection performance compared with YOLOv12n. This observation is consistent with the previously reported mAP results and confirms that the proposed method exhibits stable detection performance under different confidence thresholds.

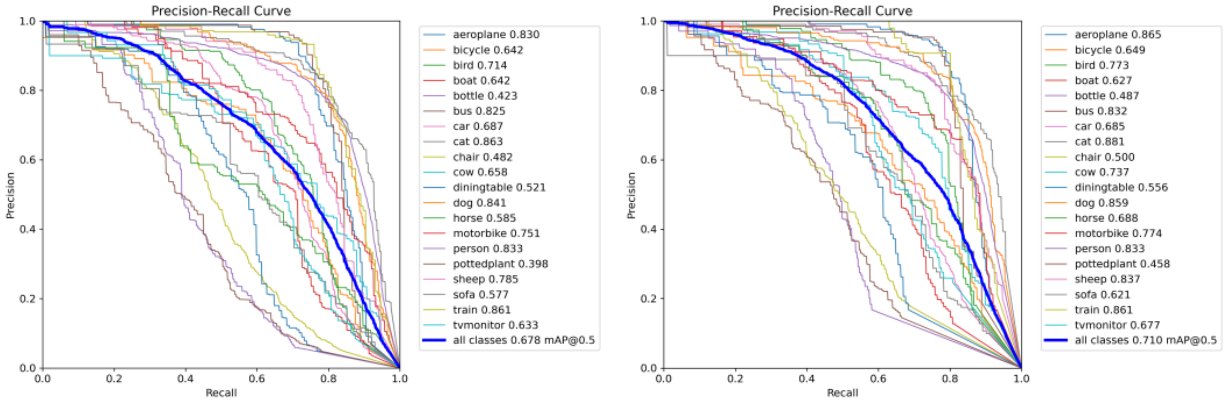

Figure 7. Comparison of PR curves between the baseline and the enhanced model

4.5. Detection results on the self-constructed dataset

As shown in Figure 8, The detection results of the improved model on the self-constructed oriented valve nozzle dataset cover typical industrial scenarios, including varying illumination conditions and multi-object coexistence. The model can stably identify three categories—valve_nozzle, valve_handle, and valve_cap—and generate relatively accurate bounding box localization. Even in cases of partial occlusion or when components are spatially close to each other, the model effectively distinguishes different structural parts without obvious category confusion, demonstrating strong engineering robustness.

From the performance curve analysis, the improved model maintains relatively stable detection accuracy on this dataset. The PR curves for each category as well as the overall PR curve exhibit favorable trends, indicating a high overall detection level. These results are consistent with the qualitative visualization outcomes and quantitative evaluation metrics, further verifying the feasibility and stability of the proposed method for oriented valve nozzle detection tasks.

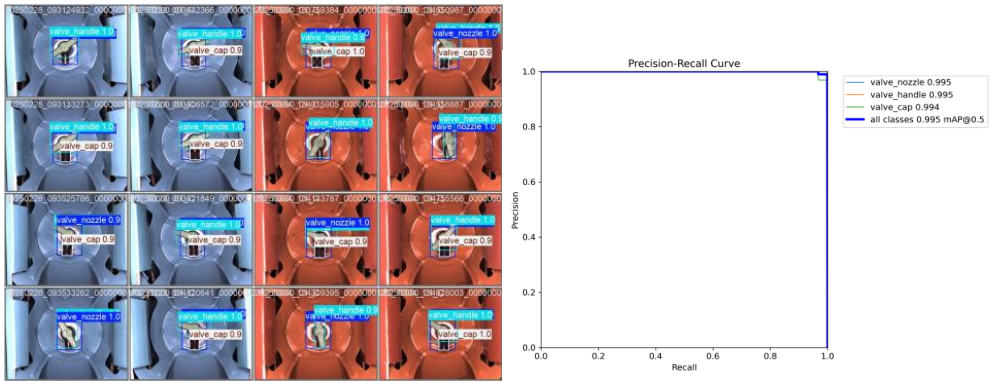

Figure 8. Detection results on the self-constructed dataset

5. Conclusion

This study proposed YOLOv12-Enhanced, a refined single-stage detection framework designed for industrial valve nozzle inspection under complex conditions. By integrating a RepViT backbone, an SPPF module with C2PSA attention, and a Global Edge Information Fusion (GEIF) module, the proposed method enhances

hierarchical feature representation, strengthens multi-scale contextual modeling, and improves boundary localization accuracy.

Experimental results on the Pascal VOC dataset demonstrate that YOLOv12-Enhanced achieves 71.0% mAP50 and 54.4% mAP50–95, outperforming the baseline YOLOv12n while maintaining comparable computational complexity. Ablation experiments confirm that each module contributes to performance improvement, and their combination yields complementary gains in both classification confidence and localization precision.

Furthermore, evaluation on a self-constructed industrial valve nozzle dataset indicates that the proposed framework maintains stable detection performance under varying illumination and partial occlusion conditions, effectively distinguishing structural components such as nozzle body, handle, and cap.

Overall, the results suggest that integrating lightweight backbone optimization, multi-scale attention enhancement, and edge-aware feature fusion provides an effective strategy for improving detection robustness in industrial-oriented object detection tasks. The proposed design can serve as a reference framework for similar component-level inspection scenarios where both accuracy and computational efficiency are required.

References

[1] Amer, G. M. H., & Abushaala, A. M. (2015). Edge detection methods. In *2015 2nd World Symposium on Web Applications and Networking (WSWAN)* (pp. 1–7). IEEE.

[2] Illingworth, J., & Kittler, J. (1988). A survey of the Hough transform. *Computer Vision, Graphics, and Image Processing, 44*(1), 87–116.

[3] Bharati, P., & Pramanik, A. (2020). Deep learning techniques—R-CNN to mask R-CNN: a survey. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019* (pp. 657–668).

[4] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. *Procedia Computer Science, 199*, 1066–1073.

[5] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I (Vol. 14, pp. 21–37). Springer International Publishing.

[6] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). *YOLOX: Exceeding YOLO series in 2021*. arXiv preprint. https://arxiv.org/abs/2107.08430

[7] Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Sun, J. (2021). RepVGG: Making VGG-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13733–13742).

[8] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). *YOLOv6: A single-stage object detection framework for industrial applications*. arXiv preprint. https://arxiv.org/abs/2209.02976

[9] Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). 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* (pp. 7464–7475).

[10] Wang, A., Chen, H., Lin, Z., Han, J., & Ding, G. (2024). RepViT: Revisiting mobile CNN from ViT perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15909–15920).

[11] Wang, K., Liu, J., & Cai, X. (2025). *C2PSA-Enhanced YOLOv11 architecture: A novel approach for small target detection in cotton disease diagnosis*. arXiv preprint. https://arxiv.org/abs/2508.12219

[12] Tang, H., Liang, S., Yao, D., & Qiao, Y. (2023). A visual defect detection for optics lens based on the YOLOv5-C3CA-SPPF network model. *Optics Express, 31*(2), 2628–2643.

[13] Tian, Y., Ye, Q., & Doermann, D. (2025). *YOLOv12: Attention-centric real-time object detectors*. arXiv preprint. https://arxiv.org/abs/2502.12524

[14] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788).

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(9), 1904–1916.

[16] Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2021). Scaled-YOLOv4: Scaling cross stage partial network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13029–13038).

[17] Zhang, H., Zu, K., Lu, J., Zou, Y., & Meng, D. (2022). EPSANet: An efficient pyramid squeeze attention block on convolutional neural network. In *Proceedings of the Asian Conference on Computer Vision* (pp. 1161–1177).

[18] Khanam, R., & Hussain, M. (2024). YOLOv11: An overview of the key architectural enhancements. arXiv preprint. https://arxiv.org/abs/2410.17725

[19] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7132–7141).

[20] Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Hassannataj Joloudari, J., Shoeibi, A., Jafari, M., Hussain, S., et al. (2021). Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomedical Signal Processing and Control, 68*, 102622.