Scallop Segmentation Using Aquatic Images with Deep Learning Applied to Aquaculture

Wilder Nina[1], Nadia L. Quispe[2], Liz S. Bernedo-Flores[3], Marx S. García[4], Cesar Valdivia[5], Eber Huanca[6]
VEOX Inc., Arequipa, Peru[1,2]
Instituto Tecnológico de la Producción - CITE pesquero Piura[4]
Departamento de Ingeniería Eléctrica y Electrónica, Universidad Católica San Pablo, Arequipa, Peru[3,5,6]

***Abstract*—This study evaluates the performance of deep learning-based segmentation models applied to underwater images for scallop aquaculture in Sechura Bay, Peru. Four models were analyzed: SUIM-Net, YOLOv8, DETECTRON2, and CenterMask2. These models were trained and tested using two custom datasets: SEG_SDS_GOPRO and SEG_SDS_SF, which represent diverse underwater scenarios, including clear and turbid waters, varying current intensities, and sandy substrates. The primary aim was to automate scallop identification and segmentation to improve the efficiency and safety of aquaculture monitoring. The evaluation showed that SUIM-Net achieved the highest accuracy of 93% and 94% on the SEG_SDS_GOPRO and SEG_SDS_SF datasets, respectively. CenterMask2 performed best on the SEG_SDS_SF dataset, with an accuracy of 96.5%. Additionally, a combined dataset was used, where YOLOv8 achieved an accuracy of 88%, demonstrating its robustness across varied conditions. Beyond scallop segmentation, the models were extended to detect six additional marine classes, achieving a maximum accuracy of 39.90% with YOLOv8. This research underscores the potential of deep learning techniques to revolutionize aquaculture by reducing operational risks, minimizing costs, and enhancing monitoring accuracy. The findings contribute valuable insights into the challenges and opportunities of applying artificial intelligence in underwater environments.**

Keywords*—Image segmentation; object detection; deep learning; computer vision; aquaculture; scallop segmentation; aquatic images

I. Introduction

Scallop cultivation has been developed in Peru since the 1970s, initially on the central coast. Due to its rapid growth, it expanded to other regions along the Peruvian coastline. By the early 1990s, scallop farming became an important economic activity, generating significant employment opportunities in aquaculture [1]. Sechura Bay, located on the northern coast of Peru (latitude: -5.742118, longitude: -80.867822) (see Fig. 1), is one of the largest semi-enclosed bays in the Peruvian sea, covering approximately 89 kilometers. It is bordered by Punta Gobernador to the north and Punta Aguja to the south, with a cultivation area of 6,752.48 hectares (Ha). The bay has depths of less than 30 meters within 10 kilometers of the coastline. Numerous scallop hatcheries have been established on the seabed in this region. During the early stages of production, scallop seeds are classified based on their valve length, which typically ranges from 2.5 cm to 4.0 cm. They are cultivated at a density of 1.5 bunches per square meter, where each bunch contains 96 seeds [2]. During hatchery cultivation, seeds that are confined or have irregular sizes are extracted and replanted in better-suited areas. At this stage, growth and mortality rates are also monitored to estimate harvest levels [3]. Scallop stock assessments in hatcheries are conducted monthly or at least once before harvesting. These assessments involve divers manually collecting population samples using a 1m² squared frame tool. Divers carefully place the frame on the seabed and harvest the scallops within its boundaries. On the support boat, the collected scallops are counted and measured using a vernier caliper and a malacometer, and the data is recorded. This process is repeated multiple times to estimate the total scallop stock in units and bunches [2]. However, these manual operations pose significant risks to divers. The average hatchery depth is 10 meters, requiring divers to spend extended periods underwater. While safety guidelines recommend a maximum working time of four hours per day, divers often exceed this limit, working up to seven or eight hours. Additionally, divers rely on inadequate compressed air systems, which can lead to nitrogen narcosis, posing a severe health risk, including fatal consequences [4]. To address these challenges, the integration of advanced methodologies and technologies for scallop production assessment is essential. Automation in data collection and analysis can significantly reduce operational risks, costs, and time associated with aquaculture monitoring.

This study explores the application of deep learning and computer vision to automate scallop segmentation, aiming to enhance accuracy, improve efficiency, and minimize risks. Two custom datasets (SEG_SDS_GOPRO and SEG_SDS_SF) were developed to evaluate the performance of four segmentation models: SUIM-Net, YOLOv8, DETECTRON2, and CenterMask2. Additionally, these models were extended to detect six marine classes, expanding their applicability in aquaculture. This research provides valuable insights into leveraging artificial intelligence for sustainable and efficient scallop farming.

This research was carried out as part of a project (Monitoring platform for non-extractive sampling of hydrobiological resources through the development of a customized underwater robot with advanced deep learning and computer vision techniques. Application case: sea floor hatchery of scallops (argopecten purpuratus) in Sechura bay) financed by the Peruvian research council FONDECYT.

This work is organized into six sections. Section II introduces related works, including the methodology for performing the datasets and SUIM-Net, YOLOv8, DETECTRON2, and CenterMask2 models; Section III describes the datasets and methods; Section IV describes experiments and analysis; Section V describes discussion of Results; finally, Section VI summarizes the work.

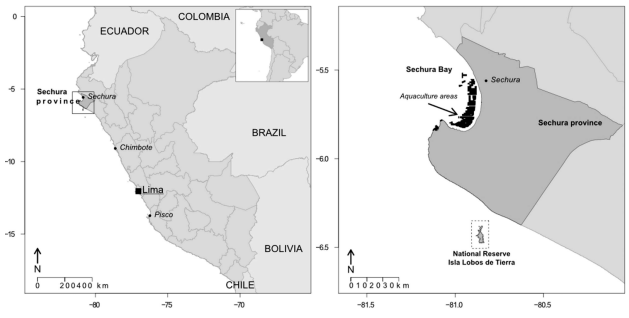


Fig. 1. Left panel: Situation the study setting Sechura Bay in Secura province in Peru. Right panel: Sechura Bay in the province of Sechura, indicating the aquaculture concession area [3].

II. Related Works

One of the most popular underwater datasets for marine species detection and classification is F4K [5]. This dataset was performed using 10 cameras between 2010 and 2013 in Taiwan and has been used for multiple detection and classification algorithms. The F4K dataset is large and contains videos and images with complex scenes and diversity of marine species. Another large data set is the Electronic Library of Deep Sea Images (J-EDI) [6] which are consistuded of videos and images of deep-sea organisms captured by remotely operated underwater vehicles (ROVs) [7]. This images are labeled at the image level and have been used to interconnect convolutional neural networks - CNN for the detection of deep-sea hydrobiological organisms. The author in [8] considers three objectives for collecting underwater images: (1) Broad diversity of underwater scenes, having different quality degradation characteristics and a wide range of image content (2) Big amount of data and (3) High quality of images. The author in [9] introduces USR-248 dataset, which is a large-scale data set of three sets of images, which were rigorously collected during ocean explorations, field experiments and some resources are publicly available online. The UFO-120 dataset is depicted in [10] which contains more than 1500 trining sample. These images were captured in oceanic explorations in many different locations and kinds of water. [9] presents the USR-248 dataset, which main characteristic is its capability for supervised training. The author in [11] presents a survey of deep learning techniques for performing the underwater image classification (see Table I).

TABLE I. Undersea Dataset [11]

Name	Object	Class	Images
Sipper	Zooplankton	81	>750K
WHOI	Plankton	70	>3.4M
ZooScan	Zooplankton	20	3,771
ZOOVIS	Zooplankton	6	>685K
ISIIS	Plankton	108	>42K
PlanktonSet	Plankton	121	>60K
ZooLake	Plankton	35	17,943
F4K	Pez tropical	23	27,370
LifeCLEF14	Pez tropical	10	19,868
LifeCLEF15	Pez tropical	15	>20K
Temperate fish	Temperate Fish	4	619
Fish-gres	Pez	8	3248
MLC	Arrecife Coral	9	2055
CADDY	Gestos de buzo	16	10,322

The underwater environment is one of the most challenging conditions for object detection using sensors. For this, sonars and cameras are mainly used. Sonars are sensitive to geometric structure and provide information in an environment of very low or no visibility, but their drawback is that they can only measure the difference in distance between objects [12]. On the other hand, underwater images allow colors, textures and edges to be easily detected as long as visibility conditions exist. However, in a real environment these conditions are altered, resulting in degradation of image quality due to wavelength-dependent absorption and dispersion, including forward and backward scattering. In addition, marine snow introduces noise and increases dispersion effects, reducing visibility, contrast and even distorting colors. Despite this, it has more potential to detect characteristics of objects, compared to sonar sensing (applied for scallop production assesment). But for this it is necessary to pre-process the images to improve their quality and extract information [6], [8]. Object detection in computer vision can be acquired with a single camera or multiple cameras (stereo vision). In [13] was developed algorithms for image detection using a stereo vision camera system for detecting objects of sea floor without previous information. The region of interest - ROI was determined by pixel similarity concentrations in *area*, *color* and *shape*. These criteria were applied to segment and validate the image recognition algorithms. For the experimental tests [13] used three underwater datasets: Garda, Portofino and Soller; each one has its own characteristics and challenges, which allowed the author to evaluate it in different underwater situations, demonstrating that the proposed algorithm is robust to changes in lighting and turbidity of the water, however it does not consider the challenges of superimposing objects intermingled on top of each other.

The UFO-120 dataset [10], it addresses the problem of simultaneous enhancement and super-resolution (SESR) for underwater robotic vision, providing an efficient solution for real-time applications. It also introduces the Deep SESR model, which is a generative model based on a residual network that can learn to restore the perceptual qualities of the image. The proposed Deep SESR model offers perceptually improved FC imaging and saliency prediction through a single efficient inference. Enhanced images restore color, contrast and sharpness at higher scales to facilitate better visual perception. For semantic segmentation [14] proposed to improve the encoder and decoder structures of the DeepLabv3+ network, in order to improve the appearance of the segmenting object and prevent its pixels from mixing with the pixels of other classes, improving the accuracy of the segmentation of the object boundaries and preserving feature information. To do this, it added the unsupervised color correction - UCM method in the encoder structure to improve the image quality, then added two upsampling layers to the decoder structure to obtain greater feature information using the (*backbone*) Xception_65. Regarding the data set, this was self-made, some images were obtained from public resources on the Internet and another part was obtained from video images taken by an underwater laboratory robot (HUBOS-2K, Hokkaido University).

Due to the absorption of light and the deeper the water, underwater images usually acquire a greenish-blue color. To counteract this effect [15] applies a combination of maximum RGB method and gray tone method to achieve underwater vision improvement. Then it proposes a method based on

CNN for the detection and classification of objects in real time, using the third version of You Only Look Once network (YOLOv3), according to the characteristics of underwater vision, two improved schemes are applied to modify the CNN structure. The proposed YOLOv3 framework is divided into three parts: feature extraction, object detection, and prediction of bounding box coordinates and object confidence. Feature extraction is performed using the Darknet-53 network. Object detection is performed by detecting objects at multiple scales, and predicting bounding box coordinates and object confidence is performed by a fully connected neural network. The database used for the tests was obtained from a video recorded by an underwater robot, with approximately 18000 images. In [16] was proposed an underwater scallop recognition algorithm using an improved version of the YOLOv5 [17] neural network. The authors first designed a new lightweight backbone model to replace the original one of YOLOv5, using group convolution and inverse residual block, which helps improve the accuracy and accelerate the detection speed. The proposal used the k-means algorithm for clustering analysis of the data set to reduce the initial prediction layer of the model and the enhancement module was used by the adaptive dark channel algorithm to improve the clarity of the blurred image. The data set used was self-constructed in a laboratory environment, with a swimming pool and a GoPro 5 camera, 2200 labeled images were acquired. The data was increased to 4400 using dark channel image enhancement. The experimental results indicate that the precision rate, recall rate, F1 and mPA of the proposed algorithm reached 90.8%. In [18] was proposed state-of-the-art real-time object detection algorithms are trained and inferred on underwater images of a hypothetical inshore aquaculture operation to investigate model selection and hyper-parameters for object detection in underwater images. The simulation results show that 54.2% mean average precision is achieved by YOLOX-m and 97.1 frames per second inference processing confirmed by YOLOX-tiny.

In [19] was proposed a detection and Classification of Subsea Objects in Forward-Looking Sonar and Electro-Optical Sensors for ROV Autonomy The objects of interest are MLOs(Mine Like Objects) that need to be located, identified, and inspected by an autonomous submersible robot. SeeByte used Deep Learning Neural Network (DNN) on both of these sensor feeds yielding a very robust detection and classification system.

state-of-the-art real-time object detection algorithms are trained and inferred on underwater images of a hypothetical inshore aquaculture operation to investigate model selection and hyper-parameters for object detection in underwater images. The simulation results show that 54.2% mean average precision is achieved by YOLOX-m and 97.1 frames per second inference processing confirmed by YOLOX-tiny.

Related work is crucial to contextualize our results within the advancement of underwater image segmentation. Previous studies have addressed segmentation in marine environments with different approaches, from CNN-based models to more advanced architectures such as Transformer-based vision models. In particular, works such as [10] and [11] have demonstrated the importance of improving image quality to optimize detection. Our study complements these efforts by evaluating models in real-world conditions of fan shell farming in Sechura Bay, an environment that presents unique challenges of visibility and environmental variability.

III. Datasets and Methods

In this study, a customized dataset of scallop images was collected from a seabed nursery in the bay of Sechura (Peru), under various underwater environmental conditions. The Fig. 2 illustrates the methodology, which is structured in three main stages: Image Acquisition, Image Preprocessing, and Scallop Training and Detection.

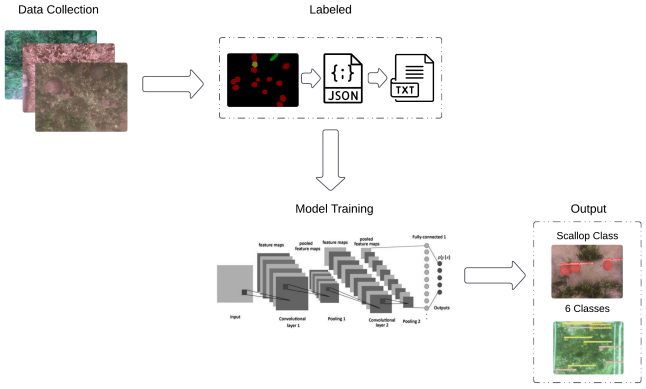


Fig. 2. Structure of the methodology used.

A. Datasets

Unlike reference datasets in underwater image segmentation, such as F4K and UFO-120, which were collected in previous decades, our study is based on recent images captured in Sechura Bay. Data collection was conducted in 2023 and 2024 using high-resolution cameras, which ensures that the dataset accurately reflects current ecosystem conditions. This update is fundamental to improve the applicability of the models in real aquaculture monitoring scenarios, allowing for more accurate and adaptive detection of environmental changes (Table II).

TABLE II. Descriptions of Datasets

DataSet name	Type	Description
SEG_SDS_GOPRO (SEGMENTATION SCALLOP DATASET WITH GOPRO)	Training and Validation	It uses a PVC sampling structure with a GoPro H9 camera for image segmentation.
SEG_SDS_SF (SEGMENTATION SCALLOP DATASET WITH SMARTFRAME)	Training and Validation	It uses a steel tube structure called smartframe v1.0 adapted Raspberry Pi V2.1 camera include enclosure.

Given the limited availability of databases featuring marine species in the reviewed state-of-the-art literature, a custom dataset was created to guarantee the presence of scallops in the images. This approach also facilitates pretraining the neural network with relevant data. Table III lists the visited locations within Sechura Bay (see Fig. 1) where scallop images were collected. These locations span the northern,

central, and southern regions, capturing a variety of underwater topographies, including sandy, rocky and mixed terrains.

TABLE III. HATCHERIES WHERE SCALLOP IMAGES WERE ACQUIRED

Zone	Scallop hatchery name in Sechura bay
North	Chulliyache - Caballero de los Mares, boarding at the Matacaballo DPA
Center	Las Delicias, boarding at the Parachique Zonal Fishery Terminal
South	Barrancos - Amigos Unidos, boarding at the Parachique Zonal Fishery Terminal
South	Sea Corp Camp, Vichayo Aquaculture Production Center (CPAV), Vichayo Zone

Scallop images were captured manually with the support of a professional diver and a support boat using two types of cameras: (1) GoPro Hero 9 and (2) Raspberry Pi V2.1. The cameras were mounted on a square frame with a 1-meter edge, elevating them 40 cm above the seabed. Fig. 3 illustrates the image acquisition process.

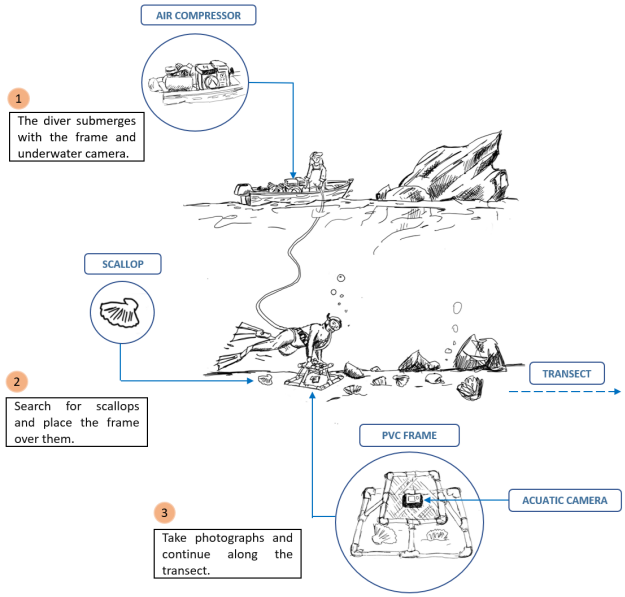


Fig. 3. Image acquisition process.

To ensure accurate scallop identification, the collected images were manually labeled with the assistance of a hydrobiology specialist from the Center for Productive Innovation and Technology Transfer (CITE) in Piura, Peru. This labeling process was critical for creating a reliable training dataset.

Fig. 4 and 5 present the underwater environment of Sechura Bay, revealing the composition of the seafloor and the diversity of its hydrobiological resources. These images provide a detailed visual representation of the natural habitat inhabited by scallops, along with other marine species, highlighting the biological richness captured in the dataset. Understanding the characteristics of this ecosystem is fundamental for the correct annotation and classification of the images, ensuring the accuracy and reliability of the dataset used in the training of deep learning models.

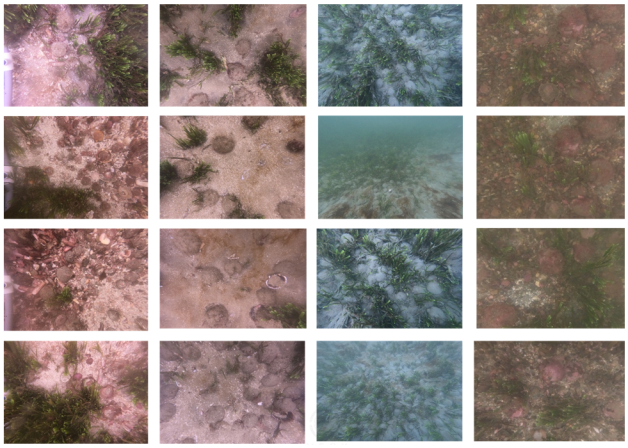


Fig. 4. Dataset SEG_SDS_GOPRO acquired in Sechura Bay

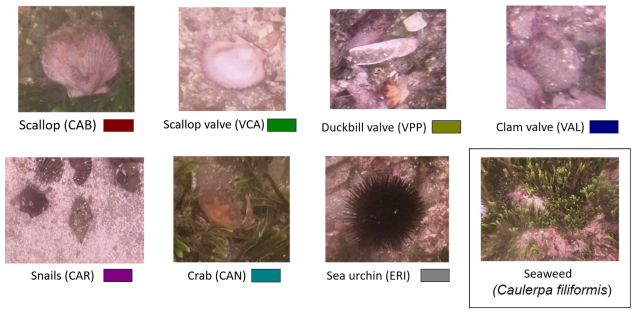


Fig. 5. Marine species considered in the labeling.

Additionally, Fig. 6 shows the SEG_SDS_SF dataset acquired in Sechura Bay.

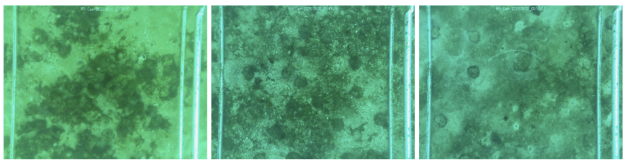


Fig. 6. Dataset SEG_SDS_SF acquired in Sechura Bay.

In Table IV, the classes with which the training is carried out and the labels of each of them from the SEG_SDS_GOPRO and SEG_SDS_SF dataset are observed.

TABLE IV. DATASET CLASSES AND LABELS

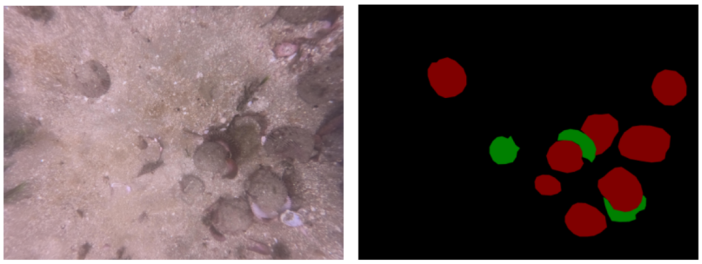
DataSet name	Classes	Labels
SEG_SDS_GOPRO	Scallop	2418
	Scallop Shell	165
	Duck Beak Shell	163
	Snail	57
	Crab	9
	Clam Shell	4
SEG_SDS_SF	Scallop	715
	Scallop Shell	42
	Duck Beak Shell	683

B. Labeling Image Process

The acquired images were labeled and reviewed by the CITE expert for validation or correction of the labeling. For

labeling images, seven species of scallops where considered: (1) Scallop - CAB, (2) Scallop valve - VCA, (3) Duckbill valve - VPP, (4) Clam valve - VAL, (5) Snails - CAR, (6) Crab - CAN, and (7) Sea urchin - ERI, Fig 7 shows each scallop specie image. For each one a different color was assigned. The seaweed (*Caulerpa filiformis*) in labeling process is considered as part of seabeed. The images were labeled with the "labelme" tool.

A validation table was created to document image characteristics, including image ID, collection day, filtering status, species present, quantity per species, and seabed type. This table facilitated coordination with the specialist for label verification and served as an inventory.



a) Original image b) Labeled image

Fig. 7. Results of labeling image process.

C. Models

The models selected for this study were SUIM-Net, YOLOv8, DETECTRON2 and CenterMask2, due to their balance between accuracy and computational efficiency in underwater image segmentation. Also, at the time of selection, these models represented some of the most recent and advanced versions of their respective architectures, which guaranteed better performance in detection and segmentation tasks.

1) SUIM-Net: The SUIM-Net is a fully convolutional semantic segmentation model (FCN) proposed by [20], which is based on an encoder-decoder architecture with skip connections between composite layers. The network employs residual learning and optional residual skip blocks to enhance performance. The encoder extracts features from input images, and the decoder generates binary pixel labels per channel for each object category. It utilizes a proprietary dataset called SUIM (Semi-supervised Underwater Image Manipulation) which is presented in Fig. 8. For training, it applies various image transformations for data augmentation. The results obtained show that SUIM-Net exhibits improved execution time compared to FCN models, SegNet, UNet, VGG-based encoder-decoders, DeepLab, and PSPNet

2) YOLOv8: YOLOv8, developed by Ultralytics, builds upon YOLOv5 with key improvements, the model is shown in Fig. 9 where it uses a similar backbone with modifications in the CSPLayer, now called the C2f module, to combine high-level features and contextual information, enhancing detection accuracy. YOLOv8 adopts an anchor-free design with a decoupled head, allowing separate processing of objectivity, classification, and regression tasks for improved accuracy.

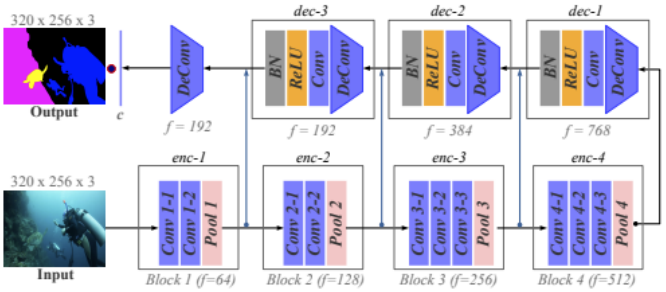


Fig. 8. SUIM-Net model [20].The end-to-end architecture of SUIM-Net$_{VGG}$: first four blocks of a pre-trained VGG-16 model are used for encoding, followed by three mirrored decoder blocks and a deconv layer.

The output layer applies the sigmoid function for objectivity scores and softmax for class probabilities [21]. YOLOv8 also includes a semantic segmentation variant, YOLOv8-Seg, which uses CSPDarknet53 as the backbone and C2f for feature extraction. It features two segmentation heads for mask prediction and detection heads with five modules for bounding box prediction [22].

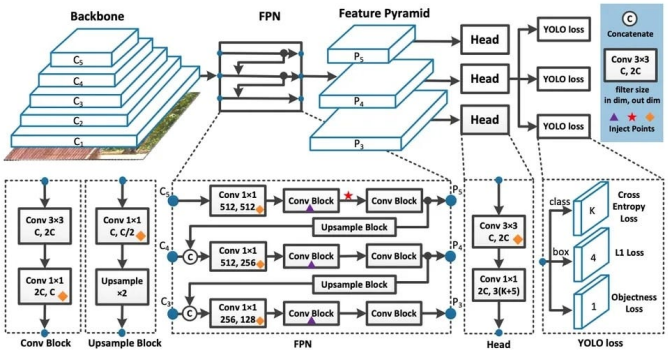


Fig. 9. YOLOv8 model [23].

3) DETECTRON2: DETECTRON2 [24] allows us to implement models of different architectures. These pretrained models in the DETECTRON2 Zoo Model have a structure that follows the meta-architecture provided by the base code. It is a framework that includes high-quality implementation of state-of-the-art instance segmentation algorithms, such as Faster R-CNN and Mask R-CNN. The architecture of the DETECTRON2 model in Fig. 10, consists of three main parts: the backbone network, region proposal network (RPN), and ROI (Region of Interest) head. Given an input image, the backbone network extracts feature maps at various scales with different receptive fields. The RPN detects object regions from feature maps and various scales by default. Finally, the box head crops the feature maps into different sizes and finds the location of the boxes and classification labels in addition to fully connected layers. In our tests, we used ResNet with transformers. ResNets are convolutional neural networks that utilize skip connections enabling a deep architecture with many layers.

4) CenterMask2: The CenterMask2 model is an instance segmentation approach that decomposes this task into two main parallel subtasks [26]. The model is shown in Fig. 11,

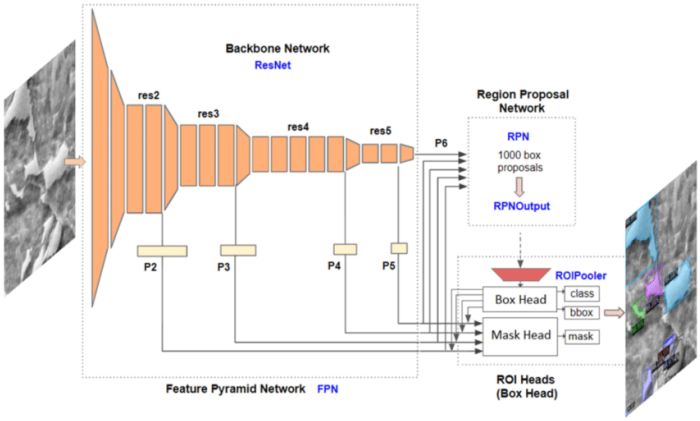


Fig. 10. DETECTRON2 model [25].

where first it performs a local prediction to separate instances even under overlapping conditions. Secondly, it carries out global prominence generation to segment instances in the complete image, pixel by pixel. Subsequently, the outputs of these two branches are assembled to form the final instance mask. It utilizes four types of backbone networks such as MobileNetV2, VoVNetV2-19, VoVNetV2-39, and ResNet-50. It is noteworthy that the accuracy of this model surpasses that of other instance segmentation methods.

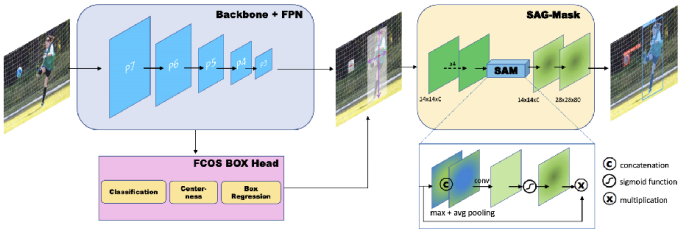


Fig. 11. CenterMask2 model [26].

IV. Experiments and Analysis

The experiments aimed to evaluate the performance of four deep learning models—SUIM-Net, YOLOv8, DETECTRON2, and CenterMask2—in scallop segmentation and multi-class detection using two custom datasets: SEG_SDS_GOPRO and SEG_SDS_SF. Each dataset represents different underwater conditions to assess model robustness. These models were selected based on their capabilities in segmentation and detection tasks in challenging environments. SUIM-Net was chosen for its specialization in underwater image segmentation, particularly in low-visibility conditions. YOLOv8 was included due to its high-speed detection and strong generalization in real-time applications. DETECTRON2, provides advanced segmentation architectures optimized for complex object detection, while CenterMask2 enhances instance segmentation with high precision in intricate visual scenes.

A. Training and Validation Scallop

The four models demonstrated varying segmentation capabilities on the two datasets. SUIM-Net performed well on SEG_SDS_GOPRO, accurately identifying scallops in challenging conditions. CenterMask2 excelled on SEG_SDS_SF, showing strong segmentation results in clearer underwater settings. YOLOv8 and DETECTRON2 provided balanced performance across both datasets, effectively handling scallop detection under diverse environmental conditions.

To provide a clear understanding of their performance, Fig. 12 presents a visual comparison of segmentation results for a representative sample across all four models. This visualization effectively highlights each model's strengths and limitations in detecting scallops, offering valuable insight into their accuracy and adaptability.

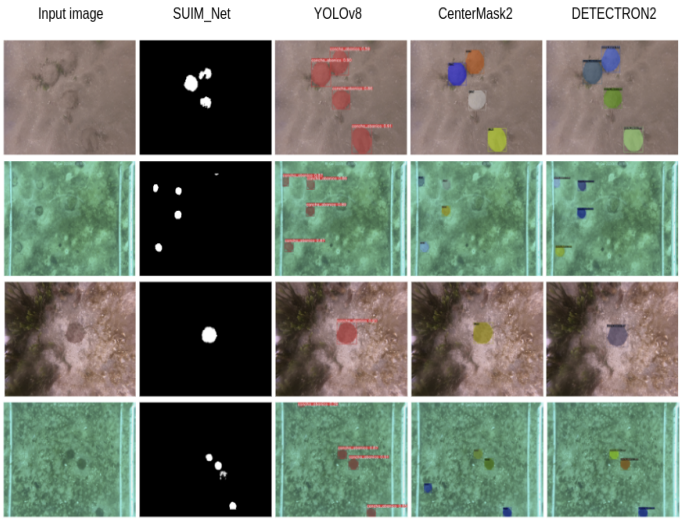


Fig. 12. Comparison of the 4 trained models.

B. Training and Validation of Six Classes

Three different models, SUIM_Net, YOLOv8 and DETECTRON2, are used for the detection of six specific marine-related classes in images. Each model employs different segmentation methodologies and detection strategies, making them suitable for different contexts and scenarios. These models have been trained to identify six marine-related classes: scallop, scallop shell, duckbill shell, snail, crab, and clam shell. Each model independently processes a set of images, predicting and delineating instances of these classes based on their unique capabilities and detection approach. Leveraging their individual strengths, the models aim to provide comprehensive and accurate identification of target objects within the images. The results obtained from each model are shown in the following images, allowing a direct comparison of their predictions. Fig. 13, 14, and 15 show the detections performed by SUIM_Net, YOLOv8, and DETECTRON2, respectively.

Fig. 13. Prediction of 6 classes using the SUIM_Net model.

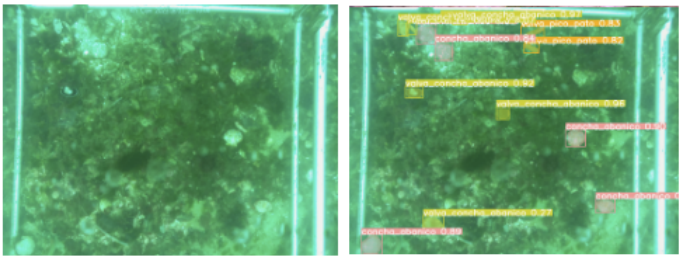


Fig. 14. Prediction of 6 classes using the YOLOv8 model.

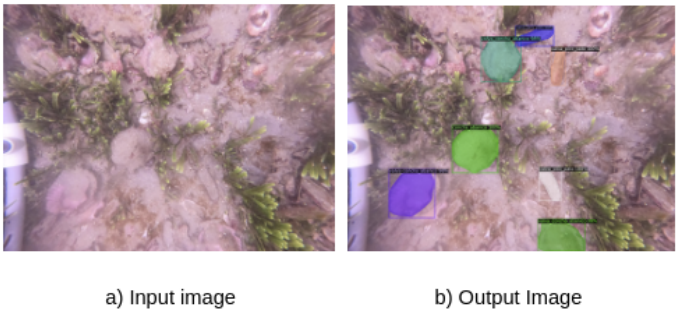


Fig. 15. Prediction of 6 classes using the DETECTRON2 model.

V. Discussion of Results

The results obtained in the SEG_SDS_GOPRO and SEG_SDS_SF datasets present significant variations due to differences in environmental conditions and characteristics of each dataset. SEG_SDS_GOPRO was captured with a GoPro Hero 9 mounted on a PVC structure, which allowed obtaining high-resolution images, although with possible optical distortions due to water refraction. In contrast, SEG_SDS_SF used a Raspberry Pi V2.1 camera integrated into a smartframe, resulting in lower-resolution images and greater variability in lighting. These differences in capture conditions directly impact the performance of the models, as they influence the accuracy of object segmentation and detection.

1) Results of the 2 datasets with the Scallop class: The SUIM-Net model was applied to the SEG_SDS_GOPRO and SEG_SDS_SF datasets for scallop segmentation, achieving accuracies of 93% and 94%, respectively. The loss plots shown in Fig. 16 illustrate the model's performance during training. In Graph (a), corresponding to the SEG_SDS_GOPRO dataset, the loss steadily decreases across epochs, with minor fluctuations, indicating consistent learning and optimization during training. Similarly, Graph (b), for the SEG_SDS_SF dataset, the loss drops sharply at first and then stabilizes, demonstrating effective learning and optimization in both cases.

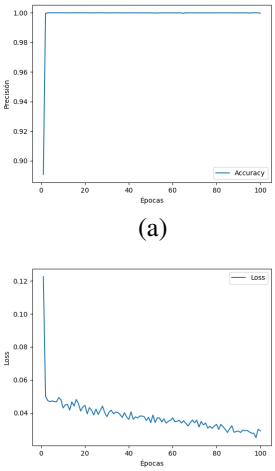


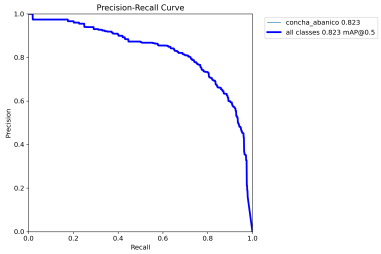
Fig. 16. Loss graphs SUIM_Net SEG_SDS_GOPRO (a) and SEG_SDS_SF (b) datasets.

While the YOLOv8 model was applied for the scallop images collected in Sechura bay. The accuracy of the a) SEG_SDS_GOPRO data set was 82%, and for the b) SEG_SDS_SF data set, its The accuracy was 95% Likewise, training was carried out with c) SEG_SDS_GOPRO + SEG_SDS_SF (Fig. 17) which obtained an accuracy of 88%. The following images refer to the experiments with each of the data sets.

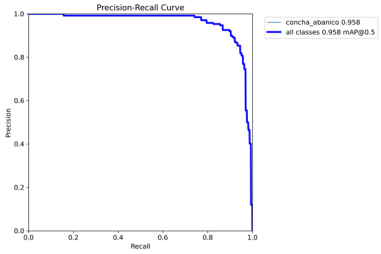
In our research, we evaluated the performance of DETECTRON2. We use different data sets: SEG_SDS_GOPRO and SEG_SDS_SF. Our goal is to understand how this model performs in different environments and imaging conditions. We start by evaluating the model trained on the (a) SEG_SDS_GOPRO dataset. He The results revealed an accuracy of 82% using DETECTRON2. This figure represents the model ability to correctly identify objects present in test images. On the other hand, when evaluating the model trained with the (b) SEG_SDS_SF data set, We see a noticeable improvement in accuracy. We achieved 96% accuracy using DETECTRON2. Likewise, the evaluation was carried out with the (c) SEG_SDS_GOPRO + SEG_SDS_SF data set, which obtained an accuracy of 80%. It is important to note that these results provide an initial view of the performance of our models with DETECTRON2.

In the following Fig. 18 you can see the loss function graphs of the datasets, where we can see that our function is decreasing which indicates that our model is efficient.

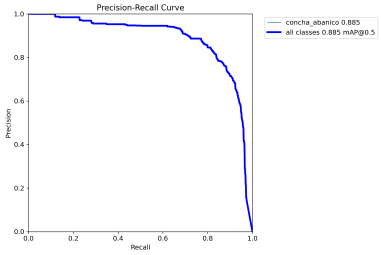
In Fig. 19, the performance of the CenterMask2 instance segmentation model was evaluated on different different data sets: SEG_SDS_GOPRO and SEG_SDS_SF. Results were visualized using average precision (mAP50) graphs, which represent the highest detection precision at 50%. The mAP50



(a)



(b)



(c)

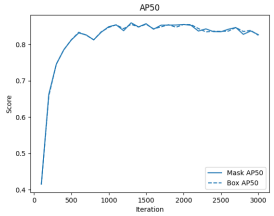
Fig. 17. Segmentation results with YOLOv8 model.

plots for the (a) SEG_SDS_GOPRO set showed an accuracy score of 81% for the CenterMask2 model. On the other hand, in the set (b) SEG_SDS_SF, CenterMask2. The model achieved a noticeably higher accuracy of 96% based on mAP50 charts. Finally, the data set of both datasets gave an accuracy of 83%.

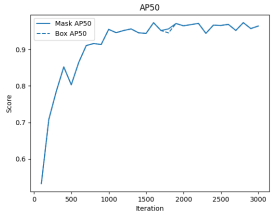
TABLE V. DATASET AND PRECISION RESULTS FOR EXPERIMENTS SCALLOP

Modelo	DataSet	F-Score	mAP
SUIM-Net	SEG_SDS_GOPRO + SEG_SDS_SF	52.45%	82.04%
	SEG_SDS_GOPRO	59.47%	93.06%
	SEG_SDS_SF	60.64%	94.05%
YOLOv8	SEG_SDS_GOPRO + SEG_SDS_SF	83.00%	88.50%
	SEG_SDS_GOPRO	76.00%	82.30%
	SEG_SDS_SF	91.00%	95.80%
DETECTRON2	SEG_SDS_GOPRO + SEG_SDS_SF	71.44%	80.84%
	SEG_SDS_GOPRO	75.62%	82.51%
	SEG_SDS_SF	85.60%	96.40%
CenterMask2	SEG_SDS_GOPRO + SEG_SDS_SF	74.46%	83.35%
	SEG_SDS_GOPRO	72.15%	81.23%
	SEG_SDS_SF	81.69 %	96.56%

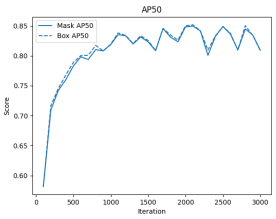
2) Results of the 2 datasets with the 6 classes: The Fig. 20 shows the precision and loss metrics of the SUIM_Net model with the SEG_SDS_GOPRO + SEG_SDS_SF dataset, which was trained with the 6 classes in the image. You can see how the precision increases, giving greater reliability to the model and the loss decreases. which indicates that the model makes more accurate predictions.



(a)

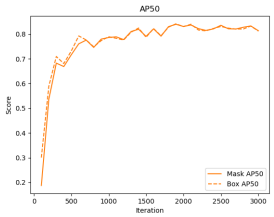


(b)

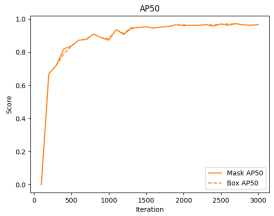


(c)

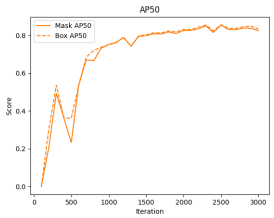
Fig. 18. Segmentation result with DETECTRON2 model.



(a)

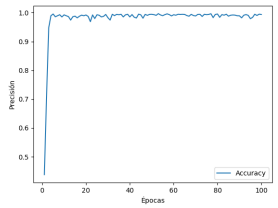


(b)

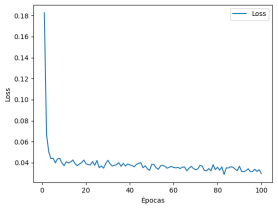


(c)

Fig. 19. Segmentation result with CenterMask2 model.



(a)



(b)

Fig. 20. Segmentation result with SUIM_Net model SEG_SDS_GOPRO + SEG_SDS_SF data set.

In Fig. 21, the performance of the model was evaluated using the Average Precision (AP) metric for six different classes. The precision results for each class are as follows: For the Crab class, the AP was 0.0, while for the Snail class it was 3.06%. The Scallop class demonstrated significantly higher accuracy, with an AP of 60.22%. Likewise, the Clam Shell class obtained an AP of 0.0, while for the Scallop Shell class it was 26.26% and for the Duck Beak Shell class it was 52.45%. These AP values provide a measure of the model's accuracy in detecting objects for each specific class, with higher values indicative of better detection.

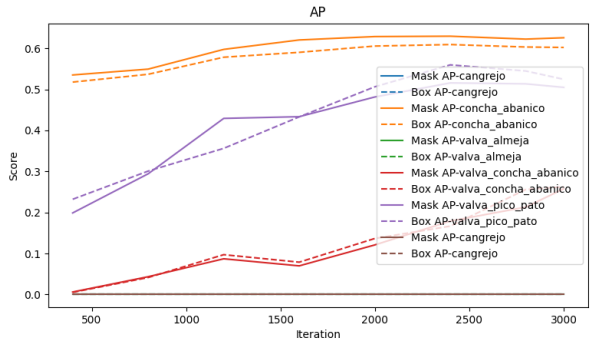


Fig. 21. Segmentation result with DETECTRON2 model SEG_SDS_GOPRO + SEG_SDS_SF data set.

In Fig. 22, the performance of the model was evaluated using the mean precision (mAP) metric for six different classes. The precision results for each class are as follows: For the Crab class, the mAP was 0.0, while for the Snail class it was 1.93%. The Scallop class demonstrated significantly higher accuracy, with a mAP of 88.50%. Likewise, the Clam Shell class obtained a mAP of 0.0, while for the Scallop Shell class it was 45.50% and for the Duck Beak Shell class it was 86.40%. These mAP values provide a measure of the model's accuracy in instance segmentation for each specific class.

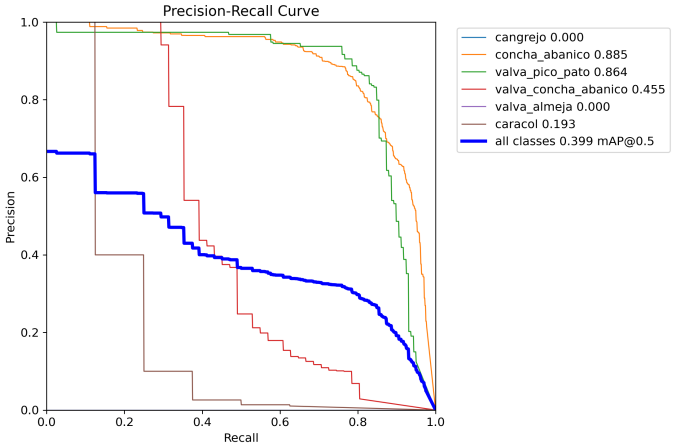


Fig. 22. Segmentation result with YOLOv8 model SEG_SDS_GOPRO + SEG_SDS_SF data set.

TABLE VI. DATASET AND PRECISION RESULTS FOR EXPERIMENTS 6 CLASSES

Modelo	DataSet	F-Score	mAP
SUIM-Net	SEG_SDS_GOPRO + SEG_SDS_SF	38.02%	35.23%
YOLOv8	SEG_SDS_GOPRO + SEG_SDS_SF	40.00%	39.90%
DETECTRON2	SEG_SDS_GOPRO + SEG_SDS_SF	37.01%	36.58%

The results obtained in our experiments reflect a high level of accuracy in the segmentation of scallops and other elements of the underwater environment. Rather than just referencing previous studies, we highlight that SUIM-Net achieved 94% accuracy on SEG_SDS_SF, while CenterMask2 reached 96.5%, which validates the effectiveness of our proposal. Compared to other works in underwater segmentation, our models demonstrated better adaptation to low visibility conditions and lighting variations, highlighting the importance of the collected dataset and the architecture of the selected models. Detailed metrics can be found in Tables V and VI.

VI. CONCLUSIONS

The SEG_SDS_GOPRO and SEG_SDS_SF datasets were designed to address the challenges of scallop segmentation in diverse underwater environments. SEG_SDS_GOPRO represents more complex conditions, including turbid waters and heterogeneous seabeds, while SEG_SDS_SF captures clearer and more structured environments. The combination of both datasets enhances model robustness by providing a balance between variability and specificity. However, the use of different camera devices (GoPro Hero 9 vs. Raspberry Pi V2.1) introduces variability in image quality, lighting conditions, and resolution, which may have impacted the model's performance.

The application of deep learning-based segmentation models demonstrated strong performance, highlighting their potential for automating aquaculture monitoring. Among the tested models, SUIM-Net achieved the highest accuracy 93% in the challenging SEG_SDS_GOPRO dataset, while CenterMask2 excelled 96.5% in the structured SEG_SDS_SF dataset, showcasing their adaptability to different underwater conditions. These results emphasize the importance of dataset-specific model selection for underwater segmentation tasks.

In addition to scallop segmentation, the models were extended to detect six additional marine classes, expanding their applicability. YOLOv8 achieved the highest accuracy 39.90%, followed closely by DETECTRON2 36.58%, while SUIM-Net had the lowest performance 35.23%. These results suggest that YOLO-based architectures, originally designed for object detection, can be effective alternatives for underwater segmentation tasks.

A key limitation was the small dataset size, constrained by logistical challenges in marine data collection and manual annotation times. Future research should integrate semi-automated labeling, data augmentation, and synthetic data generation to improve scalability.

Additionally, while this study focuses on dataset collection and segmentation performance, a more in-depth analysis of underwater image characteristics is necessary. Factors such as image noise, light distortions, and water turbidity can have a significant impact on model robustness and segmentation accuracy.

This research demonstrates the potential of deep learning models for real-time underwater aquaculture monitoring, emphasizing the importance of dataset diversity, environmental adaptability, and model selection. Further improvements in dataset size, validation methods, and robustness analysis in challenging underwater conditions will be crucial for developing more reliable AI-driven solutions for sustainable aquaculture.

ACKNOWLEDGMENT

This work was subsidized by CONCYTEC through the PROCIENCIA program within the framework of the "Special projects for the reactivation E067-2021-02" competition, according to contract or agreement 042-2021.

REFERENCES

[1] A. A. Sanchez Fernandez Baca, "The mariculture of fan shell (argopecten purpuratus) in peru and its relationship with biotrade."

[2] C. Benites Rodriguez, "The development of mariculture in peru with emphasis on fan shell (argopecten purpuratus) and shrimp (penaeus vannamei)," 1988.

[3] L. G. Yenque Moran, "Comparative analysis of the growth of the fan shell, argopecten purpuratus with respect to the suspended culture system and the bottom culture system in the company association de pescadores artesanales acuicultores chulliyachi–sechura," 2021.

[4] C. M. Salazar, R. Bandin, F. Castagnino, and B. Monteferri, "Peruvian fishing gear and methods: illustrative series," 2020.

[5] R. B. Fisher, Y.-H. Chen-Burger, D. Giordano, L. Hardman, F.-P. Lin *et al.*, *Fish4Knowledge: collecting and analyzing massive coral reef fish video data*. Springer, 2016, vol. 104.

[6] J. A. for Marine-Earth Science and Technology, "Jamstec e-library of deep-sea images," *Sensors*, 2016.

[7] J. Neira, C. Sequeiros, R. Huamani, E. Machaca, P. Fonseca, and W. Nina, "Review on unmanned underwater robotics, structure designs, materials, sensors, actuators, and navigation control," *Journal of Robotics*, vol. 2021, pp. 1–26, 2021.

[8] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[9] M. J. Islam, S. S. Enan, P. Luo, and J. Sattar, "Underwater image super-resolution using deep residual multipliers," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 900–906.

[10] M. J. Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," *arXiv preprint arXiv:2002.01155*, 2020.

[11] S. Mittal, S. Srivastava, and J. P. Jayanth, "A survey of deep learning techniques for underwater image classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[12] Z. Chen, Z. Zhang, F. Dai, Y. Bu, and H. Wang, "Monocular vision-based underwater object detection," *Sensors*, vol. 17, no. 8, p. 1784, 2017.

[13] D. L. Rizzini, F. Kallasi, F. Oleari, and S. Caselli, "Investigation of vision-based underwater object detection with multiple datasets," *International Journal of Advanced Robotic Systems*, vol. 12, no. 6, p. 77, 2015.

[14] 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.

[15] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater image processing and object detection based on deep cnn method," *Journal of Sensors*, vol. 2020, 2020.

[16] S. Li, C. Li, Y. Yang, Q. Zhang, Y. Wang, and Z. Guo, "Underwater scallop recognition algorithm using improved yolov5," *Aquacultural Engineering*, vol. 98, p. 102273, 2022.

[17] H. Wang, S. Sun, X. Wu, L. Li, H. Zhang, M. Li, and P. Ren, "A yolov5 baseline for underwater object detection," in *OCEANS 2021: San Diego – Porto*, 2021, pp. 1–4.

[18] A. Imada, T. Katayama, T. Song, and T. Shimamoto, "Yolox based underwater object detection for inshore aquaculture," in *OCEANS 2022, Hampton Roads*, 2022, pp. 1–5.

[19] V. Nguyen, L. Bezanson, and B. Kinnamman, "Detection and classification of subsea objects in forward-looking sonar and electro-optical sensors for rov autonomy," in *OCEANS 2022, Hampton Roads*, 2022, pp. 1–8.

[20] M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. S. Enan, and J. Sattar, "Semantic segmentation of underwater imagery: Dataset and benchmark," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 1769–1776.

[21] J. Terven and D. Cordova-Esparza, "A comprehensive review of yolo: From yolov1 to yolov8 and beyond," *arXiv preprint arXiv:2304.00501*, 2023.

[22] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[23] Y. b. M. M. Contributors, "https://github.com/open-mmlab/mmyolo/tree/main/," Mayo 2023.

[24] G. M. Merz, Y. Liu, C. J. Burke, P. D. Aleo, X. Liu, M. C. Kind, V. Kindratenko, and Y. Liu, "Detection, instance segmentation, and classification for astronomical surveys with deep learning (deepdisc): Detectron2 implementation and demonstration with hyper suprime-cam data," 7 2023. [Online]. Available: http://arxiv.org/abs/2307.05826

[25] M. Ackermann, D. İren, S. Wesselmecking, D. Shetty, and U. Krupp, "Automated segmentation of martensite-austenite islands in bainitic steel," *Materials Characterization*, vol. 191, p. 112091, 07 2022.

[26] Y. Lee and J. Park, "Centermask : Real-time anchor-free instance segmentation," 2020.