

Optimal Preprocessing for Enhancing Object Detection in Underwater Sonar Images

Divyabarathi G^{1*}, Baazil P Thampy¹, Judy MV¹, Samarjith Kar², KS Ravichandran³

¹Department of Computer Applications, Cochin University of Science and Technology (CUSAT), Kalamassery, Cochin-682022, Kerala, India, ²Department of Mathematics, National Institute of Technology, Durgapur-713209, West Bengal, India, ³Department of Mathematics, Amrita School of Physical Sciences, Amrita Vishwa Vidyapeetham, Coimbatore-641112, Tamil Nadu, India.
*Corresponding Author's Email: divyabarathi2108@gmail.com

Abstract

Sonar is an efficient and indispensable tool for exploring underwater environments under challenging conditions. The research has a crucial impact on domains such as marine exploration, environmental monitoring, defence, archaeological discoveries, resource exploration, and climate research. Extracting valuable insights from sonar images and detecting objects within obscure sonar images is tedious and challenging for both human experts and conventional machine learning models. This study explores state-of-the-art deep learning methods for enhanced underwater object detection in multibeam forward-looking sonar images, to determine optimal trade-off model between accuracy and complexity. A sequence of preprocessing steps is proposed in this work to mitigate noise and enhance images, improving the accuracy of benchmark object detection performance. This comprehensive framework for forward-looking sonar images integrates the preprocessing techniques and target detection thereby enhancing target visualization. The model localizes and predicts the target of each class by overcoming the challenges of target detecting in hazy images and imbalanced class distribution. The proposed approach yields a mAP of 96.3% with 3M parameters which implies a significant increase in efficiency in real time processing in comparison to other models. From the analysis, the proposed framework improves visual perception and enhances object localization of the targets in the sonar images.

Keywords: Deep Learning, Forward-Looking Sonar, Object Detection, Preprocessing, Underwater, YOLO.

Introduction

The underwater exploration research has a crucial impact in domains such as marine exploration, environmental monitoring, defence, archaeological discoveries, resource exploration, and climate research (1–3). In recent decades, there has been substantial progress in the technologies deployed for gaining insights from deep water environments beyond direct human reach. Cutting-edge Remotely Operated Vehicles (ROV) and Autonomous Underwater Vehicles (AUV), along with sophisticated sonar systems and optical technology, have revolutionized the field of underwater exploration (3–5). Sonar is an efficient and indispensable tool for exploring aquatic environments, even under challenging conditions. Sonar data collected by surveying the location of interest using a Sonar device which transmits beams of sound known as “pings” into water and record the echoes from bottom and any object in the water column reflected back to the receiver.

The acoustic reflectivity from ocean bed and objects are differentiated based on the strong contrasts from rocks or metal and low contrasts from sediments or marine life (6).

The sonar devices often employed for underwater research encompass Single Beam Sonar, Multibeam Sonar, Side-Scan Sonar, and Forward-Looking Sonar (FLS). These sonar devices have been instrumental in mapping the seafloor and identifying submerged features. For underwater exploration, sonar has clear advantages over alternative technologies, such as optical imaging (4). Sonar excels in low-light conditions, provides extended visual ranges and offers efficient 3D mapping capabilities in diverse environments, from coastal waters to the most profound ocean trenches. Extracting valuable insights from sonar images poses a formidable challenge due to non-uniform intensities, acoustic shadowing, speckle noise, limited resolution, and lack of colour and

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

(Received 25th June 2025; Accepted 15th October 2025; Published 30th October 2025)

texture information. It is essential to address these multifaceted hurdles to acquire comprehensive insights from sonar images (7, 8). The process of underwater object detection using sonar devices entails the identification, tracking and localization of entities submerged beneath the water's surface within the scanned area. Objects in sonar images appear in various sizes and aspect ratios depending on the angles of the sonar scan, resulting in different structures for the same object (9, 10). Detecting objects within obscure sonar images is more tedious and challenging for human experts (11).

Sonar object detection requires images of target objects along with their respective location labelled for training the model. Typically, a variety of sonar imaging devices are used to scan the seabed, each generating device-specific raw data. Subsequently, device modules convert the raw data into 2-dimensional images for preprocessing and post processing tasks such as detection and segmentation aiming to extract insights from the seabed. Finally, images collected from test basins and sea trials are annotated to train the model (8, 9).

Deep learning-based Detection techniques can be divided into two categories: one stage and two-stage detection techniques, based on whether or not a region proposal algorithm is needed to determine the target location. Specifically, while one-stage detection methods attempt to predict object bounding boxes and class labels in a single pass (12, 13), they usually do not necessitate an additional region proposal phase. Two-stage detection algorithms, on the other hand, use a region proposal algorithm in the first stage to identify possible regions of interest, and then

refine and classify these regions in the second stage (14). Deep learning-based detection methods commonly use Convolution Neural Networks (CNN) which is the primary building block of the technique which extracts intricate image features, accurately predicts both object class and spatial location.

Advancements in object detection based on deep learning techniques are designed for optical images captured in air medium. Recently several studies on detection of objects and classification are evolving. Based on the previous literature, A non-gaussian detection method for Synthetic Aperture Sonar (SAS) imagery, applying a multivariate extension of Laplace distribution was introduced (15). The iterative nature of this approach made it highly computational. Traditional techniques are unable to scale for vast data sets and lose their efficacy in range of environments. Furthermore, current research indicates that conventional methods imply iterative approaches, which increases the computational requirements for developing real-time systems (15, 16). Viola and Jones boosted classifiers cascade was suggested to reduce false-positive errors caused by acoustic artefacts in the Forward-Looking Sonar images (16). A CNN based model presented for localization of agent vehicles where the ROV was identified in real time when trained with two classes: small ROV and background image (11). A study by Valdenegro-Toro experimented detection of an object trained with a nine categories dataset captured in a lab water tank (4). An end-to-end system having detectors based on Haar features and boosted Cascades for Autonomous Underwater Vehicles was suggested.

Table 1: A Review of Relevant Literature on Object Detection in Sonar Images

Ref.	Author and Year	DL Method	Dataset	Metrics	Remarks
(16)	Sawas <i>et al.</i> , 2010	Viola and Jones boosted classifiers cascade	Synthetic data and semi synthetic data	ROC curve	Target detection with the local clutter
(17)	Kim <i>et al.</i> , 2016	Convolutional Neural Network	Agent vehicle dataset	99 % accuracy	Localization of agent vehicles (small ROV).
(4)	Valdenegro-Toro, 2016	Convolutional Neural Network	Ocean Systems Lab dataset	Accuracy to 85%. Recall:93%	Object detection and recognition in FLS images
(5)	Song <i>et al.</i> , 2019	Self-cascaded CNN (SC-CNN)	SSS images	mIOU:0.8401	Real-time object detection for SAILFISH AUV

(7)	Karimanzira <i>et al.</i> , 2020	Faster R-CNN	small dataset large dataset	ExAUV	AP : 0.98392 and P-R curve	a generalised solution based on AutoML to detect an underwater docking station
(18)	Zhang <i>et al.</i> , 2021	self-trained AutoDLwith DARTS detectors	FL-	SSDD and SCTD	mAP :92.8	self-trained target detection strategy using AutoDL
(19)	Yu <i>et al.</i> , 2021	TR-YOLOv5s		SSS web data	mAP :85.6% and macro-F2 :87.8%	Real-time target detection in SSS images
(8)	Liu <i>et al.</i> , 2022	Yolov5 with DCNet		SCTD1.0 dataset	precision : 92.5 Recall : 99.3 mAP@50: 98.1 mAP@5:95: 65.1	Ship and plane wreckage detection
(10)	Wang <i>et al.</i> , 2022	Multilevel feature fusion network (MLFFNet).	DL-Dataset and QD-Dataset		mAP :DL- 98.31 and QD-98.64 memory : 5.82GB, train time:3.65h, testing time 1.73s, parameters:32M	Multiscale features based Sonar image object detection
(20)	Yang <i>et al.</i> , 2023	Foreground enhancement network (FEN) with FCOS	sonar images from the Web		mAP: 0.851	Feature representation of foreground objects and detection
(21)	Zhang <i>et al.</i> , 2023	Yolov7 with BiFPN and Attention layer	SSS dataset		Precision: 95.5%, Recall:87.0%, mAP@.5 : 86.9%, mAP@.5:95 :55.1	Detecting small objects in SSS images

A module for real-time object detection used in SAILFISH AUV using Side scan sonar image patches of size 37 X 37 trained with Self-Cascaded CNN (5). A generalised underwater object detection solution based on Automated Machine Learning (AutoML) was employed to detect an underwater docking station trained with Faster R-CNN (7). Procedures for a self-trained target detection strategy using AutoML were investigated, with detectors implemented using the Differentiable Architecture Search algorithm with flexible search space and large inputs (18). Real-time target recognition was performed on SSS images retrieved from the web containing shipwreck and container as target using the TR-YOLOv5s model achieving 85.6% of mAP (19). Ship and plane wreckage detection using deep convolutional networks based on dual channel attention mechanism was presented (8). A MLFFNet which utilised feature correlation between multiple object categories was presented (10). A FEN to reduce false alarms caused by noise was evaluated (20). Further improvements were achieved utilising yolov7 with Bidirectional Feature pyramid network incorporating attention layer, trained on the SSS image dataset, where some occurrences of false detection were observed (21).

Table 1 shows that existing work contributed to the detection of objects in sonar images. The authors identified limitations in recognising extremely small objects, computationally expensive models affecting some inference speed, ubiquitous noise, and the absence of high-frequency information (16, 20). Whereas the deep learning techniques adapted for sonar images require customization to enhance accuracy, especially in low-contrast and noisy underwater sonar data (20, 22).

Generic object identification models have been successfully applied to natural images, demonstrating their usefulness in a range of applications and proving their adaptability and efficacy. However, because of the complicated ambient noise, identifying objects in underwater images is more difficult (23). From the review of literature, inferred that the performance of models may suffers in diverse object detection, requiring image refinement for distinguishing between targets having similar structural properties (21). Furthermore, scalability in resource-constrained environments can impact the robustness of the models. Advanced deep learning-driven object detection approaches can effectively tackle such challenges with reduced human resources.

The potential of advanced DL models for underwater sonar images needs further exploration. This paper explores the recent advancements in deep learning methods for underwater object detection in multibeam forward-looking sonar images. A combination of preprocessing steps is curated to mitigate noise and enhance images, thereby considerably improving the model's performance along with the precision of each target. The significant contributions of this study are developing an end-to-end framework for sonar images that integrates sonar image preprocessing with an advanced object detection method, evaluating the performance of various preprocessing techniques to overcome challenges in detecting objects in obscure underwater sonar images, aiming for enhanced visualization of targets, and analysis of advanced deep learning architectures proposed for detection of targets in underwater sonar environments.

Methodology

Proposed Framework

This paper presents an improved object identification system designed for sonar imaging to enhance visibility of objects and localization with precision. An end-to-end framework with sonar image processing to enhance the sonar image and object detection based on YOLOv8 network is proposed (24). A preliminary preprocessing module is used to lower noise and increase image clarity in order to enhance sonar image visibility. This preprocessing phase is essential to ensure the better-quality data is fed into the object detection module. This module yields the precise locations of target items inside the images in addition to the predictions. It also gives each identified object a class and computes a probability score that represents the degree of confidence in each

prediction. This procedure ensures the clarity and accuracy of the sonar images when searching for objects of interest.

Data Preprocessing

The original image in the dataset is raw data, making it visually challenging and time consuming to locate the objects manually. During signal recording and transmission, unintended signal-dependent speckle noise tends to degrade image quality (7). This section explores into the preprocessing approaches used to overcome these challenges and improve overall image quality.

Sonar blindness refers to a condition where a sonar system is unable to detect or properly identify objects due to specific acoustic challenges (25). This can occur when the acoustic signals sent by the sonar encounter conditions such as strong interference, noise, or absorption in the water, making it difficult for the system to effectively view or detect targets. The accuracy of identifying targets by manual setup is determined by the level of expertise involved. Due to varying Sonar visibility, the operator may fail to provide accurate information about underwater objects during the scan.

Preprocessing techniques were examined to address low visibility and noise in the UATD-Dataset. Initially, histogram equalization (26) was applied to balance the pixel distribution, but this resulted in visually unavailable details.

Subsequently, the data was converted from BMP to Lab space to assess its effectiveness for further preprocessing tasks. Also, in an analysis to increase image resolution, Gaussian filtering, linear interpolation (26) was experimented but found to be unsuitable for the data. Further analysis revealed, Mean Normalization technique enhances the image when combined with binary thresholding (26).

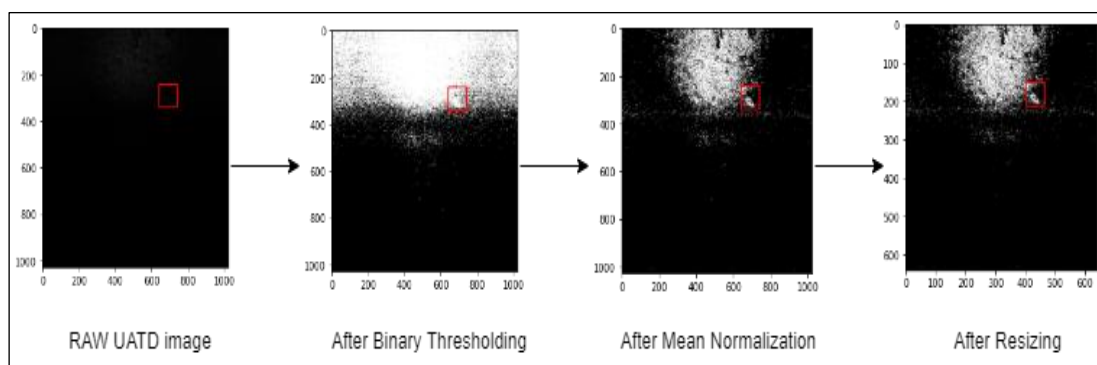


Figure 1: Preprocessing Outcomes of Sample Image Having Class Human Body from UATD Dataset

In the preprocessing module, Binary threshold is applied in the raw sonar images, the threshold is set at 1, and any pixel with a value greater than 1 is set to 255 has been identified from empirical testing as specified in equation [1]. This

$$T(I(x,y)) = \begin{cases} 255 & \text{if } I(x,y) > 1 \\ 0 & \text{otherwise} \end{cases} \quad [1]$$

Mean normalization, a statistical method is employed that takes a two-dimensional NumPy array and performs mean normalization along the specified axis. The mean normalization process involves subtracting the mean of each column (or row) from the corresponding elements in that

preprocessing step in image analysis to enhance or isolate specific features in the sonar image. It is employed to segment objects or features in an image based on their intensity levels (27).

column (or row), resulting in a new array where the mean is zero along the specified axis as shown in equation [2], where X is the original pixel value, μ is the mean pixel value, and σ is the standard deviation.

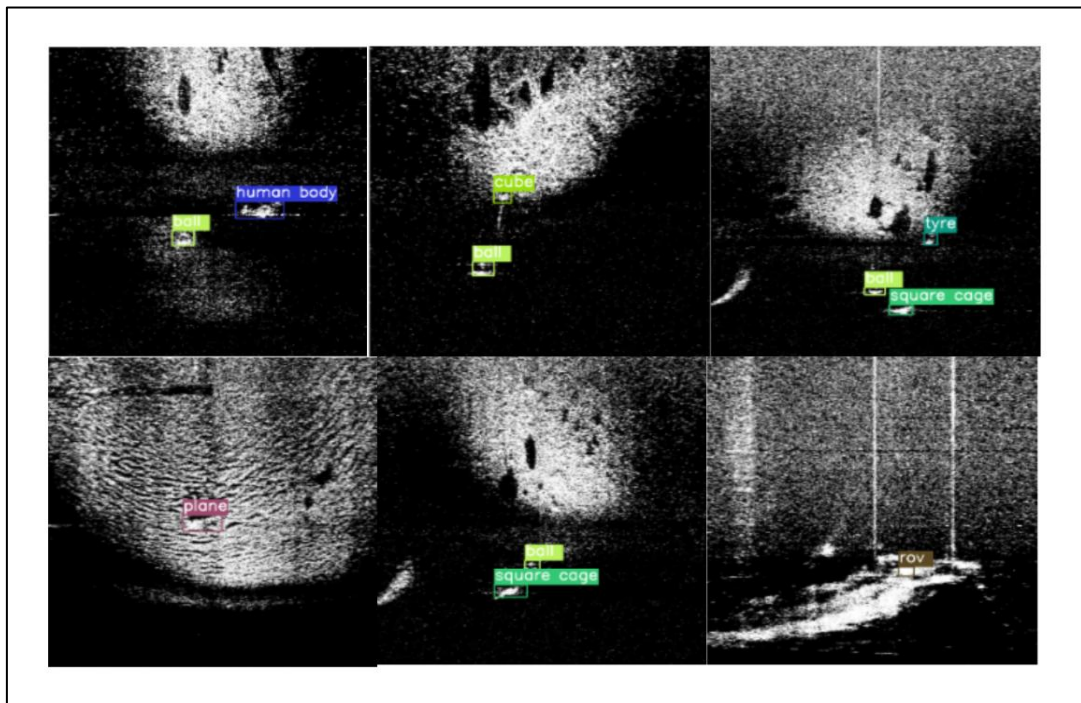


Figure 2: Sample Pre-Processed Images from the UATD Dataset with Various Class Objects

$$X_{norm} = \frac{X - \mu}{\sigma} \quad [2]$$

With our dataset, where the features had varying scales, this method performed well. The accuracy and efficiency of the analysis and modelling were enhanced by normalising them to a common scale (27).

Additionally, resize the image to 640x640 for training the model. Figure 1 explains the outcome of each stage in the Preprocessing showing the visual enhancement gradually and Figure 2 shows few sample images from Pre-processed UATD dataset with targets after the preprocessing.

SONAR Object Detection

In this section, an enhanced object detection model for sonar image is proposed based on YOLOv8 (24). The detection of objects is a critical research subject in the fields of underwater acoustics and researchers have been working to resolve it. The underwater acoustic signal is unstable due to factors such as noise, and this phenomenon considerably lowers the precision of detection.

Feature extraction is one approach to solving this challenge, and the features supplied to the classifier are critical to its effectiveness. It is inferred that when it comes to enhancing the precision of detection with localisation, the selection of the dataset, image resolution, and volume of samples may be more essential than model alteration.

YOLO (You Only Look Once) is an exceptionally rapid object identification method that exploits a single neural network. Because it examines the entire image at once rather than assessing it pixel by pixel, YOLO is usually quicker than other detection methods. Objects in an image are defined by bounding boxes which is a rectangular box specified by top-left corner coordinates and width, height to locate the target class. Each of the predicted boxes containing an object of interest assigned a confidence score indicating model's certainty. Finally, to eliminate multiple detections of the same object by retaining only the bounding box with the highest confidence score and discarding others that overlap, a post processing technique of Non- max Suppression is employed.

To detect targets from input images, YOLO divides an image into a grid, which is then divided into sections. Each area of the grid is then classified and localised, meaning that the structures and objects are identified. Subsequently, it forecasts the

location of boundary boxes. Regression-based methods are used to predict these boundary boxes. Most classification-based methods are executed in two stages: first, the Region Of Interest (ROI) is chosen, and then the CNN is trained to detect the object(s) in the areas that were selected.

Unlike previous models, YOLO's regression technique predicts the bounding boxes for the complete image concurrently. To determine which model performed better, we conducted experiments using YOLOv5 and YOLOv8.

The YOLOv5 models consist of the same three parts: Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet) in the model neck and head, and CSP-Darknet53 as the backbone (28). Sigmoid Linear Unit or SiLU as it is commonly referred to as the swish activation function, is the model's activation function. It has been applied to the hidden layers' convolution operations. In the output layer, convolution techniques have been combined with the Sigmoid activation function. Object classes, bounding boxes, and objectness scores are the three outputs that YOLOv5 outputs. Therefore, the classes loss and the objectness loss are calculated using Binary Cross Entropy (BCE). When calculating the location loss using the Complete Intersection over Union (CIoU) loss function as shown in equation [3].

$$loss = \lambda_1 L_{cls} + \lambda_3 L_{obj} + \lambda_3 L_{loc}. \quad [3]$$

Similarly, YOLOv8 performs detection of objects through the single pass whereas traditional methods of detection methods apply classifiers at multiple scales and location. Entire image is divided into grids (say, 13x13 or 19x19 cells) given as input. Each grid cell predicts the objects whose centre falls within the cell. Convolutional Neural network architecture used to extract features from Input image. The YOLO model consists of several Convolutional, pooling and fully connected layers. While last layer outputs tensor containing bounding box coordinates, class probabilities and confidence score for each grid cell. Multiple bounding boxes, namely anchors are predicted from each grid cell along with a confidence score that identifies how certain the Bounding box Prediction contains an object. The coordinates are

normalized to fall between 0 and 1. As an anchor free detection method, it predicts object bounding boxes and class probabilities based on features extracted from the input image instead of predefined anchor boxes which are priors of fixed sizes and aspect ratios.

Along with bounding boxes, each grid cell predicts the probability distribution over predefined classes. Class probabilities are calculated from the grid cell containing an object. Finally, threshold is applied to filter out low confidence prediction of the cells. Non-Max suppression is used to eliminate redundant bounding boxes while holding only the highest confidence score for each detected object. The overview of proposed approach and the outcomes at each step of object detection pipeline shown in Figure 3.

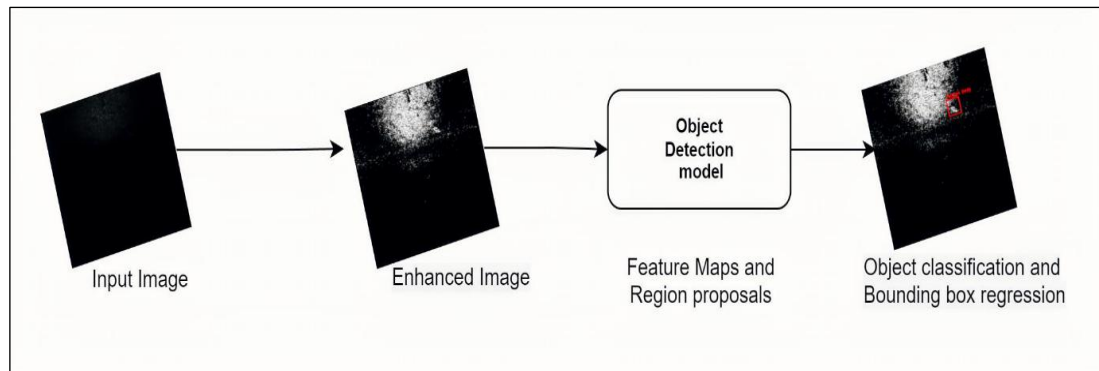


Figure 3: Outputs of the Proposed Object Detection Framework for Forward-Looking Sonar Images at Each Layer

YOLOv8 has an anchor-free detection system, whereas YOLOv5 has anchor-based approach is used in comparison. In YOLOv8, objects are predicted by their centres rather than by offsets from predefined anchor boxes. this method reduces the number of box predictions, resulting in robust non-maximum suppression. Convolutional block especially in the neck architecture, the kernel size is 1X1, which is resembles closely with ResNet block architecture. Additionally, to reduce the parameter count and tensor size by concatenating features directly in the neck without enforcing the same channel dimension [24].

The anchor-free technique to sonar image analysis improves detection accuracy in complex underwater environments by reducing false positives and enhancing object localization. The streamlined design and decreased parameter count enable faster inference times and lower computational requirements which are critical for real-time applications in AUVs and maritime surveillance systems. By tailoring for efficiency and accuracy, YOLOv8 considerably improves the capacity to detect and monitor underwater objects, which boosts operational effectiveness in sonar-based applications.

Results and Discussion

Dataset and Experimental Setup

The outcomes and feasibility of the proposed approach is illustrated with the help of a sonar image dataset containing multiple object categories named Underwater Acoustics Target Detection (UATD) dataset. It was constructed using image acquisition equipment Tritech Gemini 1200ik sonar. This is a Multibeam Forward Looking Sonar device for Underwater Object Detection experiment available publicly for researchers [29]. Data collection was performed in

a lake and shallow water environment. This dataset consists of total of 9000 MFLS images captured using Tritech Gemini 1200ik sonar that provides raw data of sonar images with annotation of 10 categories of target objects. Targets included diverse materials such as ball, circle cage, cube, cylinder, human body, metal bucket, plane, rov, square cage, tyre. Some of the sample images of UATD Dataset as shown in Figure 2 and the boxes represent the object region and target class. The domain experts have labelled the object using an annotation software for sonar images named forward-looking sonar label tool (OpenSLT) through input image stream and real time annotation [29]. The dataset used 7600 images as the training set, 800 for validation set and 800 images as the testing set. The dataset exhibits that there is an imbalance among the object in the image instances. This data set is a multi-object data set with serious seabed reverberation interference, and each image contains multiple object regions. The Windows 10 operating system and Intel(R) Core i7-7700K CPU running at 4.20GHz served as the experimental setup. The NVIDIA GeForce GTX 1050 Ti was the graphics card which was utilised. The hyperparameter settings for each model were as follows to ensure comparability. The images were resized to 640 x 640 pixels for the input size, there were eight batches, and the training period lasted for one hundred epochs.

Evaluation Metrics

Sonar object detection presents distinct challenges as compared to other types due to the inherent characteristics of underwater acoustics. This section examines a Precision, Recall and mean Average Precision(mAP) methods of evaluation for detecting objects in sonar data. Precision evaluates the accuracy of the model prediction in classifying

the instance as true positive. In this study, the measure of each target that correctly identify

actual target out of all the predicted categories. Mathematically equation [4] defines precision.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad [4]$$

Recall evaluates ability of model to correctly identify all positive instances out of the total positive instances as shown in equation [5].

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad [5]$$

Mean average precision(mAP) metric is crucial for assessing a model's ability to accurately localize objects within the image. It compares the ground-truth bounding box to the detected box and delivers a score based on equation [6]. The model determines and predicts more accurately as its mAP score increases.

$$mAP = \frac{1}{C} \sum_{i=1}^C AveragePrecision_i \quad [6]$$

Average Precision is often employed in object detection applications. It considers precision and recall across different levels of confidence thresholds.

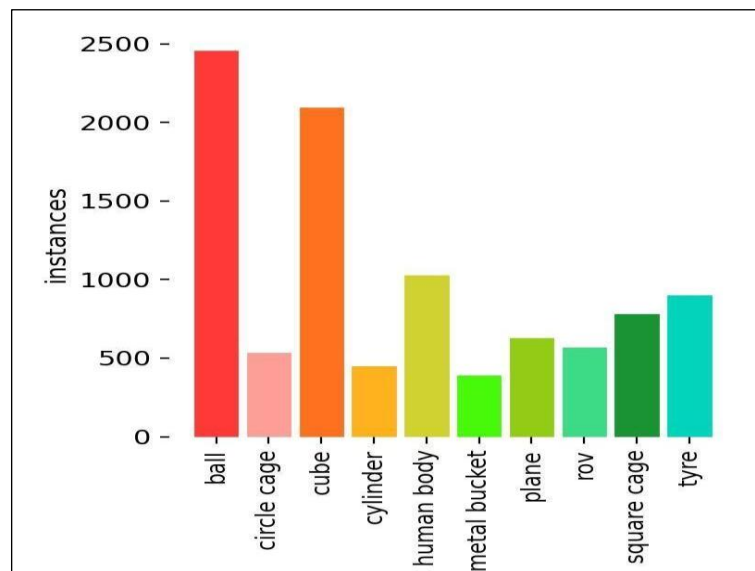


Figure 4: Bar Chart Illustrating the Distribution of Object Classes in the UATD Dataset

Training Results

In this study, deep learning models utilized the preprocessing data. Because the original sonar image contains considerable amount of nonlinear noise caused by the seabed environment and the collect device, a preprocessing module is used to minimise the noise. The variation in the number of instances of the object categories showing imbalance in the data are illustrated in Figure 4. Training the model for 100 epochs and testing our model performance using test and validation set for performance. To compare the performance, deep learning models.

Faster-RCNN, YOLOv3, YOLOv5, YOLOv8 are used. The experimental outcomes of the object detection are listed in Table 2 which shows that the yolo v3 has the lowest precision. The accuracy of yolov5 and yolov8 is similar, also it is visible that yolov8 has better GFLOPS which can bring precise prediction with speed. Furthermore, Preprocess + YOLO v8 has the highest inference speed, approximately 48.73% less than Preprocess + yolov5, and is suitable for deep learning framework based real time object detection of forward-looking sonar images. From Figure 5 its inferred that even when the training samples are imbalanced, the State-of-the-art model helps the detection capture key features with high efficiency.

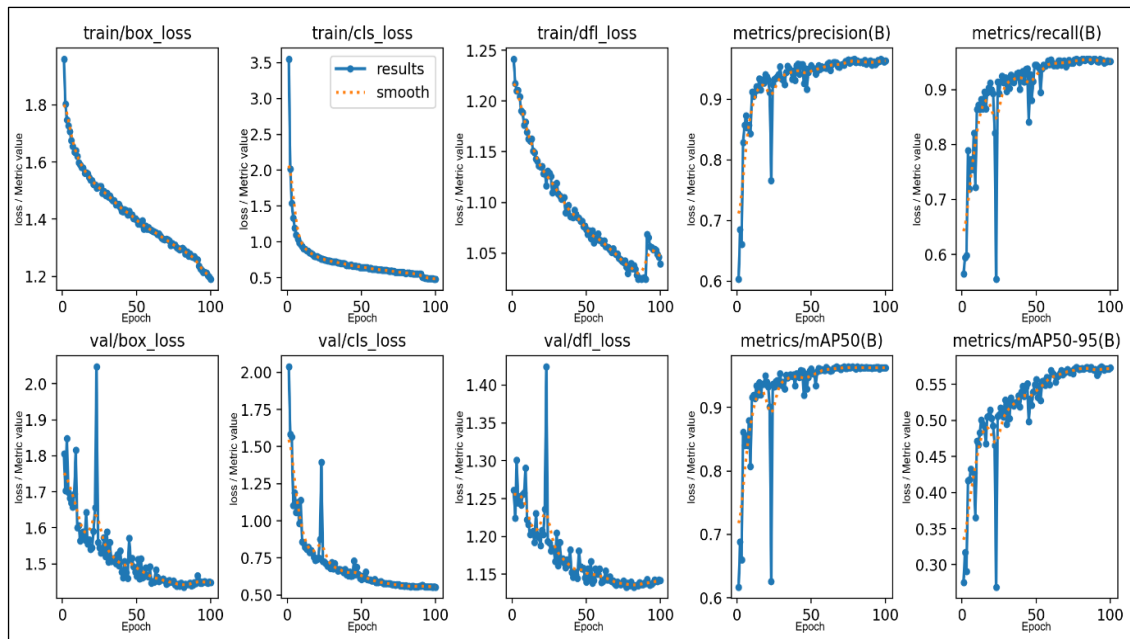


Figure 5: Training Results Plotted for the Box Loss and Class Loss

Table 2: Comparison of the Experimental Results of Various Detection Algorithms on UATD Dataset

Model	mAP	Recall	AP _{ball}	AP _{cube}	AP _{hb}	AP _{tyre}	AP _{sc}	AP _{plane}	AP _{rov}	AP _{cc}	AP _{cy}	AP _{mb}	Params	FLOPs
Faster RCNN-Resnet18	0.839	0.897	0.869	0.717	0.831	0.847	0.547	0.986	0.957	0.666	0.973	1.000	28.17M	49.78G
Faster RCNN-Resnet50	0.829	0.890	0.870	0.686	0.878	0.889	0.621	0.973	0.969	0.538	0.872	1.000	41.17M	63.29G
Faster RCNN-Resnet101	0.818	0.877	0.865	0.697	0.913	0.840	0.572	0.967	0.974	0.491	0.944	0.912	60.16M	82.77G
YOLO v3-Darknet-53	0.801	0.880	0.860	0.669	0.782	0.874	0.470	0.988	0.945	0.519	0.906	1.000	61.57M	49.67G
YOLO v3-MobilenetV2	0.787	0.868	0.790	0.573	0.808	0.738	0.518	0.992	0.986	0.498	0.963	1.000	3.68M	4.22G
YOLOv5	0.942	0.934	0.952	0.955	0.956	0.933	0.957	0.97	0.939	0.948	0.923	0.891	7.04M	15.G
YOLOv8	0.945	0.938	0.952	0.958	0.957	0.942	0.96	0.963	0.953	0.943	0.922	0.904	3M	8.1G
Preprocess +YOLOv5	0.959	0.978	0.959	0.979	0.984	0.924	0.970	0.995	0.992	0.903	0.939	0.943	7.04M	15.8G

Because of the complexity and perpetual changing of the underwater environment, environmental noise can seriously impair the effectiveness of underwater detecting devices (23). A prominent colour scheme that differs from optical and natural images poses a significant challenge to the ability to distinguish underwater sonar objects from the background.

When compared with natural images, deep learning-based detection frameworks on sonar images do not produce similar outcomes. End-to-end models designed specifically for sonar data are

necessary for optimal model performance. In addition, when the measured object is far away from the sonar, the target size is small in proportion to the entire sonar image, making it trivial to misunderstand as noise. Target detection is made more difficult by these issues. As a result, it is now very important to decrease noise interference and provide swift, precise automatic recognition of small underwater targets with a low likelihood of false alarms.

This paper proposed an approach that uses state-of-the-art detection models after pre-processing to

identify objects in forward-looking sonar imagery. As the precision recall curve and confusion matrix in Figure 6 and Figure 7 respectively shows, the training experiments have improved the ability to detect various types of objects positioned on the seabed under varied sonar operating conditions or

parameter settings. For maritime archaeological surveys looking to discover shipwreck sites and evaluate their state in relation to the surrounding environment (30), this work could become particularly relevant.

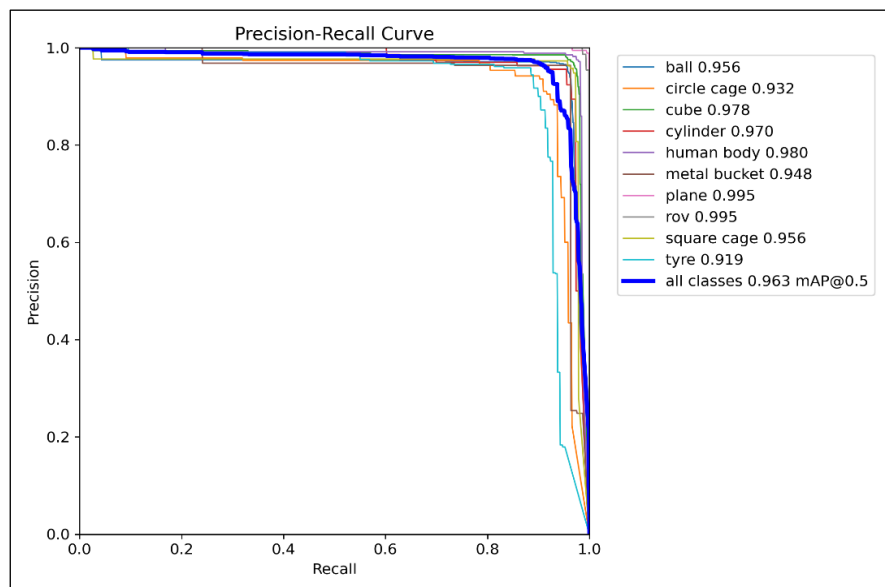


Figure 6: Precision-Recall Curve of YOLOV8 Model Trained with Pre-Processed Images

The present study provides a framework technique to combine the forward-looking sonar image with deep convolutional neural networks to improve the Mean Average Precision of underwater multiclass target detection tasks. To perform target detection, a training approach based on Preprocess + YOLOv8 model is used to construct an end-to-end preprocessing system that can automatically extract high-level features from sonar images. The effects of training techniques and network design on recognition performance

are explored through a number of experiments. The proposed approach achieves mAP of 0.96 which represents a significant enhancement in detection accuracy across the test dataset. The higher mAP indicates that our model is effective at accurately localizing objects making it a valuable tool for applications requiring high precision. The object detection result of the proposed model is shown in Figure 8 having bounding box and confidence score of the target prediction.

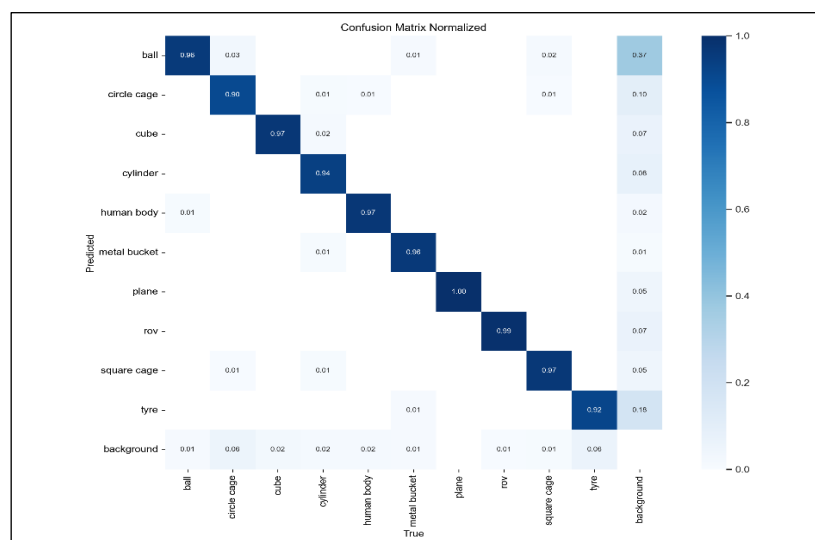


Figure 7: Confusion Matrix of YOLOV8 Model Trained with Pre-Processed Images

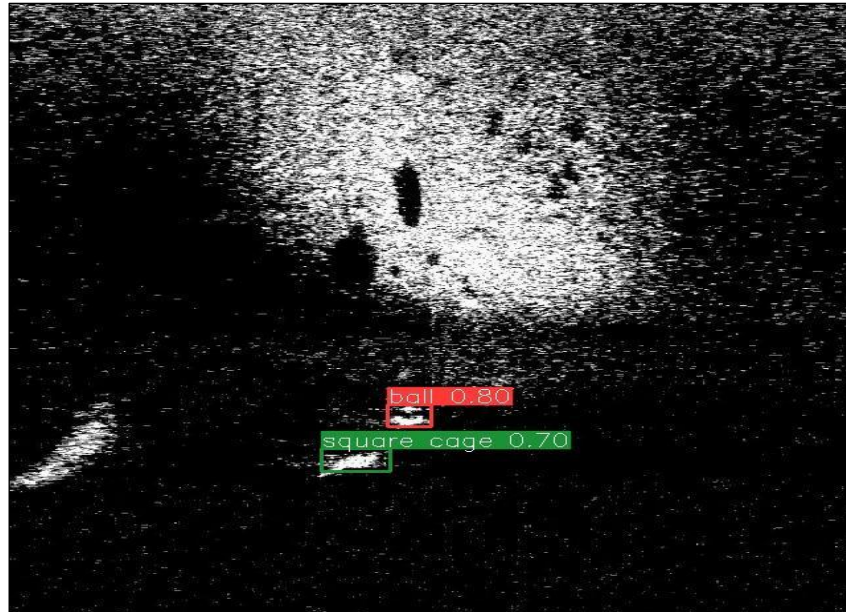


Figure 8: Detection Results Produced by the Proposed Object Detection Approach

According to results from experiments, the suggested framework outperforms the Yolov3 and Faster-RCNN models (29) in terms of real-time performance, precision, and effective anti-noise properties. On a ten-class underwater object detection challenge, our framework using preprocess+yolov5 and preprocess+ yolov8 can achieve a mAP@50 of 95.9% and 96.3%, respectively. Our study showed the significant potential of imaging sonar for underwater forward-looking sonar, which is crucial for underwater vehicles to sense their surroundings and navigate autonomously.

Enhanced inference speed enables autonomous underwater vehicles to detect objects rapidly, allowing dynamic navigation. Identifying submerged vessels or potential threats improves marine surveillance. Improved detection aids in monitoring changes in underwater ecosystems, resulting in better decisions and conservation efforts for aquatic species.

The performance of the proposed method may be hindered by the challenges of the sonar image dataset, which is limited in size and diversity, potentially limiting generalisation across varied underwater conditions. Meanwhile, the efficacy of target detection may vary in response to irregularly shaped objects, marine life occlusion, and underwater vegetation interference.

Conclusion

In this study, a preprocessing framework was proposed for object detection in multibeam forward-looking Sonar images having a low visual perception. The need for exploring specific detectors for complex environments having low to zero visibility and small targets is crucial. The outcome of proposed model includes visually enhanced image with object detection. A mAP of 96.3% with 3M parameters which implies a significant increase in efficiency in real time processing in the YOLOv8 model.

The practical relevance and effectiveness of the proposed approach are shown in the comparison study using a forward-looking sonar image dataset collected in an underwater environment. Compared with other existing models, the proposed framework has a higher precision for each object category irrespective of the imbalance of objects in the dataset. Further research may explore expanding dataset with diverse environment and object types to improve robustness and incorporating multiscale feature fusion and ensemble techniques for precise detection.

Abbreviations

AUV: Autonomous Underwater Vehicle, AutoML: Automated Machine Learning, BCE: Binary Cross Entropy, CNN: Convolutional Neural Network, FLS: Forward-Looking Sonar, mAP: mean Average Precision, ROI: Region of Interest, ROV: Remotely

Operated Vehicle, UATD: Underwater Acoustics Target Detection, YOLO: You Only Look Once.

Acknowledgement

None.

Author Contributions

Divyabarathi G: Conceptualization, Methodology, Formal analysis and investigation, Writing - original draft preparation, Writing review and editing, Resources, Baazil P Thampy: Formal analysis and investigation, Writing review and editing, Judy M V: Conceptualization, Writing review and editing, Funding acquisition, Resources, Supervision, Samarjith Kar: Writing review and editing, K S Ravichandran: Writing review and editing.

Conflict of Interest

Author Divyabarathi G declares that she has no conflict of interest. Author Baazil P Thampy declares that he has no conflict of interest. Author Judy MV declares that she has no conflict of interest. Author Samarjith Kar declares that he has no conflict of interest. Author KS Ravichandran declares that he has no conflict of interest.

Declaration of Artificial Intelligence (AI) Assistance

The authors declare no use of Artificial intelligence (AI) for the write-up of the manuscript.

Ethics Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This research was funded by Department of Science and Technology, DST/ICPS/DIGITAL POOMPUHAR/2017 under Scientific and Heritage Research Initiatives (SHRI) Division in the Department of Computer Applications, Cochin University of Science and Technology, Kochi, Kerala.

References

1. Vasiljević A, Na F, Mandić F, et al. Coordinated navigation of surface and underwater marine robotic vehicles for ocean sampling and environmental monitoring. *IEEE/ASME Trans Mechatronics*. 2017;22(3):1174–84.
2. Chen Y, Niu H, Chen H, et al. A review of underwater target recognition based on deep learning. *J Phys Conf Ser*. 2021;1881:042031. <https://iopscience.iop.org/article/10.1088/1742-6596/1881/4/042031>
3. Nayak N, Nara M, Gambin T, et al. Machine learning techniques for AUV side-scan sonar data feature extraction as applied to intelligent search for underwater archaeological sites. In: *Field and Service Robotics: Results of the 12th International Conference*. Springer. 2021:219–33. https://doi.org/10.1007/978-981-15-9460-1_16
4. Valdenegro-Toro M. End-to-end object detection and recognition in forward-looking sonar images with convolutional neural networks. In: *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*; IEEE. 2016:144–50. doi: 10.1109/AUV.2016.7778662
5. Song Y, He B, Liu P. Real-time object detection for AUVs using self-cascaded convolutional neural networks. *IEEE J Ocean Eng*. 2019;46(1):56–67. doi: 10.1109/JOE.2019.2893124
6. Gu J, Pyo J, Joe H, et al. A method for automatic detection of underwater objects using forward-looking imaging sonar. In: *OCEANS 2015 - MTS/IEEE Washington*; IEEE. 2015:1–5. doi: 10.23919/OCEANS.2015.7401920
7. Karimanzira D, Renkewitz H, Shea D, et al. Object detection in sonar images. *Electronics*. 2020;9(7):1180. <https://www.mdpi.com/2079-9292/9/7/1180>
8. Liu Y, Wang R, Cao K, et al. Sonar target detection based on a dual channel attention convolutional network. In: *2022 12th International Conference on Information Science and Technology (ICIST)*; IEEE; 2022:255–64. doi: 10.1109/ICIST55546.2022.9926829
9. Jiang L, Cai T, Ma Q, et al. Active object detection in sonar images. *IEEE Access*. 2020;8:102540–53.
10. Wang Z, Guo J, Zeng L, et al. MLFFNet: Multilevel feature fusion network for object detection in sonar images. *IEEE Trans Geosci Remote Sens*. 2022;60:1–19.
11. Kim J, Yu SC. Convolutional neural network-based real-time ROV detection using forward-looking sonar image. In: *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*; IEEE. 2016:396–400. doi: 10.1109/AUV.2016.7778702
12. Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016, Proceedings, Part I*; Springer. 2016:21–37. doi: 10.1007/978-3-319-46448-0_2
13. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:779–88. doi: 10.1109/CVPR.2016.91
14. Girshick R. Fast R-CNN. In: *2015 IEEE International Conference on Computer Vision (ICCV)*; IEEE. 2015:1440–8. doi: 10.1109/ICCV.2015.169
15. Klausner N, Azimi-Sadjadi MR. Non-Gaussian target detection in sonar imagery using the multivariate Laplace distribution. *IEEE J Ocean Eng*. 2014;40(2):452–64.
16. Sawas J, Petillot Y, Pailhas Y. Cascade of boosted classifiers for rapid detection of underwater objects. In: *10th European Conference on Underwater Acoustics 2010*:1507–16. <https://researchportal.hw.ac.uk/en/publications/c>

- ascade-of-boosted-classifiers-for-rapid-detection-of-underwater-
17. Kim J, Cho H, Pyo J, et al. The convolution neural network based agent vehicle detection using forward-looking sonar image. In: OCEANS 2016 MTS/IEEE Monterey; IEEE. 2016:1-5. doi: 10.1109/OCEANS.2016.7761384
 18. Zhang P, Tang J, Zhong H, et al. Self-trained target detection of radar and sonar images using automatic deep learning. *IEEE Trans Geosci Remote Sens.* 2021;60:1-14.
 19. Yu Y, Zhao J, Gong Q, et al. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. *Remote Sens.* 2021;13(18):3555. <https://www.mdpi.com/2072-4292/13/18/3555>
 20. Yang C, Li Y, Jiang L, et al. Foreground enhancement network for object detection in sonar images. *Mach Vis Appl.* 2023;34(4):1-14.
 21. Zhang F, Zhang W, Cheng C, et al. Detection of small objects in side-scan sonar images using an enhanced YOLOv7-based approach. *J Mar Sci Eng.* 2023;11(11):2155. <https://www.mdpi.com/2077-1312/11/11/2155>
 22. Er MJ, Jie C. Research challenges, recent advances and benchmark datasets in deep-learning-based underwater marine object detection: A review. *Authorea Preprints.* 2022. <https://www.techrxiv.org/doi/full/10.36227/techrxiv.19350389>
 23. Zhao L, Yun Q, Yuan F, et al. YOLOv7-CHS: An emerging model for underwater object detection. *J Mar Sci Eng.* 2023;11(10):1949. <https://www.mdpi.com/2077-1312/11/10/1949>
 24. Solawetz FJ. What is YOLOv8? The ultimate guide. <https://blog.roboflow.com/whats-new-in-yolov8/>
 25. Tang J, Chen Z, Xie H, et al. Unbalanced Underwater Sonar Image Target Detection Based on Weak Contrast Feature Enhancement. In: 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC) 2023;7:367-73. <https://ieeexplore.ieee.org/abstract/document/10292100/>
 26. Gonzalez RC. *Digital Image Processing.* New Delhi: Pearson Education India. 2009. https://sde.uoc.ac.in/sites/default/files/sde_videos/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%20R.%20Woods-ilovepdf-compressed.pdf
 27. Kocielek M, Strzelecki M, Obuchowicz R. Does image normalization and intensity resolution impact texture classification. *Comput Med Imaging Graph.* 2020;81:101716. <https://www.sciencedirect.com/science/article/pii/S0895611120300197>
 28. Jocher G, Stoken A, Borovec J, et al. ultralytics/yolov5: v4.0-nn. SiLU activations, weights & biases logging, PyTorch hub integration. *Zenodo.* 2021. doi: 10.5281/zenodo.4679653
 29. Xie K, Yang J, Qiu K. A dataset with multibeam forward-looking sonar for underwater object detection. *Sci Data.* 2022;9(1):739. <https://www.nature.com/articles/s41597-022-01854-w.pdf>
 30. Jin L, Liang H, Yang C. Accurate underwater ATR in forward-looking sonar imagery using deep convolutional neural networks. *IEEE Access.* 2019;7:125522-31.

How to Cite: Divyabarathi G, Thampy BP, Judy MV, Kar S, Ravichandran KS. Optimal preprocessing for enhancing object detection in underwater sonar images. *Int Res J Multidiscip Scope.* 2025; 6(4):1049-1061. doi: 10.47857/irjms.2025.v06i04.06495