Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SDS-YOLO11: An Enhanced Lightweight Model for Underwater Litter Detection in Complex Marine Environments

Yang Liu
Northeastern University
Chong Fu
fuchong@mail.neu.edu.cn
Northeastern University
Junxin Chen
Dalian University of Technology
Xingwei Wang
Northeastern University

Research Article

Keywords: Object detection, Marine litter, Underwater detection, Lightweight model, YOLO11

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8838185/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SDS-YOLO11: An Enhanced Lightweight Model for Underwater Litter Detection in Complex Marine Environments

Yang Liu[1], Chong Fu[1,2,3*], Junxin Chen[4], Xingwei Wang[1]

[1]School of Computer Science and Engineering, Northeastern University, Shenyang, 110819, China.

[2]Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang, 110819, China.

[3]Engineering Research Center of Security Technology of Complex Network System, Ministry of Education, Shenyang, 110819, China.

[4]School of Software, Dalian University of Technology, Dalian, 116621, China.

*Corresponding author(s). E-mail(s): fuchong@mail.neu.edu.cn;

Abstract

Detecting marine underwater litter remains challenging due to suspended particle interference, color distortion, feature confusion, occluded feature loss, and inadequate lightweight design, factors that severely hinder conventional detectors' performance and limit deployment on resource-constrained underwater robots. To address these issues, this paper proposes an enhanced lightweight detector named SDS-YOLO11. This model incorporates three key innovations: (1) a novel Shared-Key Latent Space Attention (SKLA) mechanism enhances the global feature consistency of targets via shared-key attention and suppresses high-frequency noise through latent space feature compression; (2) a Lightweight Dual-Path Residual Downsampling module that balances model lightweighting and refined feature extraction through a multi-branch structure; (3) an improved SEAMHead detection head that integrates channel-spatial attention with exponential normalization to enhance robustness under occlusion and complex backgrounds. Experimental results on a public Waste-9cbfm Marine Underwater Litter Dataset demonstrate that SDS-YOLO11 surpasses the YOLO11 baseline by +3.4% precision and +3.9% mAP50, while reducing parameters by 22.78% and GFLOPs by 23.81%. Our code will be available at https://github.com/liuyang0304/SDS-YOLO11.

Keywords: Object detection, Marine litter, Underwater detection, Lightweight model, YOLO11

1 Introduction

The rapid growth of global plastic production has made marine plastic pollution one of the most critical environmental challenges of the 21st century [1]. Large amounts of anthropogenic waste enter marine environments through rivers, wind transport, and direct dumping, accumulating in both coastal and deep-sea areas [2]. The long-term persistence of plastic debris not only degrades marine ecosystems but also threatens marine organisms through toxic release and physical entanglement. Therefore, developing efficient technologies for underwater litter mitigation is essential for sustainable ocean management.

Currently, marine underwater waste management still relies heavily on manual diving operations, which are inefficient and risky due to poor visibility and complex underwater conditions. With the advancement of underwater robotics, computer vision-based detection has emerged as a promising solution, offering improved scalability and automation compared to traditional approaches [2].

In recent years, deep learning-based object detection has achieved significant progress, with CNN-based methods such as Faster R-CNN [3] and the YOLO series, as well as Transformer-based approaches like DETR [4], becoming mainstream. However, directly applying these generic detectors to underwater litter detection remains challenging. Underwater litter often suffers from severe appearance degradation caused by long-term erosion, biological attachment, uneven illumination, color distortion, and water turbidity, which significantly reduces detection accuracy [5].

To address these challenges, several underwater litter detection models have been proposed, yet notable limitations persist. CEH-YOLO [6] enhances local feature extraction but struggles to maintain global feature consistency under severe degradation. AIN-YOLO [7] improves adaptive feature extraction but shows limited robustness to occlusion. YOLO-Vortex [8] suppresses background noise at the cost of high computational complexity, while YOLO-MES [9] reduces model size using a lightweight backbone but sacrifices feature representation capability.

To overcome these limitations, we propose SDS-YOLO11, a lightweight and efficient detector based on YOLO11 [10] for marine underwater litter detection. The main contributions of this work are summarized as follows:

1. We propose SDS-YOLO11, a lightweight real-time detector tailored for marine underwater litter detection.
2. We introduce SKLA, a Shared-Key Latent Space Attention module that enhances global feature consistency while suppressing underwater noise interference.
3. We design DuResDown, a Dual-Path Residual Downsampling module that balances lightweight deployment and refined feature extraction.
4. We develop SEAMHead, an improved detection head that enhances robustness under occlusion and complex backgrounds.

Experimental results on the public Waste-9cbfm Marine Underwater Litter Dataset demonstrate that the proposed method outperforms state-of-the-art detectors in both detection accuracy and computational efficiency.

2 Related Work

2.1 Object Detection

Object detection aims to localize and classify objects in images and has evolved significantly with the development of deep learning. Early methods based on handcrafted features such as HOG and Haar-like descriptors [11] exhibited limited robustness in complex environments.

With the success of AlexNet [12], modern object detectors can be broadly divided into two-stage and one-stage paradigms [13]. Two-stage detectors, represented by Faster R-CNN [3] and Mask R-CNN [14], emphasize detection accuracy through a region proposal and refinement pipeline. In contrast, one-stage detectors directly predict object categories and bounding boxes in an end-to-end manner, achieving higher inference efficiency. Among them, the YOLO series has become a dominant framework due to its favorable balance between accuracy and speed [15–17].

Recently, Transformer-based detectors have attracted increasing attention. DETR [4] introduces global self-attention for object detection, while Deformable DETR [18] improves efficiency and small-object detection through deformable attention. To better satisfy real-time requirements, hybrid CNN–Transformer architectures have been explored. Representative models such as YOLO11 adopt this hybrid design to achieve a balanced trade-off between detection performance and computational efficiency, providing a strong foundation for complex visual applications [19].

2.2 Object Detection in Underwater Environments

Underwater object detection is particularly challenging due to image degradation caused by light scattering, color distortion, and suspended particles. Early approaches relied on physics-based

image processing and handcrafted features, such as edge detection and feature matching methods [20]. Image enhancement techniques including white balance and CLAHE [21] can partially improve visual quality but generally lack robustness in complex underwater environments.

With the advancement of deep learning, learning-based methods have become the mainstream solution for underwater object detection. However, generic detectors often suffer from performance degradation when directly applied to underwater scenes. Recent studies mainly focus on enhancing feature discrimination under degraded imaging conditions and designing lightweight models for real-time underwater deployment.

For performance enhancement, attention mechanisms and structural optimizations have been widely adopted. Lv et al. [22] proposed LUOD-YOLO to improve robustness under occlusion and blur, while Ma et al. [23] incorporated area attention and multi-scale feature fusion to enhance multi-class underwater litter recognition. In parallel, lightweight model design has attracted increasing attention. Huang et al. [24] introduced YOLO-MES with a MobileNetV3 backbone, and Niu et al. [25] adopted MobileViT to balance detection performance and computational efficiency on resource-constrained underwater platforms.

3 Methods

3.1 Overall Architecture of SDS-YOLO11

The overall architecture of SDS-YOLO11 is illustrated in Fig. 1, consisting of four main components: the input layer, backbone, neck, and prediction heads. The input layer adopts multi-scale resizing and Mosaic-9 data augmentation to improve robustness under complex underwater conditions.

In the backbone, a dual-path residual structure is employed to preserve fine-grained features while maintaining model compactness. Feature refinement is further enhanced by the C2PSA-SKLA module, which strengthens global target representation and suppresses underwater noise interference. The neck network is built upon an improved PANet architecture and integrates the DuResDown module to effectively preserve multi-scale features of small debris with limited computational overhead.

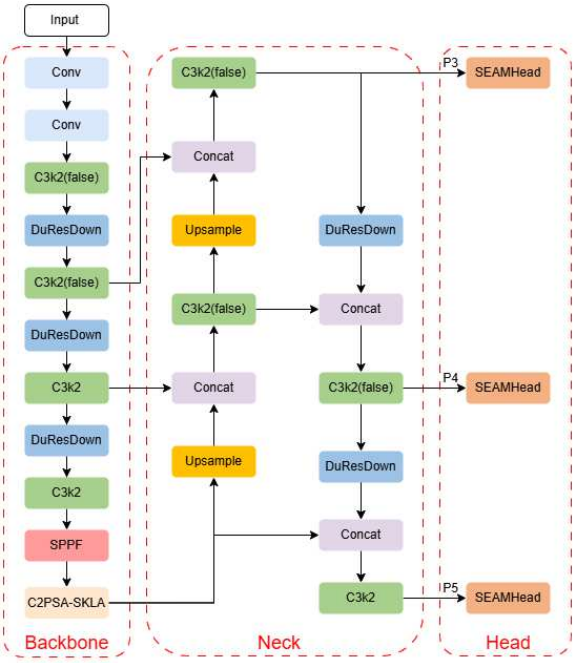

Fig. 1: Network structure of SDS-YOLO11.

Finally, detection is performed by the SEAM-Head, which enhances robustness to occlusion and complex backgrounds through adaptive feature recalibration, enabling reliable classification and localization of underwater litter.

3.2 Improved C2PSA Attention Module with SKLA

To alleviate feature confusion and noise interference in complex underwater litter detection scenarios, we propose an improved C2PSA attention module by integrating a Shared-Key Latent Space Attention (SKLA) mechanism, referred to as C2PSA-SKLA, as illustrated in Fig. 2. Compared with the original C2PSA, the proposed module enhances global feature consistency while suppressing underwater noise with reduced computational overhead.

3.2.1 Global Consistency Enhancement Mechanism Based on Shared Key

In conventional multi-head self-attention mechanisms, each attention head independently learns its own key representation, which may lead to inconsistent attention responses for the same object across different heads. This issue is particularly

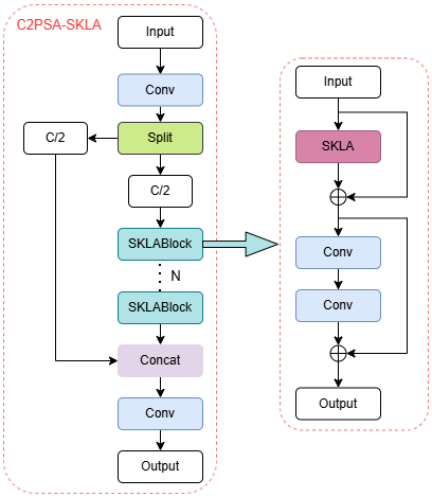

Fig. 2: The structure of C2PSA-SKLA.

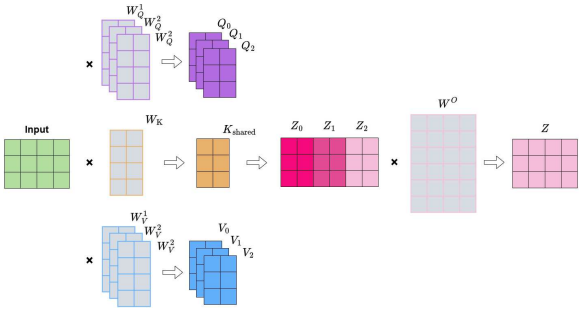

Fig. 3: Overview of Shared Key Attention Method. The shared key attention mechanism shares only one key, and all attention heads calculate the attention weights based on the same key.

detrimental in underwater environments, where targets often suffer from appearance degradation and background interference. Moreover, redundant key projections introduce unnecessary parameters and computational burden, which conflicts with lightweight deployment requirements.

To address these issues, we introduce a Shared-Key Attention (SKA) mechanism inspired by multi-query attention. As illustrated in Fig. 3, SKA enforces all attention heads to share a single Key representation, with all heads calculating the attention weights based on this identical Key within the multi-head architecture.

Given an input feature sequence

$$\mathbf{X} \in \mathbb{R}^{N \times D}, \tag{1}$$

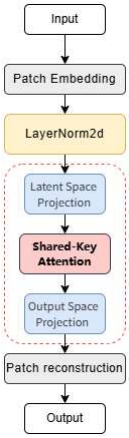

Fig. 4: The structure of SKLA.

the attention output of the h-th head is computed as:

$$\mathrm{SKA}_h = \mathrm{Softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_{\mathrm{shared}}^{\top}}{\sqrt{d}}\right)\mathbf{V}_h, \tag{2}$$

where $\mathbf{Q}_h$ and $\mathbf{V}_h$ denote the query and value matrices of the h-th head, respectively, and all heads share the same key matrix K. By enforcing a shared key across heads, the proposed SKA mechanism ensures consistent global feature alignment and reduces parameter redundancy, leading to more stable attention responses in complex underwater scenes.

3.2.2 Feature Denoising and Compression Strategy in Latent Space

Although the shared-key attention mechanism can enhance the global consistency of attention, irrelevant noise interference such as suspended particles, illumination distortion, and color aberration is prevalent in underwater environments. Directly performing attention calculations in the high-dimensional feature space tends to amplify noise interference and increase computational burden. To address this issue, we further introduce the idea of latent space feature modeling on the basis of shared-key attention and propose the Shared-Key Latent Space Attention (SKLA) mechanism. The overall workflow of SKLA is illustrated in Fig. 4.

In SKLA, the input high-dimensional features $\mathbf{X} \in \mathbb{R}^{N\times D}$ are first projected into a lower-dimensional latent space via a linear mapping:

$$\mathbf{Z} = \mathbf{X}\mathbf{W}_e, \quad \mathbf{W}_e \in \mathbb{R}^{D\times(D/2)}, \tag{3}$$

where D denotes the original feature dimension per position, and the latent dimension is set to $D/2$. This projection acts as a low-rank approximation of the input features, effectively suppressing high-frequency noise and redundancy induced by underwater scattering and illumination variations.

Subsequently, attention is computed in the latent space using the SKA mechanism, followed by a linear projection back to the original feature dimension:

$$\mathbf{Y} = \mathrm{SKA}(\mathbf{Z})\mathbf{W}_d, \quad \mathbf{W}_d \in \mathbb{R}^{(D/2)\times D}. \tag{4}$$

This design not only reduces the computational complexity but also enhances the model's robustness to noise while retaining the key semantic information.

3.3 DuResDown: A Lightweight Dual-Path Residual Downsampling Module

In marine underwater litter detection, the standard stride-2 convolutional downsampling in YOLO11 often leads to insufficient feature preservation and redundant computation under complex backgrounds and severe appearance degradation. To address this issue, we propose a *Dual-Path Residual Downsampling* module, termed *DuResDown*, which balances lightweight design and refined feature extraction.

As illustrated in Fig. 5, DuResDown adopts a dual-path parallel architecture with a residual connection. Given an input feature map $F \in \mathbb{R}^{H\times W\times C}$, average pooling is first applied to suppress high-frequency noise while preserving global spatial structure. The resulting feature map is then evenly split along the channel dimension into two subsets.

The first branch applies max pooling followed by a 1×1 convolution to emphasize locally salient features with low computational cost. The second branch employs a stride-2 3×3 convolution to perform spatial downsampling and feature transformation, enabling effective receptive field expansion.

The outputs of the two branches are concatenated and fused with the input feature map through a residual connection:

$$F_{\text{out}} = \mathrm{Concat}(F_1, F_2) + F, \tag{5}$$

where F_1 and F_2 denote the feature maps produced by the two branches. The residual connection helps preserve original feature information and stabilizes training.

By jointly integrating pooling operations and convolutional downsampling, DuResDown effectively captures both global contextual information and local discriminative features with reduced parameter redundancy, making it suitable for real-time underwater litter detection on resource-constrained platforms.

3.4 Improved YOLO11 Head with SEAM Module

Underwater litter targets are frequently affected by partial occlusion, feature misalignment, and background interference, which often lead to degraded detection performance in standard detection heads. To address these challenges, we incorporate the Separation and Enhancement Attention Module (SEAM) into the YOLO11 detection head, forming an improved SEAMHead.

As illustrated in Fig. 6, SEAM adopts a channel–spatial attention mechanism that emphasizes informative regions while suppressing irrelevant background responses. In the proposed SEAMHead, multi-scale features from different detection layers are first refined by SEAM to enhance feature discriminability under occlusion. Unlike conventional attention mechanisms that apply linear normalization, SEAM employs an *exponential normalization strategy* to amplify critical responses corresponding to occluded or low-contrast targets, which is particularly beneficial in underwater environments.

Formally, given an input feature map F, the attention weights are computed as

$$A = \exp\big(\phi(F)\big), \tag{6}$$

where $\phi(\cdot)$ denotes the channel–spatial attention function learned by SEAM. The refined feature

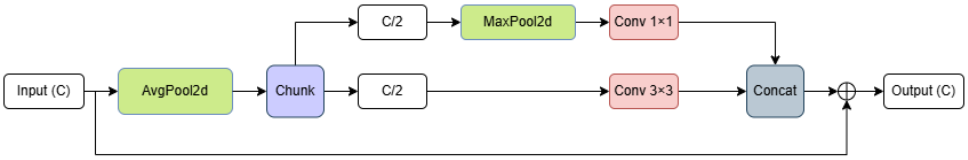

Fig. 5: The structure of DuResDown.

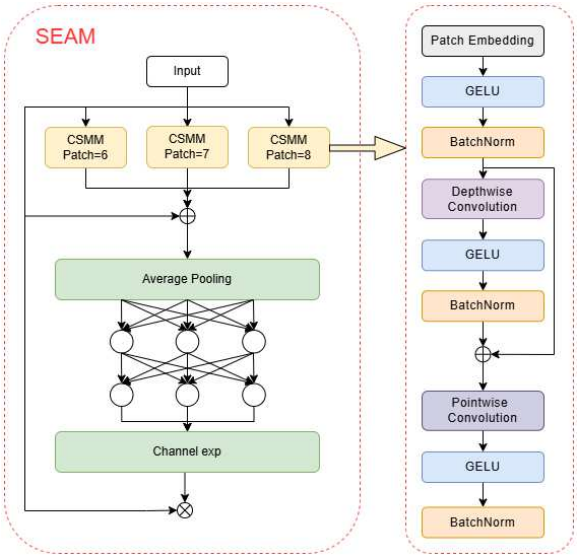

Fig. 6: Network structure of SEAM attention mechanism.

map is then obtained by element-wise multiplication:

$$F' = A \odot F. \tag{7}$$

By mapping attention weights from the range $[0, 1]$ to $[1, e]$, the exponential normalization enlarges the response gap between informative and non-informative regions, effectively enhancing the representation of partially occluded litter while suppressing background noise. Integrating SEAMHead into YOLO11 introduces minimal computational overhead while significantly improving robustness against occlusion and complex background interference, thereby enhancing detection accuracy in challenging underwater scenarios.

4 Experiments and Discussion

4.1 Dataset

The proposed SDS-YOLO11 is evaluated on the public Waste-9cbfm Marine Underwater Litter Dataset [26], which contains 4,528 underwater images annotated into 13 litter categories, including bottles, plastic bags, nets, and other common debris. The dataset is divided into training, validation, and test sets with a ratio of 70%, 20%, and 10%, respectively.

4.2 Evaluation Metrics

Model performance is evaluated using standard object detection metrics, including precision (P), recall (R), mean average precision (mAP), model parameters (Params), and floating-point operations (GFLOPs). Detection accuracy is primarily assessed by mAP_{50} and $\text{mAP}_{50:95}$, corresponding to an IoU threshold of 0.50 and the average over thresholds from 0.50 to 0.95.

4.3 Implementation Details

The experiments were conducted on a computer system running the Ubuntu 20.04 operating system. The hardware configuration comprised an Intel Core i5-9400F CPU, 16 GB of system RAM, and an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB of VRAM. The specific experimental parameters were set as follows: learning rate of 0.01, momentum coefficient of 0.937, L2 regularization coefficient of 0.0005, image size of 640×640, training epochs of 300, and the SGD optimizer was adopted.

4.4 Analysis of Experimental Results

4.4.1 YOLO11 Experiment

YOLO11 provides multiple model variants through scalable width and depth coefficients to support different deployment requirements. In this study, all five YOLO11 variants are evaluated on the Waste-9cbfm Marine Underwater Litter Dataset, with detailed results reported in Table 1. Among them, YOLO11 achieves a favorable balance between detection accuracy and computational efficiency,

Table 1: The experimental results of the five different scaled models of YOLO11.

Method	Precision(%)	Recall(%)	mAP50(%)	mAP50:95(%)	P (M)	GFLOPs
YOLO11x	80.3	76.4	80.8	51.1	184.4	54.5
YOLO11l	84.1	74.2	80.2	51.8	82.3	24.3
YOLO11m	85.4	79.4	83.6	52.9	65.5	19.5
YOLO11s	85.6	77.2	82.1	53.6	21.3	9.2
YOLO11	**85.1**	**74.3**	**80.3**	**52.3**	**2.59**	**6.3**

Table 2: The results of the ablation experiment.

Baseline	C2PSA-SKLA	DuResDown	SEAMHead	Precision(%)	Recall(%)	mAP50(%)	mAP50:95(%)	P (M)	GFLOPs
				85.1	74.3	80.3	52.3	2.59	6.3
	✓			83.7	75.9	82.6	53.3	2.57	6.3
		✓		82.9	77.7	82.1	52.8	2.11	5.3
			✓	86.9	74.4	82.3	53.0	2.50	5.8
YOLO11	✓	✓		86.2	75.6	82.5	53.6	2.09	5.3
	✓		✓	85.4	77.4	83.5	53.9	2.47	5.8
		✓	✓	81.0	77.3	82.0	53.4	2.01	4.8
	✓	✓	✓	**88.5**	**76.5**	**84.2**	**53.7**	**2.00**	**4.8**

exhibiting the smallest model size and fastest inference speed. Owing to its compact architecture and suitability for resource-constrained platforms, YOLO11 is selected as the baseline model for subsequent optimization.

4.4.2 Ablation Experiment

In this section, ablation experiments were conducted on the key components of the proposed model using the Waste-9cbfm Marine Underwater Litter Dataset. Table 2 presents the ablation experiment results, including precision, recall, mAP50, mAP50:95, parameters, and GFLOPs.

C2PSA-SKLA: Replacing the original C2PSA module in the YOLO11 backbone with the proposed C2PSA-SKLA consistently improves detection accuracy while slightly reducing model parameters. This demonstrates that the introduced SKLA module effectively enhances global feature consistency and suppresses irrelevant underwater noise.

DuResDown: Introducing DuResDown into the backbone and neck layers leads to a favorable balance between detection performance and computational efficiency. As shown in Table 2, both mAP metrics are improved while model parameters and GFLOPs are significantly reduced, validating the effectiveness of the dual-path residual downsampling design.

SEAMHead: Substituting the original detection head with SEAMHead further improves detection performance while reducing model complexity. The results indicate that SEAMHead enhances robustness under occlusion and complex backgrounds, contributing positively to overall detection accuracy.

4.4.3 Comparison of Different Detectors

To evaluate the effectiveness of SDS-YOLO11, comparative experiments are conducted against representative general-purpose and underwater-specific object detection models under identical experimental settings. The evaluation focuses on mAP_{50}, $mAP_{50:95}$, and model parameters to assess detection accuracy and lightweight performance.

The compared methods include representative YOLO-series lightweight models, classical general-purpose detectors such as Faster R-CNN [3] and RT-DETR [32], as well as specialized underwater detection models, including DSDebrisNet [33], CEH-YOLO [6], and LUOD-YOLO [22]. These models provide comprehensive baselines across general and underwater-specific scenarios.

As reported in Table 3, SDS-YOLO11 achieves superior detection accuracy compared with all baseline methods while maintaining a significantly reduced parameter count, demonstrating an improved balance between performance and

Table 3: Performance comparison of different methods.

Method	mAP50(%)	mAP50:95(%)	P (M)	GFLOPs
YOLOv5n [16]	80.6	50.4	2.51	7.1
YOLOv6n [27]	77.3	47.1	4.31	11.8
YOLOv7-tiny [28]	79.1	45.1	6.01	13.2
YOLOv8n [29]	80.8	50.8	3.01	8.1
YOLOv10n [30]	81.4	51.7	2.70	8.3
YOLO11 [10]	80.3	52.3	2.59	6.3
YOLOv12n [31]	83.6	51.6	2.51	5.8
Faster R-CNN(ResNet50) [3]	73.5	43.0	41.22	210.3
RT-DETR(ResNet50) [32]	78.5	48.9	42.0	125.7
DSDebrisNet [33]	79.1	47.4	7.05	15.9
CEH-YOLO [6]	82.0	51.6	3.44	8.4
LUOD-YOLO [22]	74.3	47.0	1.63	5.9
SDS-YOLO11 (Ours)	**84.2**	**53.7**	**2.00**	**4.8**

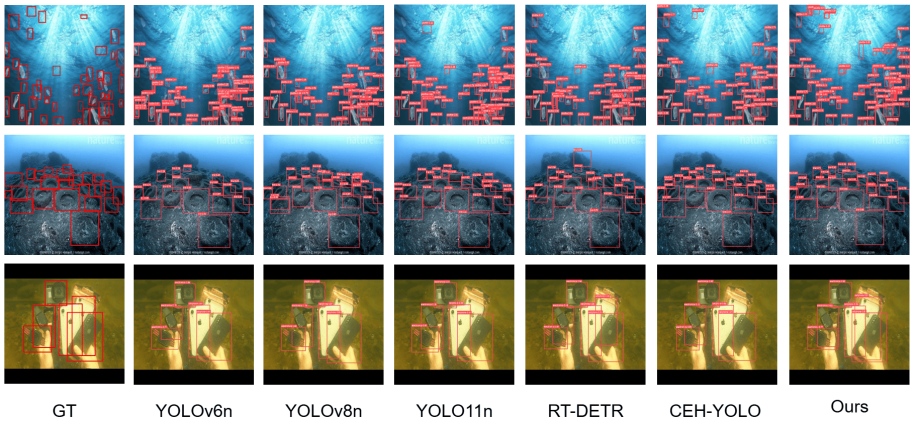

Fig. 7: The visualized detection results of different networks on Waste-9cbfm Marine Underwater Litter Dataset.

efficiency. Qualitative comparisons in Fig. 7 further illustrate that the proposed method can more accurately localize underwater litter and detect targets missed by other models, highlighting its robustness in complex underwater environments and suitability for real-time deployment on resource-constrained underwater platforms.

5 Conclusion

Marine underwater litter detection plays an important role in marine environmental protection and ecological restoration, yet remains challenging due to severe noise interference, feature degradation, and deployment constraints in complex underwater environments. To address these issues, this paper proposes SDS-YOLO11, a lightweight object detection framework specifically designed for underwater litter detection.

By enhancing global feature representation, preserving multi-scale information, and improving robustness under occlusion, SDS-YOLO11 effectively balances detection accuracy and computational efficiency. Extensive experiments on the Waste-9cbfm Marine Underwater Litter Dataset demonstrate that the proposed method outperforms existing general-purpose and underwater-specific detectors while maintaining a compact model size.

These results indicate that SDS-YOLO11 provides an efficient and reliable solution for real-time underwater litter detection and offers strong potential for practical deployment in automated underwater cleanup systems.

Author Contributions Yang Liu: Writing – original draft, Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation,

Formal analysis, Data curation, Conceptualization. Chong Fu: Writing – original draft, Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. Junxin Chen: Funding acquisition, Project administration. Xingwei Wang: Funding acquisition, Resources, Project administration.

Funding This research is supported by the National Natural Science Foundation of China (Nos. 62171114 and 62032013), the Fundamental Research Funds for the Central Universities, China (Nos. N2424010-18 and N25LPY007), and the Liaoning Provincial Science and Technology Plan Project (2023JH2/101700370).

Data Availability Data will be made available on request.

Declarations

Conflict of interest The authors declare no competing interests.

References

[1] Balamurugan, S., Kumaran, S., Sadasivuni, K., Uravakonda, V., Rao, R.B.S., Xavier, C.S.C.: Integration of deep learning in real-time monitoring and alert system for evaluating plastic pollution levels in oceanic regions. In: AIP Conference Proceedings, vol. 3259, p. 040002 (2025). AIP Publishing LLC

[2] Ma, D., Wei, J., Li, Y., Zhao, F., Chen, X., Hu, Y., Yu, S., He, T., Jin, R., Li, Z., *et al.*: Mldet: Towards efficient and accurate deep learning method for marine litter detection. Ocean & Coastal Management **243**, 106765 (2023)

[3] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)

[4] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European Conference on Computer Vision, pp. 213–229 (2020)

[5] Fulton, M., Hong, J., Islam, M.J., Sattar, J.: Robotic detection of marine litter using deep visual detection models. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 5752–5758 (2019). IEEE

[6] Feng, J., Jin, T.: Ceh-yolo: A composite enhanced yolo-based model for underwater object detection. Ecological Informatics **82**, 102758 (2024)

[7] He, X., Zhang, Y., Zhan, Q.: Ain-yolo: A lightweight yolo network with attention-based inceptionnext and knowledge distillation for underwater object detection. Advanced Engineering Informatics **66**, 103504 (2025)

[8] Chang, L., Yuan, C., Yang, Q.: Yolo-vortex: Underwater object detection based on vortex aggregation network. Computer Science and Application **15**, 57 (2025). In Chinese

[9] Huang, C., Zhang, W., Zheng, B., Li, J., Xie, B., Nan, R., Tan, Z., Tan, B., Xiong, N.N.: Yolo-mes: An effective lightweight underwater garbage detection scheme for marine ecosystems. IEEE Access (2025)

[10] Jocher, G., Qiu, J., et al.: Ultralytics YOLO11. https://github.com/ultralytics/ultralytics

[11] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886–893 (2005). Ieee

[12] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

[13] Olorunshola, O., Jemitola, P., Ademuwagun, A.: Comparative study of some deep learning object detection algorithms: R-cnn, fast r-cnn, faster r-cnn, ssd, and yolo. Nile J. Eng. Appl. Sci **1**(1), 70–80 (2023)

[14] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)

[15] 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, pp. 779–788 (2016)

[16] Jocher, G., et al.: YOLOv5. GitHub repository. [Online; accessed day-month-year] (2023). https://github.com/ultralytics/yolov5

[17] Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 (2021)

[18] Zhu, X., Su, W., Yu, L., Wang, B., Fang, Z., Yu, X., Dai, Y., Chen, Y., Yuan, J., Wei, S., *et al.*: Deformable DETR: Deformable transformers for end-to-end object detection. In: International Conference on Learning Representations (2021)

[19] Khan, A., Rauf, Z., Sohail, A., Khan, A.R., Asif, H., Asif, A., Farooq, U.: A survey of the vision transformers and their cnn-transformer based variants. Artificial Intelligence Review **56**(Suppl 3), 2917–2970 (2023)

[20] Fatan, M., Daliri, M.R., Shahri, A.M.: Underwater cable detection in the images using edge classification based on texture information. Measurement **91**, 309–317 (2016)

[21] Sanila, K., Balakrishnan, A.A., Supriya, M.: Underwater image enhancement using white balance, usm and clhe. In: 2019 International Symposium on Ocean Technology (SYMPOL), pp. 106–116 (2019). IEEE

[22] Lv, C., Pan, W.: Luod-yolo: a lightweight underwater object detection model based on dynamic feature fusion, dual path rearrangement and cross-scale integration. Journal of Real-Time Image Processing **22**(6), 204 (2025)

[23] Ma, J., Zhou, Y., Zhou, Z., Zhang, Y., He, L.: Toward smart ocean monitoring: Real-time detection of marine litter using yolov12 in support of pollution mitigation. Marine Pollution Bulletin **217**, 118136 (2025)

[24] Huang, C., Zhang, W., Zheng, B., Li, J., Xie, B., Nan, R., Tan, Z., Tan, B., Xiong, N.N.: Yolo-mes: An effective lightweight underwater garbage detection scheme for marine ecosystems. IEEE Access (2025)

[25] Niu, J., Gu, S., Du, J., Hao, Y.: Underwater waste recognition and localization based on improved yolov5. Computers, Materials & Continua **76**(2) (2023)

[26] detect: waste Dataset. Roboflow. visited on 2025-03-12 (2022). https://universe.roboflow.com/object-detect-dmjpt/waste-9cbfm

[27] Li, C., et al.: Yolov6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976 (2022)

[28] Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7464–7475 (2023)

[29] Jocher, G., et al.: YOLOv8. GitHub repository. [Online; accessed day-month-year] (2023). https://github.com/ultralytics/ultralytics

[30] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., *et al.*: Yolov10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems **37**, 107984–108011 (2024)

[31] Tian, Y., Ye, Q., Doermann, D.: Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524 (2025)

[32] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: Detrs beat yolos on real-time object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16965–16974 (2024)

[33] Huang, B., Chen, G., Zhang, H., Hou, G., Radenkovic, M.: Instant deep sea debris detection for maneuverable underwater machines to build sustainable ocean using deep neural network. Science of The Total Environment **878**, 162826 (2023)