Seeing Beneath the Surface: AI-Powered Real-Time Biofouling Detection and 3D Hull Digital Twin Generation

Final Project Report Submitted to
The Department of Computer Science
Faculty of Computer and Information Technology
Jordan University of Science and Technology
In Partial Fulfillment of the Requirements for the Degree of Bachelors of Science in Computer Science

Prepared by:

Nour Abu Beidar 166542
Hajar Alhadaris 165360
Abdalrahman Alshamaileh 165285

Supervisor:
Dr.Yaser Jararweh

January 2026

نموذج حقوق الملكية الفكرية لمشاريع التخرج في قسم علوم الحاسوب

يتم قراءة وتوقيع هذا النموذج من قبل الطلاب المسجلين لمشاريع التخرج في قسم علوم الحاسوب

تعود حقوق الملكية الفكرية لمشاريع التخرج ونتائجها (مثل براءات الاختراع أو أي منتج قابل للتسويق) إلى جامعة العلوم والتكنولوجيا الأردنية. وتخضع هذه الحقوق إلى قوانين وأنظمة و تعليمات الجامعة المتعلقة بالملكية الفكرية وبراءات الاختراع.

بناءا على ما سبق أوافق على ما يلي:

1) أن أحفظ كافة حقوق الملكية الفكرية لجامعة العلوم والتكنولوجيا الأردنية في مشروع التخرج.
2) أن ألتزم بوضع اسم جامعة العلوم والتكنولوجيا الأردنية و أسماء جميع الباحثين المشاركين في المشروع على أي نشرة علمية للمشروع كاملا أو لنتائجه. و يشمل ذلك النشر في المجلات و المؤتمرات العلمية عامة أو النشر على المواقع الإلكترونية أو براءات الاختراع أو المسابقات العلمية.
3) أن ألتزم بأسس حقوق التأليف المعتمدة في جامعة العلوم والتكنولوجيا الأردنية.
4) أن أقوم بإعلام الجهة المختصة في الجامعة عن أي اختراع أو اكتشاف قد ينتج عن هذا المشروع و أن ألتزم السرية التامة في ذلك و أن أعمل من خلال الجامعة على الحصول على براءة الاختراع التي قد تنتج عن هذا المشروع.
5) أن تكون جامعة العلوم والتكنولوجيا الأردنية هي المالك لأي براءة اختراع قد تنتج عن هذا المشروع و تشمل هذه الملكية حق الجامعة في إعطاء التراخيص و التسويق و البيع كمؤسسة راعية و داعمة لكافة الأنشطة البحثية. و يكون حق للطالب شمول اسمه على براءة الاختراع كأحد المخترعين. و في حال تم إعطاء تراخيص أو تسويق و بيع لأي من منتجات المشروع يمنح المخترعون بما فيهم الطالب نسبة من الإيرادات حسب تعليمات البحث العلمي في جامعة العلوم والتكنولوجيا الأردنية.

اسم الطالب ماجد ساعي إدريس التوقيع .

اسم الطالب نور محمد محمود أبو بيدر التوقيع .

اسم الطالب عبدالرحمن محمد هشم الشحايلة التوقيع

اسم المشرف التوقيع

تاريخ .....

Seeing Beneath the Surface: AI-Powered Real-Time Biofouling Detection and 3D Hull Digital Twin Generation

Abdalrahman Alshamaileh*, Hajar Alhadaris*,
Nour Abu Beidar*, Bayan Alfarrayeh†
*Dept. of Computer Science and Artificial Intelligence, Jordan University of Science and Technology, Irbid, Jordan
†Dept. of Cybersecurity, Jordan University of Science and Technology, Irbid, Jordan
arhashemalshamaileh22@cit.just.edu.jo, hsalhadaris22@cit.just.edu.jo, nmabubeidar22@cit.just.edu.jo,
bialfarrayeh22@cit.just.edu.jo

***Abstract*—Biofouling on submerged vessel surfaces increases hydrodynamic drag, fuel costs, and impacts marine ecosystem badly. Traditional inspection methods rely on marine companies with certified and approved technical divers for such tasks, which is time consuming, costly, and subjective. This paper proposes an AI-powered real time underwater inspection system that integrates a deep learning Computer Vision based biofouling detection model into an underwater drone (ROV) to enable automated hull inspection. The proposed system apply real time detection and classification of biofouling on submerged parts using computer vision techniques. After detection, the system generates a 3D digital twin for the hull, which is constructed from geometric parameters and inspection data, where detected biofouling regions are spatially mapped. This method allows intuitive visualization of fouling distribution, severity, and coverage across the underwater hull. The resulting digital twin provides quantitative metrics, which includes fouling coverage ratios and hotspot identification, supporting inspection reporting and maintenance planning. Experimental deployments shows the ability of the proposed method under real underwater conditions, focusing on its ability to improve inspection efficiency and reduce operational risk. The provided solution shows a scalable step toward intelligent, advanced maritime maintenance using underwater robotics and artificial intelligence.**

***Index Terms*—ROV: Remotely Operated Vehicle**

I. PROJECT GOALS AND OBJECTIVES

A. Project Goal

To develop an AI-powered underwater inspection system that detects biofouling on ship hulls safely and efficiently, replacing hazardous diver inspections with real-time analysis conducted by remotely operated vehicles (ROVs) and supported by 3D digital twin technology for data-driven maintenance.

B. Project Objectives

- Biofouling severity detecting and classifying from underwater images in real time using deep learning.
- Applying pixel level segmentation to locate and evaluate biofouling areas.
- Integrate detection and segmentation models into ROV inspections.
- Construct a 3D digital twin of the hull to visualize the distribution of biofouling.
- Provide quantitative metrics to facilitate inspection reporting and inform maintenance planning.

II. INTRODUCTION

Biofouling accumulation growth on submerged surfaces is a serious challenge for the maritime industry,it increases hydrodynamic drag, fuel consumption, and greenhouse gas emissions while accelerating material degradation and effect marine ecosystems badly. Regular inspection of vessel hulls is important to ensure operational efficiency, regulatory compliance, and sustainable maritime operations. Traditionally, hull inspections are done by certified technical divers, who spend long hours on their work sites which are weather dependent and sensitive to human variability and safety problems.

To overcome these limitations, Under Water drone (ROV) has been developed as a preferred replacement for underwater inspections, which allows for safer and more reproducible surveys while maintaining low costs of operation. Previous studies have shown the ability of ROV inspection systems in offshore and subsea environments with very limited human involvement [1]. Most existing ROV inspection operations still depend on manual visual interpretation of videos, which limits its scalability and objectivity.

Recent advances in artificial intelligence and deep learning have significantly improved underwater perception, which enabled robust object detection and classification even with challenges such as light attenuation, turbidity, and color distortion. Vision based underwater AI models especially convolutional neural networks and YOLO based architectures have shown strong performance for detecting submerged object and marine targets under real world conditions [2]. These developments

indicate strong potential for automating underwater inspection tasks however, current approaches primarily focus on frame level detection and lack spatial integration across the inspected structure.

In other hand, there are vision based 3D reconstruction techniques that gained increasing attention for underwater surveys and inspection missions. Methods based on stereo vision, visual SLAM, structure from motion, and multi-view stereo have demonstrated the feasibility of generating accurate three dimensional models of underwater structures using ROV-mounted cameras [3]. Such reconstructions enable spatial understanding, coverage assessment, and long term monitoring of submerged parts.

Existing 3D reconstruction pipelines sometimes are not connected with any semantic inspection data at all. This means that they are limited for such things as condition based maintenance work. This study aims to fill in these gaps by presenting a model with real time work flow which is an underwater inspection framework to fully integrate biofouling detection / classification with 3D hull digital twin creation on an underwater drone (ROV). While traditional frameworks consider detection and modeling as different processes, the provided system presents a condition in which it informs the 3D digital twin of the submerged hull, where biofouling detected regions have their spatial areas spatially mapped. With single unified representation, fouling distribution can be easily visualized as well as covered under a quantitative measure with better monitoring/reporting.

III. RELATED WORK

Research on vision based underwater inspection has expanded in recent years, covering a wide range of topics related to biofouling analysis, underwater perception, and robotic inspection systems. There are multiple studies that address different parts of the inspection cycle, often treating data acquisition, visual analysis, and spatial modeling as separate research problems. To better position the proposed system within this landscape, the related work is discussed across four complementary research directions: underwater data availability, segmentation driven inspection methods, integrated enhancement and detection architectures, and vision based three dimensional reconstruction of submerged structures. Reviewing these directions together helps reveal how current approaches advance underwater inspection while also exposing gaps in linking visual biofouling analysis with spatially coherent hull representations.

1) ***Underwater Data Availability and Dataset Limitations****:* A common problem in vision-based biofouling inspection is the lack of annotated underwater datasets. Getting real underwater images is not simple. It usually requires special equipment and suitable conditions, which makes the process expensive and difficult to repeat on a large scale.

Mai *et al.* [4] dealt with this issue by using synthetic images to train a semantic segmentation model for marine growth analysis. Their study discusses that collecting labeled underwater data is often not practical. According to that, many researchers depend on replacement such as synthetic data generation to support model training.

Jian *et al.* [5] also discussed this matter in their survey on underwater object detection. They noted that underwater images are affected by multiple factors, including turbidity, uneven lighting, and color distortion. Also, the authors reported that many available underwater datasets are limited in size and variety, which reduces the ability of detection models to do well under different underwater conditions.

Based on these studies, dataset limitations remain a major challenge for underwater visual inspection systems.

2) ***Segmentation-Based Approaches for Underwater Inspection****:* Approaches that uses Segmentation have been widely explored for biofouling analysis, as they enable pixel level marking of fouling regions on submerged surfaces.

Che *et al.* [6] showed a YOLOv8 based underwater image segmentation method which includes lightweight architectural enhancements to improve efficiency under challenging underwater conditions. The approach achieved real time performance of 117 FPS using pixel level annotated data prepared in a YOLOv8 compatible format, proving its suitability for deployment on hardware constrained platforms.

Liu *et al.* [7] proposed a developed DeepLabv3+ model for underwater image segmentation by integrating an unsupervised color correction module to avoid color degradation, low contrast, and turbidity. The resulting architecture improves boundary clarity and pixel level segmentation accuracy, enabling more precise marking of underwater surface regions.

Waszak *et al.* [8] proposed a pixel level semantic segmentation framework for underwater ship hull inspection and introduced the publicly available LIACI dataset, which has 1,893 annotated images spread over ten semantic classes, which includes marine growth and surface defects. The study benchmarked several state of the art segmentation architectures and identified U-Net with a MobileNetV2 backbone as a well balanced solution in terms of accuracy and computational efficiency, making it suitable for near real time inspection tasks. Although the work does not address YOLO based detection or 3D reconstruction, it provides a high quality segmentation dataset and a standardized evaluation pipeline relevant to underwater inspection research.

3) ***Integrated Enhancement and Detection Architectures****:* Pachaiyappan *et al.* [2] presented an integrated underwater perception framework, which combines image improvement and object detection to propose challenges such as color distortion, turbidity, low contrast, and noise. The architecture combines convolutional block attention mechanisms, a modified Swin Transformer that is embedded within a U-Net backbone to capture long range contextual information, and a diffusion based model for denoising that progressively removes underwater distortions. When this approach is integrated with an object detection network, it proved improved detection accuracy, highlighting the effectiveness of combining attention

based CNNs, transformer blocks, and diffusion models for robust underwater object detection.

*4) **Underwater 3D Reconstruction**:* Hong and Kim [9] proposed a stereovision based methode for dense three dimensional reconstruction of underwater ship hull surfaces using an ROV mounted camera system. The proposed method depends on controlled image acquisition, in situ camera calibration, and graph cut based dense stereo matching to generate detailed geometric reconstructions. While the approach is effective for capturing underwater surface geometry, it focuses solely on geometric mapping and does not incorporate learning based methods, biofouling detection, or publicly available benchmark datasets.

Brandou *et al.* [3] proposed a complete methodology for dense three dimensional reconstruction of natural underwater scenes using a stereovision system mounted on an ROV. The approach depends on controlled and geometrically constrained image acquisition trajectories, followed by in situ camera calibration, feature extraction and matching, structure and motion recovery, and graph cut based dense stereo matching. The study explains the effectiveness of stereovision systems for correct underwater 3D reconstruction using real underwater imagery obtained during deep sea deployments.

Ferrera *et al.* [10] introduced a vision based 3D reconstruction architecture for deep sea inspection that combines real time and offline reconstruction pipelines. The introduced framework utilizes a stereo Visual SLAM system to estimate camera trajectories and generate depth maps, which are fused using a truncated signed distance field with a voxel hash map to produce a coarse real time 3D model for navigation and coverage verification. In parallel, an offline reconstruction pipeline based on Structure from Motion and Multi View Stereo is applied to generate high resolution, textured 3D models. By integrating SLAM based real time feedback with SfM/MVS based accuracy, the approach overcomes the limitations of using either method independently and supports practical underwater survey and inspection tasks.

*5) **Integrated AI and ROV-Based Inspection Systems**:* Capocci *et al.* [1] showed the development and testing of a ROV designed for the inspection of offshore renewable devices. The study works on ROV system design, integration, and experimental validation in real inspection cases. Even though the work does not mention underwater perception or learning based analysis, it gives valuable advices into ROV platform development and deployment, foucosing on the importance of stable operation and reliable system integration for inspection tasks where visual data quality is important.

Malkawi *et al.* [11] confirm the integration of artificial intelligence components in planetary rover analog systems through field experiments done in Wadi Rum desert. The framework includes an AI based ISRU payload which work with identifying and analyzing surface samples. The study also highlights the similarities between planet based exploration and other harsh domains, which shows how integrated robotic systems evaluated under challenging field conditions can provide insights applicable to related exploration contexts.

*6) **Limitations of Existing Approaches**:* Most existing underwater inspection studies have made good progress, but there are still clear gaps. Many segmentation-based approaches only work at the image level, where biofouling is marked pixel by pixel, without showing how fouling is distributed across the full ship hull in a spatial or 3D way. Other methods focus on improving image quality or object detection accuracy, but they do not provide biofouling-specific analysis, hull-level mapping, or a clear link to 3D reconstruction. 3D reconstruction methods recover underwater geometry but lack semantic info and biofouling identification. Studies on ROVs focus on hardware, while image analysis, 3D mapping, and robotic deployment are treated separately, showing no unified biofouling-aware framework

Paper Focus	**Methodology**	**Key Results**
Underwater object detection review [5]	Comprehensive survey of detection methods and available underwater datasets	Identifies key challenges and provides a structured classification of underwater detection research
Real-time underwater image segmentation [6]	YOLOv8-based instance segmentation with lightweight and attention-enhanced modules	Improved segmentation accuracy and real-time performance (up to 117 FPS)
Ship hull inspection datasets and benchmarking [8]	LIACI dataset with pixel-level annotations and benchmarking of segmentation models	Provides a public dataset and shows efficient segmentation for near real-time hull inspection
Integrated enhancement and detection for underwater perception [2]	Combined image enhancement and object detection using attention mechanisms, modified Swin Transformer blocks, and diffusion-based denoising integrated with a YOLO-based framework	Significantly improves underwater object detection accuracy, achieving mAP@0.5 of 81.4% on the TrashCan dataset
Underwater ship hull 3D reconstruction [9]	Monocular vision-based 3D reconstruction using piecewise-planar surface modeling and SLAM	Demonstrates effective geometric reconstruction of ship hull surfaces under challenging underwater visibility conditions
Vision-based 3D reconstruction for underwater inspection [10]	Hybrid framework combining stereo visual SLAM with TSDF and SfM/MVS	Generates real-time coarse 3D models and high-quality textured reconstructions in real underwater inspection scenarios
ROV-based inspection of offshore renewable energy devices [1]	Design and development of an inspection ROV with integrated navigation and control systems	Demonstrates reliable ROV operation and system integration through experimental validation

TABLE I
OVERVIEW OF RELATED WORK IN UNDERWATER BIOFOULING INSPECTION

IV. APPROACH AND METHODOLOGY

A. Dataset Description and Annotation

1) Data Collection: Over 10,000 underwater images were gathered with the cooperation of ship maintenance companies in the city of Aqaba, Jordan. The images include those of ship hulls, propellers, underwater pipes, and underwater structures. The images were taken with varying lighting conditions and levels of biofouling. After the elimination of images with uncertain or redundant information, about 1000 images were considered to be of high quality. To allow for better training and representation, the images were evenly spread in four classes Clean, Light, Moderate, and Heavy biofouling.

The dataset includes images that are labeled for classification, which can be used to classify the levels of biofouling. In addition, pixel-level semantic segmentation provides masks for annotations, which can be used to determine the severity of fouling on the pixel-level basis. This helps to clearly demarcate boundaries of fouling, measure the area of fouling, and distinguish between clean, Light, Moderate, and Heavy fouling areas.

2) Data Annotation: The selected images were manually annotated by the research team, following labeling standards established in consultation with ship industry experts. To ensure less variability among the annotators and maintain the quality of the labels, the annotations were cross-validated by an independent team member. When some discrepancies occurred, the annotator corrected the labels based on the specified criteria. Examples of images labeled according to the different four levels of severity of biofouling are shown in Fig. 1.Each image was assigned to one of the four predefined classes :Clean, Light, Moderate, or Heavy, based on visual assessment, biofouling thickness, and coverage density:

Fig. 1. visual examples of biofouling severity classes: (a) Clean, (b) Light, (c) Moderate, and (d) Heavy.

- **Clean:** Surfaces without visible biofouling, appearing smooth.
- **Light Fouling:** Small growth of algae or thin bio layers.
- **Moderate Fouling:** Uneven growth and early formation of barnacles.
- **Heavy Fouling:** Dense growth which leads to near or complete coverage.

The annotation process was performed with a well defined threshold criteria according to biofouling thickness and surface density, as detailed in Table II.

TABLE II
ANNOTATION RULES FOR BIOFOULING BASED ON THICKNESS AND SURFACE DENSITY THRESHOLDS.

Thickness (mm)	**Density Condition (ρ)**	**Classification**
< 5	$\rho < 0.50$	Clean
	$\rho \geq 0.50$	Light
5–10	$\rho < 0.30$	Clean
	$\rho \geq 0.30$	Light
10–20	$\rho \leq 0.35$	Light
	$0.35 < \rho \leq 0.87$	Moderate
	$\rho > 0.87$	Heavy
20–35	$\rho \leq 0.30$	Light
	$0.30 < \rho \leq 0.65$	Moderate
	$\rho > 0.65$	Heavy
≥ 35	$\rho \leq 0.15$	Light
	$0.15 < \rho \leq 0.65$	Moderate
	$\rho > 0.65$	Heavy
Else	—	Unclassified

The annotations are formated according to the YOLO format, used for both classification and segmentation tasks. Each image is paired with a label for classification and a corresponding mask marking biofouling regions at the pixel level. This formatting makes sure that its fully compatibal with the YOLOv8 work flow, which supports direct training for the classification and segmentation of biofouling areas.

3) Data Preprocessing: All images were normalized to the same resolution (640x640 pixels) to be inbetween the efficiency of computation and accuracy. Also, basic manual preprocessing were performed, such as cropping and illumination adjustments, to level up visual clarity and ensure dataset consistency. Also, YOLOv8 framework automatically does additional normalization and data augmentation operations during the training process.

4) Data Splitting and Distribution: The dataset was divided into 70% training, 20% validation, and 10% testing. Table III summarizes the distribution of images, and Fig. 2 visulaizes it.

TABLE III
DISTRIBUTION OF BIOFOULING SEVERITY CLASSES OVER DATASET SPLITS.

Split	Clean	Light	Moderate	Heavy
Train (700)	175	175	175	175
Val (200)	50	50	50	50
Test (100)	25	25	25	25
Total (1000)	250	250	250	250

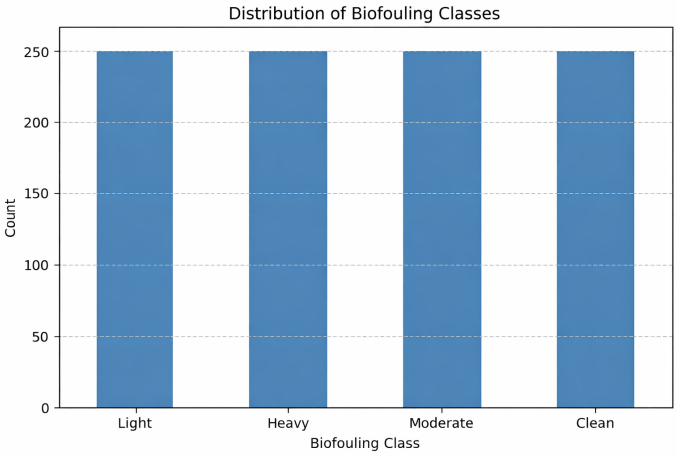

Fig. 2. Overall class distribution of the biofouling dataset.

B. Methodology

1) Model Architectures: Two deep learning models were evaluated in this study: YOLOv8-Nano (YOLOv8n) and YOLOv8-Small (YOLOv8s) [12]. These models were chosen to examine performance across different YOLOv8 variants and to find a balance between inference speed and representational capacity, with particular attention to their suitability for real-time applications in resource-limited environments.

YOLOv8 models share a common design composed of three main components: (i) a backbone network for hierarchical feature extraction, (ii) a neck that improves multi-scale feature representation by combining features from different layers, and (iii) a detection head that outputs bounding box coordinates, objectness scores, and class probabilities [13]. In this study, images were annotated with a single full-frame bounding box, effectively reducing the detection task to a four-class classification problem (*Clean, Light, Moderate, Heavy*). Consequently, the classification component of the loss function dominated the training process, while the localization term had minimal impact.

For biofouling semantic segmentation, YOLOv8 served as the main deep learning framework. Two lightweight variants were used to compare accuracy and efficiency: YOLOv8-Nano Segmentation (YOLOv8n-seg) and YOLOv8-Small Segmentation (YOLOv8s-seg) [14]. These models balance segmentation precision, speed, and computational cost, enabling practical real-time execution even on devices with limited resources, such as underwater remotely operated vehicles.

The YOLOv8 segmentation architecture extends the standard detection framework with three components: (i) a backbone for hierarchical feature extraction, (ii) a neck for enhanced multi-scale feature aggregation, and (iii) a segmentation head that incorporates a mask prototype network and per-instance mask coefficients. This design produces high-resolution pixel-level segmentation masks while maintaining spatial localization

2) Training Setup: The official Ultralytics YOLOv8 is used for implementing all the classification and segmentation models, with pretrained COCO weights for initialization [15]. Transfer learning is employed to facilitate efficient convergence and enhance generalization performance under limited annotated underwater data. Multiple input resolutions are examined during preliminary experiments, after which an input size of 640×640 pixels is selected as an effective balance between segmentation accuracy and computational efficiency.

During training, standard data augmentation techniques provided by the YOLOv8 framework are applied automatically, including random horizontal flipping, scaling, color jittering, and affine transformations. These augmentations are particularly relevant for underwater imagery, where illumination variations, color attenuation, and visual noise are common.

The dataset is organized according to the split described in Section IV-A, with 700 images for training, 200 for validation, and 100 for testing. Batch size configurations are investigated for both YOLOv8n and YOLOv8s in order to achieve stable optimization behavior. The Ultralytics framework automatically adjusts learning rates based on batch size and available GPU memory to maintain training stability.

Training is conducted for 20 epochs on an NVIDIA Tesla T4 GPU with 15 GB of memory. Model performance is evaluated on the validation set after each epoch using standard evaluation metrics. Based on empirical observations reported in related work, optimal model checkpoints emerge during the later stages of training. The model checkpoint with the highest mean Average Precision (mAP) value on the validation data is selected for evaluation on the testing data.

C. Evaluation Metrics

performance was evaluated using standard YOLO metrics, namely Precision, Recall, and mean Average Precision (mAP) [16]. Metrics were calculated on the validation set after each epoch, and the checkpoint with the highest mAP was retained for final testing. A detailed description of the evaluation methodology is provided in Section IV-D

D. Theory and Calculation

This study draws on standard evaluation metrics from object detection [17], adapted here as a four-class image classification task since each image corresponds to a single severity label. To ensure consistency with the YOLO training pipeline, Precision (P), Recall (R), and mean Average Precision (mAP) are

reported. For segmentation experiments, these metrics are also applied at the object level to provide comparable assessment of detection and pixel-level predictions.

E. Evaluation Metrics

Precision (P) and Recall (R) are defined as:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \tag{1}$$

where TP, FP, and FN denote true positives, false positives, and false negatives.

The mean Average Precision (mAP) is computed as:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{2}$$

where N as the number of biofouling classes. Following YOLO evaluation practice, mAP@0.5 was emphasized as the primary benchmark. These metrics were done in the image level classification setting (Section IV-A) and at the object level for segmentation, which provides a consistent assessment for detection and pixel wise predictions.

F. ROV Integration

The proposed system is integrated at the ROV software framework level. The ROV software platform is an open source and flexible framework, which is leveraged to include the optimized YOLOv8 model in the operational cycle, which enables direct access to the live video stream and facilitates real time processing.

As a result of this integration, biofouling is first classified according to severity levels and subsequently analyzed through pixel wise segmentation on the high definition front facing camera stream as the ROV traverses submerged surfaces. Segmentation outputs are generated alongside the video feed, enabling simultaneous visualization and detailed assessment of biofouling extent, spatial distribution, and severity characteristics

One of the main tasks was to maintain low-latency inference without sacrificing model accuracy or frame rate. To address this, we employed model quantization and image resolution scaling to sacrifice detection performance for computational resources of the onboard processor. The system processes input at about 10 FPS on YOLOv8n and 7 FPS on YOLOv8s under field conditions, providing operators with real-time visual overlays of the detected fouling regions via the ROV's tethered topside interface.

ROV-control communication is maintained through the Ethernet connection that is provided by the tethering of Ethernet cables between the ROV control system and surface control stations. Beyond real-time inspection, it is also able to perform offline video streaming and analysis results for better functionality in programming for long-duration hull inspections or biofouling observation missions.

Overall, this software-level integration enables the ROV to operate as an intelligent inspection platform, providing spatially detailed biofouling assessment capabilities within real-world underwater environments.

G. EXPECTED RESULTS/OUTPUTS

1) Initial Results: The comparative experiments were conducted using the YOLOv8 model [18], two model configurations (Nano and Small), and the same training conditions for all experiments. The aim was to compare the balance between the computational cost and the accuracy of the classification of the multi-class biofouling severity.

H. Experimental Results

Tables IV and V summarize the validation and test performance of the models under comparison. Results are reported for the YOLOv8 variants, including YOLOv8n and YOLOv8s.

TABLE IV
VALIDATION PERFORMANCE.

Model	Precision	Recall	mAP@0.5
YOLOv8-N	0.476	0.702	**0.608**
YOLOv8-S	**0.515**	0.714	**0.642**

TABLE V
TEST PERFORMANCE ON THE HELD-OUT SET.

Model	Precision	Recall	mAP@0.5
YOLOv8-N	0.400	0.800	0.586
YOLOv8-S	**0.520**	0.578	**0.642**

The validation and test results presented in Tables IV and V reveal notable differences in performance between YOLOv8-Nano (YOLOv8-N) and YOLOv8-Small (YOLOv8-S) models. On the validation set, YOLOv8-S demonstrates superior precision and mAP@0.5 values (0.515 and 0.642, respectively) compared to YOLOv8-N (0.476 and 0.608), indicating a slightly better ability to correctly identify biofouling regions while maintaining a balanced recall. In the contrary, on the test set, YOLOv8-N achieves a higher recall (0.800) than YOLOv8-S (0.578), but its precision is lower (0.400 vs. 0.520). This result shows that while YOLOv8-N is highly sensitive and able to detect a larger amount of biofouling instances, it commonly produces more false positives, even though YOLOv8-S offers more precise predictions, it is slightly less comprehensive in detection.

These results shows a fundamental trade off between model size, computational efficiency, and generalization capability. YOLOv8-N, a lightweight model, shows strong adaptability on the validation data but shows sensitivity to domain shifts which are in the held out test set. In other hand, YOLOv8-S, with more representational capacity, keeps a more consistent performance for both validation and test datasets, focusing on its robustness in real world scenarios.

The comparison also points out the importance of using both precision and recall when evaluating model suitability. A

higher recall, such as the YOLOv8-N results, could be better in applications that gives detection completeness the priority, while higher precision, such as YOLOv8-S results, is better when minimizing false alarms is most important. Overall, these outcomes show that model selection should be assisted by some specific operational requirements of biofouling monitoring and inspection tasks, to balance detection sensitivity, accuracy, and computational efficiency.

I. Per-Class Results and Confusion Matrix

The YOLOv8-S model's per class results on the test dataset are shown in Table VI. These results show different model performances for the four biofouling severity classes (Clean, Light, Moderate, Heavy). The confusion matrix for them is shown in Fig. 3.

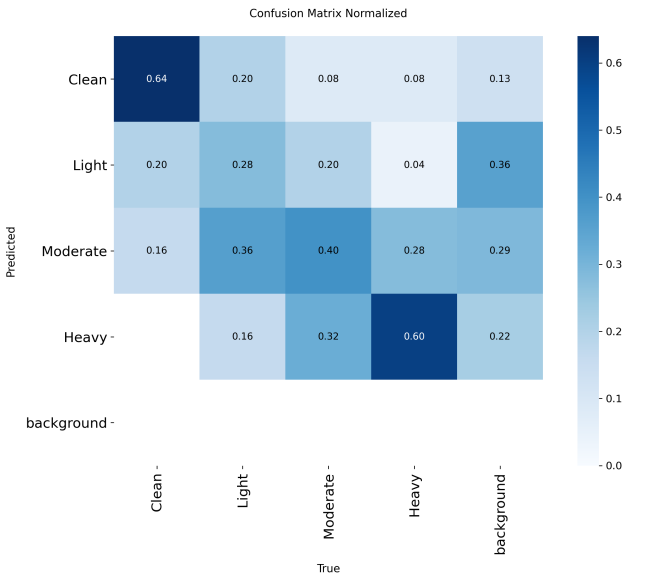

Fig. 3. Normalized confusion matrix of YOLOv8-S on the test set.

TABLE VI
PER-CLASS CLASSIFICATION PERFORMANCE OF YOLOV8-S (TEST SET).

Class	Precision	Recall	mAP@0.5
Clean	0.619	0.680	**0.754**
Light	0.432	0.440	0.462
Moderate	0.424	0.472	0.488
Heavy	0.603	0.720	0.706
Overall	**0.520**	**0.578**	**0.602**

The per class result of YOLOv8-S on the test set, shown in Table VI, shows notable difference between the four classes. The model performs best on the *Clean* and *Heavy* classes, achieving mAP@0.5 values of 0.754 and 0.706, respectively. This shows that the model is highly efficient at identifying surfaces that are either does not have biofouling or heavily fouled, likely because of their more unique characteristics compared to the other classes.

In the other hand, the *Light* and *Moderate* classes show lower precision, recall, and mAP scores (mAP@0.5 of 0.462 and 0.488, respectively).The moderate severity of biofouling in these classes is more difficult to differentiate, since the visual variations such as texture, color, and extent are not as extreme. The balanced precision and recall in these classes shows that the model can predict a big number of instances, but still not as certainly as in the extreme classes.

Overall, the average performance for all of the classes (Precision 0.520, Recall 0.578, mAP@0.5 0.602) shows that the model's ability to generalize across multiple severity levels while marking areas for possible improvement in detecting subtle biofouling patterns. This per class analysis points out the importance of considering class specific performance, especially in operational settings where identifying light or moderate biofouling correctly may be really important for early intervention and maintenance planning.

J. Qualitative Results

Qualitative comparison of ground truth and predictions on validation data. The model separates *Clean* and *Heavy* cases, cases cleanly, while the rest uncertain classes are bounded around adjacent severities (*Light* vs. *Clean*, *Moderate* vs. *Heavy*), consistent with the confusion matrix analysis.

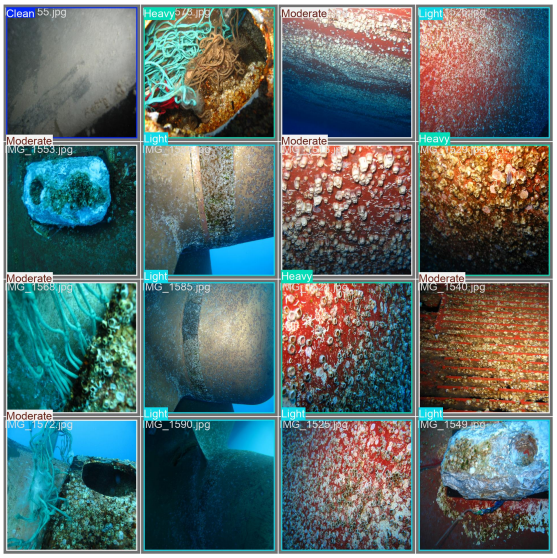

Fig. 4. Ground-truth labels for a validation batch.

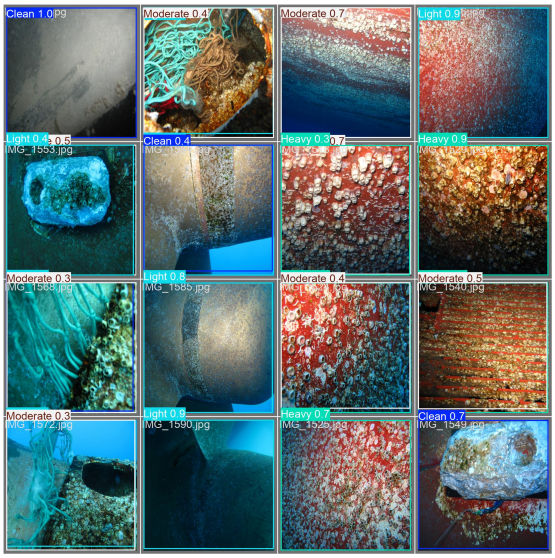

Fig. 5. Model predictions for the same batch.

REFERENCES

[1] R. Capocci, E. Omerdić, G. Dooly, and D. Toal, "Development and testing of remotely operated vehicle for inspection of offshore renewable devices," in *Technological Innovation for Resilient Systems* (L. Camarinha-Matos, K. Adu-Kankam, and M. Julashokri, eds.), vol. 521 of *IFIP Advances in Information and Communication Technology*, Springer, 2018.

[2] P. Pachaiyappan, G. Chidambaram, A. Jahid, and M. H. Alsharif, "Enhancing underwater object detection and classification using advanced imaging techniques: A novel approach with diffusion models," *Sustainability*, vol. 16, no. 17, p. 7488, 2024.

[3] V. Brandou, A.-G. Allais, M. Perrier, E. Malis, P. Rives, J. Sarrazin, and P.-M. Sarradin, "3d reconstruction of natural underwater scenes using the stereovision system iris," in *OCEANS 2007 - Europe*, IEEE, June 2007.

[4] C. Mai, J. Liniger, and S. Pedersen, "Semantic segmentation using synthetic images of underwater marine-growth," *Frontiers in Robotics and AI*, vol. 11, 2024.

[5] M. Jian, N. Yang, C. Tao, H. Zhi, and H. Luo, "Underwater object detection and datasets: a survey," *Intelligent Marine Technology and Systems*, vol. 2, no. 1, 2024.

[6] S. Che, Z. Li, Z. Shi, M. Gao, and H. Tang, "Research on an underwater image segmentation algorithm based on yolov8," in *Journal of Physics: Conference Series*, vol. 2644, p. 012013, IOP Publishing, 2023.

[7] F. Liu and M. Fang, "Semantic segmentation of underwater images based on improved deeplab," *Journal of Marine Science and Engineering*, vol. 8, no. 3, p. 188, 2020.

[8] M. Waszak, A. Cardaillac, B. Elvesaeter, F. Rødølen, and M. Ludvigsen, "Semantic segmentation in underwater ship inspections: Benchmark and data set," *IEEE Journal of Oceanic Engineering*, 2022.

[9] S. Hong and J. Kim, "Three-dimensional visual mapping of underwater ship hull surface using piecewise-planar slam," *International Journal of Control, Automation and Systems*, vol. 18, pp. 564–574, 2020.

[10] M. Ferrera, A. Arnaubec, C. Boittiaux, I. Larroche, and J. Opderbecke, "Vision-based 3d reconstruction for deep-sea environments: Practical use for surveys and inspection," in *OCEANS 2023*, pp. 1–7, IEEE, June 2023.

[11] M. Malkawi, S. Rabi, O. AlBalbaki, M. Alrifaie, A. Abuelayyan, D. Aljbour, M. Al-zaatrani, Y. Alwreikat, S. Alkalili, R. Nedal Ali, A. Allam, Z. Nader, A. Al-daamsa, and M. Al-Manaseer, "Using wadi rum desert and terrains as a space testing fields," in *74th International Astronautical Congress (IAC)*, (Baku, Azerbaijan), International Astronautical Federation (IAF), October 2023. Interactive Presentation, IAF Space Exploration Symposium (IPB).

[12] Ultralytics, "Ultralytics YOLOv8." Online, 2024. Available: https://github.com/ultralytics/ultralytics.

[13] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020. doi:10.48550/arXiv.2004.10934.

[14] Ultralytics, "Ultralytics YOLOv8 Segmentation." Online, 2024. Available: https://docs.ultralytics.com/tasks/segment/.

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[16] Ultralytics, "YOLO Model Evaluation Metrics." Online, 2024. Available: https://docs.ultralytics.com/modes/val/#metrics.

[17] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, vol. 8693 of *Lecture Notes in Computer Science*, pp. 740–755, Springer, Cham, 2014. doi: 10.1007/978-3-319-10602-1_48.

[18] Ultralytics, "YOLO docs – train mode and data augmentation." Online, 2024. Available: https://docs.ultralytics.com/modes/train.