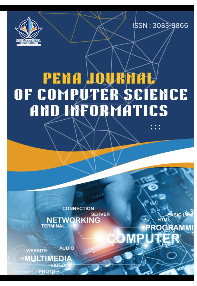
PENA CENDEKIA PUBLISHING
NURTURING IDEAS, ENRICHING KNOWLEDGE

Pena Journal of Computer Science and Informatics

Journal homepage:
https://penacendekia.com.my/index.php/pjcsi/index
ISSN: 3083-9866



MDE-YOLOv11: A Real-Time Detection Model for Small and Occluded Objects in Underwater Environments

Alaa Hussien[1,*], Sahar Magdy[2]

[1] Computer Engineering Department, Pharos University, Alexandria, Egypt

ARTICLE INFO

Article history:
Received 5 July 2025
Received in revised form 11 August 2025
Accepted 8 September 2025
Available online 18 September 2025

Keywords:
Underwater object detection; Yolov11; deep learning; RUOD dataset; URPC2020 dataset

ABSTRACT

Underwater object detection (UOD) is critical for applications ranging from marine life monitoring and pollution assessment to autonomous underwater navigation. However, detecting objects in underwater environments is highly challenging due to poor visibility, light distortion, and the presence of small, occluded, or camouflaged targets. Traditional object detection models, trained on terrestrial datasets like MS COCO, struggle to generalize to underwater environments. This paper proposes MDE-YOLOv11, an enhanced version of the YOLOv11 model tailored for UOD. The model incorporates Multi-Scale Dilated Attention (MSDA) modules to enhance contextual feature extraction, Deformable Convolutional Networks (DCN) to improve adaptability to irregular object shapes, and Exponential Linear Units (ELU) for better feature preservation in low-contrast conditions. The proposed model was evaluated on two benchmark datasets RUOD and URPC dataset. MDE-YOLOv11 achieves significant improvements, with mAP@50 of 90.6% on RUOD (10.9% over baseline YOLOv11n) and 76.8% on URPC2020 (1.4% over baseline). The model maintains near-real-time performance making it suitable for real-time AUV applications. These enhancements address UOD challenges like small object detection, occlusions, and environmental noise, demonstrating superior accuracy and robustness

1. Introduction

Oceans and seas are one of Earth's most expansive ecosystems, home to an incredible variety of life and playing a major role in shaping the planet's climate and environmental balance. Thus, Underwater Object Detection (UOD) has received attention lately [1]. Many applications nowadays depend on UOD and its precision, from studying marine life such as tracking endangered species, monitoring coral reef health, and detecting illegal fishing activities [2]. Moreover, exploring the ocean floor to detecting pollution and enabling safe navigation for autonomous underwater vehicles (AUVs), pipeline inspection, and wreckage detection for offshore industries [3]. Also, UOD could be important in Locating submerged objects or drowning victims in turbid waters [2]. Furthermore, UOD

* *Corresponding author.*
E-mail address: alaa.hussien@pua.edu.eg

https://doi.org/10.37934/pjcsi.2.1.1122

is vital in mine detection, submarine surveillance, and underwater drone reconnaissance. Without advancements in this field, many of these tasks would remain inefficient, costly, or even dangerous for human divers [4].

However, constructing a robust and accurate underwater detection system can be significantly challenging. For example, Light absorption and scattering in water lead to low contrast, blurring, and color distortion (dominant blue-green hues) which causes poor visibility to computer vision deep learning models [5,6]. Additionally, distortions like suspended particles, bubbles, and varying light conditions degrade image quality. Moreover, Marine creatures like fish and corals might appear small, or occluded especially when marine creatures overlap. Also, models trained in one underwater environment (e.g., clear tropical waters) may fail in another (e.g., murky riverbeds) [3]. Most traditional model of object detection algorithms are evaluated on benchmark datasets like MS COCO and PASCAL VOC which are primarily composed of images from terrestrial scenes, not underwater environments [7]. These general-purpose datasets cover a wide range of object types and scenarios but do not address the unique challenges posed by underwater imagery [8].

These conditions not only affect human perception but also severely impact the performance of traditional computer vision algorithms making the creation of reliable object detection systems specifically suited for underwater environments necessary [9].

In recent years, deep learning-based detectors, particularly those from the YOLO (You Only Look Once) family, have demonstrated impressive performance in real-time object detection tasks [10]. However, applying these models directly to underwater environments often leads to suboptimal results due to the domain gap between terrestrial and underwater imagery [11]. Therefore, adapting and enhancing state-of-the-art detectors like YOLOv11 for underwater scenarios is a crucial step toward improving detection accuracy and robustness in complex aquatic conditions.

In this paper, A modified architecture MDE-YOLOv11 is introduced which contains three key different modifications:

(1) integrating Multi-Scale Dilated Attention (MSDA) after the second and third C3k2 modules to enhance multi-scale feature extraction

(2) replacing standard convolutions with Deformable Convolutional Networks (DCN) in C3k2 bottlenecks to adapt to irregular object shapes

(3) substituting SiLU with ELU activation in Conv blocks to improve feature sensitivity in low-contrast scenes.

This research paper is arranged as follows, section 2 shows the methodology of the proposed model, the algorithms and techniques used in the MDE-YOLOv11 algorithm. Section 3 demonstrated the experimental results and ended with the conclusion in section 4.

2. Methodology

2.1 Proposed Model

The proposed model is using YOLOv11, which is an anchor free detection model, as a base since its architecture prioritizes speed, making it suitable for real-time underwater applications (e.g., ROVs, AUVs, or diver assistance systems). Moreover, advanced feature pyramid networks (FPN/PANet) in YOLOv11 might better detect objects at varying scales (e.g., small fish vs. large shipwrecks) [12].

2.1.2 YOLOv11

YOLOv11 is composed of three modules, backbone, neck and head. These modules work together to process input images, extract features, and generate predictions for bounding boxes, class labels, and confidence scores. Its design emphasizes efficiency, reduced parameter count, and enhanced feature extraction, which are critical for challenging environments like underwater settings with low visibility, complex backgrounds, and small objects [13].

YOLOv11's backbone is responsible for extracting multi-scale features from input images. It starts with standard convolutional downsampling layers to reduce spatial resolution and increase feature depth followed by 2D batch normalization and a Sigmoid Linear Unit (SiLU) activation function. In underwater environments, where images suffer from low contrast and color distortion, the Conv Block's robust feature extraction helps capture essential patterns despite noise or turbidity. The backbone module also contains C3k2 layer which is a redefined version of the CSP bottleneck from YOLOv8, replacing the heavier C2f block. C3k2 layers has faster computation and fewer parameters, while retaining strong representational power [13]. The C3k2 block's focus on small object detection is highly relevant for UOD, where targets like fish or marine debris are often small and occluded. Stack of C3k2 modules enhances efficient local feature extraction, followed by SPPF (Spatial Pyramid Pooling Fast) pooling and the C2PSA (Convolutional Block with Parallel Spatial Attention) attention module to refine and focus features before passing them to the neck [14]. SPPF improves robustness to objects of different sizes and crucial for detecting both small and large underwater targets. The SPPF block's ability to handle multi-scale features is vital for underwater scenarios with varied object sizes and complex textures. Moreover, the C2PSA utilizes two Partial Spatial Attention (PSA) modules on separate feature map branches, which are later concatenated. This refines the model's focus on regions of interest [12]. The C2PSA block's spatial attention is particularly effective for underwater scenarios, where targets like fish or debris may blend into complex backgrounds. The backbone generates three feature maps at different scales, corresponding to fine, medium, and coarse granularity. These maps are critical for detecting objects of varying sizes, from small marine organisms to larger underwater structures [15].

The neck aggregates feature from the backbone's multi-scale feature maps to improve detection across different object sizes and orientations, while the head generates final predictions, including bounding boxes, class labels, and confidence scores [12].

The unique architecture of YOLOv11 makes it suitable for underwater object detection, however, most YOLO models are trained on COCO or VOC datasets, which lack underwater-specific objects (e.g., corals, sonar targets). Moreover, YOLOv11 struggles with hazy, distorted, or color-shifted images (blue-green dominance) since C3k2 lacks built-in attention mechanisms to help focus on relevant regions or suppress noise (e.g., bubbles, particles, lighting distortions). Additionally, underwater scenes often have dense, overlapping objects (e.g., schools of fish) so YOLOv11 may miss small or occluded objects without specialized post-processing, since features extracted by C3k2 may be insufficiently refined for high-confidence detection, especially for small or camouflaged objects.

Thus, A modified version of YOLOv11 (MDE-YOLOv11) is proposed as shown in figure 1. The modified YOLOv11 not only enhances the YOLOv11 detection precision for underwater scenes, but also maintains the low computational cost of the YOLOv11 and the speed of the model.

2.1.2 Improved YOLOv11 Architecture (MDE-YOLOv11)

MDE-YOLOv11 proposed in figure 1 enhances the backbone and the neck modules to improve the performance of the underwater object detection of small and occluded objects. The following enhancements are made:

1) Multi-scale dilated attention mechanism (MSDA), is integrated in the backbone module to enhance features extraction at different spatial resolutions, correcting C3k2's shallow multi-scale representation.

2) Replacing Sigmoid Linear Unit (SiLU) activation with Exponential Linear Unit (ELU) activation function in the Conv layer which is much simpler and less complex compared to SiLU.
3) Improving the C3k2 layer's bottleneck by integerating Deformable Convolutional Network (DCN) module to adapt across diverse underwater conditions (lighting, turbidity, current movement). Deformable conv ensures flexible spatial understanding even after downsampling in the bottleneck layer where information is distilled.
4) Modifying the C3k2 layer by replacing SiLU activation function with SiLU activation function to preserve subtle features that SiLU might suppress due to its bounded negative region since underwater edges are often blurry or low in contrast.

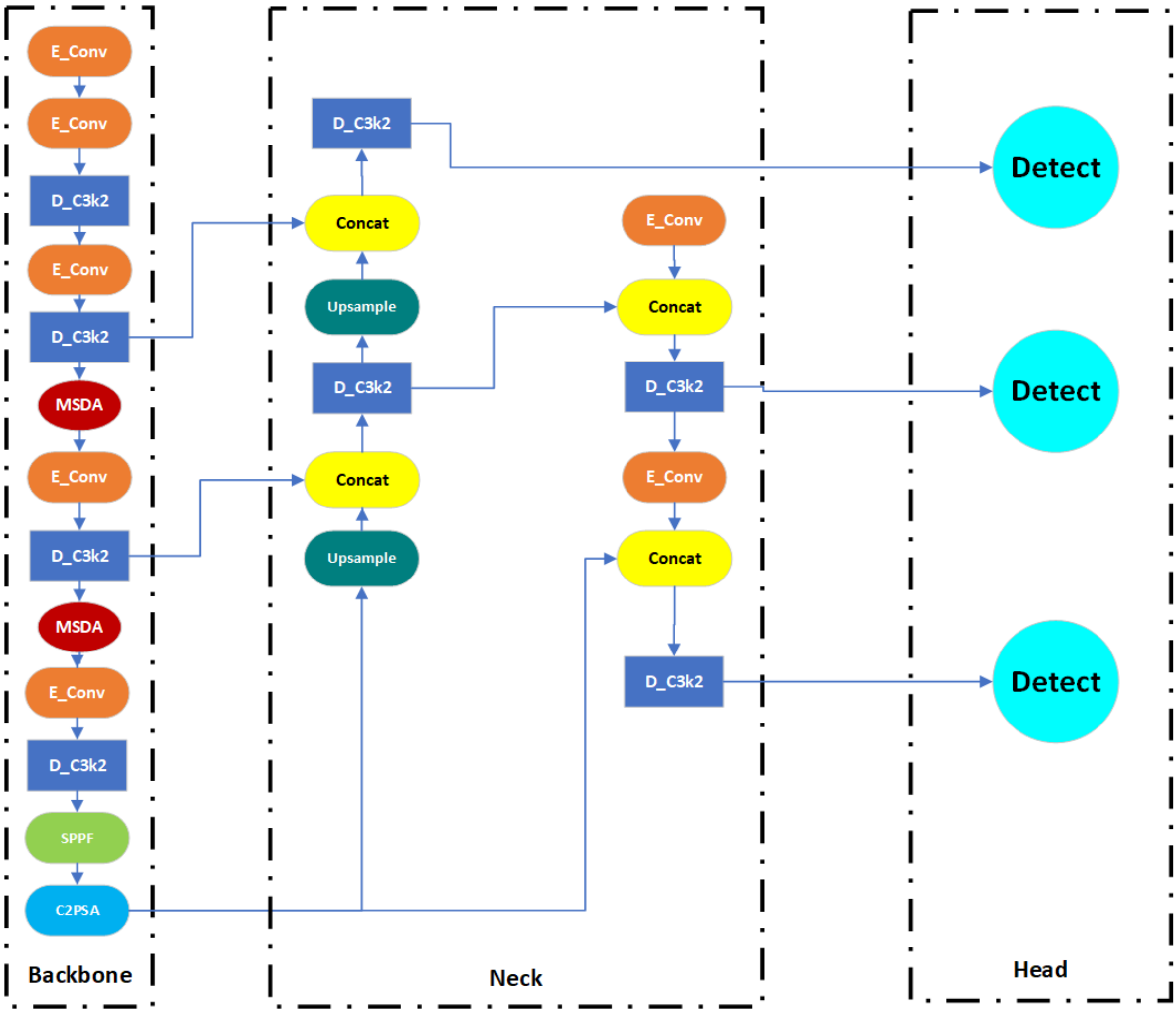


Fig. 1. Modified Yolov11 architecture

2.1.2.1 Multi-scale dilated attention mechanism (MSDA)

MSDA is designed to enhance the feature extraction mechanism by combining dilated convolutions with an attention module focusing on for capturing multi-scale contextual information. Dilated convolutions utilize varying dilation rates to expand the receptive field without increasing the number of parameters [16]. This allows the network to capture features at multiple scales, making it effective for detecting objects of different sizes, especially small ones. MSDA is particularly suited for tasks like underwater object detection, where objects vary in size, visibility is low, and backgrounds are complex due to water turbulence, lighting variations, and occlusions [17]. Moreover, it acts like a visual filter that helps the model *see through the noise* and *adapt to distortions* that traditional convolutional blocks struggle with [18].

As shown in figure 1, MSDA module is added after the second and the third C3k2 module only to not add complexity to the original YOLOv11. C3k2 is lightweight and efficient but lacks spatial flexibility and context awareness. Adding MSDA after C3k2 amplifies weak or distorted signals missed by the C3K2 layer compensating the lack of global context, especially in underwater scenes with floating particles, caustics, or occlusions. Moreover, it clarifies the spatial attention, allowing the network to “zoom in” on important regions adaptively.

MSDA is employed after the 2nd and the 3rd blocks in the YOLOv11 which are the middle and deep stages of the backbone which extract mid and high-level semantic features, so MSDA is best used at those places to enhance the feature extraction.

2.1.2.2 Deformable Convolutional Network (DCN)

DCN is an advanced convolution network that learns by adaptive sampling locations. Unlike regular convolutions, which use fixed, regular sampling grids (e.g., 3x3 kernels), deformable convolutions dynamically adjust the sampling positions of the convolution kernel based on the input features, making it particularly effective for handling geometric variations and complex object shapes [19]. In the proposed MDE-YOLOv11*,* the deformable conv replaces the conv layer in the bottleneck of the C3k2 block as shown in figure 2 which enhances the YOLOv11 ability to handle and adapt to objects at various scales or irregular shapes (e.g., jellyfish tentacles, coral branches) which standard Conv struggles with.

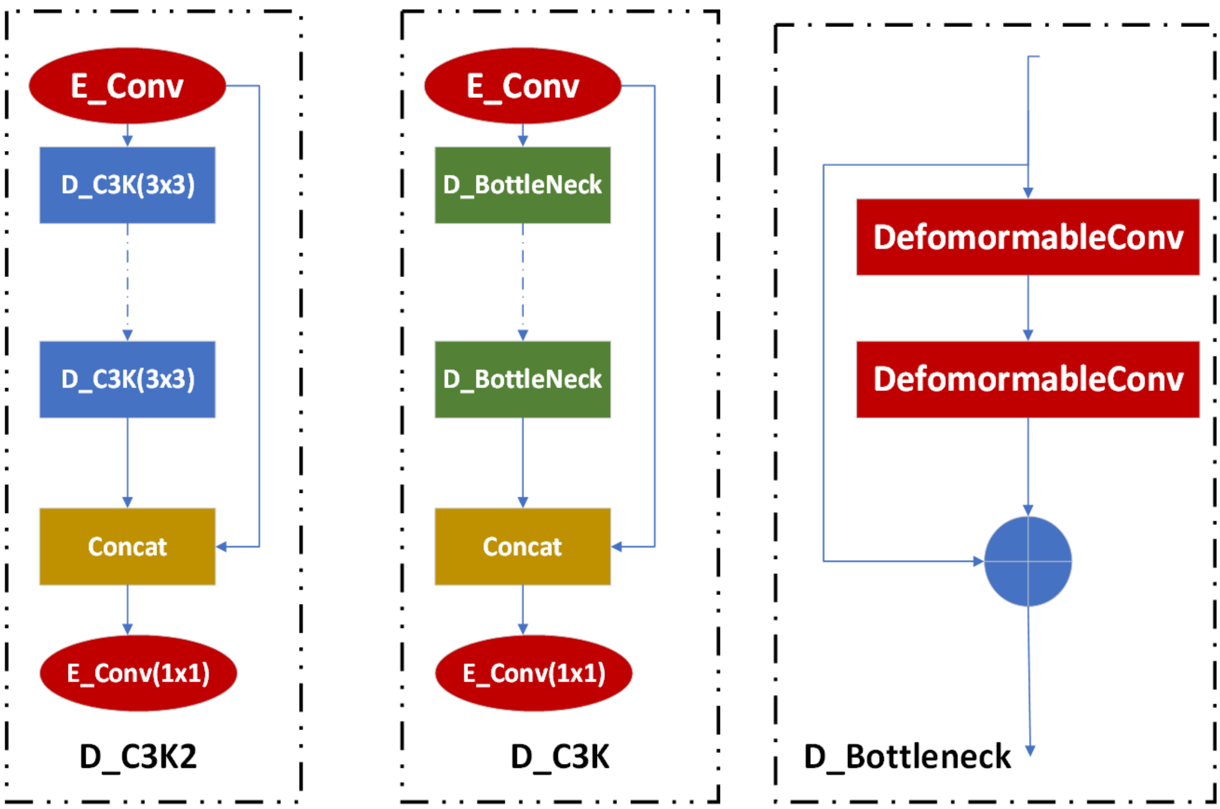


Fig. 2. Modified C3k2 architecture

2.1.2.3 Exponential Linear Unit (ELU) activation function

In challenging environments like underwater object detection with low contrast scenes and high noise, ELU perform better than SiLU presenting hight accuracy and robustness to the model [20]. As shown in figure 3, ELU replaced the SiLU in the Conv layer affecting the backbone and the neck modules. ELU preserves negative activations and positive which improves the feature extraction and helps with blurry and distorted objects presenting a more robust model. Moreover, ELU is faster and less complex than SiLU.

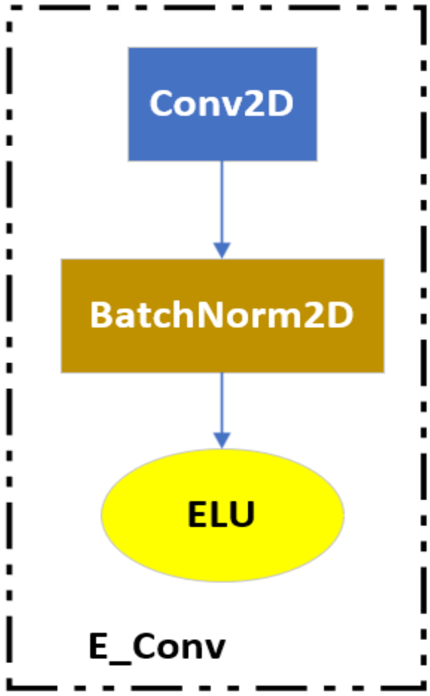


Fig. 3. Modified conv architecture

3. Evaluation and tests

3.1 Datasets

1) Real-world Underwater Object Detection (RUOD) Dataset

Is a benchmark dataset for underwater object detection, designed to address challenges like low visibility, light scattering, occlusions, and distorted shapes in marine environments [21]. It consists of 14,000 high-resolution images typically divided into 9,800 training images and 4,200 validation images. It also contains 74,903 labelled instances across 10 common aquatic categories: fish, sea urchin, coral, starfish, sea cucumber, scallop, diver, cuttlefish, turtle, and jellyfish. Unlike synthetic datasets, RUOD includes natural turbidity, lighting variations, and occlusions.

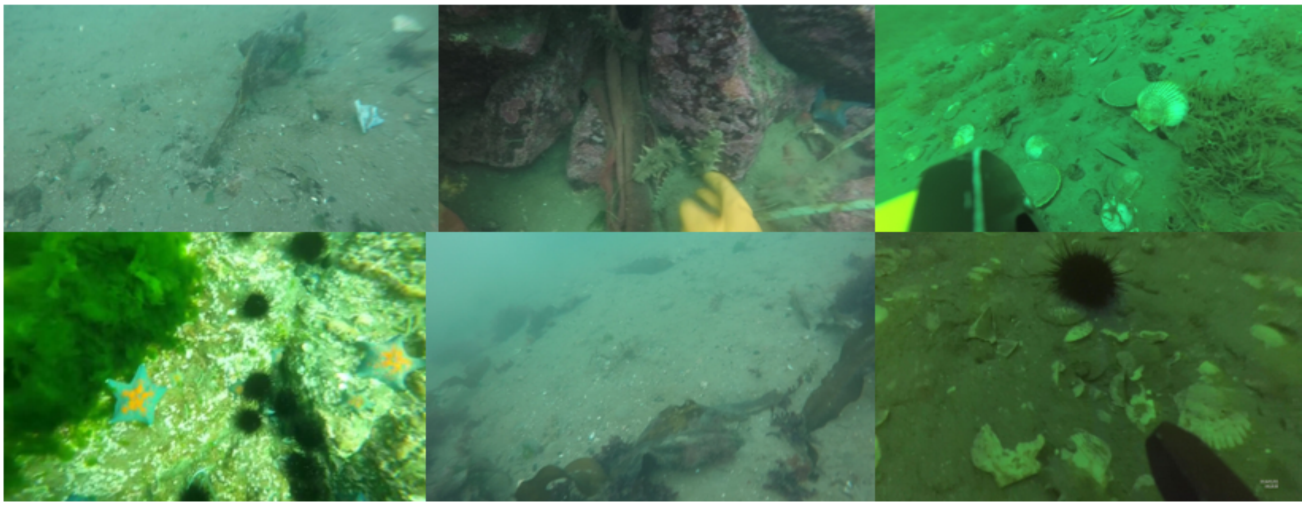


Fig. 4. Samples of RUOD dataset

2) URPC2020 Dataset

A popular benchmark dataset for underwater object detection, it focuses on four key categories of underwater organisms (Echinus (sea urchin), Starfish, Holothurian (sea cucumber), Scallop) [22]. The dataset consists of 6200 images composed to 4200 images training set, 800 validation images and 1200 test images.

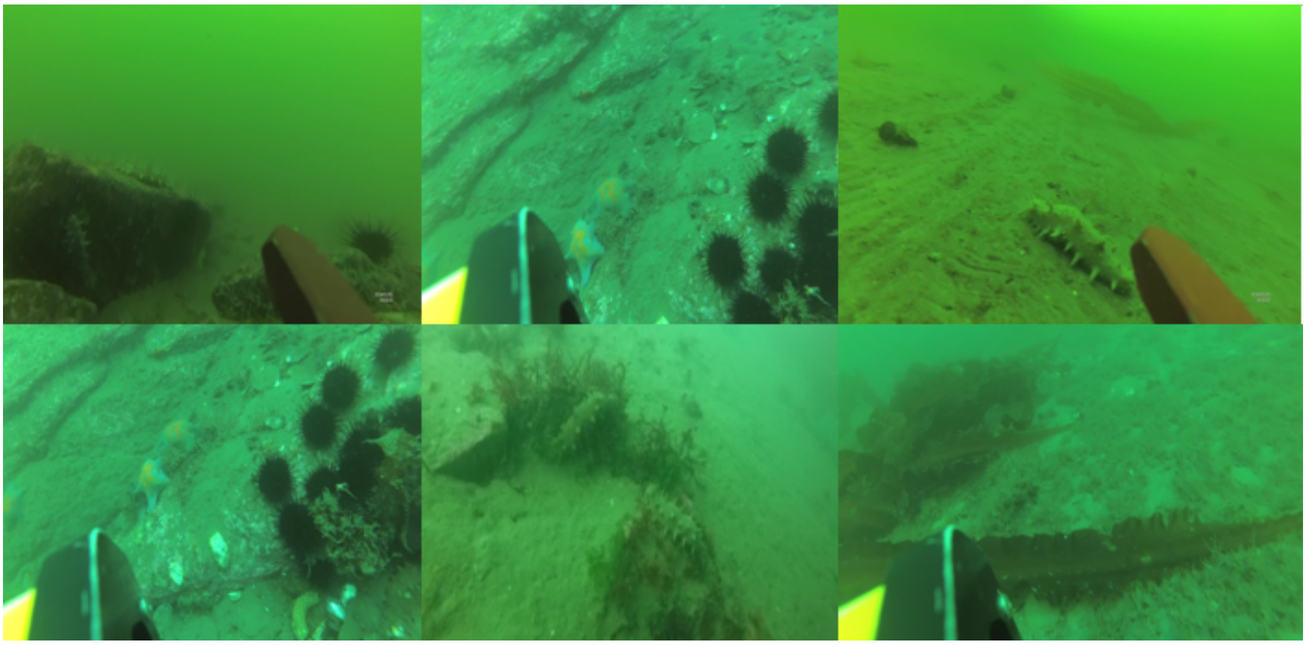


Fig. 5. Samples of URPC2020 dataset

3.2 Evaluation Metrics

The proposed model is evaluated using precision, recall and mAP (Mean Average Precision) using the following equations, where precision as shown in Eq. (1) is the percentage of correct predictions among all predicted positives of the detected underwater objects, while recall demonstrated in Eq. (2) is the percentage of correct predictions of among all ground-truth positives. The performance of the model is also tested using mAP shown in Eq. (3) and Eq. (4) which evaluates how accurate does the model local and classifies the object correctly.

$$\text{Precision } = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall } = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 \text{precision}(t) dt \tag{3}$$

$$mAP = \frac{\sum_{n=1}^{N} AP_n}{N} \tag{4}$$

3.3 Experimental Results

The proposed model MDE-YOLOv11 was evaluated through the two datasets RUOD and URPC2020 dataset to generalize the model and validate its performance and robustness in underwater object detection.

Table 1 demonstrates the ablation experiments of MDE-YOLOv11 over RUOD dataset, integrating the ELU activation function showed better performance for YOLOv11, However, the mAP was higher after the full modification of YOLOv11 presenting MDE-YOLOv11 module.

Table 1
Ablation experiments

Model	Precision	Recall	mAP50(%)
YOLOv11n	87.7	80.5	87.5
YOLOv11+ELU	88.5	80.9	88.2
YOLOV11+MSDA	88.9	81.2	88.5
YOLOv11+ELU +MSDA	89.4	81.5	89.1
YOLOv11+Deformable Conv	87.9	80.7	87.4
MDE-YOLOv11	90.2	81.7	90.6

Moreover, table 2 showed a quantitative comparison on the different versions of MDE-YOLOv11, nano, large and XL model. MDE-YOLO11X showed the highest performance according to precision, recall and mAP, however, MDE-YOLOv11x has more parameters and more computational cost than the others.

Table 2
Results on different YOLOv11 versions after modifications

Model	Precision	Recall	mAP50(%)
MDE-YOLOv11n	90.2	81.7	90.6
MDE-YOLOv11L	90.7	82.4	91.3
MDE-YOLOv11X	90.9	82.8	91.7

In order to validate the proposed model MDE-YOLOv11, the proposed model was compared to baseline state of arts. The performance of the proposed model was evaluated on RUOD dataset as shown in table 3 in terms of precision, recall, mAP, trainable parameters in million, and FLOPS. The proposed model demonstrated highest precision, recall and mAP showing an improvement rate almost 10% in terms of mAP. Despite the slight increase in the trainable parameters and the FLOPS, the performance of the proposed model was superior, demonstrating balance between accuracy and speed.

Table 3
Comparison between the proposed model and the previous state of arts on RUOD dataset

Model	Precision	Recall	mAP50(%)	Parameters (M)	FLOPS
Faster-RCNN	81.4	82.0	81.7	13.72	18.12
YOLOv3-tiny	87.6	76.8	83.5	8.6	12.9
YOLOv5n	87.3	79.6	86.4	1.7	4.2
YOLOv6	82.9	71.0	78.7	4.23	11.8
YOLOv7-tiny	86.7	79.7	86.2	6.03	13.3
YOLOv8n	88.0	79.1	86.6	3.00	8.1
YOLOv9	89.0	83.0	89.0	5.72	16.7
RT-DETR	89.1	83.8	88.3	15.5	37.6
YOLOv10	86.2	80.2	86.4	2.69	8.2
YOLOv11n	87.7	80.5	87.5	1.9	6.3
MDE-YOLOv11(Ours)	90.2	81.7	90.6	2.31	6.7

While table 4 shows the evaluation of the proposed model MDE-YOLOv11 on URPC2020 dataset to validate the model and generalize it. The proposed model achieved highest performance in terms of precision, recall and mAP with an improvement rate reaching up to 1% in terms of mAP. Moreover, figure 6 demonstrates samples of the output predicted labels achieved by MDA-YOLOv11 on RUOD dataset.

Table 4
Comparison between the proposed model and the previous state of arts on URPC2020 dataset

Model	Precision	Recall	mAP50(%)	Parameters (M)	FLOPS
YOLOv3-tiny	79.3	66.7	74.0	9.52	12.9
YOLOv5n	79.2	67.9	75.1	2.18	4.2
YOLOv6	77.5	66.9	74.1	4.15	11.8
YOLOv7-tiny	73.3	71.1	74.9	6.02	13.3
YOLOv8n	80.2	69.3	76.7	3.00	8.1
YOLOv9	79.5	68.9	76.8	5.72	16.7
RT-DETR	77.8	61.2	64.5	15.5	37.6
YOLOv10	76.3	68.8	74.9	2.68	8.2
YOLOv11n	78.2	68.8	75.7	1.9	6.3
MDE-YOLOv11(Ours)	80.1	69.1	76.8	2.31	6.7

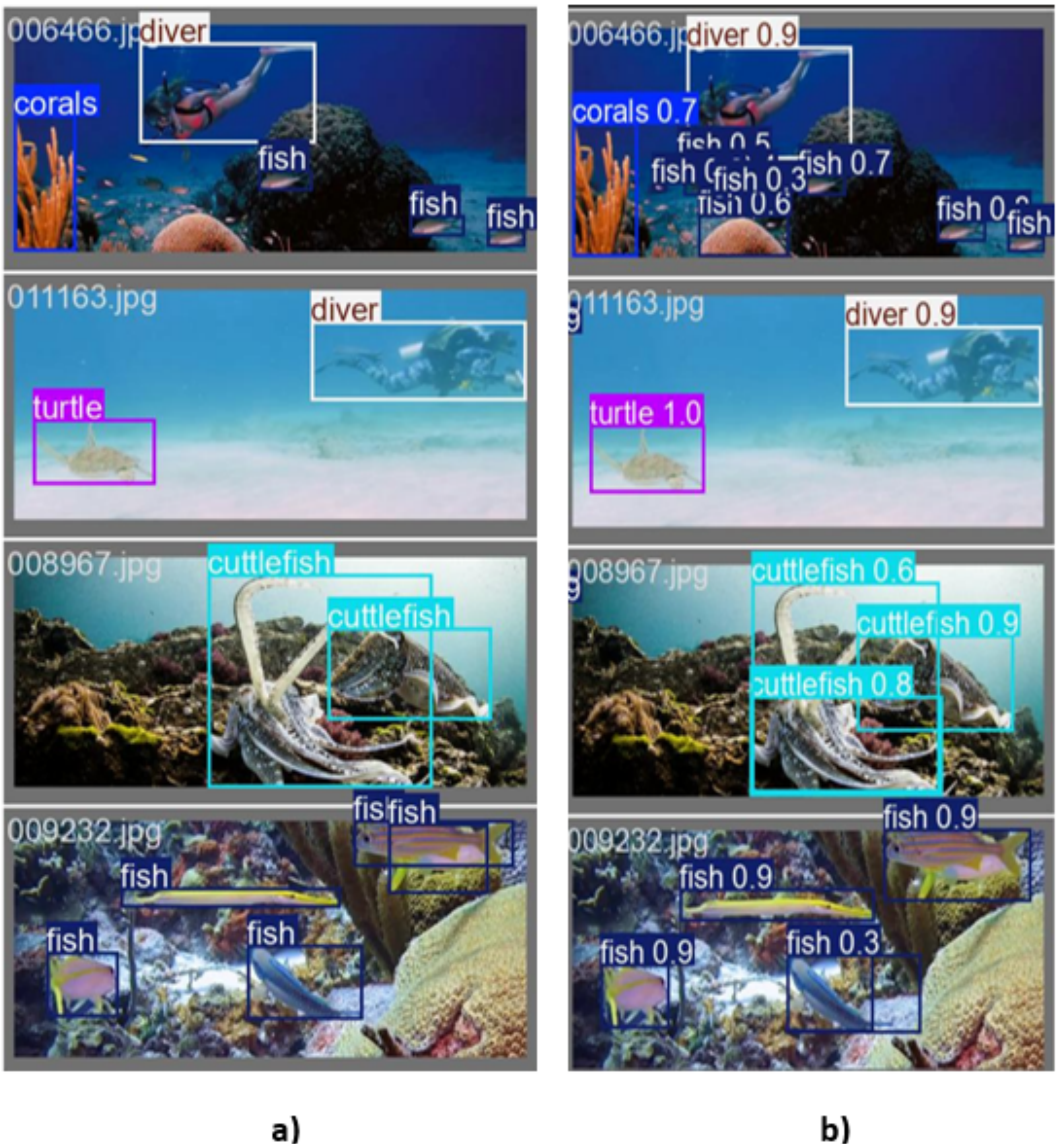


Fig. 6. Samples of prediction from the test samples of RUOD dataset where a) True labels of the underwater object detection b) Predictions made by MDA-YOLOv11

4. Conclusions

This paper presents MDE-YOLOv11, a novel adaptation of YOLOv11 optimized for underwater object detection. By integrating MSDA, DCN, and ELU activation, the model addresses critical challenges in UOD, including low contrast, occlusions, and small object detection. Experimental results on RUOD and URPC2020 datasets demonstrate superior performance, with a 90.6% mAP@0.5 and 76.8% mAP@0.5, respectively, outperforming existing baseline methods. The modifications strike a balance between accuracy and computational cost, making MDE-YOLOv11 suitable for real-time applications such as autonomous underwater vehicles (AUVs) and marine conservation.

Acknowledgement
This research was not funded by any grant

References

[1] Wang, Zhuowei, Zhukang Ruan, and Chong Chen. "DyFish-DETR: underwater fish image recognition based on detection transformer." *Journal of Marine Science and Engineering* 12, no. 6 (2024): 864. https://doi.org/10.3390/jmse12060864

[2] Zhou, Jingchun, Zongxin He, Kin-Man Lam, Yudong Wang, Weishi Zhang, Chunle Guo, and Chongyi Li. "AMSP-UOD: When vortex convolution and stochastic perturbation meet underwater object detection." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 7, pp. 7659-7667. 2024. https://doi.org/10.1609/aaai.v38i7.28599

[3] Li, Tao, Yijin Gang, Sumin Li, and Yizi Shang. "A small underwater object detection model with enhanced feature extraction and fusion." *Scientific Reports* 15, no. 1 (2025): 2396. https://doi.org/10.1038/s41598-025-85961-9

[4] Li, Xingkun, Yuhao Zhao, Hu Su, Yugang Wang, and Guodong Chen. "Efficient underwater object detection based on feature enhancement and attention detection head." *Scientific Reports* 15, no. 1 (2025): 5973. https://doi.org/10.1038/s41598-025-89421-2

[5] Chen, Long, Yuzhi Huang, Junyu Dong, Qi Xu, Sam Kwong, Huimin Lu, Huchuan Lu, and Chongyi Li. "Underwater object detection in the era of artificial intelligence: Current, challenge, and future." *arXiv preprint arXiv:2410.05577* (2024). https://doi.org/10.48550/arXiv.2410.05577

[6] Zhou, Hui, Meiwei Kong, Hexiang Yuan, Yanyan Pan, Xinru Wang, Rong Chen, Weiheng Lu, Ruizhi Wang, and Qunhui Yang. "Real-time underwater object detection technology for complex underwater environments based on deep learning." *Ecological Informatics* 82 (2024): 102680. https://doi.org/10.1016/j.ecoinf.2024.102680

[7] Jian, Muwei, Nan Yang, Chen Tao, Huixiang Zhi, and Hanjiang Luo. "Underwater object detection and datasets: a survey." *Intelligent Marine Technology and Systems* 2, no. 1 (2024): 9. https://doi.org/10.1007/s44295-024-00023-6

[8] Zhang, Jian, Jinshuai Zhang, Kexin Zhou, Yonghui Zhang, Hongda Chen, and Xinyue Yan. "An improved YOLOv5-based underwater object-detection framework." *Sensors* 23, no. 7 (2023): 3693.https://doi.org/10.3390/s23073693

[9] Chen, Jie, and Meng Joo Er. "Dynamic YOLO for small underwater object detection." *Artificial Intelligence Review* 57, no. 7 (2024): 165. https://doi.org/10.1007/s10462-024-10788-1

[10] Feng, Jiangfan, and Tao Jin. "CEH-YOLO: A composite enhanced YOLO-based model for underwater object detection." *Ecological Informatics* 82 (2024): 102758. https://doi.org/10.1016/j.ecoinf.2024.102758

[11] Pan, Wei, Jiabao Chen, Bangjun Lv, and Likun Peng. "Optimization and application of improved YOLOv9s-UI for underwater object detection." *Applied Sciences* 14, no. 16 (2024): 7162. https://doi.org/10.3390/app14167162

[12] Cheng, Shun, Yan Han, Zhiqian Wang, Shaojin Liu, Bo Yang, and Jianrong Li. "An underwater object recognition system based on improved yolov11." *Electronics* 14, no. 1 (2025): 201.https://doi.org/10.3390/electronics14010201

[13] Liu, Mingxin, Yujie Wu, Ruixin Li, and Cong Lin. "LFN-YOLO: precision underwater small object detection via a lightweight reparameterized approach." *Frontiers in Marine Science* 11 (2025): 1513740. https://doi.org/10.3389/fmars.2024.1513740

[14] Luo, Yang, Aiping Wu, and Qingqing Fu. "MAS-YOLOv11: An improved underwater object detection algorithm based on YOLOv11." *Sensors* 25, no. 11 (2025): 3433. https://doi.org/10.3390/s25113433

[15] Liu, Linlin, Chengxi Chu, Chuangchuang Chen, and Shidong Huang. "MarineYOLO: Innovative deep learning method for small target detection in underwater environments." *Alexandria Engineering Journal* 104 (2024): 423-433. https://doi.org/10.1016/j.aej.2024.07.126

[16] Zhang, Jinquan, Zhuang Luan, Lina Ni, Liang Qi, and Xu Gong. "MSDANet: A multi-scale dilation attention network for medical image segmentation." *Biomedical Signal Processing and Control* 90 (2024): 105889. https://doi.org/10.1016/j.bspc.2023.105889

[17] Li, Jinrong, Chunhua Wei, Lei Liang, and Zhisheng Gao. "Multi-Scale Dilated Attention-Based Transformer Network for Image Inpainting." *Computers, Materials & Continua* 84, no. 2 (2025). https://doi.org/10.32604/cmc.2025.063547

[18] Zhou, Yue, Yutong Jiang, Zhonglin Yang, Xingxin Li, Wenchen Sun, Han Zhen, and Ying Wang. "UAV image detection based on multi-scale spatial attention mechanism with hybrid dilated convolution." In *2024 3rd International Conference on Image Processing and Media Computing (ICIPMC)*, pp. 279-284. IEEE, 2024. 10.1109/ICIPMC62364.2024.10586685

[19] Guo, Lijia, Xiangchun Liu, Dongsheng Ye, Xuebao He, Jianxin Xia, and Wei Song. "Underwater object detection algorithm integrating image enhancement and deformable convolution." *Ecological Informatics* (2025): 103185. https://doi.org/10.1016/j.ecoinf.2025.103185

[20] Luna-Rivera, J. M., Jose Rabadan, Julio Rufo, Victor Guerra, C. A. Gutierrez, and Rafael Perez-Jimenez. "Enhancing Underwater Visible Light Communication with End-to-End Learning Techniques." In *2024 14th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 518-523. IEEE, 2024. 10.1109/CSNDSP60683.2024.10636487

[21] Fu, Chenping, Risheng Liu, Xin Fan, Puyang Chen, Hao Fu, Wanqi Yuan, Ming Zhu, and Zhongxuan Luo. "Rethinking general underwater object detection: Datasets, challenges, and solutions." *Neurocomputing* 517 (2023): 243-256.https://doi.org/10.1016/j.neucom.2022.10.039

[22] Liu, Chongwei, Haojie Li, Shuchang Wang, Ming Zhu, Dong Wang, Xin Fan, and Zhihui Wang. "A dataset and benchmark of underwater object detection for robot picking." In *2021 IEEE international conference on multimedia & expo workshops (ICMEW)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/ICMEW53276.2021.9455997