Ecological Informatics 82 (2024) 102788

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Automated species classification and counting by deep-sea mobile crawler platforms using YOLO

Luciano Ortenzi [a,b,*], Jacopo Aguzzi [c,d], Corrado Costa [b], Simone Marini [e,d,f], Daniele D'Agostino [g], Laurenz Thomsen [h], Fabio C. De Leo [i,j], Paulo V. Correa [i], Damianos Chatzievangelou [c,**]

[a] Department of Agriculture and Forest Sciences (DAFNE), Via S. Camillo De Lellis, s.n.c., 01100 Viterbo & Via Angelo Maria Ricci 35a 02100, Rieti, Italy
[b] Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) - Centro di ricerca Ingegneria e Trasformazioni agroalimentari, Monterotondo, Monterotondo, Italy
[c] Instituto de Ciencias del Mar (ICM-CSIC), Barcelona, Italy
[d] Stazione Zoologica of Napoles (SZN), Anthon Dorhn, Napoles, Italy
[e] CNR-ISMAR, National Research Council of Italy (CNR), Institute of Marine Sciences, La Spezia 19132, Italy
[f] NBFC, National Biodiversity Future Center, Palermo 90133, Italy
[g] DIBRIS, Università degli Studi di Genova, Genova, Italy
[h] University of Goteborg, 405 30 Goteborg, Sweden
[i] Ocean Networks Canada, University of Victoria, V8N 1V8 Victoria, BC, Canada
[j] Department of Biology, University of Victoria, V8P 5C2 Victoria, BC, Canada

A R T I C L E I N F O

Keywords:
Edge-computing
Ecological monitoring
Artificial intelligence
Machine learning
Cold-seep
Robotic platforms
Benthic fish

A B S T R A C T

Edge computing on mobile marine platform is paramount for automated ecological monitoring. The goal of demonstrating the computational feasibility of an Artificial Intelligence (AI)-powered camera for fully automated real-time species-classification on deep-sea crawler platforms was searched by running You-Only-Look-Once (YOLO) model on an edge computing device (NVIDIA Jetson Nano), to evaluate the achievable animal detection performances, execution time and power consumption, using all the available cores. We processed a total of 337 rotating video scans (~180°), taken during approximately 4 months in 2022 at the methane hydrates site of Barkley Canyon (Vancouver Island; BC; Canada), focusing on three abundant species (i.e., Sablefish *Anoplopoma fimbria*, Hagfish *Eptatretus stoutii*, and Rockfish *Sebastes* spp.). The model was trained on 1926 manually annotated video frames and showed high detection test performances in terms of accuracy (0.98), precision (0.98), and recall (0.99). The trained model was then applied on 337 videos.

In 288 videos we detected a total of 133 Sablefish, 31 Hagfish, and 321 Rockfish nearly in real-time (about 0.31 s/image) with very low power consumption (0.34 J/image). Our results have broad implications on intelligent ecological monitoring. Indeed, YOLO model can meet operational-autonomy criteria for fast image processing with limited computational and energy loads.

1. Introduction

Environmental monitoring of marine habitats is increasingly relying on image acquisition to describe and quantify life components of ecosystems at all depths at continental margins down to abyssal realms (Aguzzi et al., 2020c; Bicknell et al., 2016; Mallet and Pelletier, 2014; Marini et al., 2018; Shortis and Abdo, 2016). The current development of Artificial Intelligence (AI) applied to image processing is increasingly solving challenges such as underwater target detection (Malde et al., 2020). Applications in autonomous monitoring of marine ecosystems and their resources (e.g., stocks) include faunal classification and tracking by fixed and mobile robotic platforms (Aguzzi et al., 2020b; Beyan and Browman, 2020; Goodwin et al., 2022; Lopez-Marcano et al., 2021; Lopez-Vazquez et al., 2020). By processing time-stamped and geo-

* Corresponding author at: Department of Agriculture and Forest Sciences (DAFNE), Via S. Camillo De Lellis, s.n.c., 01100 Viterbo & Via Angelo Maria Ricci 35a 02100, Rieti, Italy.
** Corresponding author.
E-mail addresses: luciano.ortenzi@unitus.it (L. Ortenzi), damianos@icm.csic.es (D. Chatzievangelou).

https://doi.org/10.1016/j.ecoinf.2024.102788
Received 13 May 2024; Received in revised form 18 August 2024; Accepted 18 August 2024
Available online 20 August 2024
1574-9541/

referenced images for detection, identification, and tracking of species on board each platform, automated routines can produce time series of animal counts in real time, hence allowing for more efficient data storage and transfer, while reducing facility spaces and bandwidth demands (Aguzzi et al., 2012; Aguzzi et al., 2020b; Aguzzi et al., 2022). Therefore, coupling of AI into the control electronics of monitoring platforms and possibly connecting sensor to the cloud facilities can transform video cameras into true intelligent sensors for biodiversity estimation (Aguzzi et al., 2020b; Ferrari et al., 2024).

Such increase in automation can unlock the potential to upscale ecological monitoring in space and time, to combine different data sources for a more complete representation of marine communities, and the possibility of extracting novel insights not apparent by without treating massive amounts of data (Besson et al., 2022). In practice, this can have a positive effect on varying conservation efforts and industrial applications, from monitoring the impact of fisheries, deep-sea mining and marine litter, to marine resources stock assessment, among others (Durden et al., 2021; Naseer et al., 2022). From a technical point of view, the bias or observer's effect accompanying each case-study is also expected to be reduced as image annotating effort by humans is applied to training AI algorithms better instead of processing massive volumes of data (Cuvelier et al., 2012). In general, with the number of published studies using marine imagery growing exponentially in the last 4 decades, this step is necessary to avoid a bottleneck in uncovering new knowledge (Durden et al., 2016).

Benthic crawlers are emerging mobile robotic platforms, bearing video cameras, combined with oceanographic and geochemical sensors (Purser et al., 2013). Typically tethered to permanent seafloor infrastructures such as cabled observatories, offshore industrial rigs, surface buoys with antennas, vessel-deployed docking-garages, and mariculture assets (Aguzzi et al., 2019; Danovaro et al., 2017; Falahzadeh et al., 2023; Rountree et al., 2020; Scherwath et al., 2019), they are operated in real-time via internet connection (Internet Operated Vehicles – IOVs) (Thomsen et al., 2012). Autonomous operations both in terms of navigation and energy provision are also possible, by replacing the tether with remote connection to a Wireless Fidelity (WiFi)-capable surface buoy (Aguzzi et al., 2020a). Crawlers (just like cabled observatories in general) can record large volumes of video footage and sensor data, which are getting increasingly harder and time-consuming to analyse manually (Chatzievangelou et al., 2022). Presently, this data is stored on online data repositories (e.g., the OCEANS 3.0 of Ocean Networks Canada – ONC and the Environmental Research Division's Data Access Program (ERDDAP) of the Expandable Seafloor Observatory – OBSEA) (Aguzzi et al., 2011; Owens et al., 2022). The footage is then processed manually to detect, identify, and count animals from different species (Chatzievangelou et al., 2016; Chatzievangelou et al., 2020; Chatzievangelou et al., 2022; Doya et al., 2017), although efforts to establish automated image treatment pipelines, including preprocessing steps, are gathering pace.

In particular, Lopez-Vazquez et al. 2020 (Lopez-Vazquez et al., 2020) developed an image enhancement and classification pipeline applied to rockfish (*Sebastes* spp.) occurring in the deep-sea ($>$ 200 m) continental margin areas using the Norwegian Lofoten-Vesterålen Ocean Observatory (LoVe Ocean). (Bonofiglio et al., 2022) automatically classified and counted and (McIntosh et al., 2020; McIntosh et al., 2022) automatically detected specific behaviours of sablefish (*Anoplopoma fimbria*) using the fixed camera platforms installed at the Barkley Canyon Node of the North-East Pacific Time-series Undersea Networked Experiments (NEPTUNE) observatory in the NE Pacific. At Folger Pinnacle, a shallower coastal NEPTUNE location, Convolutional Neural Networks (CNNs) based on image colour segmentation were used to detect sponge contraction behaviour (Harrison et al., 2021). Recent efforts used deep CNNs on crawler footage to tackle poor image quality through visual enhancement, prior a three-stage animal detection and count algorithm (Lopez-Vazquez et al., 2023). Briefly, this object detection algorithm consisted in subtracting the image background and applying filters to locate the objects (first stage), followed by a feature selector algorithm (second stage) and finally, a Deep Neural Network (DNN) for animal classification.

On board processing of the acquired data is largely beneficial for underwater autonomous vehicles, transferring the data to a land station with different means (e.g., from tethered cables to acoustic modems) (Aguzzi et al., 2019). Firstly, wireless data transfer from an underwater monitoring device is quite difficult due to the narrow transmission bandwidth, as is the case in non-cabled observatories and non-tethered mobile monitoring platforms. Thus, the option to export an already processed, reduced amount of data (such as a numeric table or multidimensional array) is beneficial (Stojanovic and Beaujean, 2016). Secondly, the content-based interpretation of the acquired data allows for a dynamic adaptation of the mission plan. For example, after the detection of a relevant species, the crawler could adapt its trajectory to follow it or stop to collect a sample (Aguzzi et al., 2022). Nevertheless, DNNs have high computational costs when used in real time (He and Sun, 2015), with low versatility when embedded in autonomous platforms such as crawlers. A suitable solution is represented by the adoption of the edge-computing paradigm, where the data processing is moved closer to the data source (e.g., on board the crawler). From a hardware point of view, the key feature is represented by the fact that most edge-computing devices are based on low power system-on-a-chip architectures (SoC), enabling the design of smart and effective solutions also for battery-powered systems (Jahanbakht et al., 2021; Zhang and Tao, 2021).

From the software point of view, the You Only Look Once (YOLO) algorithm (Han et al., 2021) offers high-speed object detection with low computational costs (Cai et al., 2022; Zhang et al., 2023). Its fast computational capacity would allow neglecting tracking issues, while providing great potential to embed real-time counting (and tracking) routines into crawlers' main computers. In terms of deep-sea environments, this approach has only been used in videos from fixed cabled observatories targeting a single species (e.g., *A. fimbria*) (Bonofiglio et al., 2022). A successful application of YOLO to multiple species classification and counting in the context of "edge computing", would increase the functionalities of crawlers for environmental monitoring. In fact, the characteristics of crawler missions lead to additional challenges on top of the ones normally faced in marine imaging AI-based treatment. For instance, that moving platform results in different image backgrounds or unbalanced illumination or transitions between the seabed and the water column, while resuspended sediment can reduce visibility or prevent the use of parts of the image (Chatzievangelou et al., 2022; Lopez-Vazquez et al., 2023).

In this study, we run YOLO on a NVIDIA Jetson Nano board to demonstrate the computational feasibility of an AI-powered camera for full automated real-time video-processing on deep-sea crawler platforms. We performed a bench processing of crawler videos acquired in the Pacific deep-sea of Barkley Canyon as proxy for a real-word operational scenario. In particular, we measured the quality of automated processing by YOLO (i.e., the number of individuals from three ecologically key species) as well as the processing energy requirements to function in real-time.

2. Materials and methods

2.1. The crawler platform

The "Wally" crawler, shown in Fig. 1, is a cuboid-shaped ($\sim$1 m^3) internet operated vehicle (IOV) moving on caterpillar wheels, that has been in use for environmental monitoring for more than a decade (Chatzievangelou et al., 2022). It connects via a 70 m-long optic fibre tether to a junction box which is integrated to the nearby ONC observatory node in Barkley Canyon.

All videos used in this study were acquired via a 1280 $\times$ 720 pixels' colour camera (Wisenet XNP-6040H). The illumination was provided by two forward looking white 33 W Light Emitting Diode (LED) lamps. The

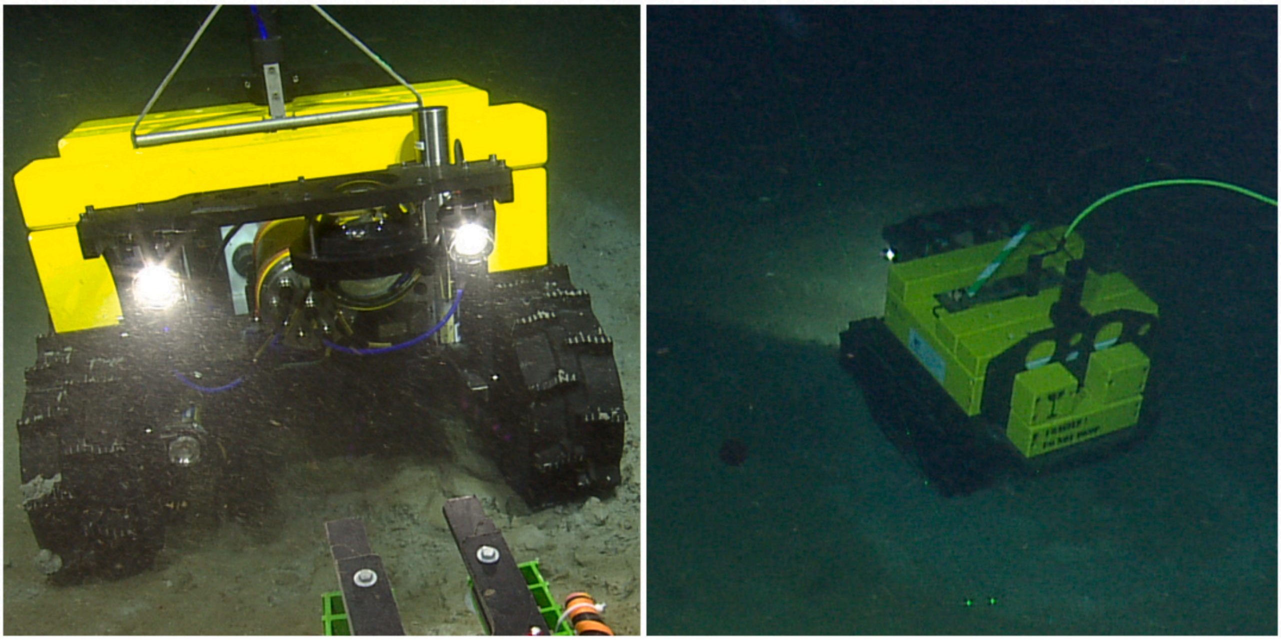


Fig. 1. The crawler inspected by the ROV Hercules during the 2021 deployment. Left panel: Front view of the crawler, with the ROV's forks and tray visible at the bottom of the image. Right panel: Rear view of the crawler from above. Images were recorded on August 21, 2021.

frame-rate for image acquisition was 60 frames per second (fps). All video recordings are stored in real-time on the ONC's data portal Oceans 3.0 (Moran et al., 2022), where they are available for downloading through the DataSearch interface (https://data.oceannetworks.ca/DataSearch). In addition, videos are also available for online viewing without downloading through the SeaTube interface (https://data.oceannetworks.ca/SeaTube) of the same portal.

2.2. Study site and video transects

The crawler is deployed at a methane hydrates site on a plateau of the west wall of Barkley Canyon (Vancouver Island; BC; Canada), at 870–890 m depth (Purser et al., 2013). Tethered to a fixed junction box of the Barkley Canyon node of the ONC observatory, it monitored a total surface of ~400 m^2 for this study. The area is characterized by soft sediment and the presence of small (~1–2 m high) hydrate mounds. A map of the area covered by the crawler's missions, including the main morphological features in the form of bathymetry contours, is presented in Fig. 2.

The crawler performed a total of 337 rotating video scans (~180°) at 9 specific spots along the same transect line (see Fig. 2), from August 24, 2021, to January 12, 2022. The timing and duration of each video varied (two typical videos are included in Supplementary Materials), but as a general rule of thumb sampling was limited by tidal and current regimes (to avoid suspended sediment clouds blocking the field of view), ONC's policy on artificial lights (maximum 2 h of lights on per day to avoid photic contamination in this deep-sea environment), and the quality of connectivity and responsiveness of the vehicle that could make operations cumbersome at times. For each spot, except for station 9 where the scans were forward-looking, the seabed on both sides (i.e., left and right) of the transect line was scanned. Table 1 summarises the information of the original video-scans.

2.3. The targeted species

Our analysis focused on the 3 most abundant vertebrate species typically observed at the Barkley Canyon site (Fig. 3) (Doya et al., 2014; Matabos et al., 2014). The sablefish or black cod (*A. fimbria*) is a large fish of grey/silver colour, which can be observed either swimming or resting/hovering above the seafloor. The Pacific hagfish (*E. stoutii*) is a long-shaped fish of characteristic purple colour, normally observed as coiled on the seafloor, or floating vertically with the head touching the sediment. The rockfish *Sebastes* spp., of characteristic orange/reddish and white striped colour and normally observed resting on the seafloor, being observed individuals of a taxonomically heterogeneous group of morphologically similar species that cannot be precisely distinguished without direct sampling (Jahanbakht et al., 2021). Hagfish and rockfish display mostly benthic lifestyle (i.e., closer association with the seabed, rarely seen in the water column), while sablefish are benthopelagic swimmers that migrate daily through the canyon (Doya et al., 2014; Doya et al., 2017; Grinyo et al., 2023). For a more detailed morphological description we refer to the guide of Barkley Canyon fauna (Gervais et al., 2012).

2.4. The overall edge-computing approach for the intelligent crawler application

The basic principle of the edge-computing approach is to move the data analysis as close as possible to the source that generated the data. This means that the crawler needs to be equipped with a Single Board Computer (SBC) capable to process the acquired images and to classify their content. The overall processing pipeline adopted for equipping the crawler with intelligent capabilities is summarized in Fig. 4 The pipeline consists of the “In Lab Activities” and of the “Edge Activities”. Within the lab activities, images extracted from historical footage, acquired by the crawler, was manually annotated to tag the relevant species to be classified. Then the tagged images are used to train an image classifier suitable to be executed by the SBC platform. Within the “Edge activities”, the trained classifier is then deployed on the SBC hosted by the crawler and used to classify the new acquired videos. For each video, frames are extracted and processed as images with the aim of classifying their content. After the classification task, the time series abundance estimate per video is obtained through a counting procedure that avoids counting the same organism more than once. All the previous steps are described in the following sections.

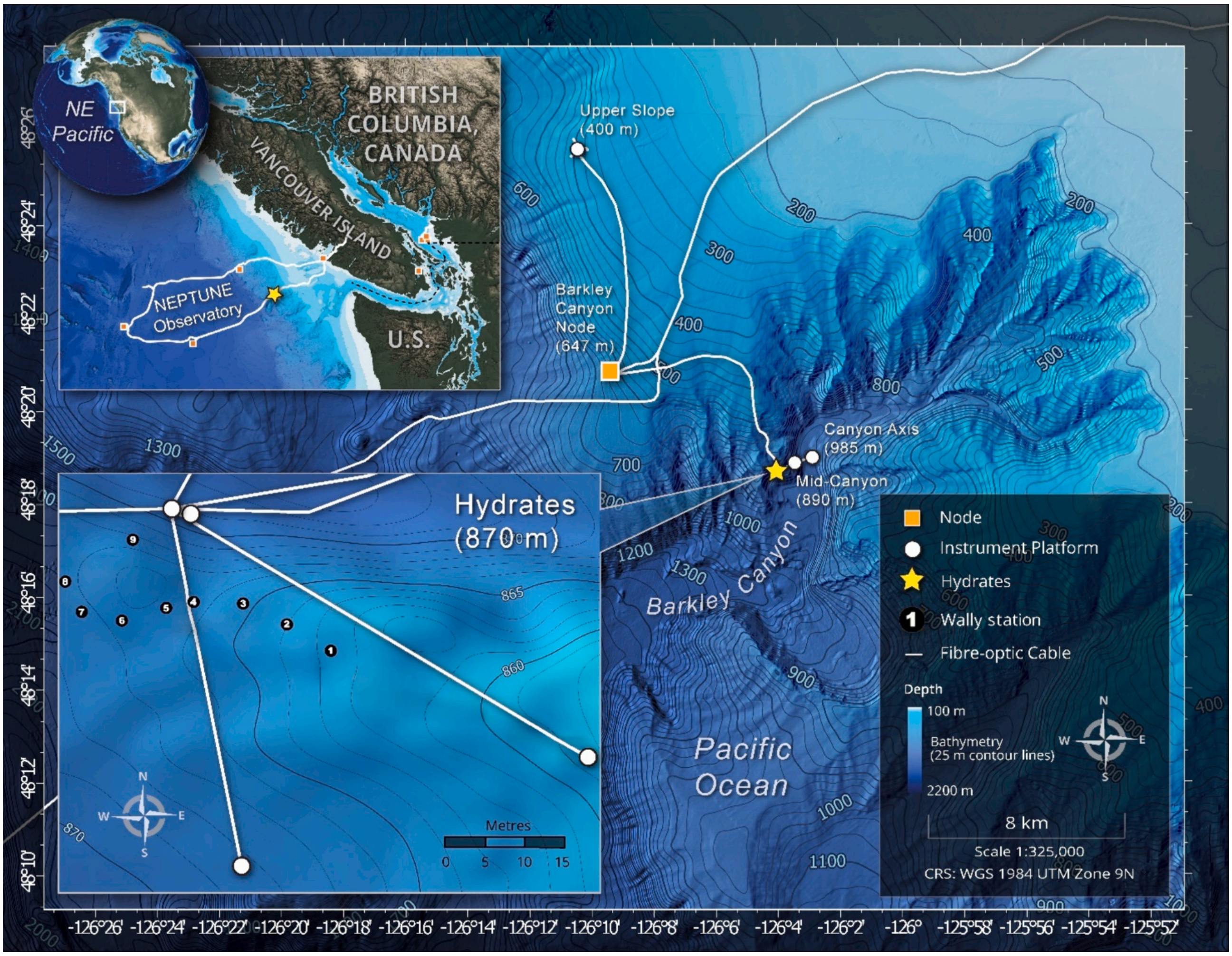


Fig. 2. The study area of Barkley Canyon (BC) hydrates. Top left: Geographic location of the site (yellow star), off Vancouver Island (BC; Canada; Northeast Pacific). Large map: Location of the site in Barkley Canyon (yellow star), among other instrumented platforms at different depths. Bottom left: Zoomed map of the hydrates site with 1 m bathymetry contours. The black numbered dots represent the stations where rotating video scans were performed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Information of the original 337 video-scans, summarized per station. Numbering starts from the eastern end of the transect line (see Fig. 2).

	Stations									
Side	1	2	3	4	5	6	7	8	9	Total
Right	20	18	21	20	21	21	21	19	–	161
Left	19	19	20	19	20	20	20	20	19	176
Total	39	37	41	39	41	41	41	39	19	337

2.5. Edge computing platform

The SBC platform used to demonstrate the edge capabilities for image analysis is the NVIDIA Jetson Nano board (Nvidia, 2024). This board is a very suitable solution for edge computing considering all together its low price, limited power consumption, and considerable compute capabilities provided by the CUDA architecture. Its main relevant hardware specifications are:

- 128-core NVIDIA Maxwell GPU @ 922Mhz;
- quad-core ARM A57 CPU @ 1.43GHz;
- 4 GB 64-bit LPDDR4;
- hardware interfaces such as GPIO and I2C 2 and MIPI CSI-2.

The platform allows for different power profile configurations to assess the performance in terms of execution time, power consumption, and potentially varying quality of the classification results. The power profiles mainly differ in term of GPU/CPU usage, peripheral use and clock limitations (Nvidia, 2024). When the GPU is used together with a central processing unit (CPU) the board exploits all the cores available on the device; i.e., the 128 Compute Unified Device Architecture (CUDA) core based on the NVIDIA Maxwell architecture and the 4 cores of the ARM[1] Cortex-A57 processor. Differently in a less power consumption profile it can exploit only the ARM cores.

2.6. The training dataset

1926 images (i.e., frames extracted from the video scans) containing animals were manually annotated with the Roboflow image analysis software (Mallet and Pelletier, 2014). To these, we added 675 null images (i.e., without animals) to reduce the false positive detection rates. Therefore, the resulting dataset consisted of 2601 images, grouped as follows: 72% training set ($N = 1348$ with animals +512 null); 19% validation set ($N = 386$ with animals +104 null); and finally, 9% test set ($N = 192$ with animals +59 null). The frames were extracted from 35 videos randomly considered among 337. The frame extraction from the selected video was randomly performed among the frames representing the same scene (i.e. a rockfish laying on the ground).

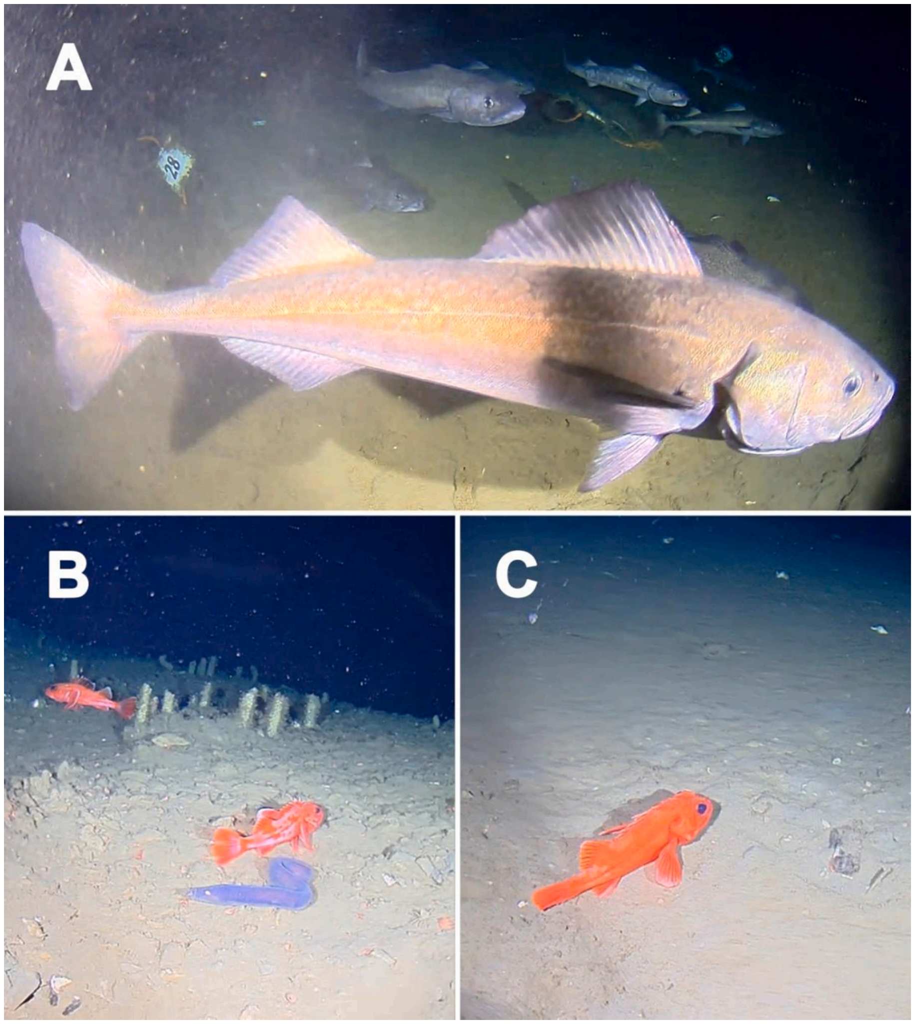


Fig. 3. Sample images of the three targeted species, extracted from the analysed footage: A) Sablefish (multiple individuals at varying distances and illumination conditions); B) Hagfish (with two rockfish also visible in the field of view); and finally, C) Rockfish.

Fig. 5 presents examples of some of the challenges of animal identification in real-world operational conditions at 900 m water depth inside a hydrodynamically active canyon system that could potentially affect the performance of YOLO. In these images, the animals can be observed obscure by uneven illumination, hidden behind other objects, seabed structures or other animals, covered by sediment (buried into the seabed or behind resuspended sediment cloud) or cut by the margins of the frame. Other issues may include morphological variability among individuals of the same species, biofouling on the camera's glass sphere, loss of focus, etc. (Yang et al., 2021).

Two augmentation steps (Random Brightness Adjustment −20 to +20%; and Random Gaussian Blur 0 to 3.5 pixels) were applied to the training set to allow for a better generalization and improve the versatility of the model (Lopez-Vazquez et al., 2020), triplicating the original images and tackling the challenges related to partial visibility. We did not perform more pre-processing augmentation steps, to maintain a realistic balance and avoid a tendency towards false positives later in the workflow. Thus, the augmented dataset consisted of 6321 images, with a ratio of training: validation: test of 88:8:4. All images were resized to 640×640 pixels, and the information of the bounding boxes was stored in separate text files in a file format compatible with the YOLOv8 model (see next section). This procedure is like the one adopted by Bonofiglio et al. (2022) (Bonofiglio et al., 2022), except for data augmentation steps and the higher number of target species in our study.

2.7. Object detection

As starting point, we used an already customized "off-the-shelf" YOLO model for the detection of species (Bonofiglio et al., 2022). This model is based on a convolutional neural network and is very popular for its accuracy and speed (Tian et al., 2019; Zheng et al., 2021). It differs from other object detection algorithms like Region-based Convolutional Neural Network (R-CNN) (Song et al., 2023; Xie et al., 2021) because it adopts an end-to-end approach. Instead of proposing a region of interest (ROI) and then classifying it in a second stage, YOLO directly outputs a 5-dimensional vector $v = (cls, x, y, w, h)$ containing the class (cls) of the object, the coordinates (x, y) of the centre of the bounding box surrounding the object, its width (w) and its hight (h). This output is associated with a confidence score defined as:

$$CS = p(object)*IoU$$

where $p(object)$ is the probability of an object of being present and the Intersection Over Union (IoU) is the area of intersection between the ground truth bounding box and the output one.

By definition, $0 \leq CS \leq 1$.

A detailed description of the algorithm and the different details differentiating its versions reported in literature, is way beyond the scope of the present work and is devoted to dedicated reviews (Hussain, 2024) (Hussain, 2023; Terven and Cordova-Esparza, 2023). The updated YOLOv8 version differs from the previous one (Bonofiglio et al., 2022)

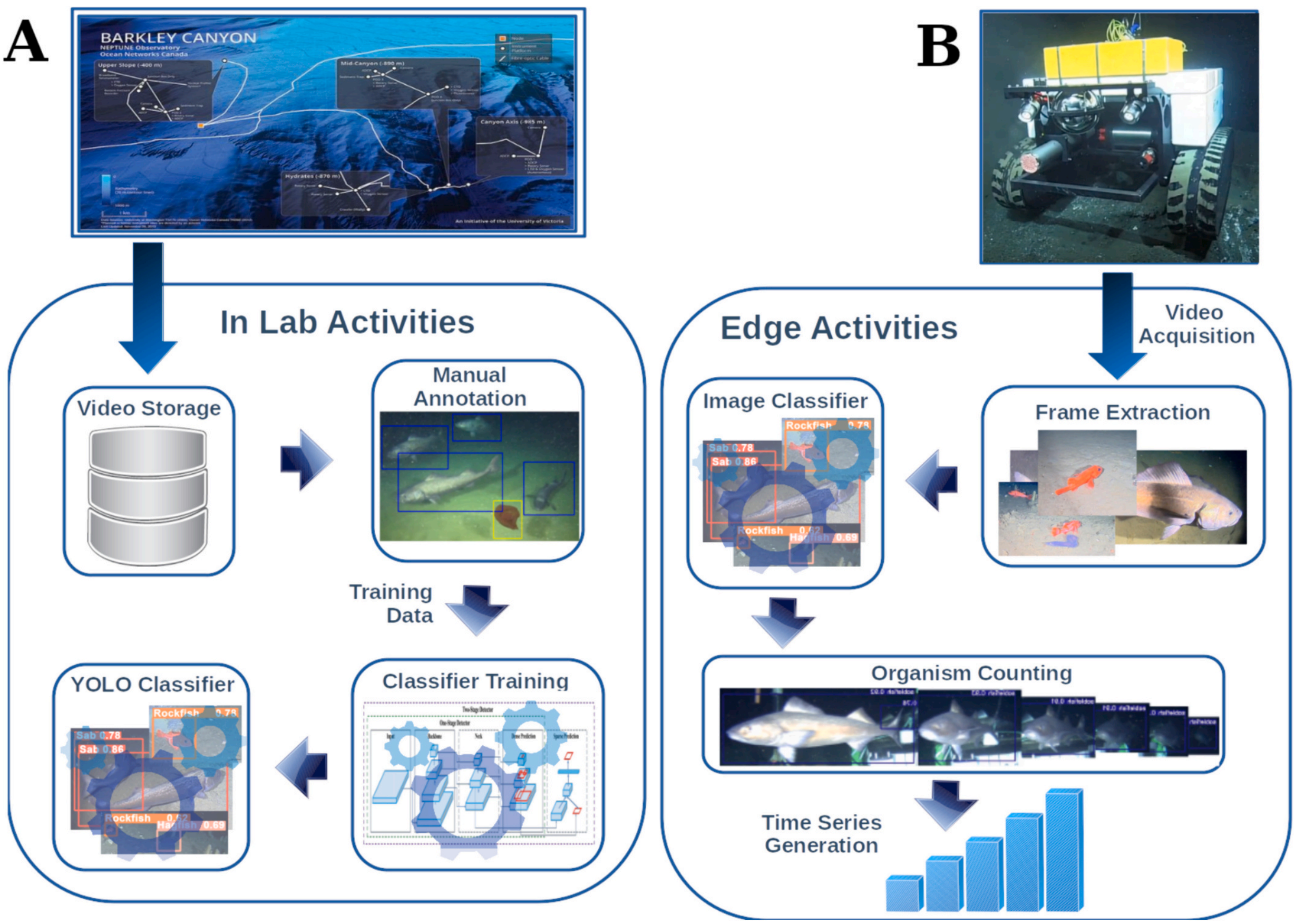


Fig. 4. The overall processing pipeline expressed as in Lab Activities and Edge Activities.

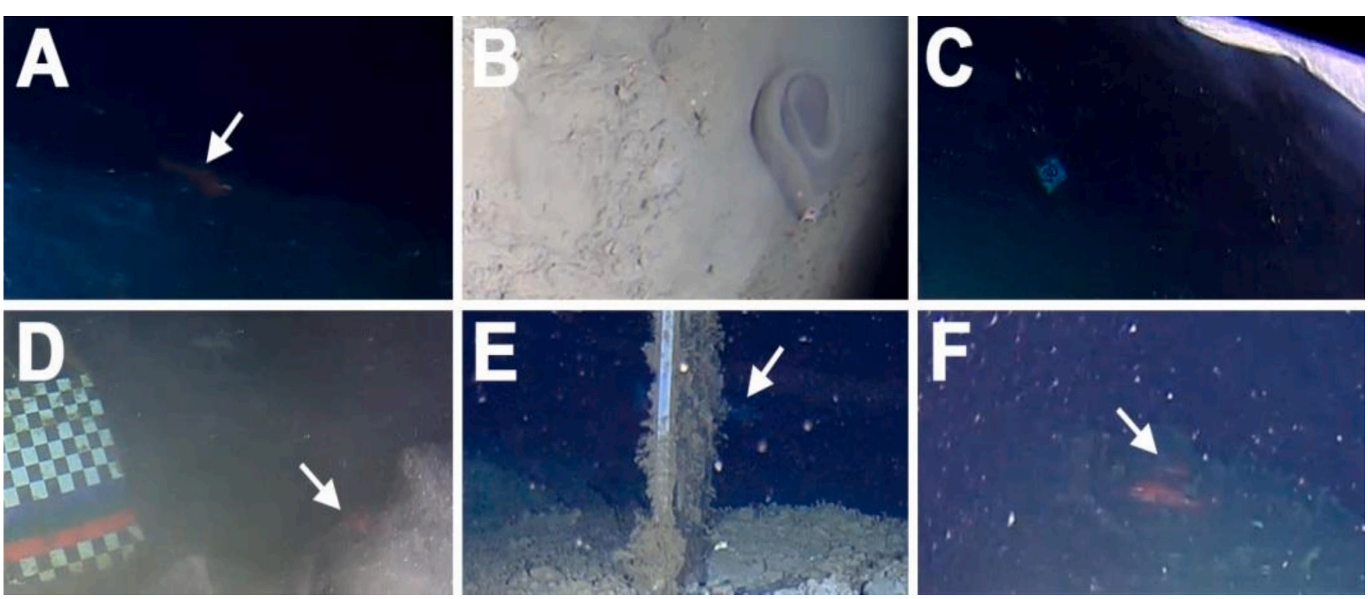


Fig. 5. Sample cropped images containing animals that can be challenging to detect (indicated by white arrows): A) Rockfish in a dark part of the image; B) Hagfish partially covered in sediment, appearing to have the same colour as the seabed; C) Sablefish only partially visible, towards the upper right margin of the image; D) Rockfish behind a suspended sediment cloud; E) Sablefish behind an artificial marker; F) Rockfish behind the bigger one in the centre on the frame, poorly visible due to distance/size and the presence of large particles.

for many details concerning the architecture and the loss functions.

We adopted the YOLOv8 (Ultralytics; https://ultralytics.com/) version because it shows the best detection performance and speed on state-of-the-art (SOTA) datasets, upgrading from version YOLOv5 (adopted in ref. (Bonofiglio et al., 2022)) as extensively described in Ref. (Terven and Cordova-Esparza, 2023). During the training phase, the algorithm converged after 150 iterations using a batch size of 16 and a learning rate of 0.01. All video-processing was run on a current Laptop, mounting the NVIDIA graphics processing unit (GPU) RTX 3070 Ti with 8 Gb video random access memory (VRAM).

2.8. Evaluation metrics of the detection algorithm

The performance of the YOLO model was evaluated through Accuracy (A), Precision (P), and Recall (R), defined respectively as (Fawcett, 2006):

- $A = (TD + TU)/(TD + TU + FD + FU)$,
- $P = TD/(TD + FD)$
- $R = TD/(TD + FU)$

Where TD, FD, TU, and FU stand for "True Detection", "False Detection", "True Undetection", and "False Undetection", respectively.

These terms are based on the detection confidence (*DC*) which is a threshold on the confidence score (defined above) on the predicted bounding box and class.

An output detection is considered as prediction if $CS > DC$.

Therefore, TD and FD refer to the detection of objects that are indeed present in (or absent from) the image, respectively (i.e., true and false positives). Similarly, TU and FU refer to correctly reporting the absence of objects and the failure to detect present objects, respectively (i.e., true and false negatives). As for TD, FD, TU, and FU, also P and R depend on the detection confidence. A small value of *DC* will cause many FD and a relatively low P and high R. Vice versa a *DC* close to 1 will enhance P reduce R to a value close to 0. The Precision and Recall trade-off is evaluated by the F_1-score defined as the harmonic average of P and R:

$$F_1 = 2\,(P^{*}R)/(P + R)$$

The Average Precision (AP) is calculated as the weighted mean of P at each value of IoU, to account for the success of the classification on top of the detection of the bounding box (important in multiclass object detection, i.e., when targeting multiple species). The weights of the mean are provided by the increase in recall from the previous threshold. The mean (m) average precision, indicated as mAP, is the AP averaged over the number of classes (i.e., animal species) to be detected. The mAP for IoU threshold, set to 0.5 ($IoU \geq 0.5$), and the IoU threshold between 0.5 and 0.95, ($0.5 \leq IoU \leq 0.95$) are indicated as mAP@0.5 and mAP@0.5:0.95, respectively.

We fixed the threshold on the IoU to a default value of 0.7 and the detection confidence *DC* to a value maximizing the F_1-score.

2.9. Time series generation and counting algorithm

The labels and bounding box coordinates of the objects detected in each photogram (i.e., analysed image frame) were recorded into a text file extracted from each video. Subsequently, a post-processing step was applied to convert these "raw" time series to estimates of animal abundances in the analysed videos. First, the raw time series were smoothed using a 30-frames moving average (corresponding to 0.5 s of video for an original frame-rate of 60 fps), to remove possible artefacts of falsely detected or undetected animals. Then, each time series was processed by a counting algorithm, which considered only positive increments of detected animals for each frame. The script implementing the counting algorithm is reported in the SUPPLEMENTARY MATERIAL and according to Huettmann and Arhonditsis (2023) (Huettmann and Arhonditsis, 2023), all data were made available in GITHUB.COM (https://github.com/Luciano-ma/crawler-wally-animal-detection), along with the converged weights of the YOLOv8 model for full results reproducibility."

3. Results

3.1. Fish detection

Out of 337 videos, 49 did not show the presence of any of the three targeted species. In the remaining 288 videos we detected 133 sablefish, 31 hagfish and 321 rockfish. During the validation phase, executed in lab, the average detection time per image was 15 ms with the selected laptop specifications and the post-processing counting algorithm only took 0.015 s per photogram. An example of detections for the three different species is reported in Fig. 6 (the original videos are provided in **Supplementary Material 2** to **4**).

The performance of YOLOv8 in detecting each of the three targeted species in the validation step is reported alongside mAP@0.5 and mAP@05:0.95 in Table 2. For completeness, Fig. 7 reports the normalized confusion matrices for training and validation. As can be seen, the accuracy is very high both in training (100%) and in validation (98%).

Once trained and validated on two separated datasets, the model was tested on a third independent dataset to establish the detection confidence defined in section 2.6, which is the threshold value above which a detection is considered as prediction. This was done by calculating the F_1-score and finding its maximum as a function of the confidence score defined in section 2.6. Fig. 8 shows the test classification performances measured by P, R and F_1. As expected, P and R show an opposite monotonic behaviour as a function of the confidence score. The former increases up to 1 and the latter decreases reaching zero for $CS = 1$. The F_1-score shows a maximum at $CS = 0.482$. This is the value of the Detection Confidence adopted in the following.

3.2. Execution time and power consumption in the edge computing application

The GPU-based version of YOLOV8, executed on board the Jetson Nano platform allowed us to process 490 images in 153 s (0.31 s/image). The CPU-only version requires 534 s (1.1 s/image) instead. The average power consumption of the classifier executed on the GPU + CPU based power consumption profile is 1.1 W, thus requiring about 168 J. The CPU-only version instead requires ~3 W, thus ~1.6 KJ (i.e., an order of magnitude higher). These results have been collected using the API provided by the jetson-stats package.

Among the Jetson Nano has power profiles, one is the Power-Efficient (PE) mode, requiring only 5 W at most. In this configuration the GPU-based YOLO requires about 0.43 s/image and an average power consumption of 0.96 W, thus 200, Joules, more than in the previous case. The CPU-based version in PE mode employs 1221 s (>20 min) to process the 490 images, requiring 0.87 W and 1062 J. We can conclude that the GPU + CPU profile (the default mode) in our case is able to exploit the GPU in a more powerful and time efficient way. Moreover, the quality of the classification results is independent by the platform and the Jetson power profile used (e.g., workstation, board, with or without GPU).

3.3. Time series generation

Implementing YOLO for object detection allows counting animals under real-time conditions (i.e., as the crawler is moving during the video acquisition). Fig. 9 presents two examples of how the time series of counts are compiled as a function of the time for each video. Starting with any individuals detected in the first frame, every time the number of detected individuals increases between consecutive photograms (as the crawler moves and slightly modifies the field of view), a new count is

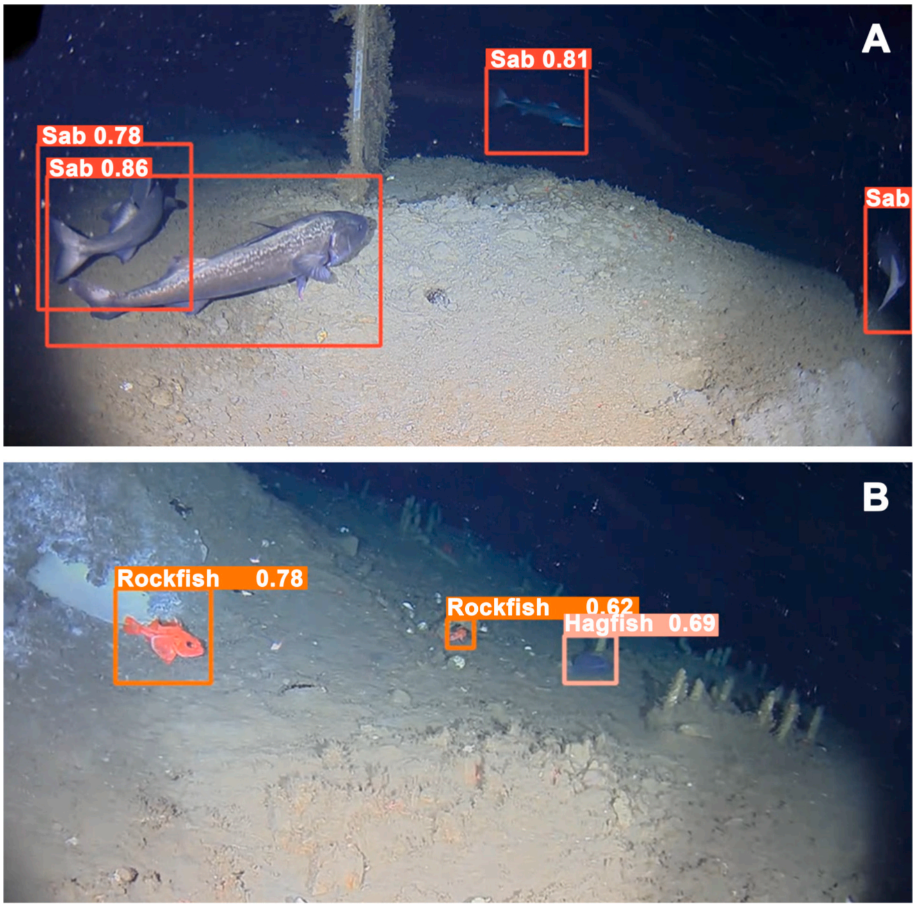


Fig. 6. Examples of crawler video-processing by YOLOv8v, according to different detection types: A) Meta-objects (i.e., different individuals of the same species) for 4 sablefish individuals (labelled "Sab"), with a numeral tag that indicates model confidence; B) Multiclass objects (i.e., different individuals of different species) for 2 individuals of rockfish and 1 hagfish together.

Table 2
Detection performance of YOLOv8 on the validation set (i.e., 490 images) in detecting sablefish, hagfish, and rockfish. mAP@0.5 and mAP@0.5:0.95 stand for mean Average Precision (mAP) for detection confidence (set to 0.5) and detection confidence (between 0.5 and 0.95), respectively.

Class	Labels	Precision	Recall	mAP@0.5	mAP@0.5:0.95
All	575	0.976	0.986	0.992	0.751
Sablefish	134	0.962	0.978	0.991	0.827
Hagfish	84	0.995	1.000	0.995	0.736
Rockfish	357	0.970	0.980	0.990	0.691

added to the time series.

4. Discussion

In this work, we presented an efficient routine for deep-sea animal detection, counting, and species classification in footage recorded by mobile Internet Operated Vehicles (IOVs), such as benthic crawlers. We also demonstrated how time-series of counts for different species can be generated by an emerging AI-based data processing routine, such as YOLOv8. Our results set the minimum edge-computing processing functionalities (i.e., image processing on board the crawler), conveying information on the hardware architecture and energy specifications, to execute the automated identification and classification of animals belonging to different species.

4.1. *Technical advancements*

We further advanced the latest works (Bonofiglio et al., 2022; Lopez-Vazquez et al., 2023) with footage from fixed cabled observatory cameras and with different footage of this same crawler (Lopez-Vazquez et al., 2023), implementing an object detection algorithm able to locate and classify three different species even when appearing simultaneously in a single frame. Instead of improving the classification performance of DNN through an image enhancement routine (Lopez-Vazquez et al., 2023), we implemented the strategy of ref. (Bonofiglio et al., 2022), using an updated version of the same processing algorithm (YOLOv8), and extending the number of targeted species. This lighter YOLOv8 model was selected based on two relevant crawler operational-autonomy criteria (Aguzzi et al., 2022): fast image processing in relation to the crawler's motion and data collection when targeting multi-class- and meta-objects, plus limited requirements for power and computational resources by a future AI-based algorithm operating in real time on the crawler's motherboard.

Thanks to convolution and anchor box selection functionalities (present from YOLOv5 version onward (Nelson and Solawetz, 2020), YOLO algorithms allow detecting multiple objects, consistently out-performing DNN in image classification (Aziz et al., 2020; Zaidi et al., 2022). We also applied image augmentation as suggested by (Lopez-Vazquez et al., 2020), improving model generalization and to eventually overcome the "concept drift" phenomenon (Ottaviani et al., 2022). Whereas image pre-processing can slow down the model development, image augmentation performed only on the training set does not introduce any detection time delays. In any case, the detrimental effect on

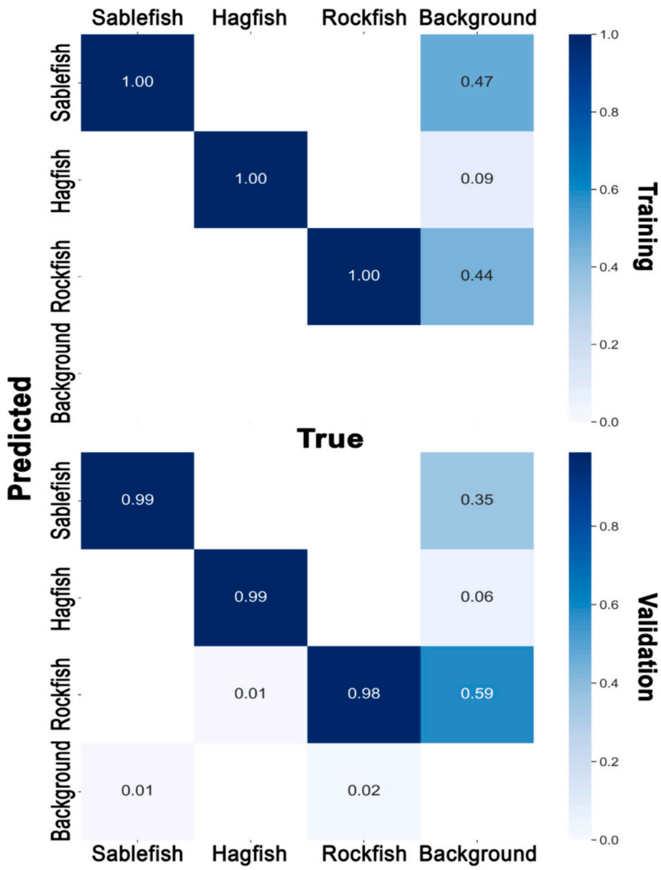


Fig. 7. Training and validation confusion matrices of YOLOv8 in detecting sablefish, hagfish, and rockfish. Values are normalized per column.

recognition performances caused by unprecedent environmental conditions can be addressed through well-established strategies like active learning (Sayin et al., 2021) and incremental learning (Delange et al., 2021; Liu et al., 2021).

Given the high accuracy of the YOLO model we used, and the workflow adopted, no further steps such as tracking (and related hyperparameters) were necessary (Bonofiglio et al., 2022). Thanks to this workflow, video-streams could be directly fed to the YOLO model for a real-time time series generation in future applications. This will allow storing light text files that can be post-processed in the lab to extract population estimates. In any case, the model is just a few megabytes, paving the way for the implementation on an AI camera on board of crawlers.

4.2. Ecological monitoring applications

Our results have broad implications on intelligent ecological monitoring. The economical architecture and set up, combined with its high-performance levels, can enhance the edge-computing capabilities of resident mobile robotic platforms (Aguzzi et al., 2020a) and turn cameras into intelligent biodiversity monitoring sensors able to perform autonomous analyses. For the same crawler, a particle camera is under development which can analyse within a few seconds thousands of particles >100 μm in diameter (including zooplankton) using a specifically designed Robot Operating System (ROS2) software package.

The impact of machine learning applications in marine ecology has witnessed significant growth in recent years (Rubbens et al., 2023), particularly for the potential of efficient analysis of large volumes of imaging data. One of the primary challenges faced in this domain is the creation and curation of high-quality training datasets for the detection of benthic and pelagic fauna. Manual analysis of marine imaging is not only time-consuming but is also prone to errors, requiring trained experts to construct annotated datasets (Matabos et al., 2017). This challenge is actively being addressed through specialized annotation tools such as BIIGLE (Langenkämper et al., 2017) and SQUIDLE+ (Sangekar et al., 2023), alongside the availability of annotated image datasets and models like in FathomNet (Katija et al., 2022). Notably, ONC possesses an extensive archive of marine videos and images obtained from multiple platforms, including fixed cameras, Remotely Operated Vehicles (ROVs), and crawlers, spanning nearly 15 years of data. The integration of annotation tools and a pipeline for model training and evaluation into Oceans 3.0 (Zhang et al., 2023) is in the current plan of ONC to empower the research community in analysing video data. The lightweight model developed here could be applied in the detection of fish on live data that are streamed by multiple cameras maintained by ONC, thus assisting the

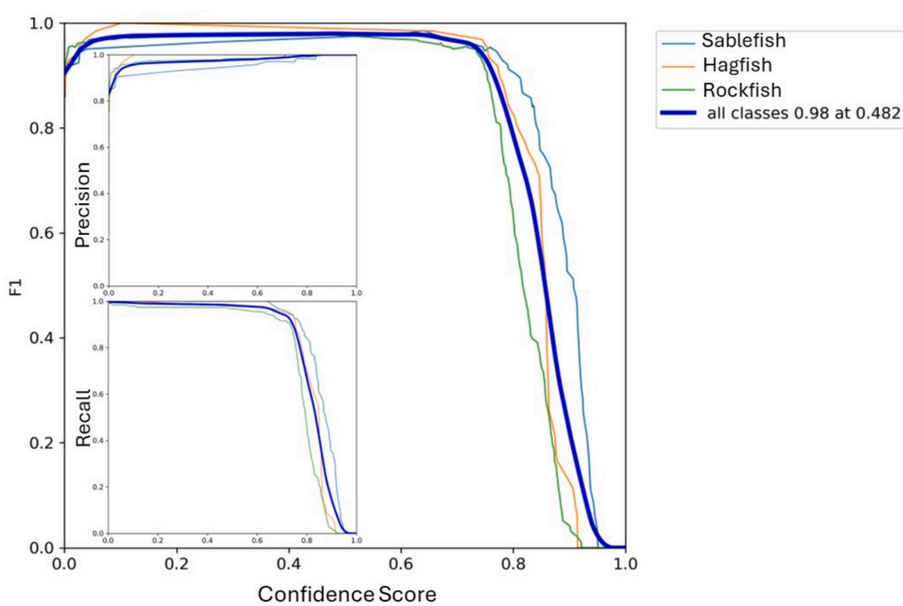


Fig. 8. Test performance curves: F_1-score (main figure), Precision (top inset) and Recall (bottom inset) as a function of the confidence score for the three classes considered.

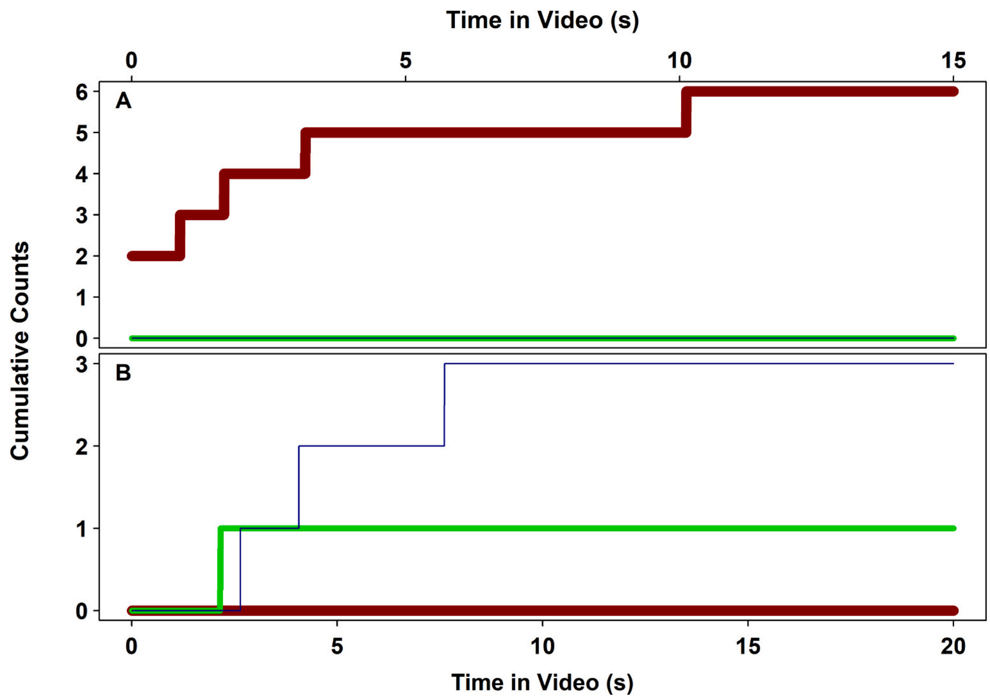


Fig. 9. Examples of the compilation of time series of animal counts in videos (see Fig. 5). Red, green and blue, correspond to Sablefish, Hagfish, and Rockfish, respectively. The lines represent the cumulative sum of counts of detected animals for the three targeted species in consecutive frames, considering only positive increments in the original time-series of detected animals to represent the entrance of a new individual in the camera's field of view. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

annotation and analysis of video at real time.

Previous studies already have shown that crawlers are capable of capturing species behaviour at day-night, tidal, and seasonal scales, and the associated biodiversity changes in deep-sea environments (Bonofiglio et al., 2022; Chatzievangelou et al., 2016; Chatzievangelou et al., 2020; Doya et al., 2017), while their use is now being extended to coastal areas as well (Falahzadeh et al., 2023; Aguzzi et al., 2015). Recently, two large datasets were also published that could be used to extend model training and allow for a better and wider description of the ecological marine environment (Francescangeli et al., 2023; Marini et al., 2022).

Assuming the crawler moves at a constant (or at least known) speed, time series of species counts generated during transect surveys can be geo-referenced and associated to different spatial sampling units (see Fig. 2), enabling an analysis of spatial distribution of species (Chatzievangelou et al., 2020). These functionalities allow crawlers to be used to consistently extend the radius of ecological monitoring of fixed cabled observatories, video-monitoring several km^2 of the seabed surface and the benthic boundary layer as a strategic transition zone between the benthic and pelagic environments (Aguzzi et al., 2019; Aguzzi et al., 2020a; Chatzievangelou et al., 2021; Rountree et al., 2020).

Increase in marine anthropogenic impacts, such as the extraction of fossil fuels and minerals along with the industrialized seabed fisheries, are creating the conditions in which the intelligent marine monitoring with robotic technologies will be indispensable and will by far outcompete typical cost intensive vessel or ROV operations (Jones et al., 2019; Simon-Lledó et al., 2019; Sutton et al., 2020). Automated, real-time image detection and counting of animals of multiple species and other objects such as particles or pollutants and eventually UXO (unexploded ordnance) will enhance the development and application of early, near-real-time warning systems for the monitoring and management of deep-sea environments, based on probabilistic modelling of animal abundances (Aguzzi et al., 2020c; Chatzievangelou et al., 2020). This development will be assisted by the consistent increase in the availability of affordable GPU resources to the scientific community.

4.3. *Limitations and methodological remarks*

Although this study advances the current state of the art in the use of AI to process marine footage on board mobile monitoring platforms, there are still certain limitations associated with data collection or processing. In particular, the need to perform the scans rotating the entire vehicle and not just the camera makes the recording process cumbersome creating delays, generating sediment clouds that reduce visibility and the possibility to repeat the scan if the video was not of adequate quality. This can inhibit the creation of maximum quality datasets. Furthermore, the limitation of moving only within the pre-defined tracks to minimize the physical disturbance on the seabed means that there is no option to record animals and objects from various angles, as in many ROV studies. An intrinsic limitation is related to the counting algorithm which is unable to count very fast objects appearing in the field of view for a time interval smaller or equal to half of the time series moving average interval (15 frames) corresponding to 0.25 s. Finally, expanding the focus on more species present in the area is challenging, either due to underrepresentation in the original dataset or due to the absence of easily discernible morphological characteristics, and is something to consider in future developments of such algorithms while keeping the computational costs low. All these challenges could be tackled with hardware updates (e.g., a rotating pan/tilt unit for the camera that increases the field of view, adding a stereo camera or an acoustic camera to complement the current HD one) and the collection and annotation of bigger datasets to serve as reference.

5. Conclusions

The YOLOv8 implemented detected and classified three benthic species, effectively and efficiently. High precision, recall and F_1-score metrics were achieved, with correct predictions even in cases where multiple individuals of all three species coexisted in a single frame.

Efficiency is highlighted by the fast processing, considering the requirements such as the crawler's motion and data collection rates, power and computational resources for real-time operations. Image augmentation performed only on the training set improved model generalization and tackled the "concept drift" phenomenon without making the detection process slower. High accuracy and the architecture of the workflow made tracking and related hyperparameters redundant, allowing for video streams to be directly fed to the model and generate real-time time series in future applications. In general, this lightweight model paves the way for the implementation of smart AI cameras on board of crawlers and other mobile autonomous platforms for real-time processing of imaging data, boosting marine ecological monitoring and conservation efforts.

Funding

Projects that supported the work were those of the Plan Estatal de Investigación Científica y Técnica y de Innovación 2017–2020 PLOME (PLEC2021-007525/AEI/10.13039/501100011033). Moreover, part of the conceptual development, falls within the framework of EU DIG-I4ECO [grant number 101112883 - GAP-101112883], REDRESS [grant number 101135492-GAP-101135492] and SUN-BIO [grant number 101157493-GAP-101157493]. ONC is funded through the Canada Foundation for Innovation-Major Science Initiative (CFI-MSI) fund 30199. The robotic operations were funded through ONC and Göteborg University.

CRediT authorship contribution statement

Luciano Ortenzi: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jacopo Aguzzi:** Conceptualization, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Corrado Costa:** Methodology, Writing – original draft, Writing – review & editing. **Simone Marini:** Methodology, Validation, Visualization. **Daniele D'Agostino:** Methodology, Validation, Visualization. **Laurenz Thomsen:** Conceptualization, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Fabio C. De Leo:** Visualization, Writing – original draft, Writing – review & editing. **Paulo V. Correa:** Visualization, Writing – original draft, Writing – review & editing. **Damianos Chatzievangelou:** Conceptualization, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Data availability

data and code are available on github repository (https://github.com/Luciano-ma/crawler-wally-animal-detection)

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102788.

References

Aguzzi, J., Manuèl, A., Condal, F., Guillén, J., Nogueras, M., del Rio, J., Costa, C., Menesatti, P., Puig, P., Sardà, F., Toma, D., Palanques, A., 2011. The new seafloor observatory (OBSEA) for remote and long-term coastal ecosystem monitoring. Sensors 11 (6), 5850–5872. https://doi.org/10.3390/s110605850.

Aguzzi, J., Company, J.B., Costa, C., Matabos, M., Azzurro, E., Mànuel, A., Menesatti, P., Sardà, F., Canals, M., Delory, E., Cline, D., Favali, P., Juniper, S.K., Furushima, Y., Fujiwara, Y., Chiesa, J.J., Marotta, L., Priede, I.M., 2012. Challenges to assessment of benthic populations and biodiversity as a result of rhythmic behaviour: video solutions from cabled observatories. Oceanogr. Mar. Biol. 50, 235–286.

Aguzzi, J., Doya, C., Tecchio, S., De Leo, F.I., Azzurro, E., Costa, C., Sbragaglia, V., Del Río, J., Navarro, J., Ruhl, H.A., Company, J.B., Favali, P., Purser, A., Thomsen, L., Catalán, I.A., 2015. Coastal observatories for monitoring of fish behaviour and their responses to environmental changes. Rev. Fish Biol. Fish. 25, 463–483. https://doi.org/10.1007/s11160-015-9387-9.

Aguzzi, J., Chatzievangelou, D., Marini, S., Fanelli, E., Danovaro, R., Flögel, S., Lebris, N., Juanes, F., De Leo, F., del Río, J., Thomsen, L., Costa, C., Riccobene, G., Tamburini, C., Lefevre, D., Gojakc, C., Poulain, P.M., Favali, P., Griffa, A., Purser, A., Cline, D., Edgington, D., Navarro, J., Stefanni, S., D'Hondt, S., Priede, I.G., Rountree, R., Company, J., 2019. New high-tech flexible networks for the monitoring of deep-sea ecosystems. Environ. Sci. Technol. 53 (12), 6616–6631. https://doi.org/10.1021/acs.est.9b00409.

Aguzzi, J., Albiez, J., Flögel, S., Rune Godø, O., Grimsbø, E., Marini, S., Pfannkuche, O., Rodriguez, E., Thomsen, L., Torkelsen, T., Valencia, J., López-Vázquez, V., Wehde, H., Zhang, G., 2020a. A flexible autonomous robotic observatory infrastructure for bentho-pelagic monitoring. Sensors 20 (6), 1614. https://doi.org/10.3390/s20061614.

Aguzzi, J., Chatzievangelou, D., Francescangeli, M., Marini, S., Bonofiglio, F., del Río, J., Danovaro, R., 2020b. The hierarchic treatment of marine ecological information from spatial networks of benthic platforms. Sensors-Basel 20, 1751. https://doi.org/10.3390/s20061751.

Aguzzi, J., Iveša, N., Gelli, M., Costa, C., Gavrilovic, A., Cukrov, N., Cukrov, M., Cukrov, Nu., Omanovic, D., Štifanić, M., Marini, S., Piria, M., Fanelli, E., Azzurro, E., Danovaro, R., 2020c. Ecological video monitoring of Marine Protected Areas by underwater cabled surveillance cameras. Mar. Policy 119, 104052. https://doi.org/10.1016/j.marpol.2020.104052.

Aguzzi, J., Flögel, S., Marini, S., Thomsen, L., Lbiez, J., Weiss, P., Picardi, G., Calisti, M., Stefanni, S., Mirimin, L., Vecchi, F., Laschi, C., Branch, A., Clark, E.B., Foing, B., Wedler, A., Chatzievangelou, D., Tangherlini, M., Purser, A., Danovaro, R., 2022. Developing technological synergies between deep-sea and space research. Elementa-Sci. Anthropoc. 10, 1–9. https://doi.org/10.1525/elementa.2021.00064.

Aziz, L., Salam, M.S.B.H., Sheikh, U.U., Ayub, S., 2020. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: a comprehensive review. IEEE Access 8, 170461–170495. https://ieeexplore.ieee.org/document/9186021.

Besson, M., Alison, J., Bjerge, K., Gorochowski, T.E., Høye, T.T., Jucker, T., Clements, C. F., 2022. Towards the fully automated monitoring of ecological communities. Ecol. Lett. 25 (12), 2753–2775. https://doi.org/10.1111/ele.14123.

Beyan, C., Browman, H.I., 2020. Setting the stage for the machine intelligence era in marine science. ICES J. Mar. Sci. 77 (4), 1267–1273. https://doi.org/10.1093/icesjms/fsaa084.

Bicknell, A.W.J., Godley, B.J., Sheehan, E.V., Votier, S.C., Witt, M.J., 2016. Camera technology for monitoring marine biodiversity and human impact. Front. Ecol. Environ. 14 (8), 424–432. https://doi.org/10.1002/fee.1322.

Bonofiglio, F., De Leo, F.C., Yee, C., Chatzievangelou, D., Aguzzi, J., Marini, S., 2022. Machine learning applied to big data from marine cabled observatories: a study case of sablefish population dynamics in the NE Pacific. Front. Mar. Sci. 9, 842946 https://doi.org/10.3389/fmars.2022.842946.

Cai, S., Li, G., Shan, Y., 2022. Underwater object detection using collaborative weakly supervision. Comput. Electr. Eng. 102, 108159 https://doi.org/10.1016/j.compeleceng.2022.108159.

Chatzievangelou, D., Doya, C., Mihály, S., Sastri, A.R., Thomsen, L., Aguzzi, J., 2016. High-frequency patterns in the abundance of benthic species near a cold-seep: an internet operated vehicle application. PLoS One 11 (10), e0163808. https://doi.org/10.1371/journal.pone.0163808.

Chatzievangelou, D., Aguzzi, J., Ogston, A., Suárez, A., Thomsen, L., 2020. Visual monitoring of key deep-sea megafauna with internet operated crawlers as a tool for ecological status assessment. Prog. Oceanogr. 184, 102321 https://doi.org/10.1016/j.pocean.2020.102321.

Chatzievangelou, D., Bahamon, N., Martini, S., del Rio, J., Riccobene, G., Tangherlini, M., Danovaro, R., De Leo, F.C., Pirenne, B., Aguzzi, J., 2021. Integration diel vertical migrations of bioluminescent deep scattering layers into monitoring programs. Front. Mar. Sci. 8, 661809 https://doi.org/10.3389/fmars.2021.661809.

Chatzievangelou, D., Thomsen, L., Aguzzi, J., Doya, C., Purser, A., 2022. Transects in the deep: opportunities with tele-operated resident seafloor robots. Front. Mar. Sci. 9, 833617 https://doi.org/10.3389/fmars.2022.833617.

Cuvelier, D., de Busserolles, F., Lavaud, R., Floc'h, E., Fabri, M.C., Sarradin, P.M., Sarrazin, J., 2012. Biological data extraction from imagery–how far can we go? A case study from the Mid-Atlantic ridge. Mar. Environ. Res. 82, 15–27. https://doi.org/10.1016/j.marenvres.2012.09.001.

Danovaro, R., Aguzzi, J., Fanelli, E., Billet, D., Gjerde, K., Jamieson, A., Ramirez-Llodra, E., Smith, C.R., Snelgrove, P.V.R., Thomsen, L., Van Dover, C., 2017. A new international ecosystem-based strategy for the global deep ocean. Science 355, 452–454.

Delange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., et al., 2021. A continual learning survey: defying forgetting in classification tasks. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/TPAMI.2021.3057446.

Doya, C., Aguzzi, J., Pardo, M., Matabos, M., Company, J.B., Costa, C., Mihaly, S., Canals, M., 2014. Diel behavioral rhythms in the sablefish (*Anoplopoma fimbria*) and other benthic species, as recorded by the deep-sea cabled observatories in Barkley canyon (NEPTUNE-Canada). J. Mar. Syst. 130, 69–78. https://doi.org/10.1016/j.jmarsys.2013.04.003.

Doya, C., Chatzievangelou, D., Bahamon, N., Purser, A., De Leo, F., Juniper, K., Thomsen, L., Aguzzi, J., 2017. Seasonal monitoring of deep-sea megabenthos in Barkley canyon cold seep by internet operated vehicle (IOV). PLoS One 12 (5), e0176917. https://doi.org/10.1371/journal.pone.0176917.

Durden, J.M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A.G., Bett, B.J., 2016. Perspectives in visual imaging for marine biology and ecology: From

acquisition to understanding. In: Oceanography and Marine Biology. CRC Press, pp. 9–80.
Durden, J.M., Hosking, B., Bett, B.J., Cline, D., Ruhl, H.A., 2021. Automated classification of fauna in seabed photographs: the impact of training and validation dataset size, with considerations for the class imbalance. Prog. Oceanogr. 196, 102612 https://doi.org/10.1016/j.pocean.2021.102612.
Falahzadeh, A., Toma, D.M., Francescangeli, M., Chatzievangelou, D., Nogueras, M., Martínez, E., Carandell, M., Tangerlini, M., Thomsen, L., Picardi, G., Le Bris, M., Dominguez, L., Aguzzi, J., del Río, J., 2023. A new coastal crawler prototype to expand the ecological monitoring radius of OBSEA cabled observatory. J. Mar. Sci. Eng. 11 (4), 857. https://doi.org/10.3390/jmse11040857.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27 (8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.
Ferrari, M., D'Agostino, D., Aguzzi, J., Marini, S., 2024. Underwater Mediterranean image analysis based on the compute continuum paradigm. Futur. Gener. Comput. Syst. 107481.
Francescangeli, M., Marini, S., Martínez, E., et al., 2023. Image dataset for benchmarking automated fish detection and classification algorithms. Sci. Data 10, 5. https://doi.org/10.1038/s41597-022-01906-1.
Gervais, F., Juniper, S.K., Matabos, M., Spicer, A., 2012. Marine Life Field Guide, First edition. NEPTUNE-Canada Publications.
Goodwin, M., Halvorsen, K.T., Jiao, L., Knausgård, K.M., Martin, A.H., Moyano, M., Oomen, R.A., Rasmussen, J.H., Sørdalen, T.K., Thorbjørnsen, S.H., 2022. Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. ICES J. Mar. Sci. 79 (2), 319–336. https://doi.org/10.1093/icesjms/fsab255.
Grinyo, J., Aguzzi, J., Kenchington, E., Costa, C., Hanz, U., Mienis, F., 2023. Occurrence and behavioral rhythms of the endangered Acadian redfish (*Sebastes fasciatus*) in the Sambro bank (Scotian shelf). Front. Mar. Sci. 10, 1158283. https://doi.org/10.3389/fmars.2023.1158283.
Han, X., Chang, J., Wang, K., 2021. You only look once: unified, real-time object detection. Proc. Comp. Sci. 183 (1), 61–72.
Harrison, D., De Leo, F.C., Gallin, W.J., Mir, F., Marini, S., Leys, S.P., 2021. Machine learning applications of convolutional neural networks and Unet architecture to predict and classify demosponge behavior. Water 13 (18), 2512. https://doi.org/10.3390/w13182512.
He, K., Sun, J., 2015. Convolutional neural networks at constrained time. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5353–5360.
Huettmann, F., Arhonditsis, G., 2023. Towards an ecological informatics scholarship that is reflective, repeatable, transparent, and sharable! Eco. Inform. 76, 102132.
Hussain, M., 2023. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines 11 (7), 677. https://doi.org/10.3390/machines11070677.
Hussain, M., 2024. YOLOv1 to v8: unveiling each variant–a comprehensive review of YOLO. IEEE Access 12, 42816–42833.
Jahanbakht, M., Xiang, Wei, Hanzo, L., Azghadi, Mostafa Rahimi, 2021. Internet of underwater things and big marine data analytics - a comprehensive survey. IEEE Communications Surveys and Tutorials 23 (2), 904–956 art. no. 9328873. Cited 137 times. https://doi.org/10.1109/COMST.2021.3053118.
Jones, D.O., Gates, A.R., Huvenne, V.A., Phillips, A.B., Bett, B.J., 2019. Autonomous marine environmental monitoring: application in decommissioned oil fields. Sci. Total Environ. 668, 835–853. https://doi.org/10.1016/j.scitotenv.2019.02.310.
Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., Boulais, O., Cromwell, M., Butler, E., Woodward, B., Bell, K.L.C., 2022. FathomNet: a global image database for enabling artificial intelligence in the ocean. Sci. Rep. 12, 15914. https://doi.org/10.1038/s41598-022-19939-2.
Langenkämper, D., Zurowietz, M., Schoening, T., T.W., 2017. Nattkemper BIIGLE 2.0 - browsing and annotating large marine image collections. Front. Mar. Sci. 4, 83. https://doi.org/10.3389/fmars.2017.00083.
Liu, S., Xue, S., Wu, J., Zhou, C., Yang, J., Li, Z., et al., 2021. Online Active Learning for Drifting Data Streams IEEE Transactions on Neural Networks and Learning Systems. IEEE. https://doi.org/10.1109/TNNLS.2021.3091681.
Lopez-Marcano, S., Jinks, E.L., Buelow, C.A., Brown, C.J., Wang, D., Kusy, B., Ditria, E. M., Connolly, R.M., 2021. Automatic detection of fish and tracking of movement for ecology. Ecol. Evol. 11 (12), 8254–8263. https://doi.org/10.1002/ece3.7656.
Lopez-Vazquez, V., Lopez-Guede, J.M., Marini, S., Fanelli, E., Johnsen, E., Aguzzi, J., 2020. Video image enhancement and machine learning pipeline for animal tracking and classification at cabled observatories. Sensors-Basel 20, 726. https://doi.org/10.3390/s20030726.
Lopez-Vazquez, V., Lopez-Guede, J.M., Chatzievangelou, D., Aguzzi, J., 2023. Deep learning based deep-sea automatic image enhancement and animal species classification. J. Big Data 10, 37. https://doi.org/10.1186/s40537-023-00711-w.
Malde, K., Handegard, N.O., Eikvil, L., Salberg, A.B., 2020. Machine intelligence and the data-driven future of marine science. ICES J. Mar. Sci. 77 (4), 1274–1285. https://doi.org/10.1093/icesjms/fsz057.
Mallet, D., Pelletier, D., 2014. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). Fish. Res. 154, 44–62. https://doi.org/10.1016/j.fishres.2014.01.019.
Marini, S., Corgnati, L., Manotovani, C., Bastianini, M., Ottaviani, E., Fanelli, E., Aguzzi, J., Griffa, A., Poulain, P.M., 2018. Automated estimate of fish abundance through the autonomous imaging device GUARD1. Measurement 126, 72–75. https://doi.org/10.1016/j.measurement.2018.05.035.
Marini, S., Bonofiglio, F., Corgnati, L.P., et al., 2022. Long-term high resolution image dataset of Antarctic coastal benthic Fauna. Sci. Data 9, 750. https://doi.org/10.1038/s41597-022-01865-7.
Matabos, M., Bui, A.O., Mihály, S., Aguzzi, J., Juniper, D.S.K., Ajayamohan, R.S., 2014. High-frequency study of benthic megafaunal community dynamics in Barkley canyon: a multidisciplinary approach using the NEPTUNE Canada network. J. Mar. Syst. 130, 56–68. https://doi.org/10.1016/j.jmarsys.2013.05.002.
Matabos, M., Hoeberechts, M., Doya, C., Aguzzi, J., Nephin, J., Reimchen, T.E., Leaver, S., Marx, R.M., Branzan Albu, A., Fier, R., Fernandez-Arcaya, U., Juniper, S. K., Expert, Crowd, 2017. Students or algorithm: who holds the key to deep-sea imagery 'big data' processing? Methods Ecol. Evol. 8 (8), 996–1004. https://doi.org/10.1111/2041-210X.12746.
McIntosh, D., Marques, T.P., Albu, A.B., Rountree, R., De Leo, F., 2020. Movement tracks for the automatic detection of fish behaviour in videos. arXiv 2011, 14070. https://doi.org/10.48550/arXiv.2011.14070 preprint arXiv.
McIntosh, D., Marques, T.P., Albu, A.B., Rountree, R., De Leo, F., 2022. Tempnet: Temporal attention towards the detection of animal behaviour in videos. In: 2022 26th International Conference on Pattern Recognition (ICPR), pp. 3318–3324. https://ieeexplore.ieee.org/document/9956609.
Moran, K., Juniper, S.K., Bligh, S., Loock, D., Kulin, I., Paulson, M., Pirenne, B., 2022. Canada's internet-connected Ocean. Front. Mar. Sci. 8, 805134 https://doi.org/10.3389/fmars.2021.805134.
Naseer, A., Baro, E.N., Khan, S.D., Vila, Y., 2022. A novel detection refinement technique for accurate identification of *Nephrops norvegicus* burrows in underwater imagery. Sensors 22 (12), 4441. https://doi.org/10.3390/s22124441.
Nelson, L., Solawetz, J., 2020. YOLOv5 Is here: State-of-the-Art Object Detection at 140 FPS. https://blog.roboflow.com/yolov5-is-here/.
Nvidia, 2024. Jetson Nano User Guide. https://developer.nvidia.com/embedded/downloads.
Ottaviani, E., Francescangeli, M., Gjeci, N., del Rio Fernandez, J., Aguzzi, J., Marini, S., 2022. Assessing the image concept drift at the OBSEA coastal underwater cabled observatory. Front. Mar. Sci. 9, 459. https://doi.org/10.3389/fmars.2022.840088.
Owens, D., Abeysirigunawardena, D., Biffard, B., Chen, Y., Conley, P., Jenkyns, R., Kerschtien, S., Lavallee, T., MacArthur, M., Mousseau, J., Old, K., Paulson, M., Pirenne, B., Scherwath, M., Thorne, M., 2022. The oceans 2.0/3.0 data management and archival system. Front. Mar. Sci. 9, 806452 https://doi.org/10.3389/fmars.2022.806452.
Purser, A., Thomsen, L., Barnes, C., Best, M., Chapman, R., Hofbauer, M., Hofbauer, M., Wagner, H., 2013. Temporal and spatial benthic data collection via an internet operated Deep Sea crawler. Met. Oceanogr. 5, 1–18. https://doi.org/10.1016/j.mio.2013.07.001.
Rountree, R., Aguzzi, J., Marini, S., Fanelli, E., De Leo, F., Del Rio, J., Juanes, F., 2020. Towards an optimal design for ecosystem-level ocean observatories. Oceanogr. Mar. Biol. 58, 79–106.
Rubbens, P., Brodie, S., Cordier, T., Destro Barcellos, D., Devos, P., Fernandes-Salvador, J.A., Fincham, J.I., Gomes, A., Handegard, N.O., Howell, K., Jamet, C., Kartveit, K.H., Moustahfid, H., Parcerisas, C., Politikos, D., Sauzède, R., Sokolova, M., Uusitalo, L., Van den Bulcke, L., van Helmond, A.T.M., Watson, J.T., Welch, H., Beltran-Perez, O., Chaffron, S., Greenberg, D.S., Kühn, B., Kiko, R., Lo, M., Lopes, R.M., Möller, K.O., Michaels, W., Pala, A., Romagnan, J.-B., Schuchert, P., Seydi, V., Villasante, S., Malde, K., Irisson, J.-O., 2023. Machine learning in marine ecology: an overview of techniques and applications. ICES J. Mar. Sci. 80 (7), 1829–1853. https://doi.org/10.1093/icesjms/fsad100.
Sangekar, M., Friedman, A., Hidaka, M., Hosono, T., Lindsay, D., 2023. SquidJam: a video annotation ecosystem. In: 2023 IEEE Underwater Technology (UT), pp. 1–5. https://doi.org/10.1109/UT49729.2023.10103435.
Sayin, B., Krivosheev, E., Yang, J., Passerini, A., Casati, F., 2021. A review and experimental analysis of active learning over crowdsourced data. Artif. Intell. Rev. 54, 5283–5305. https://doi.org/10.1007/s10462-021-10021-3.
Scherwath, M., Thomsen, L., Riedel, M., Römer, M., Chatzievangelou, D., Schwendner, J., Duda, A., Heesemann, M., 2019. Ocean observatories as a tool to advance gas hydrate research. Earth Space Sci. 6 (12), 2644–2652. https://doi.org/10.1029/2019EA000762.
Shortis, M., Abdo, E.H.D., 2016. A review of underwater stereo-image measurement for marine biology and ecology applications. Oceanogr. Mar. Biol. 47, 269–304.
Simon-Lledó, E., Bett, B.J., Huvenne, V.A., Köser, K., Schoening, T., Greinert, J., Jones, D. O., 2019. Biological effects 26 years after simulated deep-sea mining. Sci. Rep. 9 (1), 8040. https://doi.org/10.1038/s41598-019-44492-w.
Song, P., Li, P., Dai, L., Wang, T., Chen, Z., 2023. Boosting R-CNN: reweighting R-CNN samples by RPN's error for underwater object detection. Neurocomputing 530, 150–164.
Stojanovic, M., Beaujean, P.P.J., 2016. Acoustic communication. In: Springer Handbook of Ocean Engineering, pp. 359–386. https://doi.org/10.1007/978-3-319-16649-0_15.
Sutton, T.T., Frank, T., Judkins, H., Romero, I.C., 2020. As Gulf oil extraction goes deeper, who is at risk? Community structure, distribution, and connectivity of the deep-pelagic fauna. In: Scenarios and Responses to Future Deep Oil Spills: Fighting the Next War, 2020, pp. 403–418. https://doi.org/10.1007/978-3-030-12963-7_24.
Terven, J., Cordova-Esparza, D., 2023. A comprehensive review of YOLO: From YOLOv1 and beyond. arXiv. https://doi.org/10.48550/arXiv.2304.00501 arXiv:2304.00501.
Thomsen, L., Barnes, C., Best, M., Chapman, R., Pirenne, B., Thomson, R., Vogt, J., 2012. Ocean circulation promotes methane release from gas hydrate outcrops at the NEPTUNE Canada Barkley canyon node. Geophys. Res. Lett. 39 (16), L16605. https://doi.org/10.1029/2012GL052462.
Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z., 2019. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput. Electron. Agric. 157, 417–426. https://doi.org/10.1016/j.compag.2019.01.012.

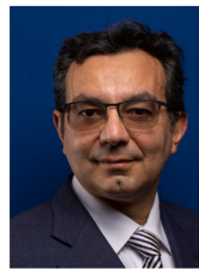
Xie, X., Cheng, G., Wang, J., Yao, X., Han, J., 2021. Oriented R-CNN for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3520–3529.
Yang, L., Liu, Y., Yu, H., Fang, X., Song, L., Li, D., Chen, Y., 2021. Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: a review. Arch. Comput. Met. Eng. 28, 2785–2816. https://doi.org/10.1007/s11831-020-09486-2.
Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M., Lee, B., 2022. A survey of modern deep learning based object detection models. Digit. Sign. Process. 126, 103514 https://doi.org/10.1016/j.dsp.2022.103514.
Zhang, J., Tao, D., 2021. Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things. IEEE Internet Things J. 8 (10), 7789–7817. https://doi.org/10.1109/JIOT.2020.3039359 art. no. 9264235.
Zhang, C., Zhang, G., Li, H., Liu, H., Tan, J., Xue, X., 2023. Underwater target detection algorithm based on improved YOLOv4 with SemiDSConv and FIoU loss function. Front. Mar. Sci. 10, 1153416. https://doi.org/10.3389/fmars.2023.1153416.
Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., Zuo, W., 2021. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Trans. Cybernet. 52 (8), 8574–8586. https://doi.org/10.1109/TCYB.2021.3095305.

Luciano Ortenzi is a Senior Assistant Professor (tenure track) in Computer Science at the Department of Agriculture and Forest Sciences (DAFNE) of Tuscia University since 2022. Artificial Intelligence and Machine Learning expert for the OBSEA seafloor observatory (www.obsea.es). Editorial Board Member of 2 Special Issues, he is co-author in more than 30 publications in condensed matter physics and machine learning applications to the agri-food sector and marine and terrestrial ecological monitoring.

Jacopo Aguzzi research focuses on changes in sampled richness and biodiversity at different temporal scales, extracting ecological indicators from imaging products by cabled observatories, crawlers, ROVs, AUVs and neutrino telescopes. He is manager for the "ecological monitoring" and "citizen science" of the EMSO Testing-Site OBSEA (www.obsea.es). Between 2019-2019, he has been Scientific Advisor for the Ocean Network Canada (ONC) for the Section "Life in the Environments of the Northeast Pacific Ocean and Salish Sea".

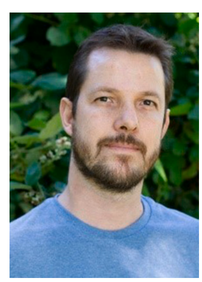
Corrado Costa is a Senior Researcher (first level) biologist at CREA-IT from 2022. Scientific responsible of two operative unit of a European and an Italian project; Coordinator of two National project and WP leader of 11 national and 5 international projects. Editorial Board Member of 10 international journals. Expert auditor for international projects. Winner of 12 prizes for his publications. He has 8 patents and is co-author of more than 380 publications among which 178 with IF (h-Index WoS = 38) mainly on Image analysis, sensors-based applications, and multivariate AI modelling in the agriculture, fishery and forestry sectors.

Simone Marini Ph.D. is researcher at the Institute of Marine Sciences of the National Research Council of Italy. His research activity deals with knowledge discovery and pattern analysis, study and design of intelligent marine observing systems, where he is first inventor of the European patent EP2863257. He leaded several research units and work packages within national and international projects, he co-authored >100 scientific publications among which 70 with impact factor. He edited several special issues, and he is associated editor of two ISI scientific journals. He acted as reviewer for different research programs, included the European Research council.



Daniele D'Agostino Ph.D., is associate professor at the University of Genoa. His main research activities are related to the development of high-performance software for 3D and Bioinformatics data processing, the resource management in heterogeneous systems and the design of science gateways for different disciplines. He acted as co-chair of the 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, co-organizers of special sessions at international conferences and co-guest editor of special issues of ISI journals. He co- authored >110 scientific papers, published in journals, book chapters and conference proceedings.

Laurenz Thomsen is professor at the University of Göteborg, Sweden. His main research activities are related to sediment transport at continental margins, the marine carbon cycle, anthropogenic impact and the development of seafloor robots. Thomsen was a member of the Senate Commission on Oceanography of DFG which evaluated project proposals and coordinated the activities of the DFG in the field of marine research. He is associate editor of the non-profit journal "Elementa" and has co-authored >100 scientific papers. His current activities focus on environmental monitoring technologies for deep sea mining and marine restoration of deep-sea habitats.



Fabio C. De Leo is a senior scientist with Ocean Networks Canada (ONC) and adjunct assistant professor in the Department of Biology at University of Victoria. He focuses on deep-sea benthic ecology and the effects of anthropogenic stressors on community structure and function. Throughout 24+ years of academic research, he participated in oceanographic expeditions in the Atlantic, Pacific, and Southern Oceans and the Mediterranean and Tasman Seas, including manned submersible dives down to 1.2 km in submarine canyons off Hawaii. Fabio lectures at various UVic departments (Biology, Earth & Ocean and Computer Science), co-supervises graduate students and has 50+ peer-reviewed publications.

Paulo V. Correa is a junior staff scientist with Ocean Networks Canada (ONC) who supports projects related to underwater imagery data and ecology. His research mainly focuses on deep-sea ecosystems and the effects of anthropogenic stressors on community structure. He has been directly responsible for annotations of deep-sea images and videos and identification of benthic and pelagic species. He has experience in multivariate earth-ocean data analysis, including studying potential drivers for community structure and species distribution on the seafloor.

Damianos Chatzievangelou is a Juan de la Cierva Formación postdoctoral fellow and member of the "ecological monitoring and indicators task force" within various national and international projects of ICM-CSIC. He got his PhD in Geosciences from Jacobs University Bremen (now Constructor University) in Germany. Between 2016 and 2022 he was the main pilot of the Internet Operated Deep-sea Crawler "Wally", integrated in ONC observatory. He is interested in the application of computational methods to integrate data from multiple sources, such as footage from mobile and stationary robotic platforms, into coherent ecological indicators.