YAN, Y., LI, Y., LIN, H., SARKER, M.M.K., REN, J. and MCCALL, J. 2024. Underwater object detection for smooth and autonomous operations of naval missions: a pilot dataset. In Ren, J., Hussain, A., Liao, I.Y. et al. (eds.) *Advances in brain inspired cognitive systems: proceedings of the 13th International conference on Brain-inspired cognitive systems 2023 (BICS 2023), 5-6 August 2023, Kuala Lumpur, Malaysia*. Lecture notes in computer sciences, 14374. Cham: Springer [online], pages 113-122. Available from: https://doi.org/10.1007/978-981-97-1417-9_11

Underwater object detection for smooth and autonomous operations of naval missions: a pilot dataset.

YAN, Y., LI, Y., LIN, H., SARKER, M.M.K., REN, J. and MCCALL, J.

2024

Underwater Object Detection for Smooth and Autonomous Operations of Naval Missions: A Pilot Dataset

Yijun Yan[1], Yinhe Li[1], Hanhe Lin[2], Md Mostafa Kamal Sarker[3], Jinchang Ren[1(✉)], and John McCall[1]

[1] National Subsea Centre, Robert Gordon University, Aberdeen, UK
{y.yan2,y.li24,j.ren,j.mccall}@rgu.ac.uk
[2] School of Science and Engineering, University of Dundee, Dundee, UK
hlin001@dundee.ac.uk
[3] Institute of Biomedical Engineering, University of Oxford, Oxford, UK
md.sarker@eng.ox.ac.uk

Abstract. Underwater object detection is essential for ensuring autonomous naval operations. However, this task is challenging due to the complexities of underwater environments that often degrade image quality, thereby hampering the performance of detection and classification systems. On the other hand, the absence of a readily available dataset complicates the development and evaluation of underwater object detection approaches, particularly for deep learning approaches. To address this bottleneck, we have created a new dataset, called National Subsea Centre Underwater Images (NSCUI). It is comprised of 243 images, divided into three subsets that are captured in bright, low-light, and dark environments, respectively. To validate the utility of this dataset, we implemented three popular deep learning models in our experiments. We believe that the annotated NSCUI will significantly advance the development of underwater object detection through the application of deep learning techniques.

Keywords: Underwater object detection · Image enhancement · Deep learning

1 Introduction

Obstacles detection and collision avoidance are the key challenges for smooth and autonomous operations of naval missions, where sonar and optics sensors are widely used [1]. Due to the complexity of the underwater environments, the image quality of both sonar and optics system can be severely degraded, resulting in inconsistent performance of target detection and classification.

Sonar has been popularly applied in many underwater inspection tasks [2], featuring long-range detection. However, its working condition is not only affected by internal factors (e.g. latency, narrow bandwidth and self-noise), but also external factors related to ocean environment (e.g. spreading loss, multipath effect, reverberation, ocean noise, target reflection characteristics and radiation noise [3]). Due to poor performance in shallow depth and noisy data [4], it makes it difficult to detect small but important objects such as nets. Another downside is the heavy cost, which has constrained its wide deployment. In addition, military sonar systems may severely affect the lives of marine mammals, leading to deafness and death of dolphins, whales and sea turtles [5].

Optic systems, an economically viable alternative, are celebrated for their substantial bandwidth, contributing rich structure, shape, and texture features, which are particularly beneficial for shallow underwater inspection [6]. Despite their advantages, light transmission in water encounters severe attenuation attributed to both internal factors, such as absorption and scattering, and external factors like turbidity and suspended particles, which collectively result in a relatively low detection accuracy.

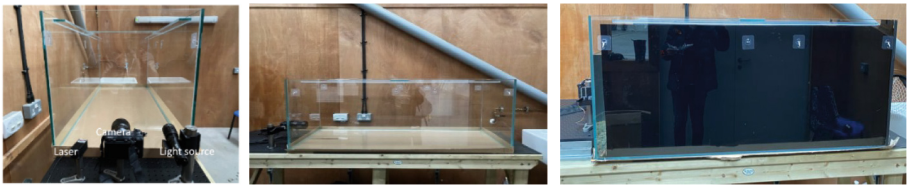
Deep learning (DL) based object detection methods offer a promising solution to address these limitations. DL models have demonstrated success in various image modalities, including video surveillance [7], aerial image [8], hyperspectral images [9], etc. Recently, the application of DL-based methods has extended to underwater object detection. Examples of applications include the usage of R-CNN [10] and its extended versions [11] for detecting and recognizing aquatic organisms, for which datasets such as UTDAC2020 and DUO have been used for modelling. Similarly, Mask R-CNN has been utilized for deep-sea litter detection [12], with the Trashcan dataset used for modelling purposes. Additionally, the YOLO series models have been employed for marine animal detection [13, 14], where the datasets UODD [13] and Blackish [14] have been conducted for modelling. However, due to lack of sufficient data and labels, the feasibility of using deep learning models for underwater obstacle recognition remains largely underexplored and presents significant opportunities for further research.

The aim of this study is to construct a new underwater image dataset and investigate the potential of deep learning techniques in underwater obstacle recognition. By integrating image enhancement techniques with deep learning-based object detection, we aspire to advance obstacle recognition under varying lighting conditions. This enhancement is expected to contribute significantly to the autonomous operation of underwater vehicles, offering improved navigation and operational efficiency of naval missions.

2 Dataset Creation

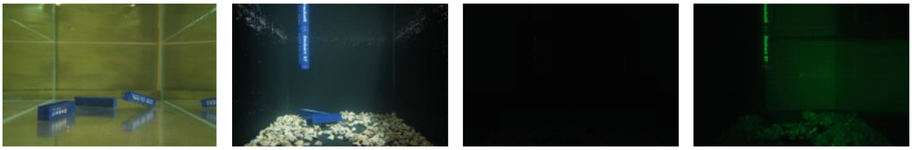
In this experiment, we captured images for simulated 5 categories of underwater targets (i.e., container, props, keel, wreck and net) in different light conditions. The water tank we used has the size of 60 cm × 60 cm × 150 cm and maximum volume of 500 L. The imaging device we used is a scotopic camera, Sony 4K video camera with full 35 mm frame Exmor CMOS sensor. It operates in the maximum 4240 × 2842 (12M) pixels for still image and 3810 × 2160 pixels for video recording with an ISO sensitivity ranges of 50 – 409600, which allows it to capture the high resolution image data in both bright and low light environments. Figure 1. (a) shows the position of imaging system, Fig. 1. (b–c) shows the experimental setting for data acquisition in bright environment, and low light/dark environment. Notably, in the low light environment, we attached black vinyl sheets on the walls of the water tank and used a black sheet to cover half of the top to reduce the incoming light. In the dark environment, the top of water tank was totally covered by the black sheet, resulting a totally black scene. However, the existing object detection models failed to perform in such kind of scene. Instead, a laser projector consisting of a laser diode with 520 nm and 50 mV output power, and a diffractive optical element with 51 × 51 dot matrix, was employed to provide consistent light source and point-cloud data for 3D measurement.

Examples of captured images in different light condition are shown in Fig. 2. Due to the uncertainty of water dynamics, it is impossible to capture the same object with the same position in different light conditions. Thus, we created three subsets separately.



(a) (b) (c)

Fig. 1. (a) Imaging system, (b) experimental setting for bright environment, and (c) dark/low light environment.



(a) (b) (c) (d)

Fig. 2. Image captured in (a) bright environment, (b) low light environment, (c) dark environment and (d)dark environment with laser light source.

where the images were captured in bright, low light and dark conditions, and the number of images in each subset is 95, 74, and 74, respectively.

After that, we made the annotation using an open-source annotation tool LabelMe. The ground-truth images annotation format is then converted to MS-COCO format. Figure 3 shows the image data containing single category and its corresponding annotation. To further simulate the complexity in the real world, image data for mixed categories of targets that were randomly distributed in the water were also captured (Fig. 4), and the ground-truth data were also carefully annotated.

3 Objective Object Detection Assessment

Deep learning-based object models are designed to replicate the mechanisms involved in visual perception, aiming to improve object recognition performance and achieve more brain-like processing capabilities. In our study, we evaluated the object detection performance using three popular deep learning methods: Mask R-CNN, Faster R-CNN, and YOLO-X. Since underwater environments often have complex lighting conditions and lower image quality, these factors can negatively impact object detection accuracy.

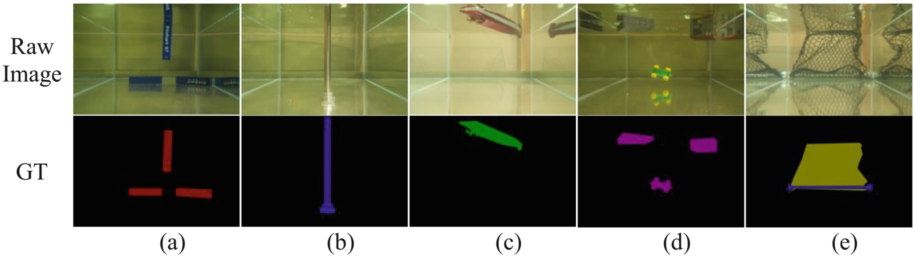


Fig. 3. Single category of data acquisition and annotation. (a) container, (b) props, (c) keel, (d) wreck, and (e) net.

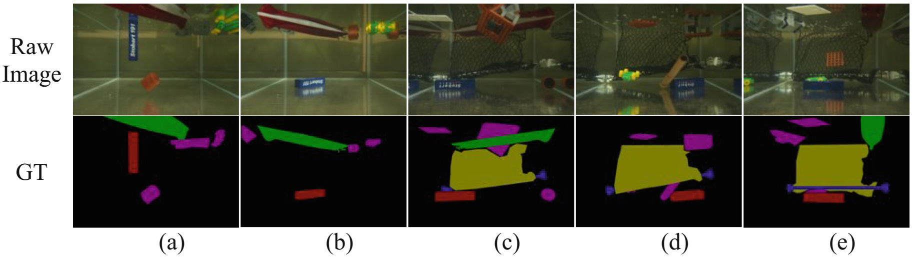


Fig. 4. Mixed categories of data acquisition and annotated ground-truth(GT).

Drawing inspiration from human visual perception, we recognized that objects with higher contrast are generally more detectable. This motivated us to further enhance the performance of these models by integrating image enhancement techniques. Specifically, we fed the enhanced images into the object detection models to evaluate their effectiveness in improving object detection accuracy. According to our previous work [15], three image enhancement methods, including two best performed traditional methods (CLAHE [16], Fusion-based [17]) and one deep learning method (WaterNet [18]), are adopted in this study. For objective evaluation, we have calculated mAP50, mAP75 and mAP to analyse the robustness of models under different Intersection over Union (IoU) thresholds, where IoU threshold is set as 0.5, 0.75 and 0.5:0.05:0.95 for mAP50, mAP75 and mAP, respectively.

4 Results and Analysis

To assess the object detection accuracy of different models under different light condition, three experiments were carried out in this paper. In the first experiment, we utilized a dataset of images captured under bright conditions. This dataset was divided into training and testing sets, consisting of 68 and 27 images, respectively. For the second and third experiments, we focused on different lighting conditions. In the second experiment, we used a dataset of images captured under a specific low-light condition. This dataset was divided into training and testing sets, with 49 images allocated for training and 25 images for testing. Similarly, in the third experiment, we considered a specific dark condition. The dataset of images captured under this condition was also divided into training and testing sets, with 49 images for training and 25 images for testing.

4.1 Object Detection in Bright Environment

Several key findings can be drawn from the results presented in Table 1. YOLOX consistently outperforms Faster-RCNN and Mask-RCNN in terms of mAP50 score, regardless of whether image enhancement is applied. The combination of YOLOX with WaterNet achieves the highest mAP50 score of 0.961. Additionally, Fig. 5 reveals that YOLOX exhibits superior object recognition capabilities, producing tidier and more precise bounding boxes compared to Faster-RCNN and Mask-RCNN.

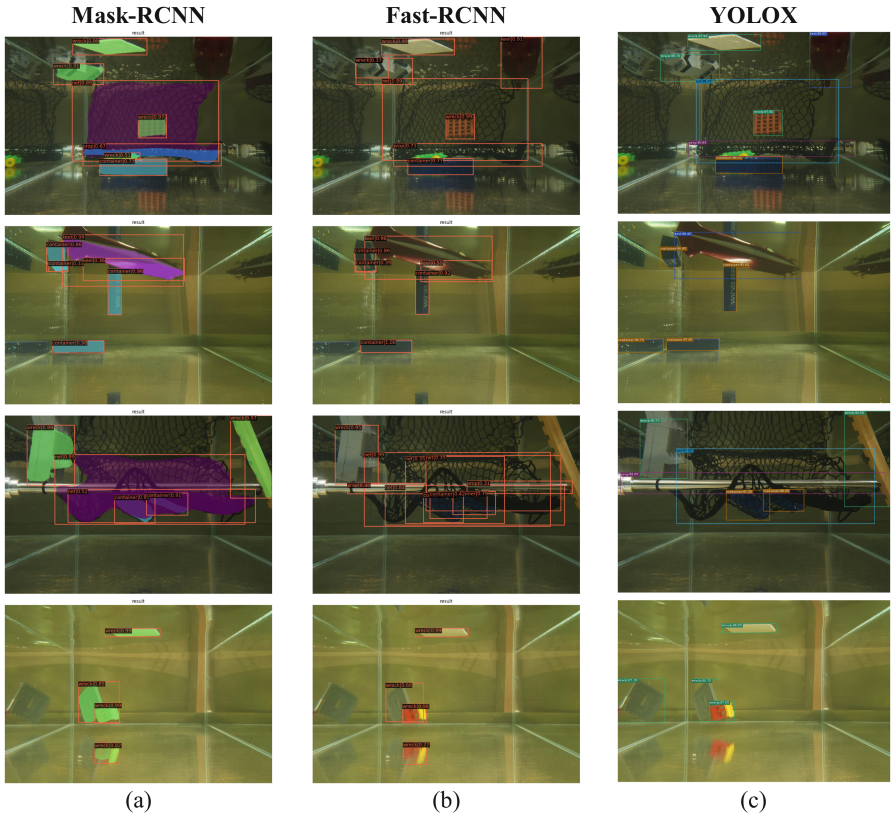


Fig. 5. Comparison of three deep learning methods for underwater object detection on the original images

Furthermore, the impact of image enhancement on object detection performance depends on the integration strategy and the specific selection of enhancement and detection methods. Specifically, CLAHE enhances the detection accuracy of Faster-RCNN,

Table 1. Evaluation of deep learning methods with or without image enhancement in terms of mAP50, mAP75 and mAP in bright enviroment.

mAP50/mAP75/mAP	Faster-RCNN			Mask-RCNN			YOLOX		
Original images	0.899	0.780	0.688	**0.901**	0.778	0.688	0.938	0.915	**0.844**
With CLAHE	**0.926**	0.782	**0.695**	0.866	**0.803**	**0.710**	0.945	0.846	0.809
With Fusion	0.903	**0.798**	0.623	0.881	0.789	0.679	0.955	0.889	0.832
With WaterNet	0.883	0.763	0.652	0.884	0.791	0.692	**0.961**	**0.921**	0.807

Table 2. Evaluation of deep learning methods with or without image enhancement in terms of mAP50, mAP75 and mAP in low light environment.

mAP50/mAP75/mAP	Faster-RCNN			Mask-RCNN			YOLOX		
Original images	0.882	0.591	0.56	0.786	0.511	0.519	0.936	0.757	**0.730**
With CLAHE	0.862	**0.660**	0.555	0.864	**0.744**	0.597	0.398	0.237	0.245
With Fusion	**0.897**	0.491	**0.560**	**0.886**	0.725	**0.622**	0.930	**0.795**	0.729
With WaterNet	0.860	0.453	0.504	0.872	0.583	0.576	**0.948**	0.774	0.724

Table 3. Object detection in the dark environment with laser light source

Methods	mAP50	mAP75	mAP
Faster-RCNN	**0.769**	0.415	0.448
Mask-RCNN	0.737	0.458	**0.474**
YOLOX	0.563	**0.488**	0.468

while WaterNet benefits the performance of YOLOX. However, it should be noted that the selected image enhancement methods do not yield improvements in Mask-RCNN.

When evaluating the detection accuracy with more restrictive criteria such as mAP75 and mAP, YOLOX consistently demonstrates superior performance. Under the mAP75 criteria, the integration of Fusion, CLAHE, and WaterNet enhances the precision of Faster-RCNN, Mask-RCNN, and YOLOX, respectively. When considering the mAP criteria, CLAHE proves to be effective when combined with Faster-RCNN and Mask-RCNN. However, none of the image enhancement methods yield improvements in the detection accuracy of YOLOX.

4.2 Object Detection in Low-Light Environment

In general, object detection performance tends to be inferior in low light conditions compared to bright conditions. The reduced contrast between objects of interest and the background in low light conditions poses challenges for object detection methods to accurately recognize objects. However, YOLOX consistently outperforms other object detection methods when applied to both original images and images enhanced using Fusion and WaterNet methods (Table 2).

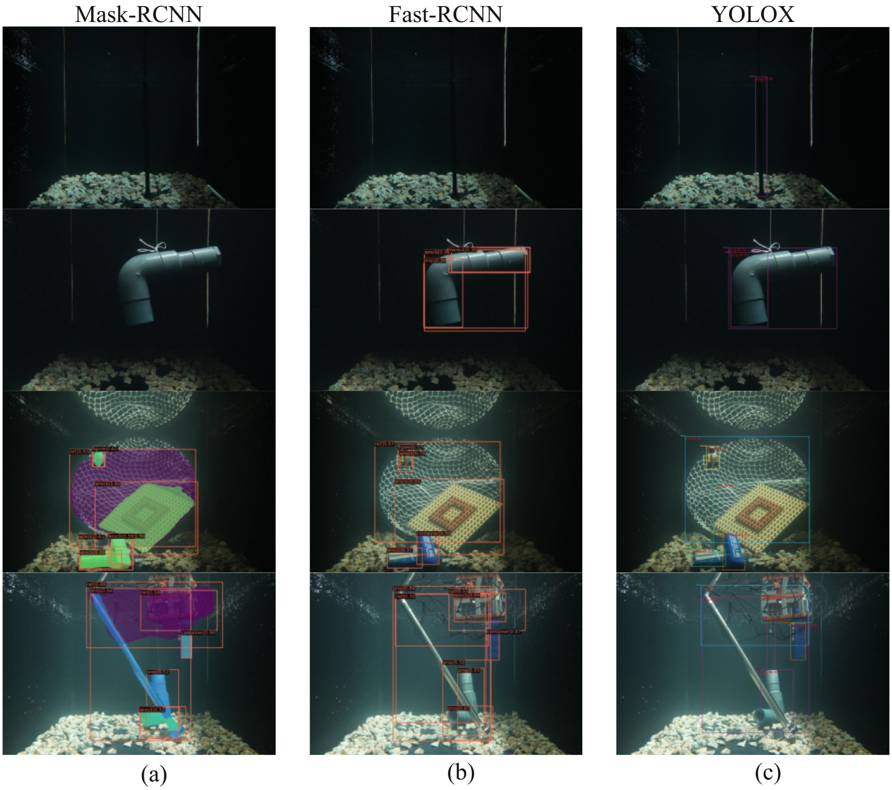


Fig. 6. Visualised object detection results generated by (a) Mask-RCNN, (b) Fast-RCNN, and (c) YOLOX.

An example of applying three DL-based object detection methods on original images is given in Fig. 6. As seen Faster-RCNN and Mask-RCNN either mis-detect the objects or make redundant detections, while YOLOX can precisely recognize the objects. It is surprised that CLAHE dramatically reduces the detection accuracy of YOLOX, and the possible reason is that CLAHE is based on histogram equalization which brings the distortion to the low light image and mislead the object detection methods to make decision. Also, similar to the experimental results in Sect. 4.1, image enhancement can more or less improve the object detection accuracy in the low light condition, though it depends on the selection of image enhancement methods and the combination strategies of image enhancement and object detection methods.

4.3 Object Detection in Dark Environment

Additionally, we extended this experiment by applying the object detection methods to the image acquired in a totally dark environment. The motivation of this extended experiment is to further investigate how the existing object detection methods work in the extremely harsh environment. As seen in Figs. 7 & 8, the three object detection methods are unable to identify the object in the dark environment. However, with the support of laser light source, all three methods are able to make the better detection despite the fact that the detection accuracy (as shown in Table 3) is worse than that of prior experiments. This main reason is that there are insufficient color attributes and only shape and potential texture attributes when capturing the image under a laser light source. Consequently, the detection accuracy will inevitably decline.

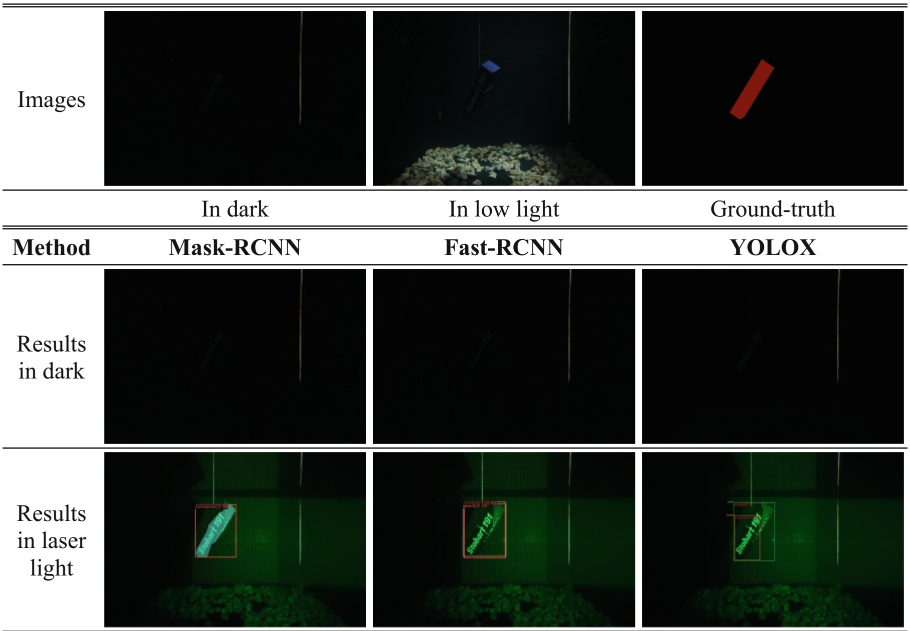


Fig. 7. Object detection in dark condition with laser as light source - Example 1.

To overcome this issue, depth information should be obtained from laser-based triangulation system and then integrated with image data for improved detection performance. Due to the page limitation, we didn't include this work in this paper. However, it will be our future focus.

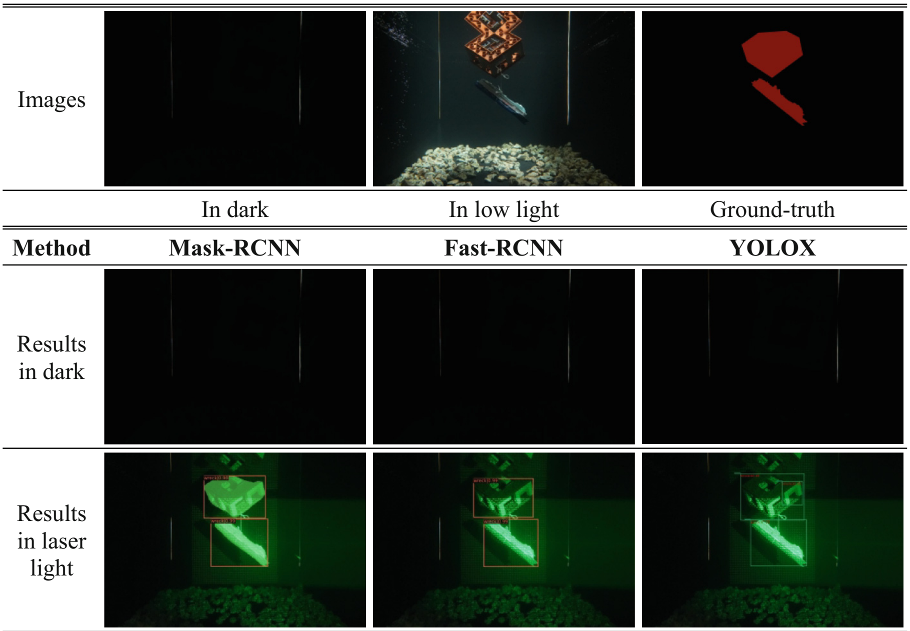


Fig. 8. Object detection in dark condition with laser as light source - Example 2.

5 Conclusion

In this project, a scotopic imaging system was built up to generate unique underwater object detection dataset consisting of 243 images with five obstacle-alike objects. Comprehensive assessment on this dataset has been carried out by three step-by-step experiments. From the experimental results, some interesting findings can be concluded as follows.

- Current experimental setting can support data acquisition in different experimental settings (bright, low-light and dark);
- Image enhancement can somehow enhance visibility of underwater images and prediction accuracy of the objects in various light conditions, but still has big room to improve;
- Existing deep-learning models can produce good object detection results in the bright and low light underwater environment;
- There is a big challenge for object detection models to make right predication in the dark environment;
- Laser can be useful for object detection as an additional light source in the dark environment and potentially provide supplementary information of depth.

Acknowledgement. This work was partially funded by the Office of Naval Research.

References

1. Braginsky, B., Guterman, H.: Obstacle avoidance approaches for autonomous underwater vehicle: Simulation and experimental results. IEEE J. Oceanic Eng. **41**, 882–892 (2016)
2. Neves, G., Ruiz, M., Fontinele, J., Oliveira, L.: Rotated object detection with forward-looking sonar in underwater applications. Expert Syst. Appl. **140**, 112870 (2020)
3. Pranitha, B., Anjaneyulu, L.: Review of research trends in underwater communications—a technical survey. In: Presented at the 2016 International Conference on Communication and Signal Processing (ICCSP) (2016)
4. Ostashev, V.E.: Sound propagation and scattering in media with random inhomogeneities of sound speed, density and medium velocity. Waves Random Media **4**, 403 (1994)
5. Life, M.: Mitigation of underwater anthropogenic noise and marine mammals: the 'death of a thousand'cuts and/or mundane adjustment? Mar. Pollut. Bull. **102**, 1–3 (2016)
6. Dairi, A., Harrou, F., Senouci, M., Sun, Y.: Unsupervised obstacle detection in driving environments using deep-learning-based stereovision. Robot. Auton. Syst. **100**, 287–301 (2018)
7. Yan, Y., Zhao, H., Kao, F.-J., Vargas, V.M., Zhao, S., Ren, J.: Deep background subtraction of thermal and visible imagery for pedestrian detection in videos. In: Presented at the Advances in Brain Inspired Cognitive Systems: 9th International Conference, BICS 2018, Xi'an, China, July 7–8, 2018, Proceedings 9 (2018)
8. Fang, Z., Ren, J., Sun, H., Marshall, S., Han, J., Zhao, H.: SAFDet: A semi-anchor-free detector for effective detection of oriented objects in aerial images. Remote Sens. **12**, 3225 (2020)
9. Li, Y., et al.: CBANet: an end-to-end cross band 2-D attention network for hyperspectral change detection in remote sensing (2023)
10. Song, P., Li, P., Dai, L., Wang, T., Chen, Z.: Boosting R-CNN: reweighting R-CNN samples by RPN's error for underwater object detection. Neurocomputing **530**, 150–164 (2023)
11. Liu, C., et al.: A dataset and benchmark of underwater object detection for robot picking. In: Presented at the - 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (2021). https://doi.org/10.1109/ICMEW53276.2021.9455997
12. Hong, J., Fulton, M., Sattar, J.: TrashCan: a semantically-segmented dataset towards visual detection of marine debris (2020)
13. Jiang, L., et al.: Underwater species detection using channel sharpening attention. In: Presented at the Proceedings of the 29th ACM International Conference on Multimedia (2021)
14. Pedersen, M., Bruslund Haurum, J., Gade, R., Moeslund, T.B.: Detection of marine animals in a new underwater dataset with varying visibility. In: Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
15. Lin, H., Men, H., Yan, Y., Ren, J., Saupe, D.: Crowdsourced quality assessment of enhanced underwater images - a pilot study. In: Presented at the - 2022 14th International Conference on Quality of Multimedia Experience (QoMEX) (2022). https://doi.org/10.1109/QoMEX55416.2022.9900904
16. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. 38, 35–44 (2004)
17. Saleem, A., Beghdadi, A., Boashash, B.: Image fusion-based contrast enhancement **2012**, 1–17 (2012)
18. Li, C., et al.: An underwater image enhancement benchmark dataset and beyond. IEEE Trans. Image Process. **29**, 4376–4389 (2019)