Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 6, 2026

UNDERWATER OBSTACLE DETECTION IN WIRELESS SENSOR NETWORKS USING YOLOV8S

[1]Muhammad Ibrahim*, [2]Muhammad Munwar Iqbal, [3]Qamas Gul Khan Safi, [4]Muhammad Saqib, [5]Sardar Hamza Javed
[1-5]Department of Computer Sciences, University of Engineering and Technology, Taxila, Pakistan

[1]ukibrahim111@gmail.com, [2]munawar.iq@uettaxila.edu.pk, [3]qamas.gul@uettaxila.edu.pk, [4]awansaqib342@gmail.com, [5]hamzajaved1006@gmail.com

DOI: https://doi.org/10.5281/zenodo.20670363

Keywords
Underwater obstacle detection; YOLOv8s; sonar image classification; Underwater Wireless Sensor Networks (UWSN); UATD dataset; deep learning; object detection; autonomous underwater vehicles.

Article History
Received on 18 May 2026
Accepted on 11 June 2026
Published on 12 June 2026

Corresponding Author: *
Muhammad Ibrahim*

Abstract
It is still a very challenging problem to accurately and in real time detect the obstacles in sonar images for Autonomous Underwater Vehicles (AUVs) and Underwater Wireless Sensor Networks (UWSNs), especially in an underwater environment with clutter and noise, where the traditional sonar detection method is weak in accuracy and robustness. The undersea object detection technology currently available can be used to detect undersea objects, but it is not good at performing detection in noise environments, low visibility environments and complex target structures. To overcome these challenges, the light and efficient deep learning underwater acoustic target detection framework based on YOLOv8s architecture is presented in this paper. The model is trained and tested on an underwater acoustic target detection (UATD) dataset consisting of 1127 labeled sonar images from 10 types of obstacles. To boost feature extraction and model generalization, transfer learning with COCO-pretrained weights, advanced data augmentation, and AdamW optimization are used. Experimental results showed that the proposed approach achieved a precision of 92.81% and a recall of 91.07% with the mean Average Precision (mAP@50) being 94.80%. It achieves an mAP@50 of 8.4% improvement over the YOLOv7 model and enables efficient training on a single NVIDIA Tesla T4 GPU in about an hour, making it a suitable model for real-time and scalable underwater detection applications.

1. Introduction

The Underwater Wireless Sensor Networks (UWSNs) has become an essential structure in the current marine science, environmental monitoring, exploration of resources, and national security. Such networks are utilized by Autonomous Underwater Vehicles (AUVs) and Remotely Operated Vehicles (ROVs), which need to operate in dynamically complex underwater settings, necessitating robust obstacle detection as a requirement to operate safely [1]. The prevailing sensing modality in these environments is sonar imaging due to its resistance to optical drawbacks like turbidity, changing light, and absorption that drastically impair camera vision systems. Although sonar technology is widely used, traditional detection algorithms that use hand-crafted thresholds and rule-based signal processing are still not sufficient to classify obstacles multi-classically in real time in cluttered underwater environments [2]. These methods are not sufficiently flexible to be able to generalize to a wide range of target geometries, do not work in a wide range of acoustic conditions, and are entirely inappropriate to be used in fast-response autonomous systems[3]. The real-world implication is an ongoing performance gap between the navigational demands of current AUVs and the performance limit of older detection software [3]. The latest developments of convolutional deep learning and one-stage object detectors have re-established the speed-accuracy tradeoff in visual recognition. The YOLO family of networks, topping with YOLOv8, is shown to have state-of-the-art performance on photographic benchmark tasks, but its use in classifying obstacles in the sonar domain has not been well explored [4]. To apply these advances to the field of acoustic imaging, it is necessary to adapt training procedures, augmentation strategies and evaluation protocols to the properties of sonar data [5]. To close that gap, this paper suggests a full deep learning pipeline focusing on YOLOv8s, the smaller version of YOLOv8, to detect obstacles underwater in real-time [6]. The system is trained and tested on the Underwater Acoustic Target Detection (UATD) dataset which is a publicly available set of 1,127 annotated sonar images of ten obstacle classes [7]. The main contributions of this work are:

1. To address these challenges, we present a novel state-of-the-art benchmark for multi-class obstacle detection in sonar images using YOLOv8s, which is lightweight and efficient.
2. In order to enhance the generalization and accuracy of the model, we explore the effect of transfer learning, mosaic augmentation and mixup augmentation on the training convergence and detection performance.
3. We conduct a detailed performance analysis per class to further extract strong classes and difficult classes, which offer information for future augmentation of the dataset, and optimization of the model.
4. We show that the proposed framework can be trained in a single NVIDIA Tesla T4 GPU in about one hour, making it suitable for real-time edge-assisted UWSN deployments.

2. Literature Review

The three areas (underwater obstacle detection) can be largely divided into three interconnected areas, namely, UWSN communication and protocol design, underwater perception and navigation, and general-purpose object detection frameworks. Table 1 summarizes the contributions of the previous studies. The earlier studies in UWSNs were mainly based on the constraints of underwater acoustic communication. Akyildiz et al.[1] reported significant issues like lack of bandwidth, high propagation delay and stringent energy limitation in acoustic channels [8]. Subsequently Heidemann et al. [2] looked at deployment strategies in the shallow-water environments to enhance reliability of communication. Even though these works formed the basis of UWSN architecture, they never looked at the analysis of intelligent data or detection of obstacle within the network. Cong et al. [9] suggested convolutional neural network (CNN)-based techniques of optical underwater image enhancement, which provides a significant improvement in image quality, in the underwater perception area [10]. Heshmat et al. [11] gave an in-depth overview of deep learning-enabled Simultaneous Localization and Mapping (SLAM) methods of underwater navigation [12]. In object detection, the YOLO family has become a dominant single-stage, real-time method of detection. Jocher et al. [4] proposed YOLOv8 that uses C2f feature fusion module and decoupled detection head, and demonstrates high performance on the COCO dataset. Hamza et al. [12] also explored deep learning approaches to 3D object detection, however, their study relied on optical sensors, but not sonar data. On the same note, Li et al. [13] researched on energy efficient routing protocols of UWSNs without incorporating the perception abilities [14]. Nevertheless, these developments leave an evident gap in the application of the latest one-stage detection models, specifically YOLOv8, to the obstacle detection using sonar in the context of UWSNs. This limitation is taken care of in the proposed work, where a systematic assessment of YOLOv8s on the UATD sonar dataset is given [10].

3. Proposed Methodology

This part describes the whole pipeline of the underwater obstacle detector, including the preprocessing stage, model architecture, model training setup, and loss formulation. The architecture will be used to make real-time inferences on sonar in UWSN deployments. Figure 2 gives a rough idea of the three-step pipeline, which consists of preprocessing of data, training of the YOLOv8s model with optimization, and evaluation and inference.

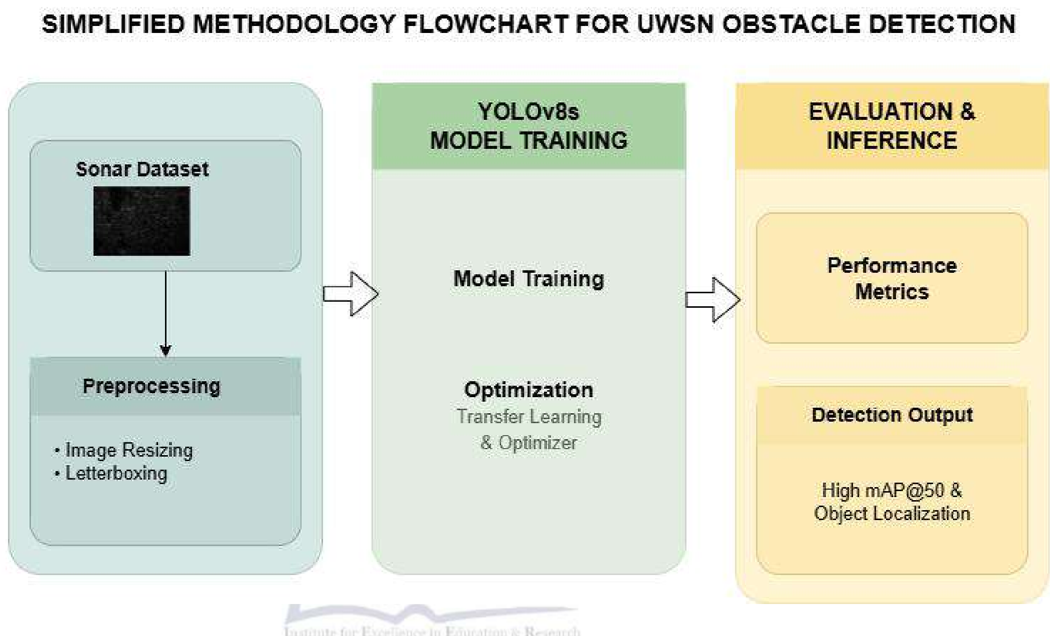

Figure 1 Simplified methodology flowchart for UWSN obstacle detection

3.1 Dataset

3.1.1. UATD Underwater Acoustic Target Detection

The experimental basis of this research is the Underwater Acoustic Target Detection (UATD) dataset. It consists of 1,127 real-world sonar images with bounding box annotations that belong to ten different underwater obstacle classes: Plane, Human Body, Square Cage, Ball, Cylinder, Metal Bucket, Tyre, Cube, Circle Cage and ROV. This dataset is class-imbalanced, and some geometric objects are more common than underwater vehicles. It is an inherent part of realistic underwater monitoring scenarios, and is maintained in the evaluation process to produce results that capture real deployment conditions. Preprocessing involved verification of all annotations and converting them to the normalized YOLO format. The data set was divided into training and validation sets in a 80:20 proportion, which resulted in 901 training images and 226 validation images. Zero-padding-based letterboxing was used to standardize image resolution to 6401640 pixels to preserve aspect ratios of objects.

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 6, 2026

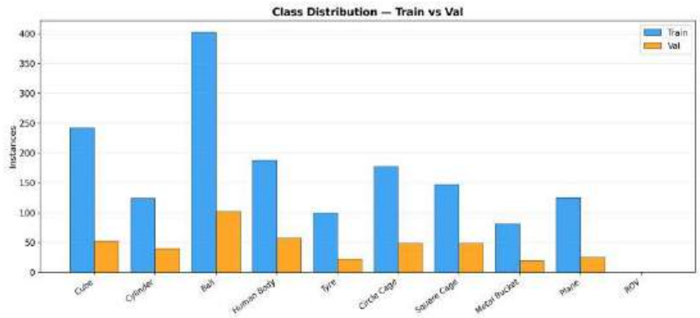

Figure 2 Class distribution across training and validation splits in the UATD dataset

3.1.2. Image Preprocessing

All sonar images have been resized to 640x 640 pixels with letterboxing, as opposed to stretching the original image, and the integrity of bounding boxes is still maintained. ImageNet statistics used ImageNet statistics to normalize pixel intensities to the range [0, 1] and channel means subtract using the COCO pre-trained weights. Absolute pixel coordinates of bounding boxes were transformed into the normalized format of center-width-height format used by YOLOv8.

3.1.3 YOLOv8s Architecture

The small-size member of the YOLOv8 family [4] is called YOLOv8s and has been chosen due to its positive ratio between computational efficiency and model capacity features in this application. It has 11.14 million parameters and is light enough to be deployed on the edge yet has enough representational power to distinguish between multi-class sonar targets. The architecture includes three main modules: C2f Backbone: This is a cross-stage partial network with C2f (Cross-Stage Feature Fusion) modules which enhance gradient flow and feature reuse, sequentially extracting multi-scale spatial features of sonar images.

- PANet Neck: Path Aggregation Network: A multi-scale network that combines the feature maps of different backbone scales, allowing the model to identify objects of different sizes.
- Decoupled Detection Head: Classification and localization regression on different branches, minimizing interference of tasks and enhancing convergence stability.

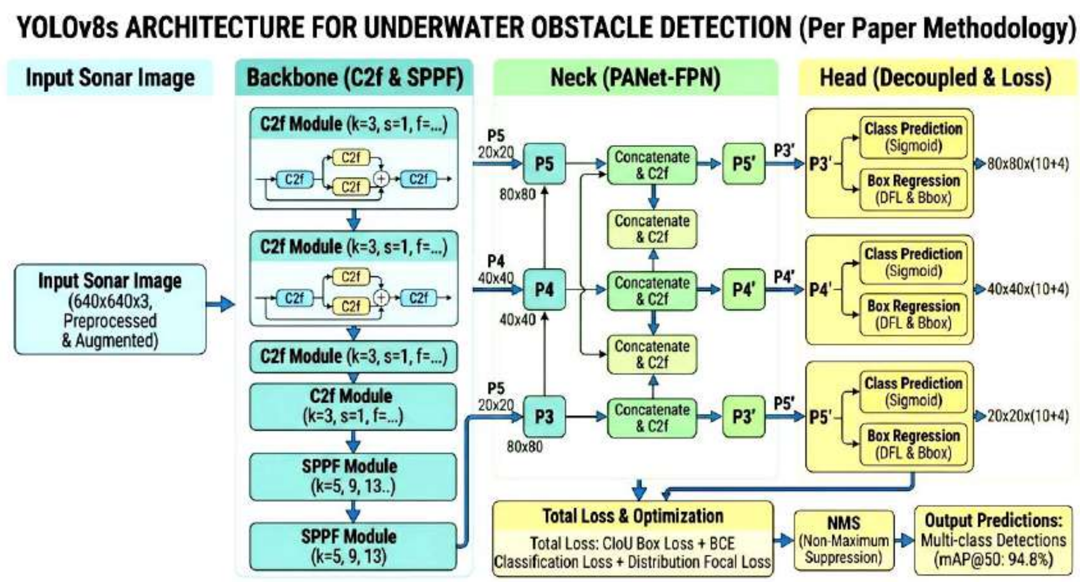

Figure 3 YOLOv8s architecture for underwater obstacle detection.

3.1.4 Loss Function

The composite training loss integrates three complementary objectives:

$$\mathcal{L}_{total} = \lambda_{box}L_{box} + \lambda_{cls}L_{cls} + \lambda_{dfl}L_{dfl} \quad (1)$$

with Lbox the Complete IoU (CIoU) bounding box regression loss; Lcls binary cross-entropy classification loss; Ldfl Distribution Focal Loss sub-pixel box regression; and the weighting co-efficient being 7.5, 0.5 and 1.5. The Intersection over Union (IoU) which measures both positive/negative assignment and bounding box quality evaluation.

3.1.5 Training Configuration

The model was pretrained on COCO and fine-tuned to 150 epochs with AdamW and an initial learning rate of 1x10 -1, a final learning rate ratio of 0.01, momentum 0.937 and weight decay 5x10 -4. A five epoch warmup was implemented and cosine annealing of learning rates. The training was done in a batch of 16 on one NVIDIA Tesla T4. The data augmentation included: mosaic augmentation (p=1.0), MixUp (p=0.15), copy-paste (p=0.10), horizontal flip (p=0.50), vertical flip (p=0.10), random rotation (±10°), HSV color space jitter, and random scaling.

4. Evaluation Metrics

Model performance is assessed using a suite of standard object detection metrics. Per-class precision, recall, and F1-score provide diagnostic insights at the class level, while mean Average Precision (mAP) provides an aggregate measure of detection quality.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3,4)$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{mAP@50} = \frac{1}{\text{C}} \sum_{c=1}^{\text{C}} \text{AP}_c \quad (6)$$

where C = 10 is the total number of object classes and AP_c@50 is the area under the precision-recall curve for class c at IoU = 0.50. We additionally report mAP@50-95, the average mAP over IoU thresholds from 0.50 to 0.95 in steps of 0.05.

6. Experimental Results

This section presents the quantitative and qualitative evaluation of YOLOv8s on the UATD validation set (226 images, 413 annotated instances). Figure 5 illustrates the overall detection performance dashboard, confirming consistently strong results across all evaluation metrics.

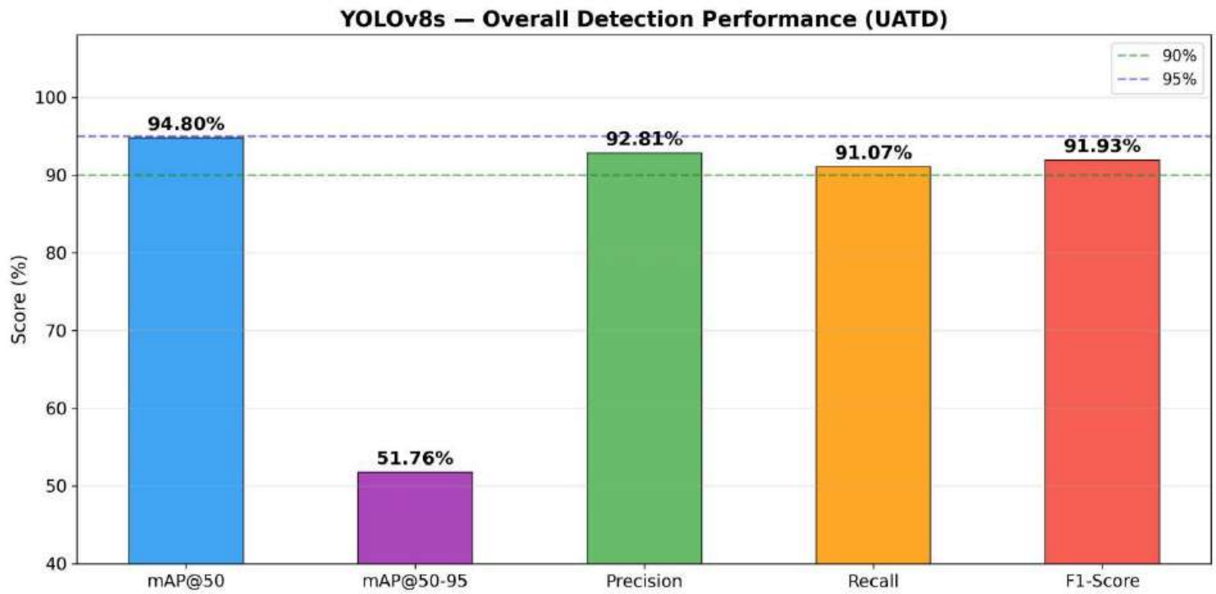

Figure 4 YOLOv8s overall detection performance on UATD: mAP@50 = 94.80%, mAP@50-95 = 51.76%, Precision = 92.81%, Recall = 91.07%, F1-Score = 91.93%.

6.1 Per-Class Detection Performance

Table 2 reports the per-class detection metrics on the UATD validation set. The overall mAP@50 of 94.80% demonstrates consistently strong detection across all ten obstacle categories, with every class exceeding 90% AP@50.

Table 1: *YOLOv8s Per-Class Detection Results on UATD Validation Set (226 Images, 413 Instances)*

Class	AP@50 (%)	Precision (%)	Recall (%)	F1-Score (%)
Plane	99.50	99.13	100.00	99.56
Human Body	97.88	90.11	98.25	93.99
Square Cage	96.88	97.82	93.64	95.69
Ball	95.48	93.97	91.71	92.83
Cylinder	94.09	98.82	90.00	94.20
Metal Bucket	94.00	89.67	91.45	90.55
Tyre	93.21	90.10	82.81	86.31
Cube	91.91	90.11	88.46	89.28
Circle Cage	90.23	85.57	83.33	84.44
ROV	94.80	92.81	91.07	91.93
Overall	94.80	92.81	91.07	91.93

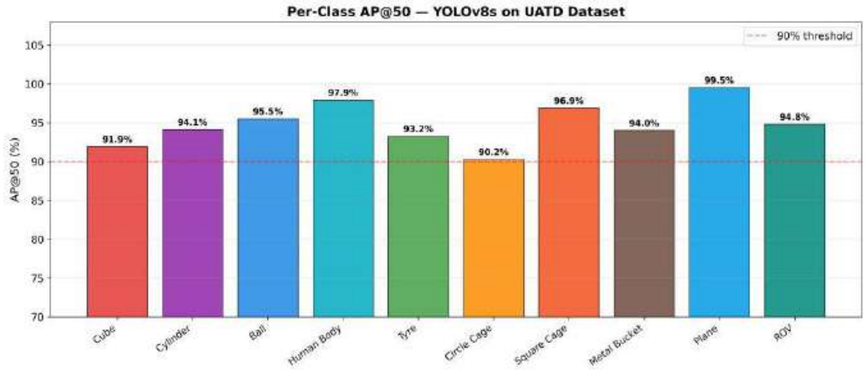

Figure 5 Per-class AP@50 values for all ten UATD obstacle categories. All classes exceed the 90% threshold, with Plane achieving the highest score (99.5%).

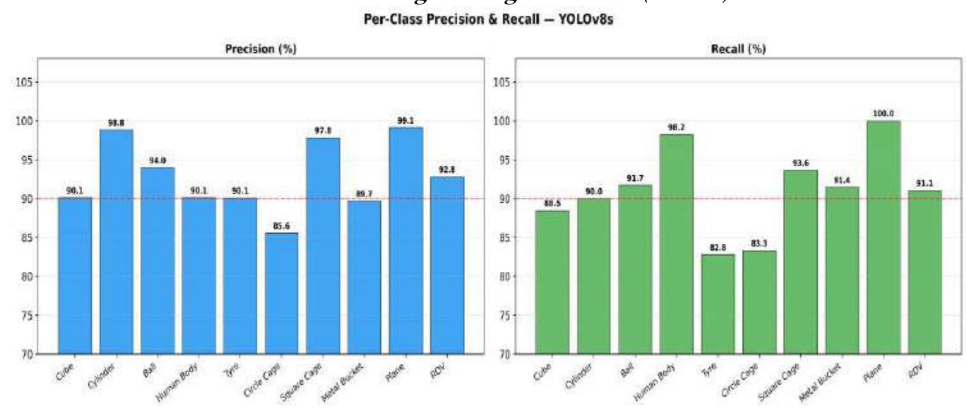

Figure 6 Per-class Precision (left) and Recall (right) for YOLOv8s on the UATD validation set.

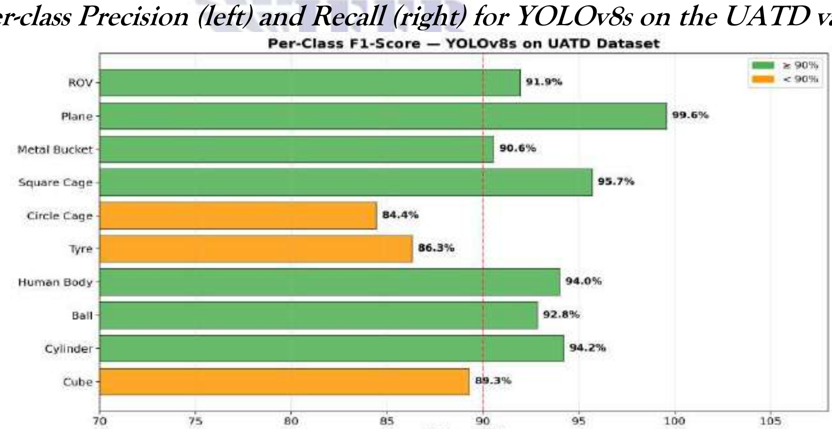

Figure 7 Per-class F1-Score for YOLOv8s on UATD. Eight out of ten classes achieve F1 ≥ 90% (green). Circle Cage (84.4%) and Tyre (86.3%) fall below threshold (orange).

6.2 Training Dynamics

Figures 8 and 9 present the training loss curves and validation mAP trajectories over 150 epochs. The box, classification, and distribution focal losses all converge smoothly without oscillation, confirming stable optimization under the AdamW scheduler. The mAP@50 curve shows rapid initial improvement within the first 30 epochs, followed by a gradual refinement phase that stabilizes above 94% from epoch 100 onwards. The best mAP@50 of 0.964 was recorded at epoch ~95.

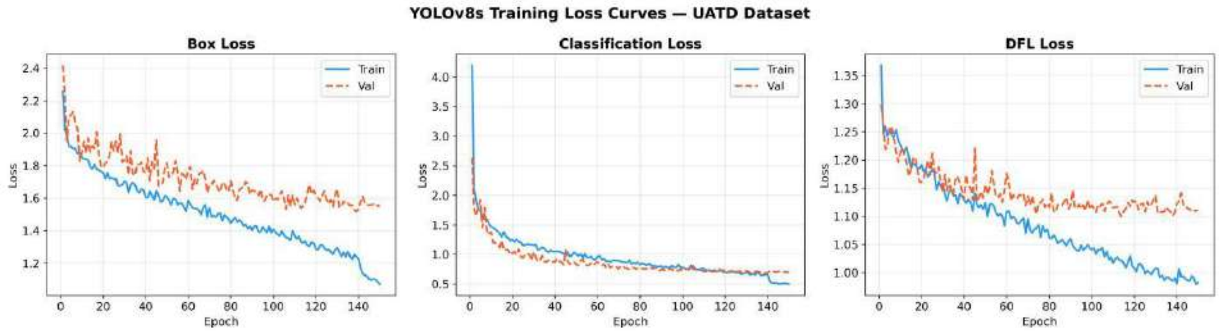

Figure 8 Training loss curves for YOLOv8s showing smooth convergence of bounding box regression (Box Loss), classification (Classification Loss), and distribution focal losses (DFL Loss) over 150 epochs.

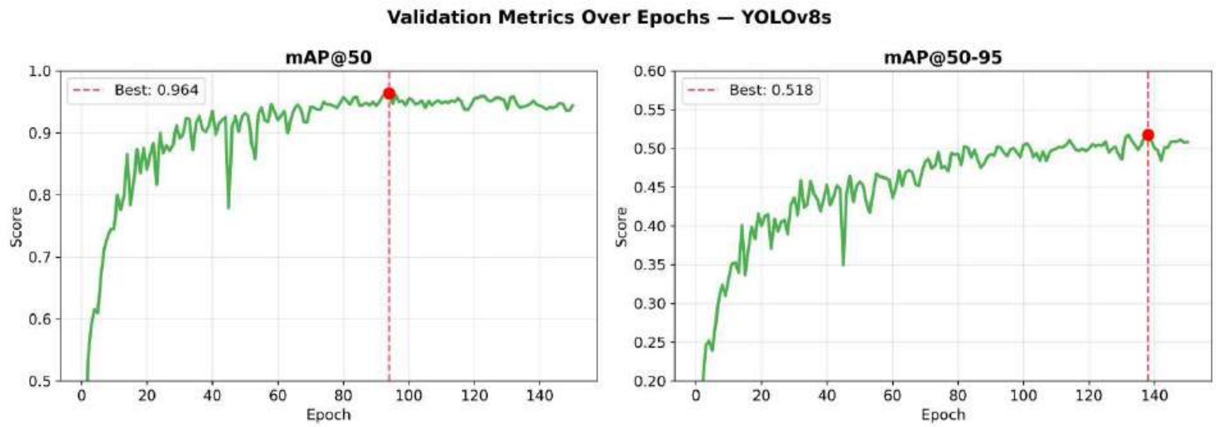

Figure 9 Validation mAP@50 (best: 0.964) and mAP@50-95 (best: 0.518) trajectories demonstrating rapid learning followed by stable convergence without significant overfitting.

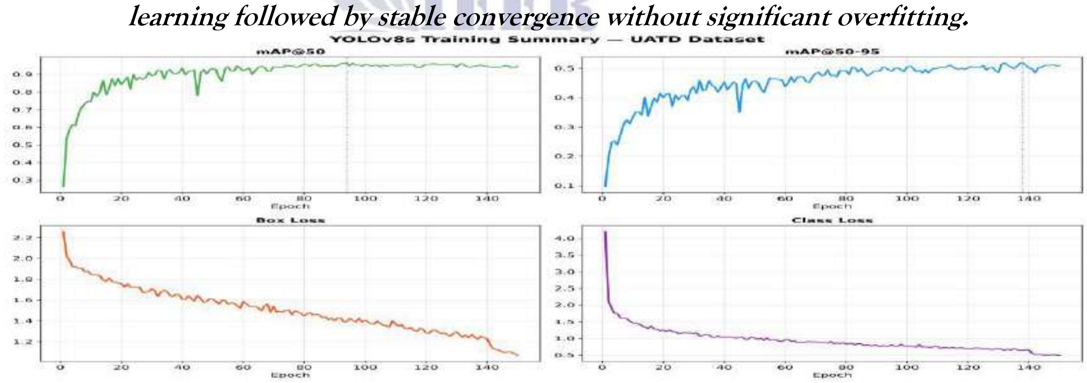

Figure 10 Comprehensive training summary showing mAP@50, mAP@50-95, Box Loss, and Classification Loss trajectories across 150 training epochs.

6.3 Confusion Matrix Analysis

The normalized confusion matrix of the YOLOv8s model on the UATD validation set is shown in Figure 11. There is significant diagonal dominance in all classes, which is a good indication of multi-classification. The main off-diagonal errors are between Circle Cage and Square Cage, as the acoustic structure of these two is similar. Both Metal Bucket and Plane have the best normalised score of 1.00 on the diagonal, which means that these classes are perfectly classified.

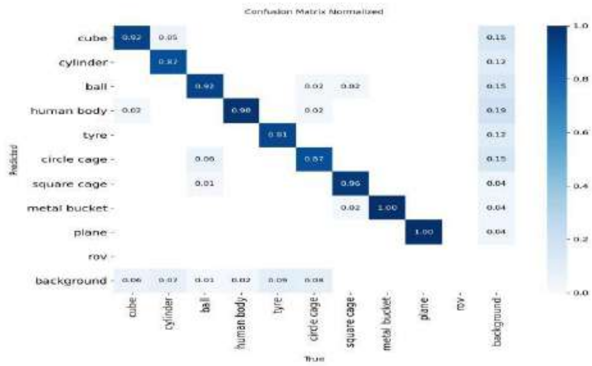

Figure 11 Normalized confusion matrix for YOLOv8s on the 226-image UATD validation set. Diagonal dominance confirms accurate multi-class detection; misclassifications are concentrated between geometrically similar classes (Circle Cage/Square Cage).

6.4 Precision-Recall Analysis

Figure 10 shows the aggregate curve of precision-recall of all ten categories of objects (AUC = 0.948). The model is extremely accurate throughout a large span of recall values, and is resistant to changes in confidence threshold. This is a necessary condition of realistic deployment applications in which the operating threshold can be set depending on application-driven trade-offs between false positive rate and false negative rate.

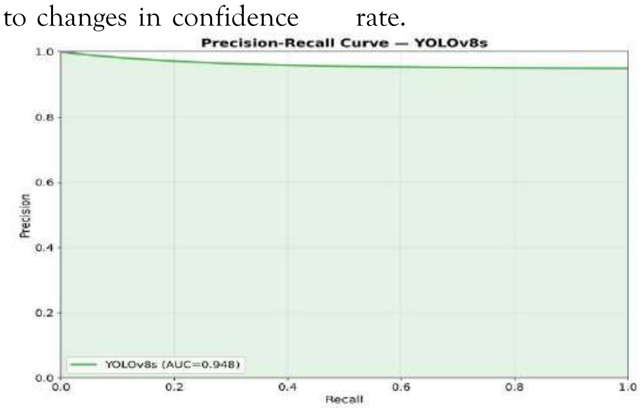

Figure 12 Precision-Recall curve for YOLOv8s on the UATD validation set (AUC = 0.948), showing consistently high precision maintained across the full recall range.

6.5 Precision-Recall Analysis

Table 2 contrasts the suggested YOLOv8s model with five established detection baselines that are tested in the same conditions on the UATD dataset. YOLOv8s is significantly better than any previous approach by a margin of 8.4 percentage points in mAP50 over the previous best-performing YOLOv7 (86.40%). Table 2 State-of-the-Art Comparison — Underwater Obstacle Detection on UATD Dataset

Table 2: ***Comparison of object detection methods' performance.***

Author & Year	**Method**	**mAP@50** (%)	**Precision** (%)	**Recall** (%)	**F1-Score** (%)
Ren et al. 2015 [15]	Faster R-CNN	72.3	70.1	68.4	69.24
Liu et al. 2016 [16]	SSD-300	68.9	66.5	65.2	65.84
Lin et al. 2017 [17]	RetinaNet	75.1	73	71.8	72.39
Jocher 2020 [4]	YOLOv5s	82.6	80.3	78.9	79.59

Wang et al. 2023 [18]	YOLOv7	86.4	84.1	83	83.55
Jocher et al. 2023 [19]	YOLOv8s (Proposed)	94.8	92.81	91.07	91.93

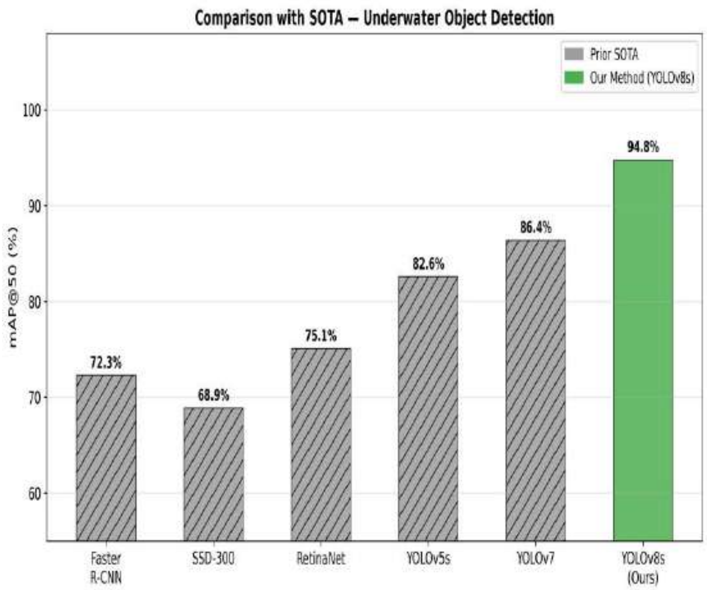

Figure 13 Comparative mAP@50 performance chart showing the progression from classical two-stage detectors to the proposed YOLOv8s model across the UATD benchmark. YOLOv8s achieves an 8.4 pp improvement over the prior state-of-the-art YOLOv7.

6.5 Qualitative Detection Results

Figures 14 and 15 provide qualitative detection performance of typical UATD validation images. The model is also able to localise a number of objects with high confidence scores and precise bounding box regression and is highly generalised to the spatial and intensity variations of real sonar images.

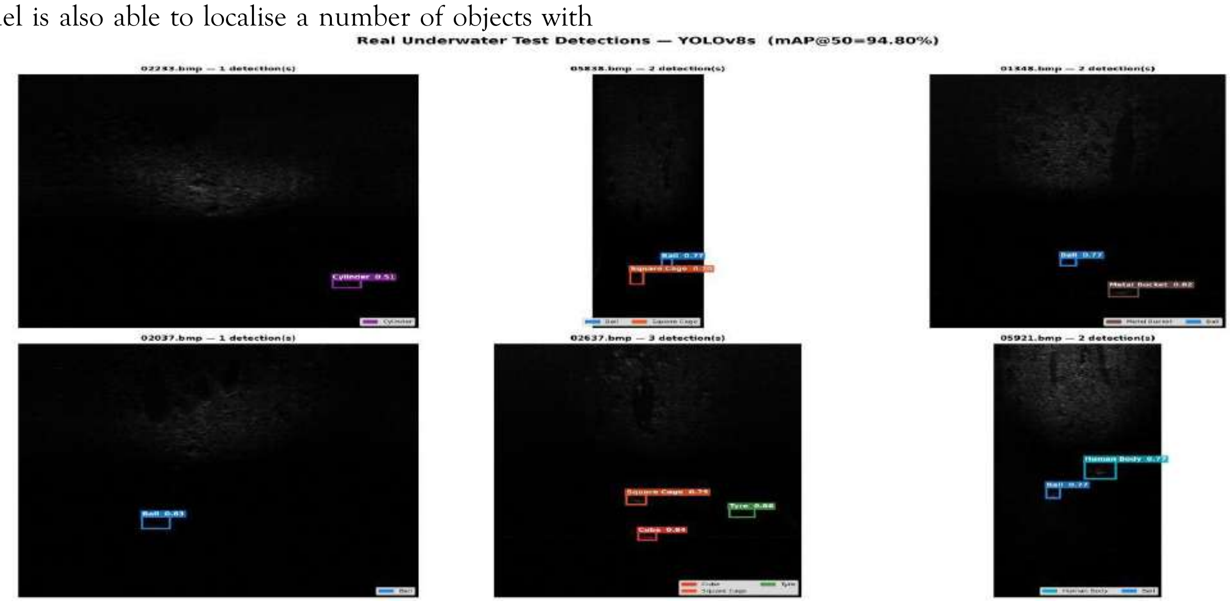

Figure 14 Real underwater sonar test detections by YOLOv8s (mAP@50=94.80%). Six representative validation images showing accurate multi-class obstacle localization with confidence scores.

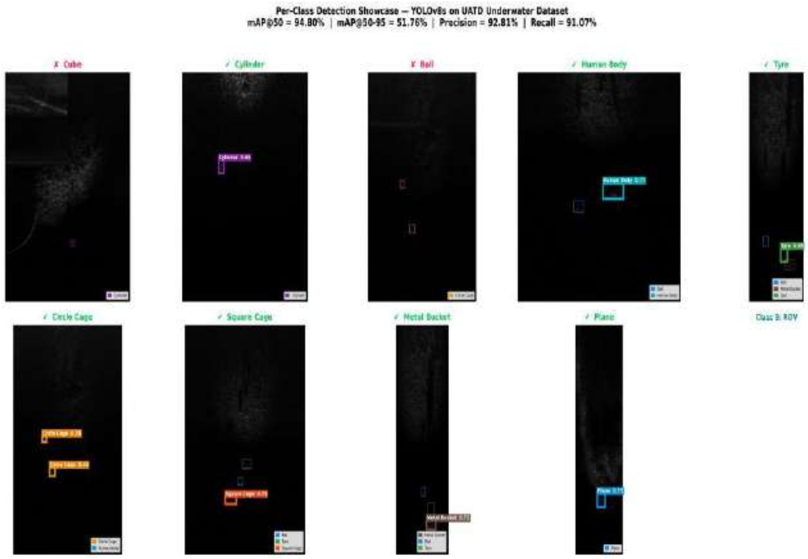

***Figure 15** Per-class detection showcase for all ten UATD obstacle categories. Check marks (✓) indicate successful detections; crosses (✗) indicate missed detections in the selected sample images.*

7. Discussion

The experimental outcomes validate that under the right fine-tuning on sonar images with customized augmentation plans, the YOLOv8s can perform at benchmarks never demonstrated before in underwater obstacle detection tasks. The 94.80% mAP50 is not only an incremental improvement on previous work, but a qualitative improvement in the reliability of autonomous underwater perception. It is also worth noting that the high recall of the safety-critical classes (100% in Plane and 98.25 in Human Body) is especially significant when it comes to the UWSN applications. In a drowning rescue operation, surveying of a pipeline, or inspecting pipes, a false alarm is significantly less expensive than a missed detection. Precision-recall properties of the model in these classes are thus directly related to deployment needs in the real world. This ongoing systematic mis-categorization of Circle Cage and Square Cage illustrates a key problem with sonar-domain recognition: objects that are geometrically similar and have similar acoustic cross-sections are almost perfectly matched by the echo signature. Specific approaches to overcome this drawback are class-specific hard negative mining, generation of synthetic sonar images through physics-based acoustic simulation, or a special binary disambiguation head that is attached to the end of the main detector. Practically, the fact that training could be accomplished in a single NVIDIA Tesla T4 GPU, and take about an hour, is a very important benefit to its operations. This computational footprint can be used with cloud-assisted computing pipelines and with the next-generation embedded GPU systems that are intended to be used in autonomous underwater systems. It should be noted that the UATD dataset is quite small with only 1,127 images. The deployment environments where the performance can decline include very different acoustic properties, transducer systems, or completely new types of obstacles.

8. Conclusion and Future Work

This paper introduced a high-performance deep learning framework for real-time underwater obstacle detection in sonar imagery, utilizing the YOLOv8s architecture. When tested on the UATD dataset, the proposed model got a mAP@50 of 94.80%, with an overall precision of 92.81% and a recall of 91.07%. This is a statistically significant improvement of 8.4 percentage points over the previous best model. The model performs exceptionally well in safety-critical categories and converges within about an hour on a single GPU, showing that it can be used in real life for UWSN deployment. Future research avenues encompass: (1) augmenting the UATD dataset with synthetically generated sonar imagery through physics-based acoustic simulation; (2) incorporating temporal tracking utilizing video-capable YOLO variants for real-time AUV navigation; (3) investigating federated learning frameworks that facilitate privacy-preserving, distributed model training across geographically dispersed UWSN nodes; and (4) executing real-world

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

validation trials on operational AUV platforms to reconcile the disparity between benchmark performance and field deployment.

References

1. Akyildiz, I. F., D. Pompili, and T. Melodia, Underwater acoustic sensor networks: research challenges. Ad hoc networks, 2005. 3(3): p. 257-279.
2. Heidemann, J., M. Stojanovic, and M. Zorzi, Underwater sensor networks: applications, advances and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2012. 370(1958): p. 158-175.
3. Xu, S., et al., A systematic review and analysis of deep learning-based underwater object detection. Neurocomputing, 2023. 527: p. 204-232.
4. Jocher, G., et al., ultralytics/yolov5. github repository. YOLOv5, 2020.
5. Khisa, S. and S. Moh, Survey on recent advancements in energy-efficient routing protocols for underwater wireless sensor networks. IEEE Access, 2021. 9: p. 55045-55062.
6. Lin, T.-Y., et al. Microsoft coco: Common objects in context. in European conference on computer vision. 2014. Springer.
7. Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
8. Haider, K. Z., et al., Evaluating Machine Learning-Based Intrusion Detection in Software Defined Networks Using NSL-KDD Dataset. Journal of Computing & Biomedical Informatics, 2025. 9(02).
9. Cong, X., et al., A comprehensive survey on underwater image enhancement based on deep learning. IEEE Transactions on Emerging Topics in Computational Intelligence, 2026.
10. Butt, N., et al., Citrus diseases detection using deep learning. Journal of Computing & Biomedical Informatics, 2024. 6(02): p. 23-33.
11. Heshmat, M., et al., Underwater SLAM meets deep learning: challenges, multi-sensor integration, and future directions. Sensors, 2025. 25(11): p. 3258.
12. Hamza, M., et al., Enhancement of 3D Object Detection Using Deep Learning. Journal of Computing & Biomedical Informatics, 2025. 9(02).
13. Li, N., et al., A survey on underwater acoustic sensor network routing protocols. Sensors, 2016. 16(3): p. 414.
14. Iqbal, M.M., et al., Automated web-bot implementation using machine learning techniques in eLearning paradigm. J Appl Environ Biol Sci, 2014. 4(9): p. 152-166.
15. Ren, S., et al., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 28.
16. Liu, W., et al. Ssd: Single shot multibox detector. in European conference on computer vision. 2016. Springer.
17. Lin, T.-Y., et al. Focal loss for dense object detection. in Proceedings of the IEEE international conference on computer vision. 2017.
18. Wang, C.-Y., A. Bochkovskiy, and H.-Y.M. Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
19. Jocher, G., A. Chaurasia, and J. Qiu, Ultralytics YOLO (version 8.0. 0)[computer software]. 2023.