



OPEN ACCESS

EDITED BY

Haiyong Zheng,
Ocean University of China, China

REVIEWED BY

Duane Edgington,
Monterey Bay Aquarium Research Institute
(MBARI), United States
Peng Ren,
China University of Petroleum, China

*CORRESPONDENCE

Ignacio A. Catalán
✉ Ignacio@imedea.uib-csic.es

[†]These authors share first authorship

SPECIALTY SECTION

This article was submitted to
Ocean Observation,
a section of the journal
Frontiers in Marine Science

RECEIVED 26 January 2023

ACCEPTED 20 March 2023

PUBLISHED 05 April 2023

CITATION

Catalán IA, Álvarez-Ellacuría A,
Lisani J-L, Sánchez J, Vizoso G,
Heinrichs-Maquilón AE, Hinz H, Alós J,
Signarioli M, Aguzzi J, Francescangeli M
and Palmer M (2023) Automatic detection
and classification of coastal Mediterranean
fish from underwater images: Good
practices for robust training.
Front. Mar. Sci. 10:1151758.
doi: 10.3389/fmars.2023.1151758

COPYRIGHT

© 2023 Catalán, Álvarez-Ellacuría, Lisani,
Sánchez, Vizoso, Heinrichs-Maquilón, Hinz,
Alós, Signarioli, Aguzzi, Francescangeli and
Palmer. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Automatic detection and classification of coastal Mediterranean fish from underwater images: Good practices for robust training

Ignacio A. Catalán^{1*†}, Amaya Álvarez-Ellacuría^{1†},
José-Luis Lisani^{2†}, Josep Sánchez², Guillermo Vizoso¹,
Antoni Enric Heinrichs-Maquilón², Hilmar Hinz¹, Josep Alós¹,
Marco Signarioli¹, Jacopo Aguzzi^{3,4}, Marco Francescangeli⁵
and Miquel Palmer¹

¹Marine Ecology Department, Mediterranean Institute for Advanced Studies (IMEDEA) Spanish National Research Council-University of the Balearic Islands (CSIC-UIB), Esporles, Spain,

²Mathematics and Computer Science Department, University of the Balearic Islands (UIB), Palma, Spain, ³Department of Renewable Marine Resources, Institut de Ciències del Mar (ICM-Spanish National Research Council), Passeig Marítim de la Barceloneta, Barcelona, Spain,

⁴Department of Research Infrastructures for Marine Biological Resources, Anton Dohrn Zoological Station, Naples, Italy, ⁵SARTI Research Group, Electronics Department, Universitat Politècnica de Catalunya (UPC), Vilanova i la Geltrú, Spain

Further investigation is needed to improve the identification and classification of fish in underwater images using artificial intelligence, specifically deep learning. Questions that need to be explored include the importance of using diverse backgrounds, the effect of (not) labeling small fish on precision, the number of images needed for successful classification, and whether they should be randomly selected. To address these questions, a new labeled dataset was created with over 18,400 recorded Mediterranean fish from 20 species from over 1,600 underwater images with different backgrounds. Two state-of-the-art object detectors/classifiers, YOLOv5m and Faster RCNN, were compared for the detection of the 'fish' category in different datasets. YOLOv5m performed better and was thus selected for classifying an increasing number of species in six combinations of labeled datasets varying in background types, balanced or unbalanced number of fishes per background, number of labeled fish, and quality of labeling. Results showed that i) it is cost-efficient to work with a reduced labeled set (a few hundred labeled objects per category) if images are carefully selected, ii) the usefulness of the trained model for classifying unseen datasets improves with the use of different backgrounds in the training dataset, and iii) avoiding training with low-quality labels (e.g., small relative size or incomplete silhouettes) yields better classification metrics. These results and dataset will help select and label images in the most effective way to improve the use of deep learning in studying underwater organisms.

KEYWORDS

deep learning, mediterranean, fish, pre-treatment, YOLOv5, EfficientNet, faster RCNN

Introduction

Underwater marine images are widely used to study fish abundance, behavior, size structure, and biodiversity at multiple spatial and temporal scales (Aguzzi et al., 2015; Díaz-Gil et al., 2017; Follana-Berná et al., 2022; Francescangeli et al., 2022). In recent years, advances in artificial intelligence and computer vision, specifically deep learning (DL), have enabled the reduction of the number of hours required for manually detecting and classifying species in images. Studies have demonstrated the capabilities of these techniques, particularly deep convolutional networks (CNN; LeCun et al., 1998; Lecun et al., 2015) in detecting and classifying fish in underwater images or video streams (Salman et al., 2016; Villon et al., 2018, see reviews in Goodwin et al., 2022; Li and Du, 2022; Mittal et al., 2022; Saleh, Sheaves and Rahimi Azghadi, 2022). These studies have utilized different types of image databases and have faced similar unresolved questions, such as the number of fish needed for training (Marrable et al., 2022), the need for color image pre-processing (e.g., Lisani et al., 2022), the need for transfer learning from large databases (e.g., Imagenet or coco), improving results when working with small image areas or limited computing power (Paraschiv et al., 2022), whether to use segmentation of bounding boxes and how well a trained set will perform for different habitats (backgrounds). In particular, the detection and classification of multiple species using different combinations of backgrounds (the “domain-shift” phenomenon: Kalogeiton et al., 2016; Ditria et al., 2020), number of species, and labeling quality, is an area that requires further investigation. In general, it is believed that a greater volume of training data and a greater variety of backgrounds can improve the performance of DL datasets (Moniruzzaman et al., 2017; Sarwar et al., 2020; Ditria et al., 2020). Highly varied backgrounds are typical in coastal areas, where non-invasive video-based automatic fish censusing methods are increasingly needed for conservation and fisheries sustainability issues (Aguzzi et al., 2020; Connolly et al., 2021; Follana-Berná et al., 2022). However, these types of exercises are limited, and the need for a high number of labeled individuals from many species can be challenging in areas or laboratories with limited resources.

The Mediterranean Sea is an example of a scarcity of approaches in the field of DL for fish detection. A recent search in the Web of Science for papers on “Deep Learning”, “Fish” and “Mediterranean” (conducted in December 2022) yielded only seven results, with only one of them taking into account the variation of background (seasonal variation over time, in a fixed station) in a multispecific dataset of Mediterranean fish (Ottaviani et al., 2022). The Mediterranean is a highly diverse sea (Coll et al., 2010) where underwater video monitoring exercises are primarily semi-supervised (Aguzzi et al., 2015; Díaz-Gil et al., 2017; Marini et al., 2018b; Follana-Berná et al., 2019, Follana-Berná et al., 2022) and monitoring is essential due to the high impact of invasive species and climate change (Azzurro et al., 2022). In this context, the main objective of this work is to evaluate, for newly generated Mediterranean fish datasets of over 20 species, the relative importance of combining backgrounds in the detection (of “fish”) and classification (of species), how these combinations interact with

the balance/unbalance in fish labeling, and how the labeling quality affects the quality of fish detection. Additionally, we compare, as a function of matrix size, the classification performance of a single-step classifier (i.e., objects are classified into specific categories) versus a classifier requiring a two-step procedure (objects are first classified into a generic fish category, and then classified into more specific categories).

Material and methods

Four different underwater image datasets were constructed for analysis (Table 1). Datasets A through C are newly generated images and are available in a free repository (Zenodo, <https://doi.org/10.5281/zenodo.7534425>). Dataset A (Figure 1) was created using images from an underwater cabled camera located in a wreck inside Andratx Bay on the western coast of Mallorca Island (Subeye, https://imedea.uib-csic.es/sites/sub-eye/home_es/). The camera (SAIS-IP-bullet cam, 2096 x 1561 pixels) was situated within the wreck (6 m depth) and has been sending still images every 5 mins since 2019 to our research center. Dataset B was obtained from various underwater video surveys in Palma Bay on the southern coast of Mallorca Island. The cameras were used either in drop-down surveys (Go-pro Hero 3, 1920 x 1440) or were operated by scuba divers (Go-pro Hero 7, 1920 x 1440). The obtained images included depths ranging from 5 to 20 meters, and balanced backgrounds, including sand, seagrass meadows and rocks were selected (Figure 1B). In both A and B datasets, more than 20 object classes (species/genus) were observed (Table 2) and labeled by an expert using bounding boxes. The number of observations of each species ranged from 2 to more than 3000; this imbalance forced us to reduce the bulk of the main analyses to 9 fish classes with a higher number of observations, although some species with a low number of labels were included for comparison (Table 2). Subsets of the main datasets A and B were used as validation sets, as detailed in Table 1. Training and validation (approx 20% of the images) were conducted using an NVIDIA QUADRO GV100 32 GB GPU. Four small test sets (images not belonging to the validation or training sets) were also used, both from datasets A and B and from two external datasets. The first external dataset (dataset C, Table 1, Figure 1) consisted of images from a second fixed camera located at 4 m from the wreck (8 m depth, Sony Ipela SNC-CH210 2048 x 1536 pix). Additionally, a small set (dataset D, Table 1; see Figure 1) from the OBSEA cabled observatory located in Catalonia, NE Spain (Aguzzi et al., 2011) was also used as a test set (Francescangeli et al., 2023).

Datasets pre-processing and scenarios

Underwater images often exhibit low contrast, color cast, noise and haze due to depth-dependent attenuation of light wavelengths and the scattering effect (Hsiao et al., 2014; Wang et al., 2019; Zhou et al., 2020; Wang et al., 2023). To improve the dataset images, we employed the Multiscale Retinex Model (MSR, Land and McCann, 1971), which has been identified as one of the best methods for

TABLE 1 Combination of images and number of fish for each of the scenarios (E0-5) used to detect fish.

		Scenarios					
Train and Validation datasets	Fish or image (train/validation/test)	E0	E1	E2	E3	E4	E5
		Imb; all A & B	Imb; reduced A, no B	Imb; All B	Bal; A & B	Imb; All A	Bal; reduced and selected A & B
A	FISH (train)	12096	3074	0	3074	12096	1462
	FISH (validation)	2422	892	0	892	2431	140
B	FISH (train)	3032	0	3032	3032	0	1716
	FISH (validation)	892	0	892	892	0	388
	TOTAL FISH (train)	15128	3074	3032	6106	12096	3178
	TOTAL FISH (validation)	3314	892	892	1784	2431	528
A	IMAGES (train)	762	196	0	196	762	168
	IMAGES (validation)	184	69	0	70	184	24
B	IMAGES (train)	576	0	576	576	0	305
	IMAGES (validation)	143	0	143	142	0	58
	TOTAL IMAGES (train)	1338	196	576	772	762	473
	TOTAL IMAGES (validation)	327	69	143	212	184	82
Test datasets							
A	FISH (test)	235	235	235	235	235	235
	IMAGES (test)	15	15	15	15	15	15
B	FISH (test)	290	290	290	290	290	290
	IMAGES (test)	13	13	13	13	13	13
C	FISH (test)	369	369	369	369	369	369
	IMAGES (test)	43	43	43	43	43	43
D	FISH (test)	103	103	103	103	103	103
	IMAGES (test)	21	21	21	21	21	21

A and B datasets were split into training, validation and test sets. Further, test sets C and D were obtained from different areas and backgrounds. For classification, see further in the text. Imb, imbalanced scenario. Bal, balanced scenario.

detecting fish for labeling purposes using different backgrounds (Lisani et al., 2022).

After image enhancement, labeling was conducted using the free online software Supervisely (<https://supervise.ly/>). Six training datasets were created to evaluate the relevance of the type of background and number of fish within the images for neural network training (Table 1).

Scenario E0 included all the training images available from both datasets A and B (15128 objects and 1338 images for training, 3314 objects and 327 images for validation). Scenario E1 was a reduced subsample of dataset A, comprising 196 images and 3074 fishes. Scenario E2 included all of dataset B, comprising 546 images and 3032 fishes. Scenario E3 was a balanced scenario, containing around 3000 fish for each dataset A and B. Scenario E4 contained all the training images of dataset A (12096 objects and 762 images for training, 2442 objects and 184 images for validation). Finally, scenario E5 consisted of a selected group of images from both datasets A and B (approx 1500 fish each), avoiding images that

appeared to disturb the training, particularly those that did not include small fish (<100 pixels², Figure 2) or overlapping fish.

Fish detection and classification were compared in two steps. First, two state-of-the-art CNNs (Faster R-CNN and YOLOv5M) were compared across scenarios for single-class detection (fish/no fish). Second, classification metrics were compared between the best-performing network in classifying fish/no fish, which was then used as both a detector and classifier, and a pure classifier network (the latter using only the bounding boxes previously classified as “fish”). For classification training, fish were pre-classified to the lowest taxonomical category possible (species, genus, or family).

Models metrics

Model comparison and evaluation (see below) on validation or test sets was conducted through the analysis of the interaction of two standard metrics: precision (P) and recall or sensitivity (R) (Everingham et al., 2010). For a given fish class, precision is defined

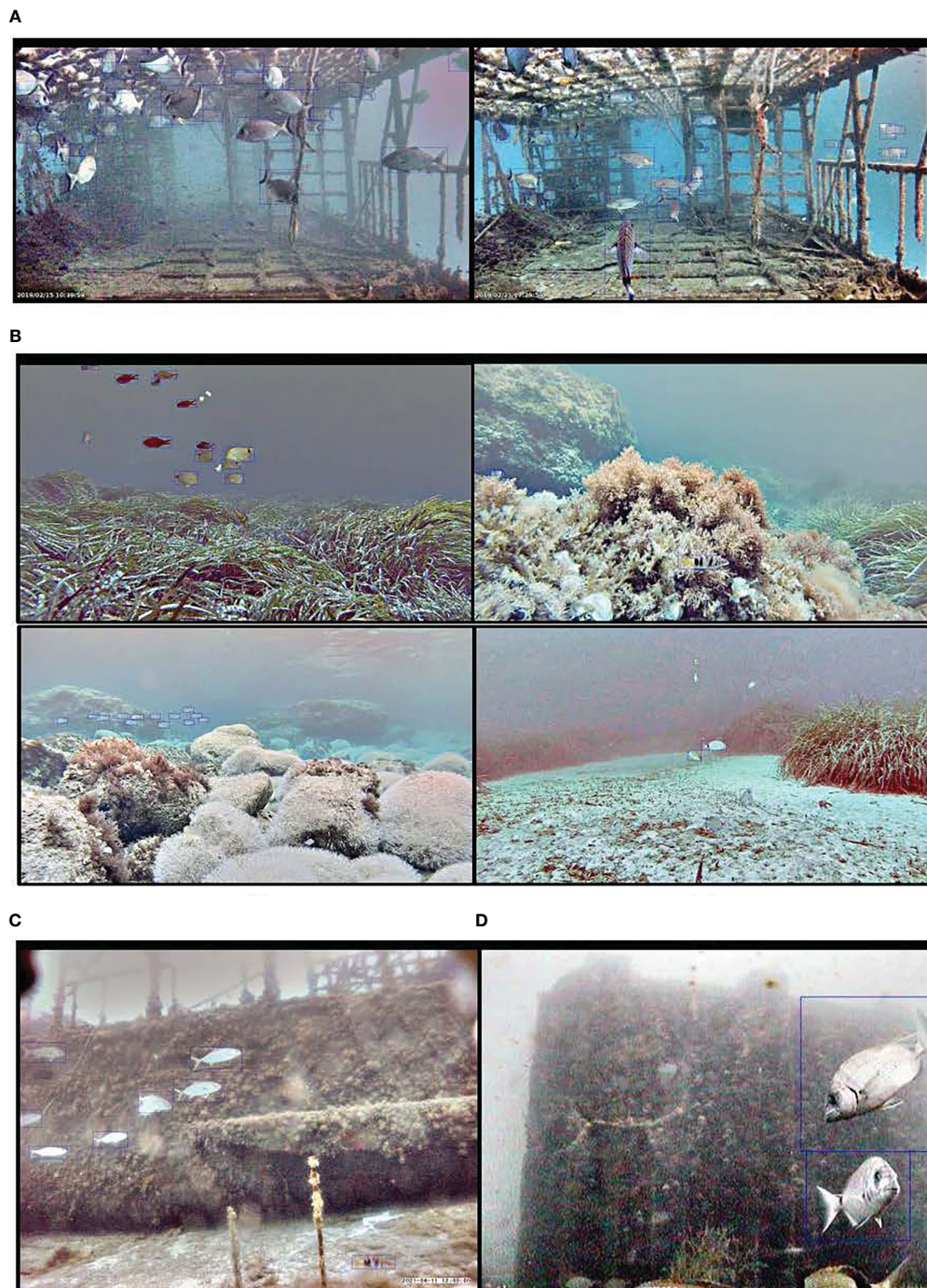


FIGURE 1

Example images from the main coastal Mediterranean datasets (A) fixed observatory, (B) varied coastal bottoms, and two other test datasets (C) fixed observatory in Mallorca, (D) fixed observatory in Catalonia.

















as the fraction of relevant fish among all retrieved fish, whereas recall is the fraction of retrieved and relevant fish among all relevant fish. They are defined as:

$$P = \frac{TP}{TP + FP}; R = \frac{TP}{TP + FN}$$

where TP=true positive, FP=false positive and FN=false negative.





Neither P nor R provide a full picture of the model performance. To attain a more global metric for comparisons, we calculated the F1 score and the mean average precision (mAP). The F1 score will only be high if both P and R are high and is calculated as:

TABLE 2 Count and image example (after MSR model pre-processing) of the main fish classes appearing in datasets A and B.

Class name	Example Image	Occurrence in A	Occurrence in B	Total
Unidentified fish	–	3309	771	4080
<i>Chromis chromis</i>		2788	1357	4145
<i>Coris julis</i>		7	572	579
<i>Dentex dentex</i>		5	0	5
<i>Diplodus annularis</i>		121	637	758
<i>Diplodus puntazzo</i>		2	5	7
<i>Diplodus sargus</i>		3301	12	3313
<i>Diplodus sp.</i>		1090	8	1098
<i>Diplodus vulgaris</i>		1155	379	1534
<i>Epinephelus costae</i>		2	0	2
<i>Epinephelus marginatus</i>		2	0	2
<i>Lithognathus mormyrus</i>		395	0	395
<i>Mugilidae (prob Chelon)</i>		483	0	483
<i>Mullus surmuletus</i>		3	12	15
<i>Oblada melanura</i>		972	68	1040
<i>Pomatosus saltatrix</i>		234	0	234
<i>Sarpa salpa</i>		20	75	95
<i>Seriola dumerilii</i>		1256	0	1256

(Continued)

TABLE 2 Continued

Class name	Example Image	Occurrence in A	Occurrence in B	Total
<i>Serranus scriba</i>		17	203	220
<i>Sparus aurata</i>		80	0	80
<i>Sphyaena viridis</i>		27	0	27
<i>Symphodus</i> sp.		22	257	279

Some species were aggregated to a genus level if species could not be recognized, or it was a genus with many species appearing in low abundances (e.g., *Symphodus*).

$$F1 \text{ score} = \frac{2 \cdot P \cdot R}{P + R}$$

The P and R values from the nets were obtained so that they maximized the F1 score, thus achieving their best balance. The mAP is often used for global model comparison and is calculated as the area under the precision vs recall curve, at all levels of intersection over union (<http://cocodataset.org/>). Here, we calculated mAP@0.5, meaning that true positives are defined as detections whose bounding boxes have at least a 50% overlap with the ground truth bounding boxes. This overlap is measured in terms of the Intersection over Union (IoU), which ranges from 0 to 1, as the ratio between the area of their intersection and the area of their union.

Object detection

For object (class “fish”) detection, we first compared the performance of Faster RCNN (Ren et al., 2015) and several configurations of the fifth version of the You Only Look Once (YOLO) algorithm (first described by Redmon et al., 2016), using the implementation from Ultralytics (<https://github.com/ultralytics/yolov5>). YOLOv5 has been shown to work particularly well in underwater environments (Wang et al., 2021). The medium pre-trained model from YOLOv5, YOLOv5m (pre-trained on COCO image database, <http://cocodataset.org/>) was selected after training on the E0 scenario with the l, m and x pre-trained models (Supplementary Table S1).

Excluded from E5 training Included in E5 training

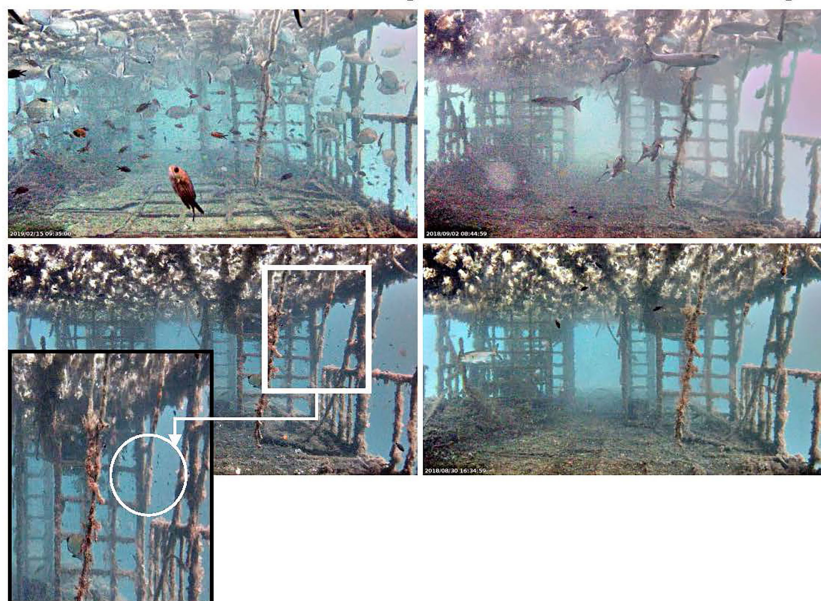


FIGURE 2

Left panel, images excluded from scenario E5 due to small fish and abundant overlap (note inset in the lower-left picture for small fish). Right panel: types of images included in scenario E5, selected for clearly identifiable species.

YOLOv5m (hereafter referred to as YOLO) produced the best compromise between metrics ($mAP@0.5 = 0.84$, precision=0.83, recall=0.78) and computation time and was selected for subsequent analyses. For Faster RCNN we used the implementation for object detection from the TensorFlow API (https://www.tensorflow.org/api_docs), with the ResNet50 configuration, pre-trained on ImageNet (<https://www.image-net.org/>). Object detection performance was evaluated on each training scenario using the aforementioned metrics.

Classification

Fish can be classified in a single step using the YOLO algorithm, which scans the entire image, identifies fish, and classifies them. Alternatively, a classifier that only operates on pre-defined bounding boxes of fish can also be used among other possibilities. We compared the results from a state-of-the-art classifier, EfficientNet V2 (here forth EfficientNet) (Tan and Le, 2021) implemented with the TensorFlow API, with those from the best-performing YOLO model. The EfficientNet was trained on the Google Colab platform (<https://colab.research.google.com/>), while the YOLO network was trained locally on an NVIDIA GPU. An initial comparison was conducted using two sets of increasing fish object classes (4 and 8 classes) to observe the effect of the number of classes and instances on classification success. Given the superior performance of YOLO on classification (see corresponding section), it was used to further compare the effect on increasing the number of fish categories with more than 50 individuals (4, 8, 14 species) in expanded class sets. Each trial was trained using only the selected classes in each set. Confusion matrices are provided for selected results, and specific variations in the species composition were made, re-training the network to illustrate the confounding effect of including new fishes at the genus level that could not be classified to species level but were morphologically similar. Direct comparisons between YOLO and EfficientNet performance using mAP cannot be

made due to structural differences in the networks, so F1 score (mean \pm SD) was used to compare equal sets of species datasets.

Results

Fish detection

The comparison of the two networks, YOLO and Faster RCNN, across six scenarios revealed that YOLO performed notably better than Faster RCNN both in validation and test sets in most cases (Tables 3, 4) with $mAP@0.5$ values over 0.8 in most scenarios in the validation datasets (Table 3). The inferior performance of Faster RCNN was primarily attributed to lower R values. In general, using a larger number of fish resulted in slightly better results. However, it was noteworthy that E5 achieved nearly as good results using one-tenth the number of objects for training, but only considering images without small fish and using a balanced set of backgrounds. Comparing YOLO results in the test sets across scenarios, the following patterns were apparent (Table 4, see Supplementary Figure S1 for examples): the evaluation of scenarios that were not trained with either A or B datasets performed poorly on the test sets from datasets not used in training, but not necessarily with other never-before-seen datasets (C and D, scenarios E2, E4). The best results across test sets were obtained using a YOLO network trained in scenario E0 (high number of fish but unbalanced background), followed by E3 (trained on approximately half the objects but with balanced datasets) and E5 (half the images than E3, balanced datasets and selected images). These three training scenarios yielded $mAP@0.5$ values ranging from 0.70-0.84 across all test scenarios.

Species classification

As expected, classification metrics tended to improve with an increasing number of objects. On average, YOLO performed better

TABLE 3 Performance metrics for each scenario computed over the training datasets (see Table 1).

Training Scenario	Model	P	R	$mAP@0.5$
E0	Faster RCNN	0.80	0.48	0.60
	YOLO	0.83	0.78	0.84
E1	Faster RCNN	0.66	0.45	0.37
	YOLO	0.75	0.75	0.77
E2	Faster RCNN	0.76	0.52	0.83
	YOLO	0.84	0.73	0.80
E3	Faster RCNN	0.81	0.45	0.70
	YOLO	0.82	0.73	0.80
E4	Faster RCNN	0.71	0.53	0.42
	YOLO	0.81	0.79	0.84
E5	Faster RCNN	0.78	0.53	0.83
	YOLO	0.88	0.71	0.83

See Table 4 for test sets. Noticeably, E5 yielded relatively good results with a low number of training objects (by eliminating fish that are only dots or very difficult to recognize at the species level).

TABLE 4 Results of the application of YOLO to the four test datasets (never seen by the trained DL nets, see Table 1).

Training Scenario	Model	mAP@0.5 value for each test dataset				Number of training objects	
		A	B	C	D	Training objects	Backgrounds
E0	Faster RCNN	0.34	0.34	0.48	0.63		
	YOLO	0.84	0.83	0.80	0.78	15128	Unbalanced
E1	Faster RCNN	0.35	0.16	0.47	0.57		
	YOLO	0.82	0.34	0.78	0.83	3074	Unbalanced
E2	Faster RCNN	0.15	0.35	0.39	0.42		
	YOLO	0.34	0.64	0.49	0.43	3032	Unbalanced
E3	Faster RCNN	0.32	0.30	0.47	0.55		
	YOLO	0.82	0.81	0.80	0.76	6106	Balanced
E4	Faster RCNN	0.35	0.18	0.48	0.74		
	YOLO	0.86	0.45	0.79	0.82	12096	Unbalanced
E5	Faster RCNN	0.32	0.29	0.48	0.69		
	YOLO	0.79	0.80	0.76	0.76	3178	Balanced

Balanced and unbalanced scenarios and the number of training objects are specified.

than EfficientNet when using eight species, although both networks performed similarly on four species (Table 5). The average F1 score for both networks was around 0.75 for four species. For eight species, YOLO showed around 14% higher values than EfficientNet, with a standard deviation one order of magnitude lower. In some cases, EfficientNet had high precision and F1 score for classes with a low number of objects (e.g., *S. scriba*) when the number of classes was low. Overall, YOLO was considered a more convenient tool, providing reasonable results in an integrated detection and classification process. A test for confounding species showed that if a class that could contain two similar species was included (*Diplodus* sp.), YOLO confused it with *D. sargus* at the same proportion as the generic *Diplodus* sp. (Figure 3). The category “background” (Figure 3, see also Figure S2) comprises different objects depending on the matrix size. In small matrices (e.g., four sp.), wrongly classified information is included in the background category, which in fact contains general categories like “fish”, plus others (See Supplementary Figure S2). When the category “fish” is included, most of the information previously attributed to background is, in many instances, attributed to this “fish” category (see Figure S2). This general category is comprised by fish that were unidentifiable at a higher taxonomic resolution. Additionally, a large proportion of true *Diplodus* sp was inferred to be background, likely due to initial labeling issues: the contour of these *Diplodus* sp. could not be fully determined due to partial overlap with other fish. Using YOLO in a larger dataset (Table S2, Supplementary Figure S2) showed that, although the average classification power decreased, i) increasing the number of species did not necessarily decrease the classification success for the species with large numbers (e.g., *C. chromis*, *D. sargus*, *D. vulgaris*) or conspicuous shape differences with respect to the others (e.g., Mugilidae) (See Figure S1 for an example), ii) several other species with a low number of labels were reasonably classified

(e.g., *L. mormyrus*, *P. saltatrix*). These well-detected species were conspicuous and largely different in shape or color from the rest (see Table 2).

Discussion

In this paper, we present a new labeled dataset of underwater images of coastal Mediterranean fishes and investigate the best dataset combinations for obtaining optimal deep learning (DL)-based classification results that can be applied to various habitats. Firstly, we compared two popular architectures, Faster RCNN and YOLO, in terms of their object detection capabilities. Results indicate that YOLO significantly outperforms Faster RCNN in detecting the category “fish” and performs better than EfficientNet in many cases, without the need for pre-defining bounding boxes. However, in some instances, such as classifying conspicuous species in scenarios of limited training data, directly utilizing bounding boxes may yield better results, as observed in other studies (Knausgård et al., 2022).

Using YOLO, we addressed specific areas that required further investigation, particularly the “domain shift” phenomenon (Kalogeton et al., 2016; Ditria et al., 2020) characterized by a decrease in classification performance with varying habitat backgrounds and fish species assemblages. Automatic fish classification often involves the use of relative or absolute (e.g., Campos-Candela et al., 2018) abundance estimators that utilize underwater baited cameras (Connolly et al., 2021) or cabled observatories (Bonofiglio et al., 2022) to count, classify or track fish (Saleh et al., 2022). These underwater images differ significantly from typical free datasets that contain single individuals; these images contain a high diversity of species and large variability in abundance, resulting in reduced classification success. However, as

TABLE 5 Results of comparable classification metrics between YOLO and EfficientNet using either 4 or 8 classes.

4 classes	Training objects	Validation objects	YOLO				EfficientNet		
			P	R	F1 score		P	R	F1 score
<i>C. chromis</i>	2730	854	0.80	0.65	0.72	<i>C. chromis</i>	0.86	0.97	0.91
<i>D. sargus</i>	2281	492	0.79	0.73	0.76	<i>D. sargus</i>	0.81	0.85	0.83
<i>D. vulgaris</i>	1011	251	0.83	0.61	0.70	<i>D. vulgaris</i>	0.74	0.37	0.50
<i>S. scriba</i>	152	48	0.87	0.75	0.80	<i>S.scriba</i>	0.94	0.71	0.81
Av F1 score					0.75	Av F1 score			0.76
Sd F1 score					0.05	Sd F1 score			0.18
8 classes									
<i>C.chromis</i>	2730	854	0.79	0.67	0.73	<i>C.chromis</i>	0.62	0.96	0.75
<i>D.sargus</i>	2281	492	0.73	0.75	0.74	<i>D.sargus</i>	0.73	0.78	0.76
<i>D.vulgaris</i>	1011	251	0.75	0.64	0.69	<i>D. vulgaris</i>	0.71	0.16	0.27
<i>S. scriba</i>	152	48	0.81	0.75	0.78	<i>S.scriba</i>	0.97	0.60	0.74
<i>S.dumerilii</i>	870	172	0.89	0.89	0.83	<i>S. dumerilii</i>	0.88	0.66	0.75
<i>D.annularis</i>	434	195	0.85	0.72	0.78	<i>D. annularis</i>	0.87	0.51	0.65
<i>O.melanura</i>	691	184	0.80	0.80	0.62	<i>O.melanura</i>	0.93	0.14	0.24
<i>C. julis</i>	368	132	0.78	0.78	0.82	<i>C. julis</i>	0.87	0.82	0.84
Av F1 score					0.75	Av F1 score			0.63
Sd F1 score					0.07	Sd F1 score			0.23

Bounding boxes are extracted from 343 images.

previously identified (e.g., Saleh et al., 2022), it is necessary to develop models that can generalize their learning and perform well on new, unseen data samples, bridging the gap between DL and the requirements of image-based ecological monitoring (e.g. MacLeod et al., 2010; Christin et al., 2019; Aguzzi et al., 2020; Goodwin et al., 2022).

Related to the above, another common problem in classification is the imbalance of objects per class, as the DL model tends to weigh more heavily on the more abundant classes. Class-aware approaches have been proposed for fish classifications (Alaba et al., 2022). Beyond confirming that balancing improved classification in our datasets, we found that comparable results to an imbalanced dataset with an order of magnitude more training images could be obtained by carefully selecting images. Additionally, our results showed that avoiding training with images containing many small bounding boxes yields better precision and recall values on validation and test sets. The relation between object size and classification properties has been described previously, and it is recommended to separate the classification analyses as a function of object size (e.g., Connolly et al., 2022). However, to our best knowledge, this practice is not commonly used in fish ecology studies applying DL algorithms to underwater images. Overall, the fact that a model trained with a limited number of images performs relatively well across multiple test scenarios is a promising result for applications in ecological studies.

Recent reviews (e.g., Goodwin et al., 2022; Saleh et al., 2022) have concluded that for the application of DL methods to fish ecology research, transparent and reproducible research data and tools are necessary. This paper aims to contribute to this goal. There have been few studies on Mediterranean fish that have been experimental in nature (e.g., testing new network developments on a reduced number of species, such as Paraschiv et al., 2022 for a

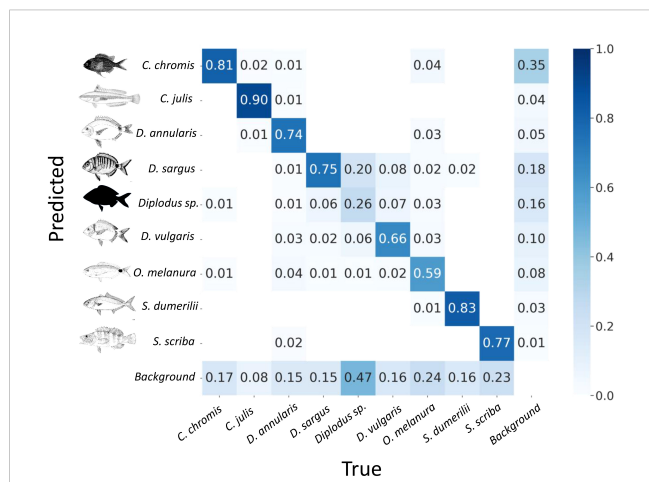


FIGURE 3 Confusion matrix based on YOLO results for eight species, background, and one genus (*Diplodus* sp) that could not be attributed to other congeneric species. Refer to the text for further explanations.

few pelagic species). To increase the use of DL in this field, we concur with other authors that not only should common databases and reproducible methods be made available (e.g., [Francescangeli et al., 2023](#)), but also that more integrated engineers-ecologists interactions are institutionally needed ([Logares et al., 2021](#)). Additionally, statistical corrections to DL estimates must be developed ([Connolly et al., 2021](#)) and the use of lighter networks (e.g., [Paraschiv et al., 2022](#)) should become more common, as computer power may be a significant limitation for unplugged underwater devices (e.g., [Lisani et al., 2012](#)).

In summary, our research has discovered or reinforced several key findings that have important implications for fish ecology. Firstly, we found that using fast, single-step classifiers like YOLOv5, we can classify fishes in entire images cost-effectively, without the need for a two-step approach. Secondly, while having a large number of labeled fish images is important, a better approach may be to use a variety of backgrounds with a smaller, more carefully selected set of images. When selecting images, it is important to ensure that the bounding box fully captures the fish, and that the bounding box is not too small relative to the image. Lastly, we found that increasing the number of classes in the training dataset may lower overall classification metrics, but it may not significantly affect species with a high number of labels and can improve the identification of less abundant species.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

IC and AA-E conceptualized the paper. AA-E, J-LL and JS ran the models. MP, HH, GV, MF and IC provided images. All authors contributed to the writing and interpretation of the results, lead by IC. IC and J-LL contributed to funding. All authors contributed to the article and approved the submitted version.

References

- Aguzzi, J., Chatzievangelou, D., Company, J. B., Thomsen, L., Marini, S., Bonfiglio, F., et al. (2020). The potential of video imagery from worldwide cabled observatory networks to provide information supporting fish-stock and biodiversity assessment. *ICES J. Mar. Sci.* 77, 2396–2410. doi: 10.1093/icesjms/fsaa169
- Aguzzi, J., Doya, C., Tecchio, S., De Leo, F. C., Azzurro, E., Costa, C., et al. (2015). Coastal observatories for monitoring of fish behaviour and their responses to environmental changes. *Rev. Fish Biol. Fish.* 25:463–83. doi: 10.1007/s11160-015-9387-9
- Aguzzi, J., Mánuel, A., Condal, F., Guillén, J., Noguera, M., del Rio, J., et al. (2011). The new seafloor observatory (OBSEA) for remote and long-term coastal ecosystem monitoring. *Sensors* 11, 5850–5872. doi: 10.3390/s110605850
- Alaba, S. Y., Nabi, M. M., Shah, C., Prior, J., Campbell, M. D., Wallace, F., et al. (2022). Class-aware fish species recognition using deep learning for an imbalanced dataset. *Sensors* 22, 8268. doi: 10.3390/s22128268
- Azzurro, E., Smeraldo, S., and D'Amen, M. (2022). Spatio-temporal dynamics of exotic fish species in the Mediterranean Sea: Over a century of invasion reconstructed. *Glob. Change Biol.* 28, 6268–6279. doi: 10.1111/gcb.16362
- Bonfiglio, F., De Leo, F. C., Yee, C., Chatzievangelou, D., Aguzzi, J., and Marini, S. (2022). Machine learning applied to big data from marine cabled observatories: A case study of sablefish monitoring in the NE Pacific. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.842946
- Campos-Candela, A., Palmer, M., Balle, S., and Alós, J. (2018). A camera-based method for estimating absolute density in animals displaying home range behaviour. *J. Anim. Ecol.* 87, 825–837. doi: 10.1111/1365-2656.12787
- Christin, S., Hervet, É., and Lecomte, N. (2019). Applications for deep learning in ecology. *Methods Ecol. Evol.* 10, 1632–1644. doi: 10.1111/2041-210X.13256

Funding

Project DEEP-ECOMAR. 10.13039/100018685-Comunitat Autònoma de les Illes Balears through the Direcció General de Política Universitària i Recerca with funds from the Tourist Stay Tax law ITS 2017-006 (Grant Number: PRD2018/26).

Acknowledgments

We thank Juan José Enseñat for his help in the acquisition and storage of images. The present research was carried out within the framework of the activities of the Spanish Government through the “María de Maeztu Centre of Excellence” accreditation to IMEDEA (CSIC-UIB) (CEX2021-001198-M) and the “Severo Ochoa Centre Excellence” accreditation to ICM-CSIC (CEX2019-000928-S) and the Research Unit Tecnoterra (ICM-CSIC/UPC).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmars.2023.1151758/full#supplementary-material>

- Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Lasram, F. B. R., Aguzzi, J., et al. (2010). The biodiversity of the Mediterranean Sea: Estimates, patterns, and threats. *PLoS One* 5(8):e118. doi: 10.1371/journal.pone.0011842
- Connolly, R. M., Fairclough, D. V., Jinks, E. L., Ditria, E. M., Jackson, G., Lopez-Marcano, S., et al. (2021). Improved accuracy for automated counting of a fish in baited underwater videos for stock assessment. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.658135
- Connolly, R. M., Jinks, K. I., Herrera, C., and Lopez-Marcano, S. (2022). Fish surveys on the move: Adapting automated fish detection and classification frameworks for videos on a remotely operated vehicle in shallow marine waters. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.918504
- Díaz-Gil, C., Smees, S. L., Cotgrove, L., Follana-Berná, G., Hinz, H., Martí-Puig, P., et al. (2017). Using stereoscopic video cameras to evaluate seagrass meadows nursery function in the Mediterranean. *Mar. Biol.* 164:137. doi: 10.1007/s00227-017-3169-y
- Ditria, E. M., Lopez-Marcano, S., Sievers, M., Jinks, E. L., Brown, C. J., and Connolly, R. M. (2020). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00429
- Everingham, M., Van Gool, L., Williams, C., Winn, K., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* 88, 303–338. doi: 10.1007/s11263-009-0275-4
- Follana-Berná, G., Arechavala-Lopez, P., Ramirez-Romero, E., Koleva, E., Grau, A., and Palmer, M. (2022). Mesoscale assessment of sedentary coastal fish density using vertical underwater cameras. *Fish. Res.* 253:106362. doi: 10.1016/j.fishres.2022.106362
- Follana-Berná, G., Palmer, M., Campos-Candela, A., Arechavala-Lopez, P., Díaz-Gil, C., Alós, J., et al. (2019). Estimating the density of resident coastal fish using underwater cameras: Accounting for individual detectability. *arXiv* 615:177–88. doi: 10.3354/meps12926
- Francescangeli, M., Marini, S., Martínez, E., Del Río, J., Toma, D. M., Noguera, M., et al. (2023). Image dataset for benchmarking automated fish detection and classification algorithms. *Sci. Data* 10, 1–13. doi: 10.1038/s41597-022-01906-1
- Francescangeli, M., Sbragaglia, V., Del Río, J., Trullols, E., Antonijuan, J., Massana, I., et al. (2022). Long-term monitoring of diel and seasonal rhythm of dentex dentex at an artificial reef. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.837216
- Goodwin, M., Halvorsen, K. T., Jiao, L., Knausgård, K. M., Martin, A. H., Moyano, M., et al. (2022). Unlocking the potential of deep learning for marine ecology: Overview, applications, and outlook. *ICES J. Mar. Sci.* 79, 319–336. doi: 10.1093/icesjms/fsab255
- Hsiao, Y. H., Chen, C. C., Lin, S. L., and Lin, F. P. (2014). Real-world underwater fish recognition and identification, using sparse representation. *Ecol. Inform.* 23, 13–21. doi: 10.1016/j.ecoinf.2013.10.002
- Kalogeiton, V., Ferrari, V., and Schmid, C. (2016). Analysing domain shift factors between videos and images for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 2327–2334. doi: 10.1109/tpami.2016.2551239
- Knausgård, K. M., Wiklund, A., Sordalen, T. K., Halvorsen, K. T., Kleiven, A. R., Jiao, L., et al. (2022). Temperate fish detection and classification: a deep learning based approach. *Appl. Intell.* 52, 6988–7001. doi: 10.1007/s10489-020-02154-9
- Land, E. H., and McCann, J. J. (1971). Lightness and retinex theory. *Josa* 61, 1–11. doi: 10.1364/JOSA.61.000001
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2323. doi: 10.1109/5.726791
- Li, D., and Du, L. (2022). *Recent advances of deep learning algorithms for aquacultural machine vision systems with emphasis on fish* (Netherlands: Springer). doi: 10.1007/s10462-021-10102-3
- Lisani, J. L., Petro, A. B., Sbert, C., Álvarez-Ellacuría, A., Catalán, I. A., and Palmer, M. (2022). Analysis of underwater image processing methods for annotation in deep learning based fish detection. *IEEE Access* 10:130359–72. doi: 10.1109/ACCESS.2022.3227026
- Logares, R., Alós, J., Catalán, I. A., Solana, A. C., del Campo, J., Ercilla, G., et al. (2021). “Oceans of big data and artificial intelligence,” in *Ocean science challenges for 2030* (Madrid: CSIC), 163–179.
- MacLeod, N., Benfield, M., and Culverhouse, P. (2010). Time to automate identification. *Nature* 467, 154–155. doi: 10.1038/467154a
- Marini, S., Corgnati, L., Manotovani, C., Bastianini, M., Ottaviani, E., Fanelli, E., et al. (2018a). Automated estimate of fish abundance through the autonomous imaging device GUARD1 126, 72–75. doi: 10.1016/j.measurement.2018.05.035
- Marini, S., Fanelli, E., Sbragaglia, V., Azzurro, E., Del Rio Fernandez, J., and Aguzzi, J. (2018b). Tracking fish abundance by underwater image recognition. *Sci. Rep.* 8:13748. doi: 10.1038/s41598-018-32089-8
- Marrable, D., Barker, K., Tippaya, S., Wyatt, M., Bainbridge, S., Stowar, M., et al. (2022). Accelerating species recognition and labelling of fish from underwater video with machine-assisted deep learning. *Front. Mar. Sci.* 9, 944584. doi: 10.3389/fmars.2022.944582
- Mittal, S., Srivastava, S., and Jayanth, J. P. (2022). A survey of deep learning techniques for underwater image classification. *IEEE Trans. Neural Networks Learn. Syst.* 1–15. doi: 10.1109/TNNLS.2022.3143887
- Moniruzzaman, M., Islam, S. M. S., Bennamoun, M., and Lavery, P. (2017). Deep Learning on Underwater Marine Object Detection: A Survey. *Adv. Concepts Intell. Vis. Syst. ACIVS 2017 Lect. Notes Comput. Sci.*, 10617. doi: 10.1007/978-3-319-70353-4_13
- Ottaviani, E., Aguzzi, J., Francescangeli, M., and Marini, S. (2022). Assessing the image semantic drift at coastal underwater cabled observatories. *Front. Mar. Sci.* 9, 840088. doi: 10.3389/fmars.2022.840088
- Paraschiv, M., Padrino, R., Casari, P., Bigal, E., Scheinin, A., Tchernov, D., et al. (2022). Classification of underwater fish images and videos via very small convolutional neural networks†. *J. Mar. Sci. Eng.* 10, 1–21. doi: 10.3390/jmse10060736
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). *You only look once: Unified, real-time object detection* in CVPR.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). “Faster r-CNN: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems* (United States: Microsoft Research), 91–99. Available at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84960980241&partnerID=40&md5=18aaa500235b11fb99e953f8b227f46d>.
- Saleh, A., Sheaves, M., and Rahimi Azghadi, M. (2022). Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish Fish.* 23:977–99. doi: 10.1111/faf.12666
- Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., et al. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnol. Oceanogr. Methods* 14, 570–585. doi: 10.1002/lom3.10113
- Sarwar, S. S., Ankit, A., and Roy, K. (2020). Incremental learning in deep convolutional neural networks using partial network sharing. *IEEE Access* 8, 4615–4628.
- Tan, M., and Le, Q. V. (2021). Smaller models and faster training. *arXiv* 2104: arXiv:2104.00298v3. doi: 10.48550/arXiv.2104.00298
- Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., et al. (2018). A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecol. Inform.* 48, 238–244. doi: 10.1016/j.ecoinf.2018.09.007
- Wang, Y., Song, W., Fortino, G., Qi, L. Z., Zhang, W., and Liotta, A. (2019). An experimental-based review of image enhancement and image restoration methods for underwater imaging. *IEEE Access* 7, 233–251. doi: 10.1109/ACCESS.2019.2932130
- Wang, H., Sun, S., Bai, X., and Wang, J. (2023). A reinforcement learning paradigm of configuring visual enhancement for object detection in underwater scenes. *IEEE J. Ocean. Eng.*, 1–19. doi: 10.1109/JOE.2022.3226202
- Wang, H., Sun, S., Wu, X., Li, L., Zhang, H., Li, M., et al. (2021). A YOLOv5 baseline for underwater object detection. *Ocean. Conf. Rec.*, 2021–2024. doi: 10.23919/OCEANS44145.2021.9705896
- Zhou, J. C., Zhang, D. H., and Zhang, W. S. (2020). Classical and state-of-the-art approaches for underwater image defogging: a comprehensive survey. *Front. Inf. Technol. Electron. Eng.* 21, 1745–1769. doi: 10.1109/JOE.2018.2863961