RHCNet: Residual-Guided Hierarchical Calibration Network for Robust Underwater Object Detection

Supplementary Material

6. Supplement to The Text

In this supplementary material, we present additional visual results and extended discussions that were not included in the main text. We provide a detailed overview of our visual analysis of characteristic distributions for different types of underwater degradation, along with a comprehensive qualitative comparison, which further substantiates the effectiveness of the proposed RHCNet.

7. Feature Distribution Visualization Analysis

As illustrated in Fig. 6, compared with the baseline RetinaNet, RHCNet exhibits stronger feature consistency and higher object sensitivity within object regions under complex underwater scenarios, demonstrating its superior robustness in handling background interference and semantic ambiguity.

To further evaluate the discriminative feature representation capability of the model, Fig. 7 presents the feature space distributions of the baseline and our proposed method over one thousand test images from the DUO [24] dataset, where the baseline features appear widely scattered. In contrast, our method demonstrates markedly higher semantic compactness, reflecting its superior ability to capture and represent discriminative features.

8. Comprehensive Qualitative Comparison

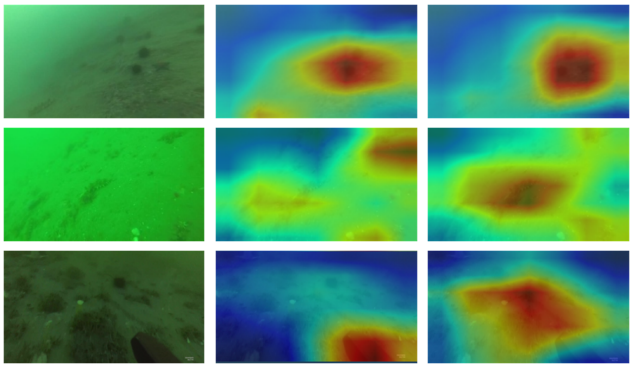
Fig. 8, Fig. 9, and Fig. 10 present qualitative comparison results for three typical complex underwater scenarios on the DUO [24] dataset.

1) Similar Foreground and Background Scenes. RHCNet effectively suppresses false activations in background regions during qualitative comparisons, resulting in more accurate object localization.

2) Object boundary degradation Scenes. Detection boxes generated by RHCNet more accurately cover object areas, markedly reducing instances of missed detections outside the box or redundant coverage within it caused by boundary blurring.

3) Low-Visibility Underwater Scenes. RHCNet maintains stable object coverage even under low-light and highly turbid conditions, demonstrating superior detection confidence and localization consistency compared with other state-of-the-art methods.

Fig. 11, Fig. 12, and Fig. 13 illustrate the corresponding results on the UTDAC dataset, encompassing degradation types consistent with those in the DUO dataset, where RHCNet consistently surpasses six state-of-the-art counterparts in terms of object localization accuracy and boundary compactness, thereby further validating its strong generalization capability under diverse and challenging underwater conditions.



(a) Raw image (b) Baseline (c) Ours

Figure 6. Visual comparison of feature heatmaps for the raw image, Baseline and Ours in underwater images.

9. More Implementation Details

9.1. Data Augmentation

During training, a concise yet effective data augmentation strategy was adopted, where images were randomly rescaled across multiple resolutions with the shorter side sampled between 800 and 1920 pixels while maintaining the original aspect ratio, thereby substantially improving the model's capacity to capture multi-scale object variations particularly beneficial in underwater scenarios with significant object size diversity; moreover, no additional preprocessing techniques such as histogram equalization, CLAHE contrast enhancement, or style transfer were utilized, ensuring that all observed performance gains arise primarily from the proposed architecture and its enhanced feature representation capability.

9.2. Evaluation Metrics

To ensure fairness and rigor in the evaluation process, this paper uniformly employs the official COCO object detection metrics for performance assessment across all experiments. Details are as follows.

Average Precision (AP): This metric represents the average accuracy achieved across a range of IoU thresholds

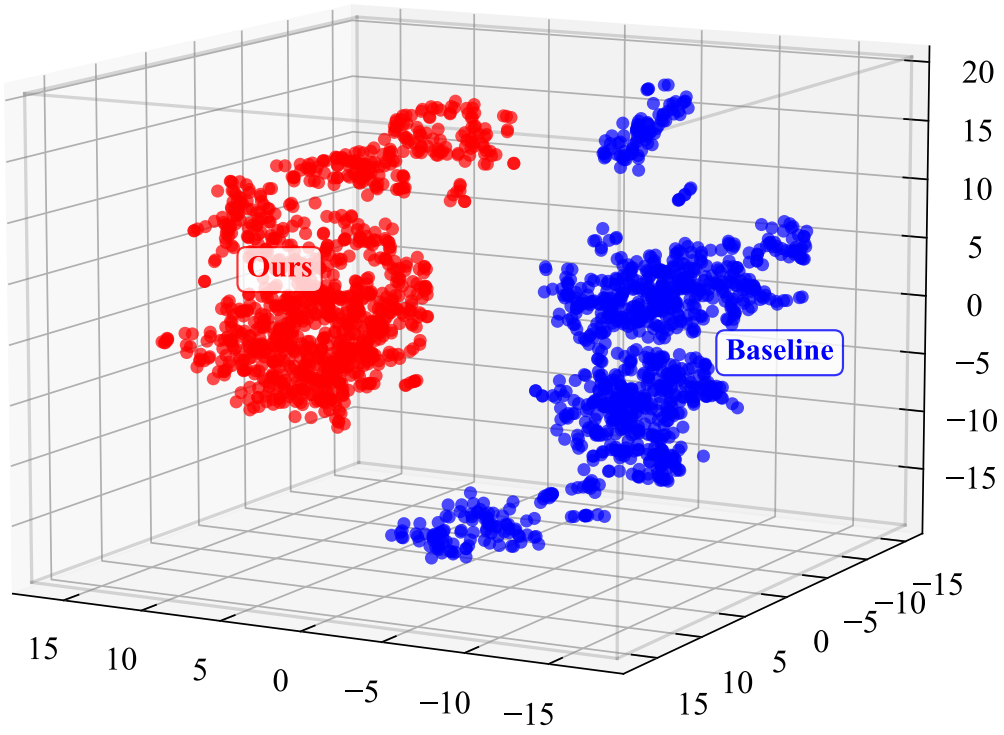


Figure 7. Scatter plot comparing feature distributions of our RHCNet and baseline methods in three-dimensional feature space.

from 0.50 to 0.95, incremented by 0.05. It serves to evaluate a model's overall detection capability.

AP_{50}/AP_{75}: These denote the accuracy under conditions of IoU=0.50 and IoU=0.75, respectively. The former measures coarse-grained localization capability, while the latter emphasizes precise detection.

$AP_S/AP_M/AP_L$: These denote detection accuracy for small object areas (less than 32^2 pixels^2), medium object areas (between 32^2 pixels^2 and 96^2 pixels^2), and large object areas (greater than 96^2 pixels^2), respectively. They are employed to analyses a model's robustness across different object scales.

All the aforementioned metrics were uniformly tested on datasets such as DUO [24], UTDAC [28], and COCO [21] to validate the model's robustness and stability in complex underwater environments and general detection scenarios.

10. Additional Ablation Study

To validate the effectiveness of each component within RHCNet, we provide extended ablation results and visualized performance trend analyses based on Tab. 2 in the main text. As shown in Fig. 14, the metrics AP, AP_{50}, AP_{75}, AP_S, AP_M, and AP_L on the DUO [24] dataset exhibit a stable upward trend with increasing module stacking.

To further validate the effectiveness of each module, we conducted additional ablation experiments on the UTDAC [28] dataset, with results presented in Tab. 4. The overall findings demonstrate that the complete model (Case 7) achieves optimal performance across all evaluation metrics, thereby corroborating the positive contributions of the LAM, RGFE, PAM, and CGCA modules in enhancing the model's perceptual capabilities and detection robustness.

As shown in Fig. 15, our method exhibits a stable upward trend across the AP, AP_{50}, AP_{75}, AP_S, AP_M, and AP_L metrics on the UTDAC dataset as modules are stacked.

11. Future Work and Limitation

Although the proposed RHCNet demonstrates outstanding performance across multiple evaluation metrics and exhibits strong robustness in complex underwater scenarios, several areas warrant further investigation alongside existing limitations. Firstly, RHCNet may still exhibit insufficient feature representation when processing severely

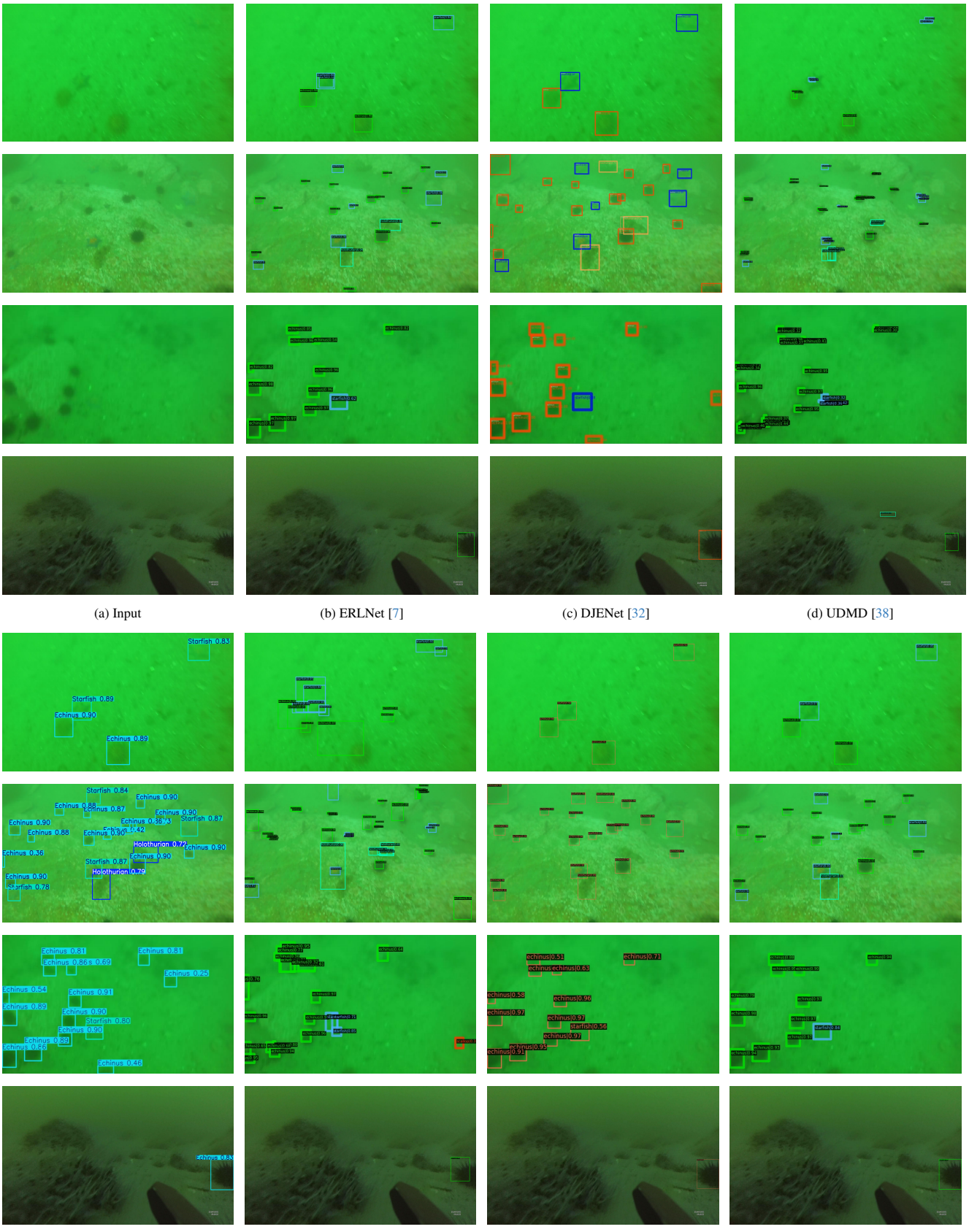
Table 4. In the UTDAC [28] dataset, the ablation experiments of the key modules in RHCNet (✓= MODULE RETAINED, BLANK = MODULE ABLATED).

Cases	LAM	RGFE	PAM	CGCA	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
RetinaNet					46.02	70.53	48.36	19.57	40.34	51.23
1		✓	✓	✓	48.96	83.42	53.76	23.20	45.07	54.71
2	✓		✓	✓	48.45	83.11	53.64	22.68	44.82	54.19
3			✓	✓	47.13	82.62	52.49	22.07	43.90	53.37
4	✓	✓		✓	48.84	83.26	53.31	22.75	44.83	54.08
5	✓	✓	✓		47.67	82.81	52.64	22.18	44.01	53.69
6	✓	✓			47.54	81.75	52.13	21.08	43.87	53.34
7	✓	✓	✓	✓	**50.87**	**85.86**	**55.37**	**24.82**	**46.23**	**56.42**

degraded images (e.g., intense light refraction, significant suspended particle occlusion), indicating that the model's generalization capability under extreme conditions requires enhancement. Secondly, the K-means cluster-guided feature calibration module depends on unsupervised clustering priors. It remains sensitive to the predefined number of clusters, indicating a potential benefit from incorporating adaptive or learnable clustering strategies. Thirdly, the model's performance relies heavily on large-scale, well-annotated training datasets, yet obtaining high-quality underwater labels is both costly and labor-intensive. Therefore, future work should investigate strategies such as self-supervised representation learning, semi-supervised fine-tuning, or synthetic-to-real domain adaptation to enhance performance in data-scarce or weakly supervised settings.

12. Broader Impacts

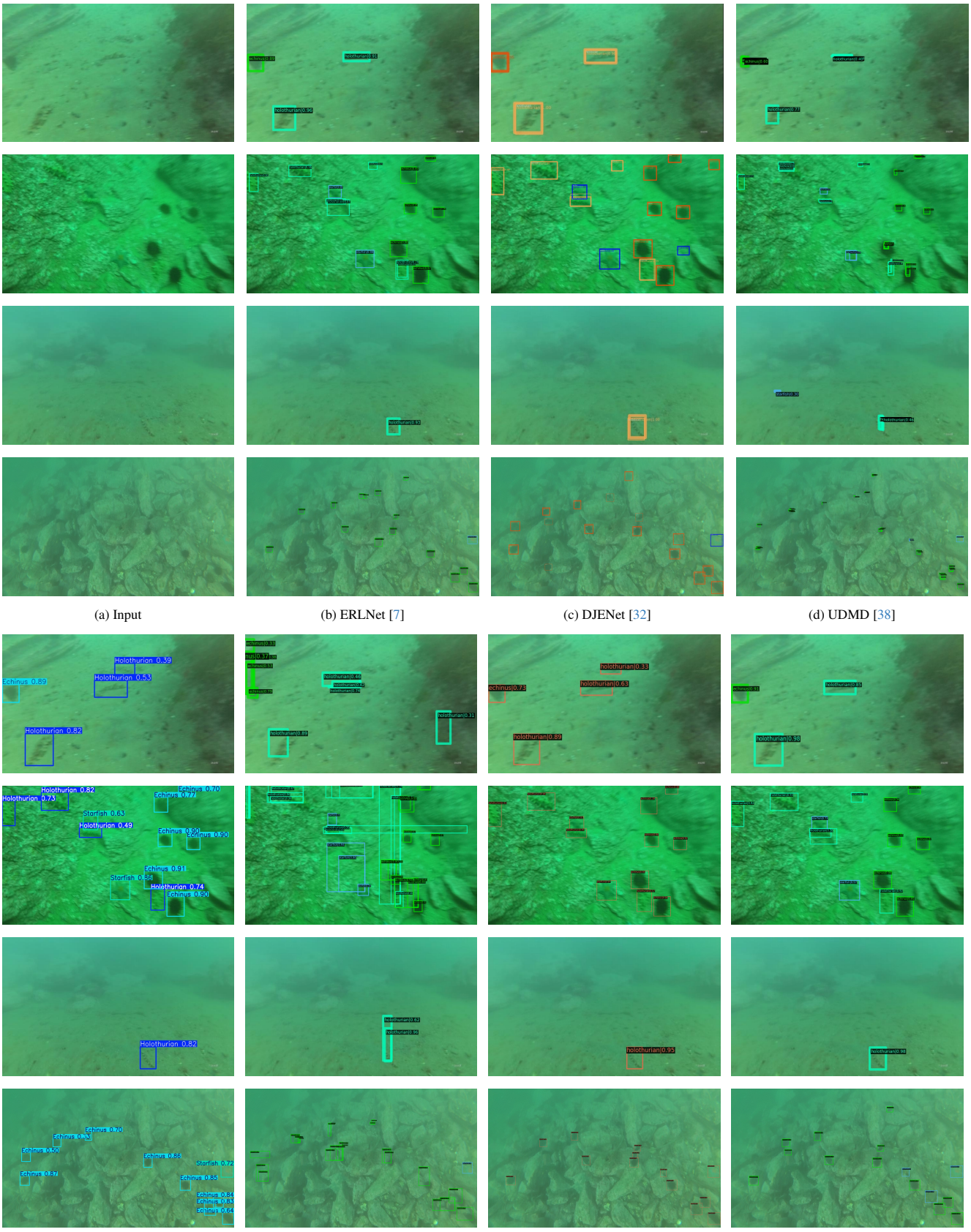
The proposed RHCNet framework holds substantial significance beyond underwater target detection, demonstrating strong potential for broader visual perception and cross-domain detection tasks. In marine resource exploration, RHCNet enhances automated recognition of marine organisms and submerged infrastructure, thereby improving both the efficiency and precision of oceanographic surveys. Within ecological conservation and environmental monitoring, it enables accurate tracking of coral reef dynamics, fish populations, and endangered species, offering essential technological support for the sustainable development of marine ecosystems. In military and engineering contexts, RHCNet enables autonomous perception and positioning for underwater unmanned systems, thereby enhancing mission reliability and operational safety. More broadly, its residual attention mechanism and hierarchical calibration strategy provide a generalized solution to visual recognition challenges under ambiguous conditions, extending its applicability to low-light surveillance, medical imaging analysis, and remote sensing interpretation. Nevertheless, the large-scale deployment of intelligent underwater perception systems may raise ethical and privacy concerns, particularly in defense and security domains; hence, future research should not only emphasize technical optimization but also incorporate considerations of social responsibility and sustainability, ensuring that technological advancement promotes scientific innovation while upholding environmental ethics and human values.



(a) Input (b) ERLNet [7] (c) DJENet [32] (d) UDMD [38]

(e) YOLOv11 [15] (f) SqNet [3] (g) CIDNet [42] (h) Ours

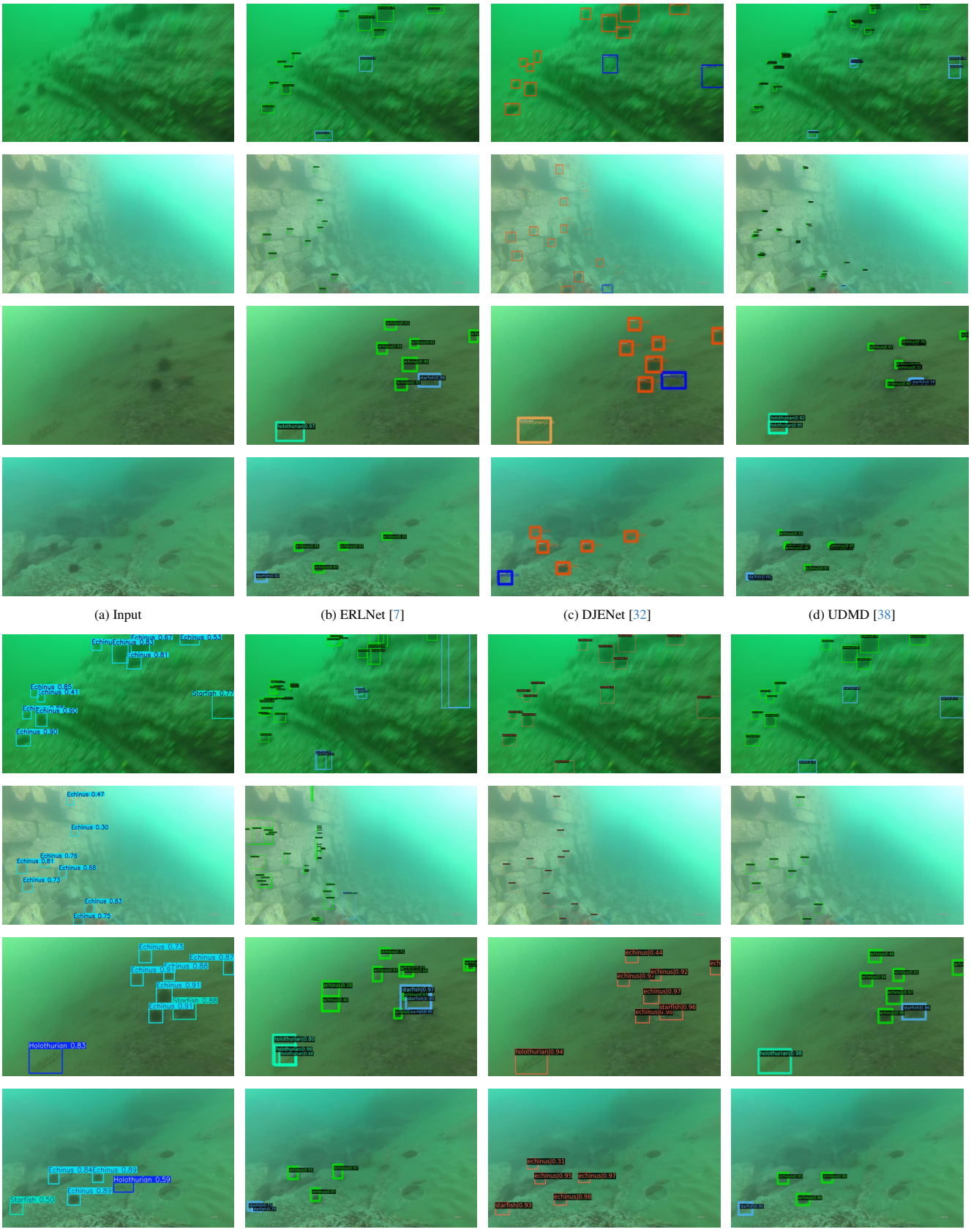
Figure 8. Qualitative evaluation of underwater scenarios with similar foreground and background in the DUO [24] dataset.



(a) Input (b) ERLNet [7] (c) DJENet [32] (d) UDMD [38]

(e) YOLOv11 [15] (f) SqNet [3] (g) CIDNet [42] (h) Ours

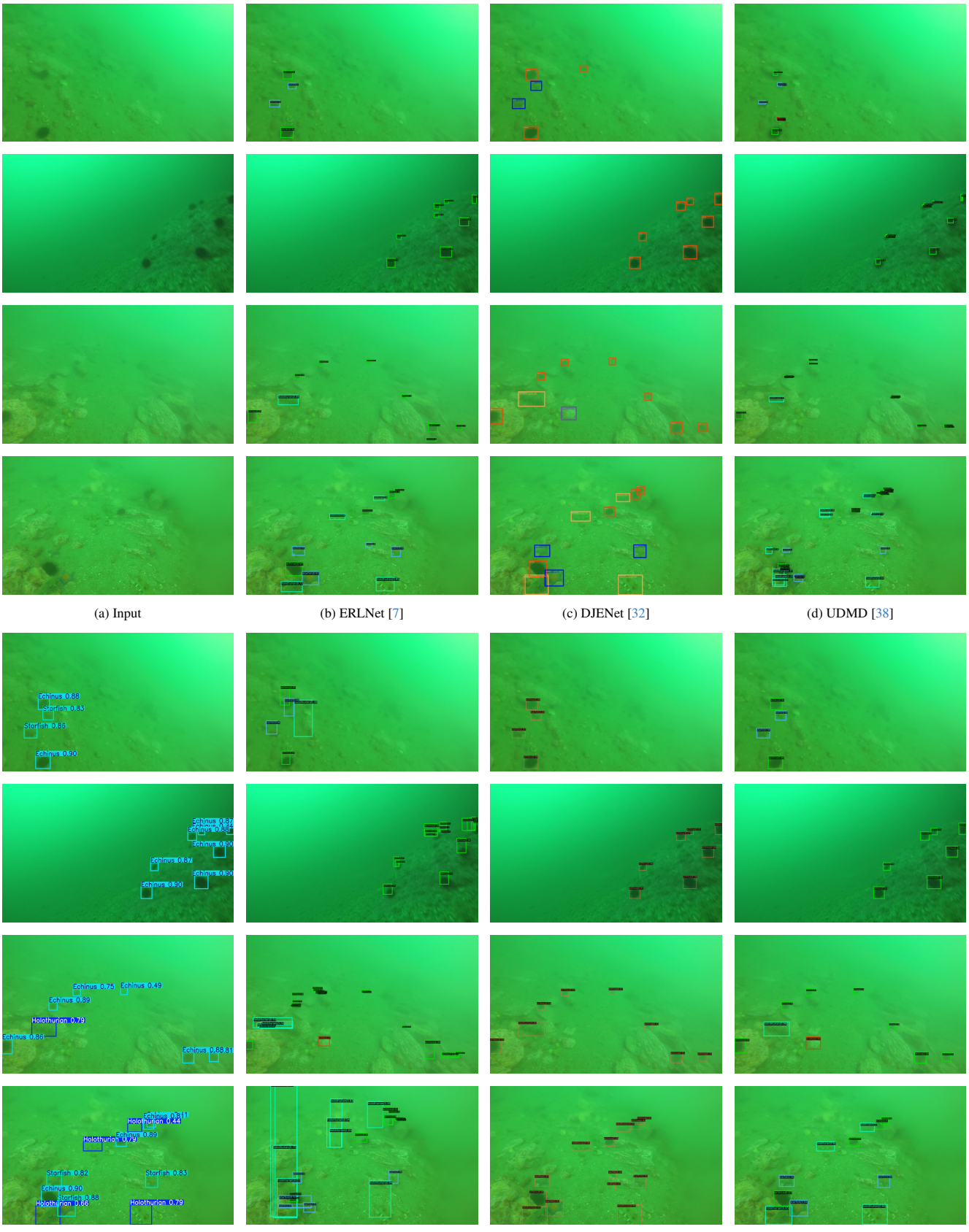
Figure 9. Qualitative evaluation of underwater scenarios with object edge degradation in the DUO [24] dataset.



(a) Input (b) ERLNet [7] (c) DJENet [32] (d) UDMD [38]

(e) YOLOv11 [15] (f) SqNet [3] (g) CIDNet [42] (h) Ours

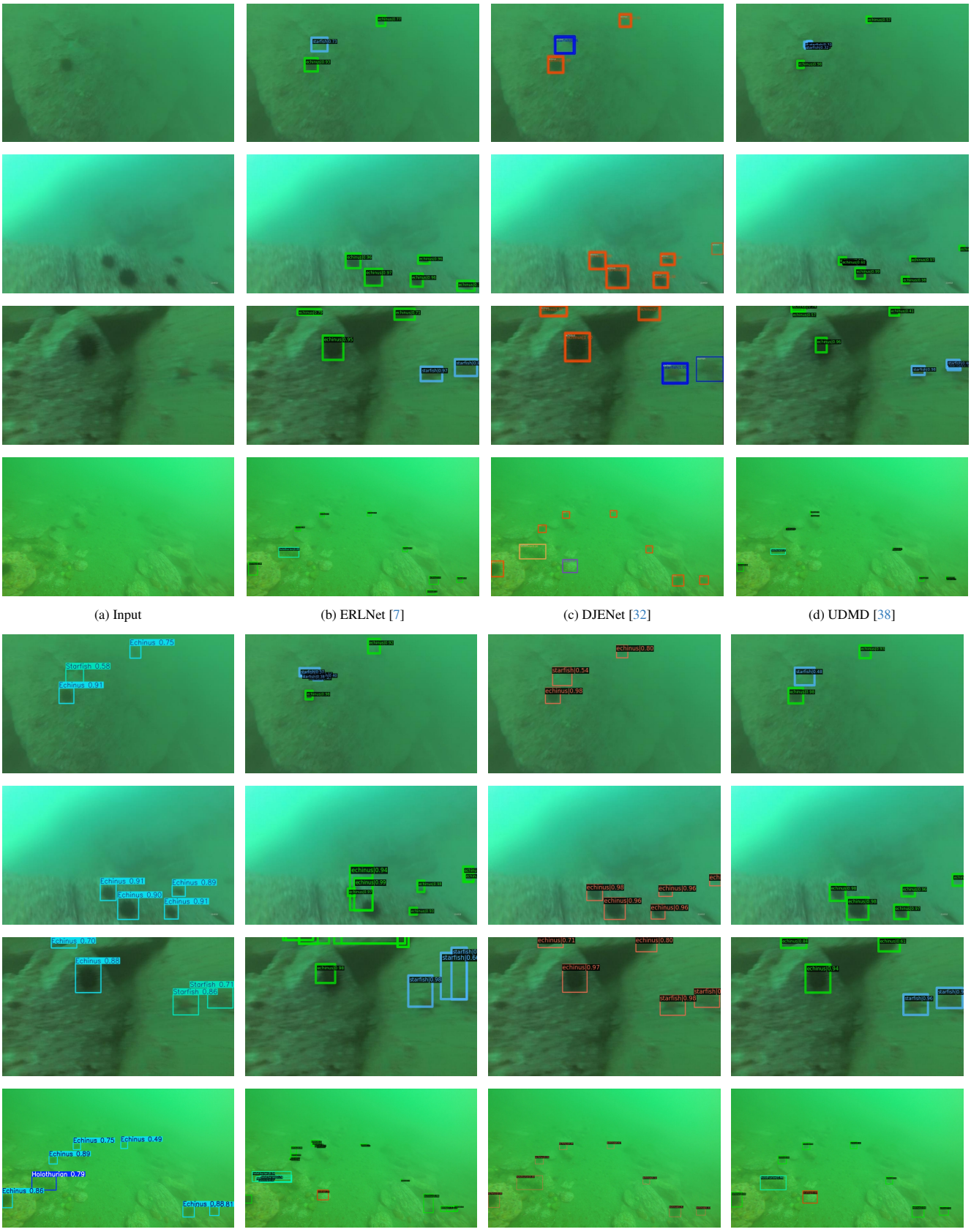
Figure 10. Qualitative evaluation of low-visibility underwater scenes in the DUO [24] dataset.



(a) Input (b) ERLNet [7] (c) DJENet [32] (d) UDMD [38]

(e) YOLOv11 [15] (f) SqNet [3] (g) CIDNet [42] (h) Ours

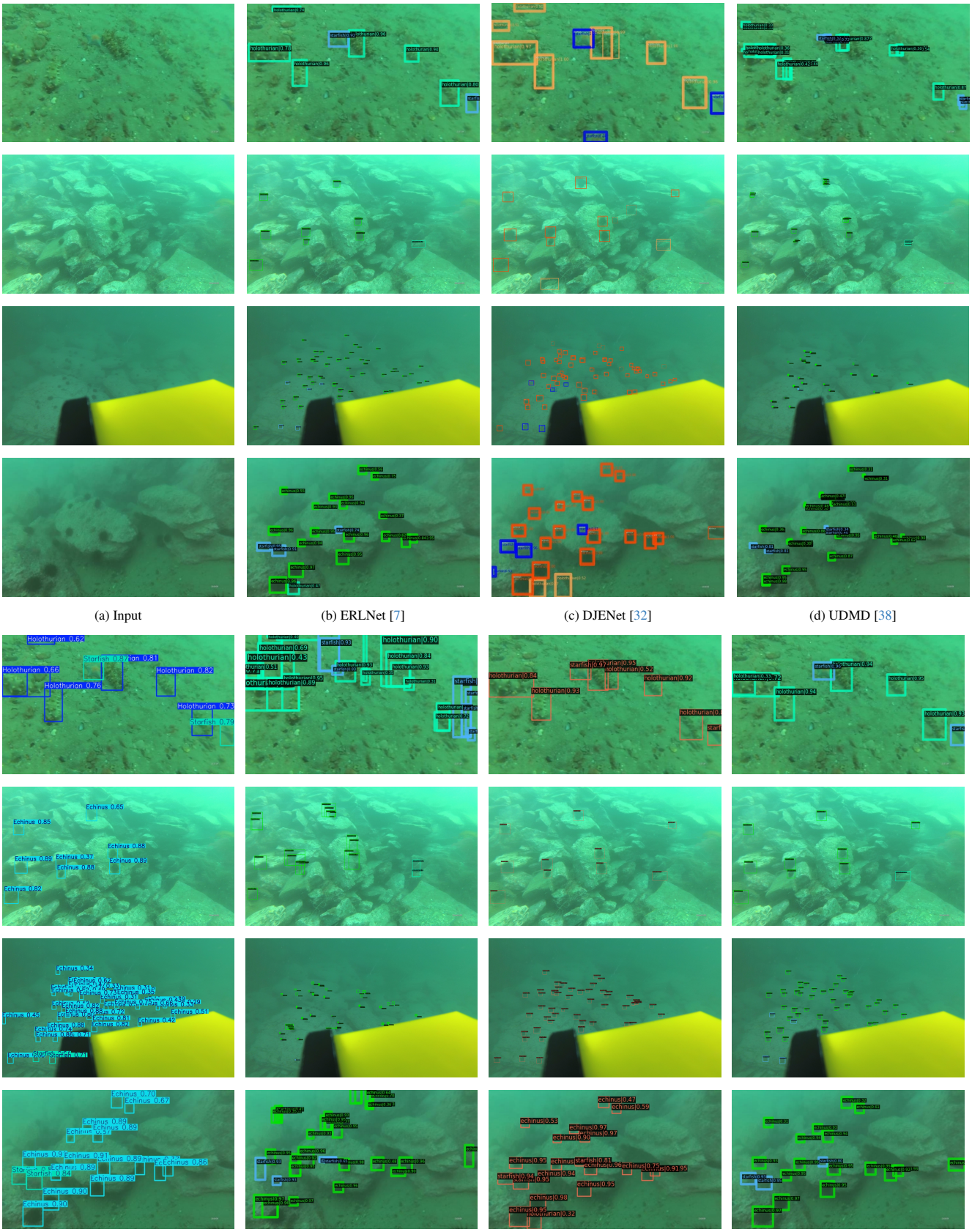
Figure 11. Qualitative evaluation on underwater scenes with similar foreground and background in the UTDAC [28] dataset.



(a) Input (b) ERLNet [7] (c) DJENet [32] (d) UDMD [38]

(e) YOLOv11 [15] (f) SqNet [3] (g) CIDNet [42] (h) Ours

Figure 12. Qualitative evaluation of underwater scenarios with object edge degradation in the UTDAC [28] dataset.



(a) Input (b) ERLNet [7] (c) DJENet [32] (d) UDMD [38]

(e) YOLOv11 [15] (f) SqNet [3] (g) CIDNet [42] (h) Ours

Figure 13. Qualitative evaluation of low-visibility underwater scenes in the UTDAC [28] dataset.

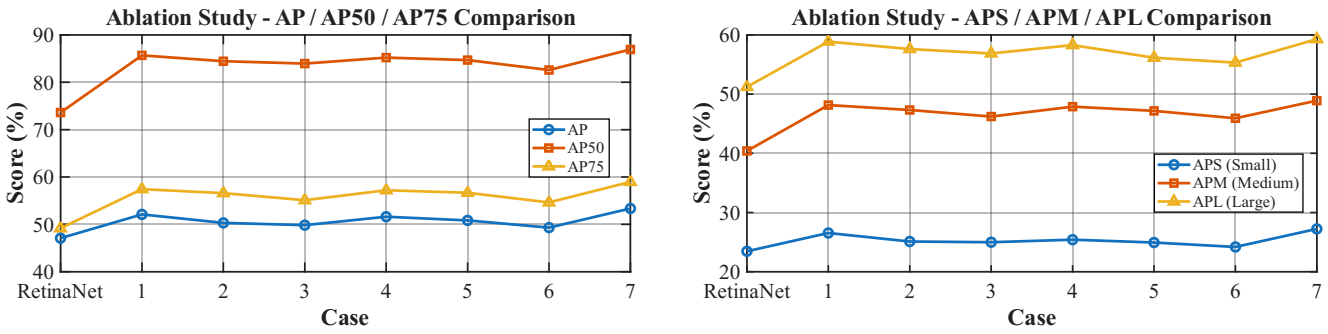


Figure 14. Line chart of ablation experiments on the DUO [24] dataset

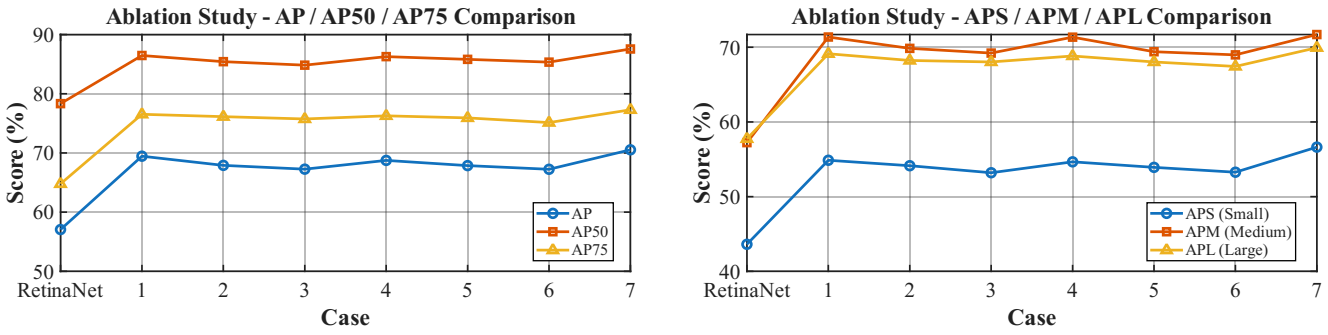


Figure 15. Line chart of ablation experiments on the UTDAC [28] dataset