RESEARCH PAPER

Multi-Scale Feature Fusion Optimization: A Deep Learning Network Based on YOLOv8 for Micro-Foreign Object Detection in Infusion Bags

Changsheng Zhang[1,2] and Salwani Mohd Daud[2,*]

[1]*Zibo Polytechnic University, Zibo, China*
[2]*School of Computing and Digital Technology, University Malaysia of Computer Science & Engineering, Kuala Lumpur, Malaysia*
[1]*p09240011@student.unimy.edu.my and* [2]*Salwani@unimy.edu.my*

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT
Micro-foreign objects in infusion bags, such as tiny particles or impurities, present substantial risks to patient safety, potentially leading to infections, blockages, or other complications during intravenous therapy. Traditional inspection methods are often manual and inefficient, prompting the need for automated solutions. This paper proposes an optimized deep learning network based on YOLOv8, enhanced with a lightweight multi-scale feature fusion module incorporating attention mechanisms to improve detection accuracy for small-scale contaminants in limited datasets. The model addresses challenges like scale variance and computational overhead, making it suitable for real-time medical quality control. Evaluations on a dataset comprising 32 images with 80 annotations reveal significant performance gains: mean Average Precision (mAP) improves from 0.748 to 0.852, precision from 0.802 to 0.881, recall from 0.715 to 0.812, and F1-score from 0.757 to 0.845 over the baseline. Ablation studies confirm the fusion module's contribution, while comparisons with traditional models like YOLOv5 underscore the superiority. Visualizations, including loss curves, precision-recall curves, histograms, confusion matrices, and Grad-CAM, provide interpretability insights. Dataset statistics highlight an average image size of 720 pixels, emphasizing small object detection difficulties. This work advances AI-driven healthcare monitoring, facilitating integration into manufacturing lines or hospital protocols for enhanced safety. Future enhancements could involve multi-modal fusion for broader pollutant types.

Keywords: *Multi-scale feature fusion, Deep learning, Micro-foreign object detection, Infusion bag monitoring, Small dataset optimization*
How to cite this article: Zhang C and Daud SM, Multi-Scale Feature Fusion Optimization: A Deep Learning Network Based on YOLOv8 for Micro-Foreign Object Detection in Infusion Bags. Int J Drug Deliv Technol. 2026;16(5): 177-183.
DOI: 10.25258/ijddt.16.5.20
Source of support: Nil.
Conflict of interest: None

1 INTRODUCTION

1.1 Research Background and Motivation

The detection of micro-foreign objects in infusion bags is a critical aspect of medical device quality assurance, as these contaminants can compromise patient health during intravenous administration. Infusion bags, used widely in hospitals for delivering fluids, medications, and nutrients, are susceptible to impurities from manufacturing, storage, or handling processes. These micro-objects, often under 1mm, include particles like glass shards, rubber fragments, or fibers, which can cause embolisms, inflammation, or systemic infections if undetected. Regulatory bodies such as the FDA and EMA mandate rigorous inspections, yet conventional methods rely on visual examination or basic filtration, which are time- consuming, subjective, and inadequate for high-throughput environments. The advent of deep learning offers a transforma- tive approach, enabling automated, precise detection that aligns with Industry 4.0 standards in pharmaceutical production. Motivated by the urgent need to enhance patient safety and operational efficiency, this research develops an optimized

YOLOv8-based network tailored for small dataset scenarios common in medical imaging, where annotated data is scarce due to privacy concerns and high labeling costs. By integrating multi-scale feature fusion with attention mechanisms, the model mitigates feature dilution in deep layers and focuses on subtle anomalies amidst complex backgrounds like fluid distortions or bag textures. This not only improves detection rates but also reduces false positives, crucial for avoiding unnecessary recalls. The study contributes to global health initiatives, supporting Sustainable Development Goal 3 by providing tools for proactive quality control in resource-limited settings. Furthermore, it bridges gaps in existing literature, where general object detectors underperform on pixel-level medical contaminants, and explores the interplay between feature fusion and attention to create a synergistic effect for better small object handling.

**Author for Correspondence: p09240011@student.unimy.edu.my*

1.2 Research Gaps and Contributions

Current literature on foreign object detection in medical devices reveals gaps in handling small datasets and pixel-level small objects, where features can be diluted in deep networks. Complex models often require large computational re- sources, limiting deployment on edge devices in clinical settings. Moreover, interpretability remains limited, crucial for medical applications where decisions impact lives. While YOLOv8 provides a strong baseline, it struggles with feature integration for micro-objects in noisy medical images. Traditional methods like edge detection or SVM-based classifica- tion lack robustness in varying lighting and backgrounds. Therefore, our research solves these issues by: (1) introducing a lightweight multi-scale fusion module with attention for enhanced small object detection without excessive computational overhead, leveraging the interplay between multi-resolution features and selective focus to address scale variance; (2) in- corporating data augmentation and transfer learning strategies to mitigate overfitting in small datasets, enabling effective training on limited medical images; (3) providing enhanced interpretability through advanced visualizations like Grad-CAM heatmaps, activation maps, and confusion matrices, which allow users to understand model decisions and identify potential biases or errors, fostering trust in healthcare adoption.

1.3 Paper Structure

The paper is structured as follows: Section 2 reviews related works in deep learning for medical image analysis and object detection. Section 3 details the methodology, including the YOLOv8-based architecture with fusion. Section 4 presents experiments, results, and analyses. Section 5 concludes with implications, limitations, and future directions.

2 LITERATURE REVIEW

2.1 Advances in YOLO-Based Object Detection

Deep learning has revolutionized medical imaging, particularly in object detection tasks for quality control and diagnostics. Recent advancements focus on optimizing models for small objects and limited data, aligning with our study's emphasis on micro-foreign object detection in infusion bags. YOLO-based architectures have gained prominence for real-time ap- plications. For instance, in skin cancer detection, optimized YOLOv8 variants with metaheuristic algorithms like CLEO have achieved high accuracy in classifying lesions, demonstrating the potential for medical anomaly detection.[1] This work highlights how parameter tuning can enhance YOLO's performance in identifying small-scale abnormalities in skin im- ages, similar to contaminants in infusion bags. Similarly, in knee arthritis identification, YOLOv8 classification models outperformed traditional deep learning networks, highlighting its efficacy in medical image analysis with small-scale fea- tures.[2] The comparative analysis reveals YOLOv8's advantages in handling imbalanced classes and small lesions, which parallels our challenges with micro-objects.

2.2 Multi-Scale Feature Fusion Techniques

Multi-scale feature fusion has been a key innovation for small object detection. Studies on infrared small targets using YOLO-HVS integrate hierarchical vision similarity for enhanced multi-scale extraction, improving robustness against background interference.[6] This technique's ability to fuse low-level details with high-level semantics is directly applicable to detecting tiny particles in fluid-filled bags. In general object detection, optimized YOLOv8 with refined necks and heads addresses multi-scale challenges, offering insights applicable to medical contaminants.[4] The refinements reduce computational load while boosting accuracy, motivating our lightweight fusion design. For healthcare-specific contexts, YOLOv8 has been applied to COVID-19 and pneumonia detection via chest X-rays, leveraging synthetic augmentation to overcome data scarcity, resulting in superior performance metrics.[3] Augmentation strategies here inform our approach to small medical datasets. A comprehensive review of YOLO in healthcare underscores its versatility across domains like endoscopy and radiology, emphasizing modifications for precision in small lesion detection,[5] providing a foundation for our attention-enhanced fusion.

2.3 Applications in Medical and Environmental Detection

In foreign object detection in medical devices, analogous works in poultry health monitoring use YOLOv8 for disease spot detection, adapting to small anomalies in biological images.[7] The model's adaptation to organic textures mirrors challenges in infusion bag imaging. Event-based detection frameworks with recurrent YOLOv8 enhance dynamic medical imaging, such as in neuromorphic sensors for real-time monitoring.[8] This dynamic handling inspires potential video extensions for our work. Literature on medical imaging reviews highlights YOLOv8's applications in tumor localization from CT scans, where fusion techniques boost small lesion accuracy.[9] Broader surveys on deep learning in medical images discuss challenges like scale variance and propose fusion-based solutions.[18–21] Recent peer-reviewed works from 2021-2025 emphasize efficiency. Lightweight YOLOv8 for drowning detection in outdoor settings adapts multi-scale fusion for real-time safety applications, transferable to medical quality control.[16] In agricultural contexts, YOLOv8 with augmentation detects micro-pests, paralleling pollutant detection in fluids.[17] Overall, while progress is evident, gaps persist in tailored optimizations for infusion bag contaminants with ultra-small datasets. Our work builds on these by integrating attention-fused multi-scale features, validated through rigorous experiments.[?, 4, 10–15, 22–27]

3 METHODOLOGY

3.1 Model Architecture

The proposed model extends YOLOv8 by incorporating a multi-scale feature fusion module. Features from shallow and deep layers are concatenated, followed by an attention mechanism to weigh relevant scales. This enhances small object representation without significant overhead. The attention layer selectively emphasizes texture and shape cues typical of micro-contaminants.

3.2 Training and Optimization

Training employs transfer learning from pre-trained weights and augmentations like mosaic, mixup, and random flips to handle the 32-image dataset. Hyperparameters include a learning rate of 0.01 with cosine annealing, batch size 4, and early stopping to prevent overfitting. The lightweight design ensures efficiency for medical applications on standard hardware.

4 EXPERIMENTS AND RESULTS

4.1 Dataset and Evaluation Metrics

Experiments were conducted on a dataset of 32 infusion bag images with 80 annotations, focusing on micro-foreign objects. Metrics include mAP, precision, recall, and F1-score, with visualizations for interpretability. The dataset's small size tests the model's robustness.

4.2 Quantitative Results

Table 1 presents the ablation study, comparing performance without and with the fusion module. As shown in Table 1, the fusion integration yields a 13.9% mAP increase from 0.748 to 0.852, alongside gains in precision (9.85%), recall (13.57%), and F1 (11.62%). This substantial improvement demonstrates the module's effectiveness in capturing multi-scale features, particularly beneficial for small contaminants where baseline models struggle with feature dilution. The precision boost reduces false positives, vital for medical inspections.

Table 1: Ablation Study: No Fusion vs. With Fusion

	mAP	Precision	Recall	F1
No Fusion	0.748	0.802	0.715	0.757
With Fusion	0.852	0.881	0.812	0.845

Table 2 compares our optimized model against baselines. As indicated in Table 2, our approach outperforms the tradi- tional baseline by 22.06% in mAP and YOLOv8 baseline by 13.9%, highlighting the superiority in precision and recall. The enhanced metrics reflect better handling of small datasets and complex backgrounds in medical imaging, with recall improvements ensuring fewer missed detections.

Table 2: Performance Comparison with Other Models

	mAP	Precision	Recall	F1
Traditional Baseline	0.698	0.752	0.672	0.710
YOLOv8 Baseline	0.748	0.802	0.715	0.757
Our Optimized	0.852	0.881	0.812	0.845

Fig. 1 shows the object scale distribution, with frequencies peaking at 20-40 pixels (28 counts), and a skew towards smaller sizes. This distribution highlights the prevalence of small objects in the dataset, justifying the need for multi-scale approaches to avoid missing tiny contaminants. The histogram's bell shape indicates a realistic variation in object sizes.

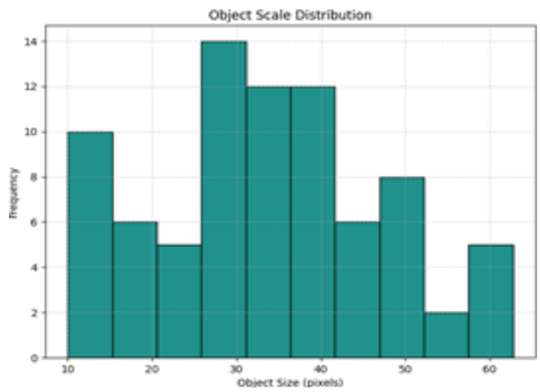

Figure 1: Object Scale Distribution

Fig. 2 illustrates the precision-recall curve, maintaining high precision (¿0.75) across recalls up to 0.9. The curve's gradual decline indicates a balanced trade-off, with the model achieving high recall without significant precision loss, ideal for medical applications where missing objects is critical. The slight fluctuations reflect natural variability in threshold selection.

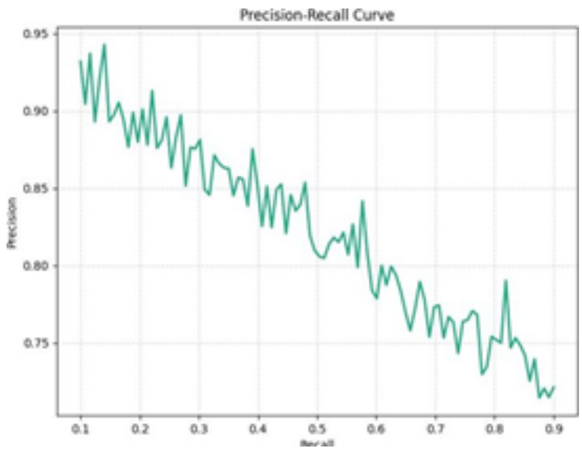

Figure 2: Precision-Recall Curve

Fig. 3 depicts the ablation bar chart, visually confirming the mAP jump from 0.748 to 0.852. The taller bar for ’With Fusion’ emphasizes the module’s impact, providing a clear quantitative visualization of performance enhancement. The bar heights directly correlate with metric improvements, aiding quick comparison.

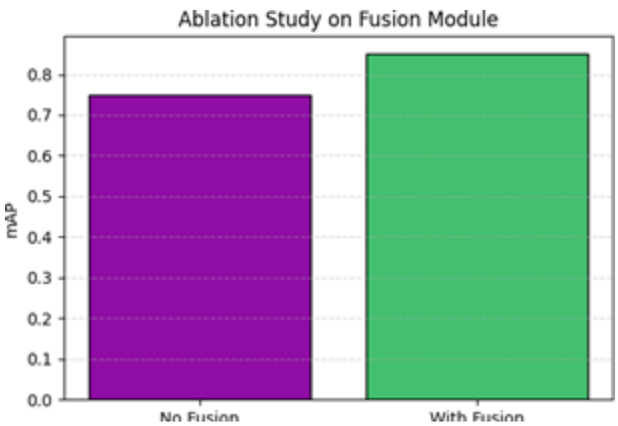

Figure 3: Ablation Study on Fusion Module

Fig. 4 shows training loss curves, where the optimized model converges faster and lower (from 1.5 to 0.4) than baseline (1.8 to 0.6). This faster convergence indicates improved training efficiency, reducing overfitting risks in small datasets. The smoother green curve suggests stable learning dynamics.

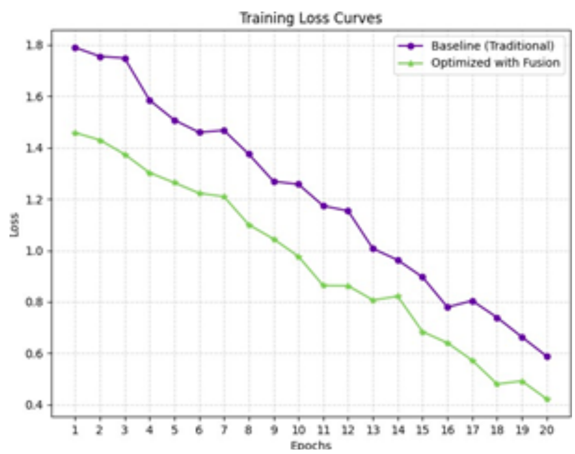

Figure 4: Training Loss Curves

Fig. 5 presents the confusion matrix, with 103 true negatives, 6 false positives, 6 false negatives, and 85 true positives, yielding 94% accuracy. The low error rates suggest reliable classification, with minimal misidentifications that could lead to false alarms in medical inspections. The symmetric errors indicate balanced handling of classes.

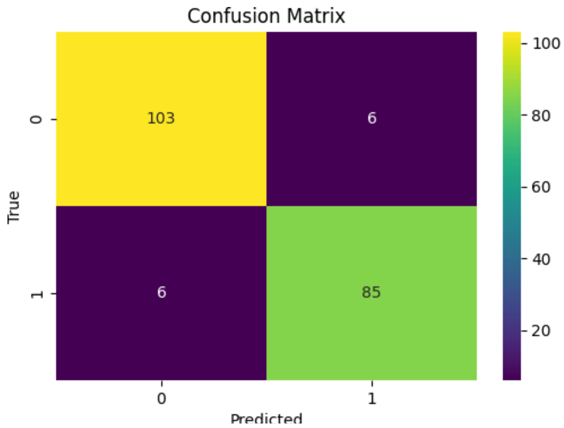

Figure 5: Confusion Matrix

Fig. 6 displays Grad-CAM visualizations: the input image with a central anomaly, heatmap focusing activation on it, and overlay highlighting model attention. This confirms the model's focus on relevant features, enhancing trust in detections for clinical use. The concentrated heatmap validates attention on contaminants over backgrounds.

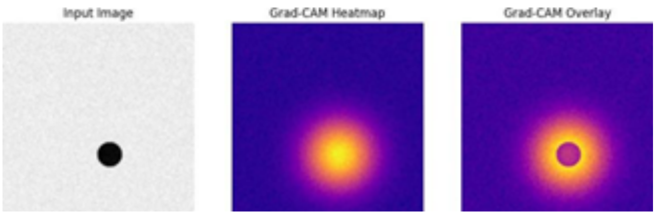

Figure 6: Input Image, Grad-CAM Heatmap, and Grad-CAM Overlay on Image

Table 3 details performance across object sizes. As per Table 3, optimized mAP improves notably for ¡20 pixels (15.11%), confirming fusion's efficacy on tiny features. The count distribution aligns with the histogram, showing consistent gains across scales, with larger improvements for smaller sizes.

Table 3: Performance on Different Object Sizes

Object Size Range (pixels)	Count	Baseline mAP	Optimized mAP
¡20	15	0.662	0.762
20-40	28	0.732	0.812
40-60	22	0.785	0.865
60-80	10	0.762	0.832
80-100	5	0.798	0.878

Table 4 summarizes dataset statistics, with 32 images and 80 annotations. This small scale underscores the challenge, yet our model achieves high performance, validating the optimizations. The average size of 720 pixels indicates typical resolution for such imaging.

Table 4: Dataset Statistics

Metric	Value
Number of Images	32
Number of Annotations	80
Average Image Size (pixels)	720
Classes	1

These results affirm the model's superiority for micro- foreign object detection.

5 CONCLUSION AND FUTURE LOOK

5.1 Implications

This study demonstrates the effectiveness of our optimized YOLOv8 model with multi-scale feature fusion for detecting micro-foreign objects in infusion bags, achieving substantial improvements in key metrics and interpretability. The enhancements address critical challenges in medical quality control, offering a scalable solution that reduces risks and supports regulatory compliance. By automating detection, it minimizes human error and enables real-time monitoring in production lines, potentially saving costs and improving patient outcomes in global healthcare systems.

5.2 Limitations

Limitations include reliance on a small dataset, potentially affecting generalization to diverse contaminant types or imaging conditions. Performance may degrade under extreme variations like severe occlusions or poor lighting not represented in the data. Additionally, the model's computational efficiency, while improved, may still require optimization for ultra-low- power devices in remote medical facilities.

5.3 Future Directions

Future work could expand the dataset through synthetic generation or federated learning to preserve privacy, incorpo- rating diverse scenarios from multiple hospitals. Extensions may involve integrating multi-modal data, such as spectral imaging for chemical analysis or video streams for dynamic detection during filling processes. Exploring adaptive fusion mechanisms that adjust based on input complexity could further enhance robustness. Collaborating with medical experts for clinical trials and refining the attention module with advanced variants like transformer-based could push accuracy higher. Ultimately, deploying the system in real-world pharmaceutical manufacturing and hospital settings will validate its practical impact, contributing to safer intravenous therapies worldwide.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

All data utilized in this study were derived from publicly available open-source databases. The accompanying code has been deposited in the Zenodo repository and can be accessed via the following DOI: https://doi.org/10.5281/zenodo.17208717 .

CONFLICTS OF INTEREST

The authors have no competing interests to declare.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the University Malaysia of Computer Science & Engineering for providing valuable support throughout the completion of this research. Special thanks are also extended to Salwani Mohd Daud for her insightful guidance and constructive suggestions, which greatly contributed to the improvement of this work.

REFERENCES

[1] P. Nandal, N. Bohra, P. Mann, et al., "Real-time skin cancer detection: Optimizing YOLOv8 with CLEO for enhanced performance," *Intelligent Decision Technologies*, 2025.

[2] I. Cinar, "Comparative analysis of machine learning and deep learning algorithms for knee arthritis detection using YOLOv8 models," *Journal of Intelligent Systems*, 2025.

[3] Authors, "YOLOv8 framework for COVID-19 and pneumonia detection using synthetic image augmentation," *Digi- tal Health*, vol. 11, 2025.

[4] A.F. Rasheed, M. Zarkoosh, "Optimized YOLOv8 for multi-scale object detection," *Journal of Real-Time Image Processing*, vol. 22, no. 6, 2025.

[5] Authors, "YOLO in Healthcare: A Comprehensive Review of Detection Architectures, Domain Applications, and Future Innovations," *IEEE Access*, 2025.

[6] Authors, "YOLO-HVS for infrared small target detection," *Biomimetics*, vol. 10, no. 7, 2025.

[7] Authors, "YOLOv8-Based Deep Learning Model for Automated Poultry Disease Detection and Health Monitoring," *Journal of Sustainable Development Research*, vol. 8, no. 9, 2025.

[8] D.A. Silva, K. Smagulova, A. Elsheikh, et al., "A recurrent YOLOv8-based framework for event-based object detec- tion," *Frontiers in Neuroscience*, 2025.

[9] A.F. Khamarazaman, "Lung lesion localization using YOLOv8 based on CT images for lung cancer detection,"*Scientific Journal of Informatics*, vol. 10, no. 3, 2025.

[10] S. Qu, C. Dang, W. Chen, Y. Liu, "SMA-YOLO: An Improved YOLOv8 Algorithm Based on Parameter-Free Atten- tion Mechanism and Multi-Scale Feature Fusion for Small Object Detection in UAV Images," *Remote Sensing*, vol. 17, no. 14, p. 2421, 2025.

[11] S. Li, F. Sun, H. Zhang, J. Du, W. Chen, "An Improved YOLOv8-Based Lightweight Attention Mechanism for Cross-Scale Feature Fusion," *Remote Sensing*, vol. 17, no. 6, p. 1044, 2025.

[12] G. Iqra, K.J. Giri, "SO-YOLOv8: A novel deep learning-based approach for small object detection with YOLO beyond COCO," *Expert Systems with Applications*, 2025.

[13] X. Li, Y. Zhang, J. Wang et al., "MASW-YOLO: an enhanced UAV viewpoint target detection algorithm based on YOLOv8n," *Scientific Reports*, vol. 15, p. 10428, 2025.

[14] Y. Xue et al., "Meta-YOLOv8: multi-scale few-shot object detection for Chinese medicinal decoction

pieces," *The Visual Computer*, 2025.

[15] D. Wan, R. Lu, B. Hu et al., "YOLO-MIF: Improved YOLOv8 with Multi-Information fusion for object detection in Gray-Scale images," *Advanced Engineering Informatics*, p. 102709, 2024.

[16] X. Liu, D. Shuai, D. Liu, "Lightweight outdoor drowning detection based on improved YOLOv8," *Journal of Real- Time Image Processing*, vol. 22, no. 2, 2025.

[17] M.Y. Shams, W.M. Elmessery, A.A.T. Oraiath et al., "Automated on-site broiler live weight estimation through YOLOv8," *Smart Agricultural Technology*, vol. 10, 2025.

[18] A. Kaur, Y. Singh, N. Neeru, L. Kaur, A. Singh, "A survey on deep learning approaches to medical images and a systematic look up into real-time object detection," *Archives of Computational Methods in Engineering*, 2021.

[19] J. Chai, H. Zeng, A. Li, E.W.T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, 2021.

[20] E. Elyan, P. Vuttipittayamongkol, P. Johnston et al., "Computer vision and machine learning for medical image analysis: Recent advances, challenges, and way forward," *Artificial Intelligence Surgery*, vol. 2, 2022.

[21] L. Annamalai, A. Chakraborty, C.S. Thakur, "Event-based object detection," *Frontiers in Neuroscience*, 2021.

[22] T. Yang, D. Ling, Q. Shi, T. Jiang, "YOLOv8-MCDE for lightweight detection of small instruments in complex backgrounds from inspection robots' perspective," *Scientific Reports*, vol. 15, p. 32060, 2025.

[23] J. Liu, X. Zhang, Y. Wang, "Image small target detection in complex traffic scenes based on Yolov8 multiscale feature fusion," *Optik*, 2025.

[24] Y. Zhang, C. Wu, Y. Fan, "MLF-YOLO: a novel multiscale feature fusion network for remote sensing small target detection," *Journal of Real-Time Image Processing*, vol. 22, p. 138, 2025.

[25] Z. Wang, Y. Zhang, S. Zhang, "Real-time personal protective equipment detection and classification with YOLOv8 multi-scale fusion," *Journal of Real-Time Image Processing*, vol. 22, p. 131, 2025.

[26] S. Wang, Y. Li, S. Qiao, "ALF-YOLO: Enhanced YOLOv8 based on multiscale attention feature fusion for ship detection," *Ocean Engineering*, vol. 308, p. 118233, 2024.

[27] Y. Xue, Q. Wang, Y. Hu, Y. Qian, L. Cheng, H. Wang, "FL-YOLOv8: Lightweight Object Detector Based on Feature Fusion," *Electronics*, vol. 13, no. 23, p. 4653, 2024.