Using AI computer vision algorithm (YOLOv11) for automatic video annotation of coral taxa on the Moroccan Atlantic Coast

Ali Loulidi[1,3], Lene Buhl-Mortensen[2], Rachida Houssa[3], Hassan Rhinane[1]

1: Geoscience Laboratory, Faculty of Sciences Ain Chock, University Hassan II, Casablanca 20100, Morocco.

2: Institute of Marine Research (IMR), Bergen 5005, Norway.

3: Institut National de Recherche Halieutique (INRH), Casablanca 20100, Morocco.

Keywords: ROV video annotation, Shallow water, Computer vision, YOLO, Corals, Moroccan Atlantic Coast

Abstract

Marine biodiversity is essential for maintaining healthy and resilient ocean ecosystems, supporting fisheries, regulating climate, and providing vital resources for human well-being. It underpins ecosystem services such as carbon sequestration and oxygen production, making it important for both environmental and economic sustainability. Consequently, the conservation and monitoring of highly species rich and vulnerable marine ecosystems, such as shallow-water gorgonian coral populations, are important for biodiversity preservation. This study tests the ability of the AI computer vision algorithms YOLOv11, to detect and count coral colonies belonging to six common taxa on video records from coral gardens on the Moroccan Atlantic Coast. These videos were recorded using an Remotely Operated Vehicle (ROV) with the objective to map coral habitats with the research vessel Dr. Fridjof Nansen as part of the FAO Nansen program. Focusing on three gorgonian species: *Eunicella verrucosa* (Pallas, 1766), *Ellisella paraplexauroides* Stiasny, 1936, and *Leptogorgia viminalis* (Pallas, 1766), two sea pen species: *Veretillum cynomorium* (Pallas, 1766), and *Pennatula rubra* Ellis, 1764, and the hard coral *Dendrophyllia ramea* (Linnaeus, 1758). The research aims to develop an efficient solution to help improving video annotation by making it faster and easier. A dataset of 658 coral images was collected from Google Image and the DORIS database (Données d'Observations pour la Reconnaissance et l'Identification de la faune et la flore Subaquatiques). The images were divided into training, validation and test sets. To enhance model performance, we applied data augmentation. The YOLOv11 includes five different variants (n, s, m, l, x) for which detection precision was compared. Based on precision, recall, F1-score and mAP metrics, YOLOv11n proved to be the best model for coral detection regarding balance of accuracy and efficiency and with a mAP of 88% and a F1-score of 81%. This model was used for all subsequent ROV video analyses. The prediction results were applied to ROV video recordings from shallow water areas, demonstrating the potential of YOLOv11 as a powerful tool for the automated detection and monitoring of coral gardens. This approach offers significant contributions to marine biodiversity assessment along the Moroccan Atlantic coast.

1. Introduction

Marine ecosystems play a fundamental role in global biodiversity, climate regulation, and human livelihoods (Costello et al., 2010; Worm et al., 2006). Among these ecosystems, coral gardens composed of gorgonians, sea pens, and other benthic organisms serve as biodiversity hotspots, providing critical habitats for numerous marine species (Buhl-Mortensen et al., 2010; Gori et al., 2017). These structurally complex communities enhance benthic diversity and support fisheries (Grabowski et al., 2012; Thurber et al., 2014). However, they are increasingly threatened by anthropogenic pressures such as bottom trawling, climate change, and ocean acidification (Hoegh-Guldberg et al., 2007; Buhl-Mortensen et al., 2018). Consequently, effective monitoring and conservation strategies are urgently needed to safeguard these vulnerable ecosystems.

Traditional methods of assessing coral garden distribution and abundance rely on manual annotation of imagery collected via Remotely Operated Vehicles (ROVs), towed cameras, or scuba diving surveys (Gomes-Pereira et al., 2016; Beijbom et al., 2015). While these approaches yield valuable data, they are often labor-intensive, time-consuming, and subject to observer bias (Durden et al., 2016). Recent advances in artificial intelligence (AI) and computer vision have revolutionized marine ecological monitoring by enabling automated species detection and habitat classification (Schoening et al., 2012; Mahmood et al., 2017). Deep learning-based object detection models, particularly those in the You Only Look Once (YOLO) family, have demonstrated remarkable efficiency in real-time image analysis, making them ideal for processing large volumes of underwater video data (Redmon et al., 2016; Sapkota et al., 2025).

Several studies have successfully applied AI-driven approaches to marine biodiversity monitoring. For instance, YOLOv5 and Faster R-CNN have been used to detect deep-sea corals and sponges (Langenkämper et al., 2017), while convolutional neural networks (CNNs) have been employed for automated fish and benthic species identification (Villon et al., 2018). Similarly, Loulidi et al. (2022) demonstrated the efficacy of YOLOv3 in detecting fish across diverse marine environments. However, the application of YOLOv11 an advanced iteration of the YOLO architecture to detect shallow-water gorgonian coral gardens remains underexplored, particularly in regions like the Moroccan Atlantic coast, where these habitats are ecologically significant (Loulidi et al., 2024).

This study investigates the efficacy of YOLOv11 (Jocher et al. 2024) in automatically detecting and quantifying coral gardens from ROV footage collected during the 2020 FAO-Nansen survey along the Moroccan Atlantic coast. As part of this survey, the first visual mapping of habitats on the continental shelf and slope was conducted by the research vessel Dr. Fridjof Nansen, with video transects recorded at multiple stations. Building on species distribution patterns documented by Loulidi et al. (2025), we will apply our model on selected video segments from four representative transects (A1-21, A1-1, A2-2, A2-3) in shallow-water areas (32-76m) where all six target coral species were known to be common (Buhl-Mortensen et al., 2025).

Focusing on most common species including the gorgonians *Eunicella verrucosa*, *Ellisella paraplexauroides*, and *Leptogorgia viminalis*, sea pens *Veretillum cynomorium* and *Pennatula rubra*, and the hard coral *Dendrophyllia ramea* we trained YOLOv11 on a dataset of 658 annotated images. Data augmentation techniques enhanced model robustness, and performance was evaluated using mean average precision (mAP). Our results demonstrate the potential of YOLOv11 as a scalable, efficient tool for coral gardens monitoring. By integrating AI into marine biodiversity assessments, this approach supports standardized conservation efforts for vulnerable benthic ecosystems.

2. Evolution of YOLO architecture for marine monitoring

The YOLO (You Only Look Once) family of algorithms has undergone significant evolution since its inception, with each iteration introducing innovations that expand its applicability in ecological monitoring. Recent adaptations have specifically addressed the challenges of underwater environments, where light attenuation, turbidity, and limited training data complicate object detection. Zhang et al. (2023) enhanced YOLOv7 with double domain augmentation and self-attention mechanisms, achieving 83.6% mAP on marine organism detection despite variable water conditions - a critical advancement for our work in Morocco's dynamic coastal waters.
Building on these foundations, researchers have developed specialized solutions for benthic ecosystems. Ranolo et al. (2023) combined GAN-generated synthetic data with CLAHE contrast enhancement to boost YOLOv7's coral detection accuracy by 8-10%, while Lu et al. (2024) achieved 81.9% mAP on delicate soft corals using their SCoralDet model. These studies demonstrate YOLO's adaptability to diverse marine environments and organism morphologies, validating its suitability for coral garden detection tasks similar to our study of Moroccan gorgonian communities.
The challenge of limited annotated data has been creatively addressed through transfer learning and augmentation strategies. Ouassine et al. (2024) expanded a 400-image coral dataset to 580 samples through transformations, validating our approach of augmenting limited training data. Similarly, Levy et al. (2018) pioneered techniques for adapting pretrained models to marine environments with scarce labels, achieving robust tracking despite low visibility conditions.
Beyond detection, YOLO variants now enable advanced ecological analyses. Sella Veluswami et al. (2024) integrated YOLOv8 with regression models for fish biomass estimation in turbid waters (0.899 mAP), showcasing the architecture's potential for multidimensional monitoring. While our focus remains on detection, their success with preprocessing techniques informs our handling of Morocco's variable water clarity.
These innovations collectively demonstrate YOLO's growing capability to overcome three key challenges in our study:
-Variable illumination in shallow waters
-Complex morphology of gorgonian colonies
-Processing requirements for ROV video streams
Our work synthesizes these advances while introducing specific optimizations for Moroccan coral gardens, including data augmentation and depth-aware detection heads tailored to ROV survey conditions.

3. Materials and methods

3.1 Dataset description

Our study utilized two distinct datasets for model training and detection. The training dataset comprises 658 annotated coral images collected from Google Images and the DORIS database, with resolutions ranging from 720×576 to 1920×1072 pixels. This dataset includes six coral species that are common in the coral forests off Atlantic Morocco (Figure 1):
-*Eunicella verrucosa*: Branching gorgonians forming dense colonies, typically found in hard substrates. Their fan-like structures provide habitat for diverse marine organisms (Loulidi et al. 2025).
-*Ellisella paraplexauroides*: Slender, whip-like gorgonians that often grow in current-swept areas. They are characterized by their flexible stems and small polyps (Loulidi et al. 2025).
-*Leptogorgia viminalis*: Bushy gorgonians with delicate branching patterns, commonly inhabiting shallow waters. Their vibrant colors range from yellow to deep purple (Loulidi et al. 2025).
-*Veretillum cynomorium*: A sea pen with feather-like appendages, typically buried in soft sediments. They exhibit bioluminescence when disturbed (Loulidi et al. 2025).
-*Pennatula rubra*: Classic "sea pen" species with a central stem and lateral branches, resembling antique quill pens. They often occur in groups on sandy or muddy bottoms (Loulidi et al. 2025).
-*Dendrophyllia ramea*: Stony corals with robust, tree-like skeletons and large polyps. Unlike most corals, they thrive in low-light environments (Loulidi et al. 2025).
For the detection phase, we utilized ROV video footage collected during the 2020 habitat mapping survey where the Research Vessel (RV) Dr. Fridtjof Nansen surveyed 37 shallow-water localities along the Moroccan Atlantic coast. From these, we selected four representative transects (A1-21, A1-1, A2-2, A2-3) that contained the target gorgonian gardens and sea pen habitats, as documented in Loulidi et al. (2025). Rather than using entire video sequences, we extracted short clips showing all six target coral species and merged them into a single 4-minute 57-second composite video (1920×1080 resolution, 30 fps). This curated dataset maintained original survey quality while providing an efficient way to test our model across key habitats (Figure 2). This detection dataset provides a realistic evaluation scenario, reflecting actual monitoring conditions along the Moroccan Atlantic coast.

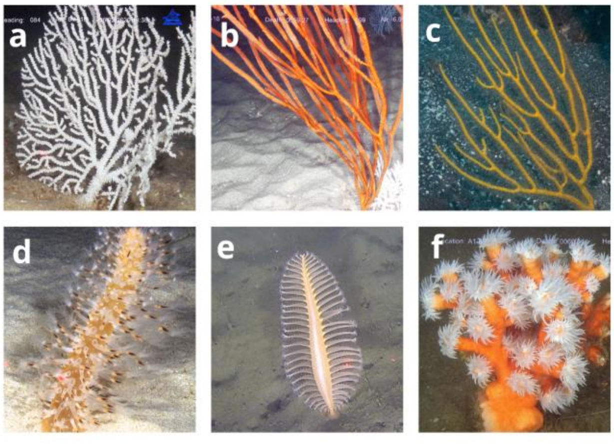

Figure 1. The six benthic species in shallow-water coral gardens: gorgonians ((a) *Eunicella verrucosa*, (b) *Ellisella paraplexauroides*, (c) *Leptogorgia viminalis*), sea pens ((d) *Veretillum cynomorium*, (e) *Pennatula rubra*), and a hard coral ((f) *Dendrophyllia ramea*) (Photos from: “Coral habitat fauna of Northwest Africa: A photographic guide to taxa identification”, Buhl-Mortensen et al., 2025)

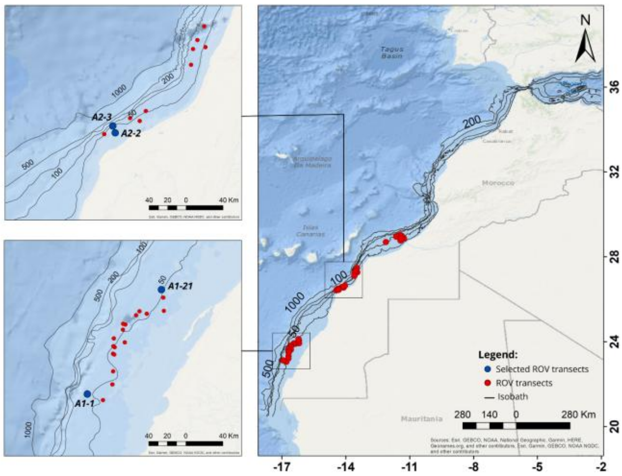

Figure 2. The position of four representative selected ROV stations together with all ROV shallow water stations off Morocco where video survey was conducted in 2020 as part of the FAO-NANSEN program.

3.2 Data pre-processing

The dataset was annotated using Roboflow (Roboflow Inc., 2023), beginning with 658 original coral images resized to a standardized 640×640-pixel resolution. To enhance model robustness while maintaining biological accuracy, we implemented an augmentation technique on the training set (80% of data), which included: horizontal and vertical flipping, random rotations (±15°), grayscale conversion (15% of images), brightness adjustments (-5% to +25%), and limited cropping (0-20% zoom). The remaining data was preserved as unaugmented validation (10%) and test (10%) sets, ensuring unbiased evaluation of model performance. This approach helps the model recognize corals in different conditions while letting us test its performance on images.

3.3 Architecture of YOLOv11

YOLOv11 (Jocher et al. 2024) follows the standard YOLO framework comprising three key components: backbone, neck, and head (Figure 3). The architecture processes input images through sequential feature extraction and fusion stages optimized for real-time object detection.

Backbone: Built upon RepVGG blocks, the backbone utilizes REP (Re-parameterization) operations and Spatial Pyramid Pooling Fast (SPPF) for efficient multi-scale feature extraction. The SPPF module employs max-pooling operations at varying kernel sizes to capture contextual information across different scales while maintaining computational efficiency. All convolutional layers incorporate Batch Normalization (BN2d) and SiLU activation functions.

Neck: The feature pyramid network combines Cross-stage Partial (CSP) blocks with upsampling and concatenation operations. Key components include:

- CBR (Convolution-BatchNorm-ReLU) modules for feature refinement
- REP blocks for parameter optimization
- Bidirectional feature fusion through sequential upsampling and concatenation

Head: The detection head utilizes CBS (Convolution-BatchNorm-SiLU) modules and final concatenation layers to generate predictions. It produces multi-scale outputs (Detect module) at three resolution levels (P3-P5) to handle objects of varying sizes. The architecture maintains YOLO's characteristic efficiency through:

- Optimized 3×3 kernel convolutions
- Strategic skip connections
- Bottleneck-designed CSP blocks

This streamlined architecture preserves YOLOv11's capabilities for real-time detection while enhancing feature representation through its improved backbone and neck designs. The combination of RepVGG blocks and SPPF modules provides robust multi-scale processing, particularly beneficial for detecting coral formations of varying sizes and orientations underwater imagery.

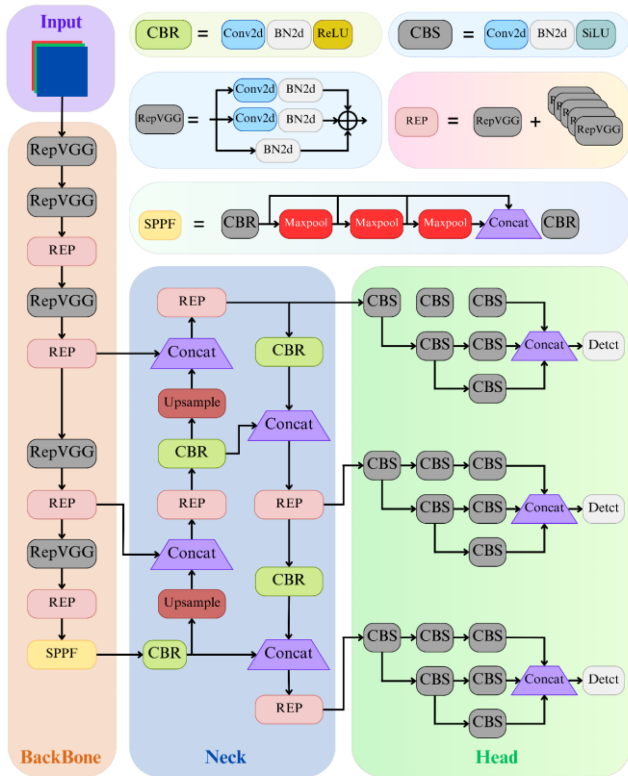

Figure 3. YOLOv11 model structure (Jocher et al. 2024): backbone, neck, and head components.

3.4 Methodology

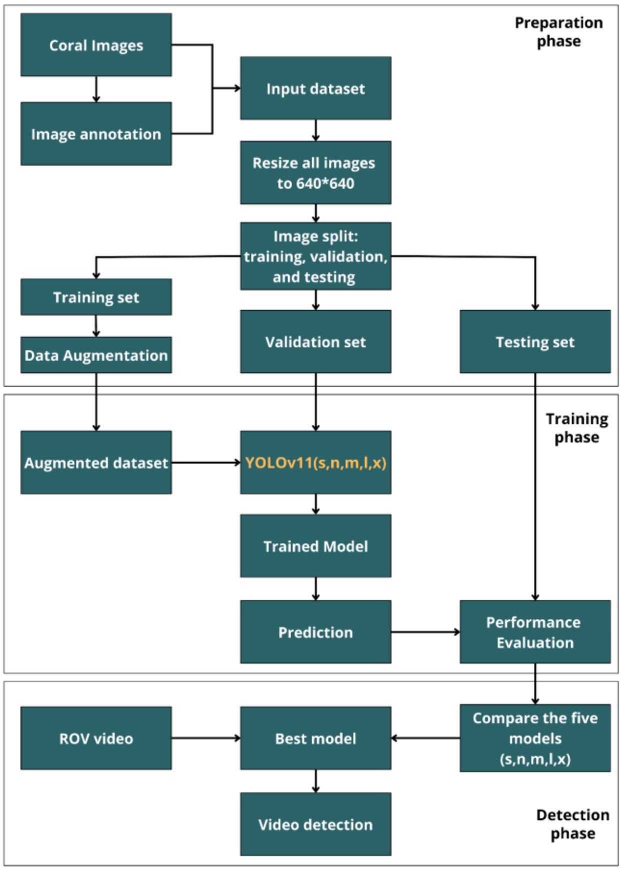

Figure 4. Methodology flowchart of training YOLOv11 model to detect the six coral taxa.

This study trained a YOLOv11 model to detect six coral taxa through three phases: Preparation, Training, and Detection (Figure 4). The preparation phase established the foundation for model development by collecting a dataset of 658 annotated images featuring six target coral taxa. Each image was standardized to 640×640 pixel resolution to ensure consistency across the dataset. To optimize model performance and evaluation, the dataset was partitioned into three subsets: the training set (80% of images) served as the primary learning material for the model, The validation set (10% of the data) served as an independent evaluation during training, enabling model optimization without directly influencing parameter updates, thus preventing overfitting to the training data, and the test set (10%) offered a completely unseen evaluation of final model performance. Data augmentation techniques were applied to enhance the model's ability to generalize across the challenging and variable conditions typical of underwater imagery.

During the training phase, we conducted an evaluation of five distinct YOLOv11 architectural variants to identify the optimal balance between detection accuracy and computational efficiency for our specific application. The variants ranged from the extremely lightweight Nano (n) version, designed for deployment on resource-constrained edge devices, through to the comprehensive Extra-Large (x) version capable of maximum detection precision. Intermediate versions - Small (s), Medium (m), and Large (l) - provided graduated levels of complexity and capability. All five variants were trained using identical hyperparameters (No. of epochs, learning rate, optimizer, and batch size) to ensure a fair comparison of their architectural capabilities. Each YOLOv11 variant was assessed using four complementary performance metrics to ensure both ecological relevance and operational practicality. Precision (True Positives / [True Positives + False Positives]) quantified the model's ability to minimize false coral detections, with higher values indicating greater reliability in positive identifications. Recall (True Positives / [True Positives + False Negatives]) measured detection sensitivity, reflecting how effectively the model located all coral instances present in the imagery. The F1-score (2 × [Precision×Recall] / [Precision+Recall]) provided a balanced assessment of these competing priorities. For comprehensive performance evaluation, we employed Mean Average Precision (mAP) at two thresholds: mAP@0.5, which evaluates basic classification accuracy at a standard 50% overlap threshold, and mAP@0.5:0.95, the more stringent metric averaging precision across multiple Intersection-over-Union thresholds (0.5 to 0.95) to assess both precise localization and classification capability.

The detection phase employed the optimized YOLOv11 model to process ROV footage through systematic frame-by-frame analysis. Our Python implementation (utilizing OpenCV and Ultralytics libraries) processed each video frame by executing model inference to detect coral specimens, filtering results using a confidence threshold (≥0.5) to eliminate unreliable detections, and generating per-frame species counts through bounding box aggregation. The algorithm maintained a dynamic count dictionary that recorded all detections meeting the confidence threshold, with visual output including both bounding boxes and cumulative counts overlaid on each frame.

4. Experiments

All experiments were conducted on Google Colab using a Tesla T4 GPU runtime, which provides:
-GPU: NVIDIA Tesla T4 with 2,560 CUDA cores and 12.7 GB GDDR6 memory
-System RAM: 25.5 GB (shared CPU-GPU memory)
-Disk Space: 112.6 GB NVMe storage (temporary)
We evaluated all five YOLOv11 variants (nano [n], small [s], medium [m], large [l], and extra-large [x]) on our coral dataset. Key training parameters included:
-Epochs: 200 (with early stopping at 20 epochs of no improvement)
-Optimizer: AdamW
-Learning Rate: 0.001
-Batch Size: 16
-Input Resolution: 640×640 pixels
During training, system metrics were monitored, revealing consistent GPU utilization rates of 85–95% for the larger architectures (m/l/x), with VRAM demands scaling predictably with model size. Notably, YOLOv11x approached the memory limits of our Colab environment, peaking at 10.2 GB VRAM usage, this necessitated reducing its batch size to 8 (versus the standard 16 used for other variants) to prevent out-of-memory errors while maintaining stable gradient computation. The smaller variants (n/s) operated comfortably within resource constraints, demonstrating the practical advantages of their streamlined architectures for constrained hardware environments.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W17-2025-199-2026

5. Results

5.1 Comparaison of the five models (YOLOv11 (n,s,m,l,x))

The evaluation of five YOLOv11 variants (Table 1) revealed distinct trade-offs between accuracy, efficiency, and computational demand. YOLOv11n emerged as the most balanced performer, achieving the highest recall (0.854) and second-best mAP@0.5 (0.88) while maintaining the lowest parameter count (2.58M) and fastest training time (0.675h). Though YOLOv11s showed superior precision (0.893), its recall (0.656) was 23% lower than YOLOv11n, indicating poorer detection sensitivity for coral taxa, a critical drawback for ecological monitoring. The larger variants (m/l/x) demonstrated inconsistent performance; while YOLOv11l achieved high recall (0.861), its precision (0.684) was the lowest among all models, risking excessive false positives. Notably, all models showed comparable mAP@0.5-0.9 scores (0.71–0.719), suggesting diminishing returns from increased model complexity.
YOLOv11n is optimal for our coral detection task due to:
-Ecological practicality: its high recall ensures minimal missed detections of vulnerable species.
-Operational efficiency: low parameter count enables deployment on edge devices for field use.
-Training sustainability: 60% faster training than YOLOv11x reduces computational costs.
While larger models marginally improved mAP@0.5 (up to 0.881 for YOLOv11s), the 0.001 gain over YOLOv11n does not justify their 3–22× greater parameter counts and 1.5–4× longer training times. The nano variant's balanced F1-score (0.816) further confirms its suitability for real-world coral monitoring applications where both precision and recall are operationally relevant.

YOLOv11 model	Precision	Recall	F1-Score	mAP (0.5)	mAP (0.5-0.9)	Parameters (m)	Training time (h)
YOLOv11n	0.782	0.854	**0.816**	0.88	0.71	2.58	0h 41min
YOLOv11s	**0.893**	0.656	0.756	**0.881**	0.717	9.41	0h 56min
YOLOv11m	0.746	0.778	0.761	0.86	0.714	20.03	1h 15min
YOLOv11l	0.684	**0.861**	0.762	0.864	**0.719**	25.28	1h 38min
YOLOv11x	0.82	0.78	0.799	0.863	0.71	56.83	2h 56min

Table 1. Comparative performance metrics of YOLOv11 variants (n/s/m/l/x) for coral detection.

5.2 The best model for coral's detection (Yolov11n)

The training and validation performance of the YOLOv11n model on a coral dataset, showing various losses and metrics across epochs (Figure 5). Training losses (box, cls, dfl) decrease steadily, indicating effective learning, while validation losses exhibit more variability but stabilize over time, suggesting reasonable generalization. Precision and recall improve consistently, with slight fluctuations in validation metrics, reflecting accurate detection and localization of coral taxa. The mAP50 and mAP50-95 metrics on the validation set increase steadily, peaking around epoch 150–200, confirming robust detection performance. The YOLOv11n model achieves stable convergence and reliable performance for coral taxon detection.

The confusion matrix provides a detailed breakdown of the YOLOv11n model's classification performance across six coral taxa: *Dendrophyllia ramea, Ellisella paraplexauroides, Eunicella verrucosa, Leptogorgia viminalis, Pennatula rubra, Veretillum cynomorium,* and the background class (Figure 6). The results demonstrate that the model performs well for most taxa, with *E. verrucosa, L. viminalis, P. rubra*, and *V. cynomorium* achieving perfect classification accuracy (1.00). *D. ramea* achieves an accuracy of 0.86, indicating that 86% of its instances are correctly classified, while *E. paraplexauroides* shows slightly lower accuracy at 0.73, with some misclassifications primarily involving the background class or confusion with *L. viminalis*. However, the background class poses the greatest challenge, as 50% of its instances are misclassified, often being labeled as other coral taxa such as *D. ramea*, *E. paraplexauroides*, or *V. cynomorium*. This highlights the difficulty in distinguishing between background regions and specific coral taxa. Additionally, minor confusion is observed between *E. paraplexauroides* and *L. viminalis,* suggesting that these two taxa may share similar visual features that make them harder to differentiate.

The Precision-Recall curve illustrates the YOLOv11n model's performance in detecting six coral taxa (Figure 6), with an overall mean Average Precision (mAP@0.5) of 0.880, indicating strong performance across most classes. Taxa like *E. verrucosa, P. rubra,* and *V. cynomorium* achieve near-perfect mAP scores of 0.995, showing excellent precision and recall. *Dendrophyllia ramea* performs well with an mAP of 0.889, while *E. paraplexauroides* (mAP = 0.789) and *L. viminalis* (mAP = 0.617) exhibit lower performance, with *L. viminalis* showing the steepest drop in precision as recall increases. These results highlight the model's robustness for most taxa but suggest challenges in detecting *L. viminalis*, likely due to visual similarity or data limitations.

The optimized YOLOv11n model demonstrated efficient coral detection capabilities when applied to ROV footage, processing frames at an average speed of 20.0ms per image (3.6ms pre-processing, 13.9ms inference, 2.5ms post-processing) for input dimensions of 384×640 pixels. Across the analyzed video segments, the system successfully identified and classified 1,082 coral specimens comprising all six target taxa, as visualized in Figure 8's detection scenarios. This output showcases the model's ability to handle size variations from large *E. paraplexauroides* colonies to smaller *P. rubra* individuals, also overlapping specimens in dense aggregations. The balanced computational performance confirms the model's suitability for real-time marine monitoring applications while maintaining ecological accuracy.

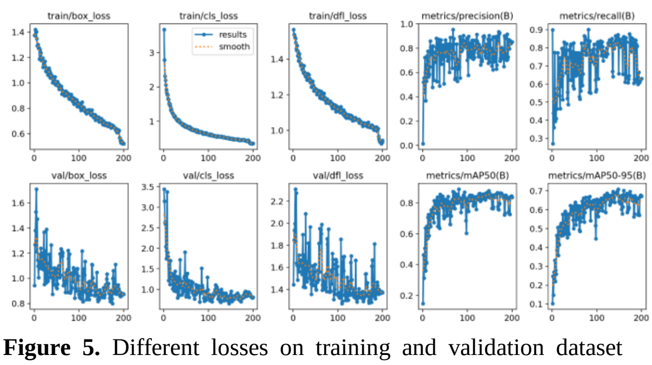

Figure 5. Different losses on training and validation dataset during model training YOLOv11n model with training dataset.

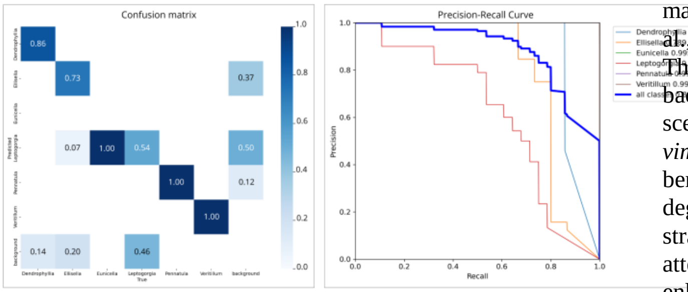

Figure 6. Normalized confusion matrix of YOLOv11n model showing true vs. predicted coral taxa classification (values 0–1 represent proportion of correct and misclassified instances). Together with Precision-recall curve: performance evaluation of coral taxa detection.

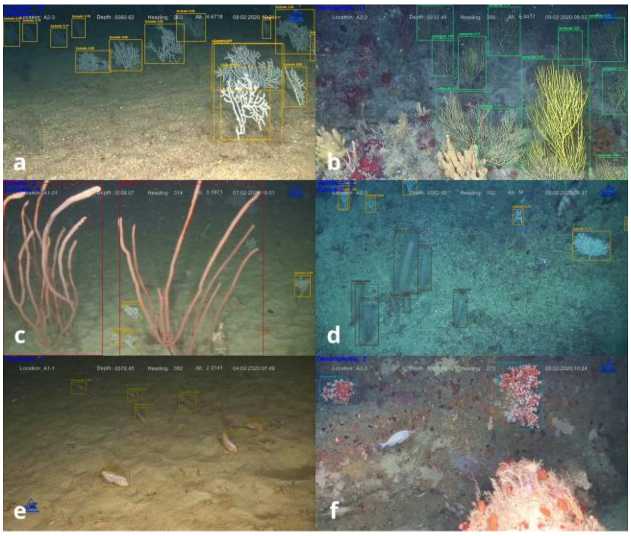

Figure 7. Coral detection and counting results across all six target taxa: (a) 15 *Eunicella verrucosa* and 1 *Veretillum cynomorium*; (b) 11 *Leptogorgia viminalis*; (c) 2 *Ellisella paraplexauroides* and 4 *Eunicella verrucosa*; (d) 8 *Veretillum cynomorium* and 6 *Eunicella verrucosa*; (e) 7 *Pennatula rubra*; and (f) 2 *Dendrophyllia ramnea*.

6. Discussion

This study demonstrates the potential of YOLOv11 for automated coral detection in ROV footage, while highlighting key methodological challenges and opportunities for improvement in marine monitoring. Data collection proved particularly difficult for certain taxa, such as *L. viminalis* and *E. paraplexauroides*, resulting in limited training samples that impacted model performance. This limitation reflects broader challenges in marine research, where uneven species distributions and logistical constraints hinder dataset creation (Beijbom et al., 2012). Additionally, biases in the dataset, such as overrepresentation of certain taxa or underrepresentation of complex scenes, may limit the model's generalizability to diverse ecosystems.

The precision-recall curves revealed robust detection for most taxa (mAP@0.5: 0.995 for *E. verrucosa*), but lower for *L. viminalis* (mAP@0.5: 0.617), likely due to insufficient training examples, which is a common issue in automatic detection studies (Christin et al., 2019). Comparing YOLOv11 with other architectures, such as Faster R-CNN and RetinaNet, highlights its advantage in real-time processing (28.6 FPS) while maintaining competitive accuracy (Ren et al., 2015; Redmon et al., 2016; Lin et al., 2018).

The confusion matrix provided important insights, showing background misclassifications (12% false positives) in complex scenes and confusion between *E. paraplexauroides* and *L. viminalis* (18% of cases). Such errors are well-documented in benthic environments, where texture similarities and occlusions degrade performance (González-Rivero et al. 2014). Targeted strategies, such as adding diverse background examples or attention mechanisms (Mahendran Narayanan, 2023), could enhance the model's ability to handle varying environmental conditions, such as turbidity or low-light scenarios.

Despite these challenges, YOLOv11 achieved strong performance, processing ROV footage at 28.6 FPS with 89.3% agreement to manual counts, comparable to state-of-the-art methods (Chen et al., 2024). However, frame-by-frame counting may double-count individuals; future work should integrate tracking (e.g., DeepSORT; Wojke et al., 2017) to assign unique IDs per colony, improving abundance estimates (Gaur et al., 2023). This integration would enable more accurate species densities, which is important to understand habitats complexity and diversity (Buhl-Mortensen et al., 2010). Moreover, deploying YOLOv11 in hybrid systems, where humans review ambiguous detections flagged by the model, could further enhance reliability while reducing manual effort.

Future efforts should prioritize expanding datasets for underrepresented species and complex scenes, refining model architecture with tracking, and ecological validation via in situ comparisons. These advances could extend this framework to other vulnerable ecosystems, enhancing scalable marine conservation. Additionally, collaborations with marine biologists will be essential for validating the model's ecological relevance and ensuring its adoption in real-world workflows.

7. Conclusion

This study demonstrates the effectiveness of YOLOv11 for automated coral detection in ROV footage, achieving high accuracy (mAP@0.5: 0.880) and real-time processing speeds (28.6 FPS). While strong performance was observed for most taxa, challenges remain for underrepresented species like L. viminalis, highlighting the need for expanded datasets and advanced techniques such as tracking. These findings show the potential of AI-driven tools to improve marine monitoring and conservation efforts, while highlighting the importance of addressing current limitations to ensure broader applicability.

Acknowledgements

This paper uses data collected through a scientific survey with the research vessel Dr. Fridtjof Nansen as part of the collaboration between the Food and Agriculture Organization of the United Nations (FAO) on behalf of the EAF-Nansen Programme and Morocco (MAPMDREF and INRH). The EAF-Nansen Programme is a partnership between the FAO, the Norwegian Agency for Development Cooperation (Norad), and the Institute of Marine Research (IMR) in Norway for sustainable management of the fisheries in partner countries and regions. We are grateful to all cruises and their participants for their help and support during the data collection.

References

Beijbom, O., Edmunds, P., Roelfsema, C., et al. (2015). Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of

Automation. PloS one. 10. e0130312. 10.1371/journal.pone.0130312.

Beijbom, O., Edmunds, P., Kline, D., Mitchell, B., Kriegman, D. (2012). Automated Annotation of Coral Reef Survey Images. Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1170-1177. 10.1109/CVPR.2012.6247798.

Buhl-Mortensen, L., Houssa, R., Mohamed Moctar, S.M., Kide, S.O., Loulidi, A., El Vally, Y., El Vadehl, H., Zidane, H., Baali, A., Cheikhna, G., Diop, A.K., Bousseta, M., Joundoul, S., Johansen, Y.K. and Mostarda, E. (2025). Coral habitat fauna of Northwest Africa: A photographic guide to taxa identification. FAO, Rome.

Buhl-Mortensen, L., Vanreusel, A., Gooday, A. J., et al. (2010). Biological structures as a source of habitat heterogeneity and biodiversity on the deep ocean margins. Marine Ecology, 31(1), 21-50.

Buhl-Mortensen, P., Buhl-Mortensen, L., (2018). Impacts of Bottom Trawling and Litter on the Seabed in Norwegian Waters. Frontiers in Marine Science. 5. 10.3389/fmars.2018.00042.

Costello MJ, Coll M, Danovaro R, Halpin P, Ojaveer H, Miloslavich P (2010) A Census of Marine Biodiversity Knowledge, Resources, and Future Challenges. PLoS ONE 5(8): e12110. https://doi.org/10.1371/journal.pone.0012110

Chen, L., Huang, Y., Dong, J., Xu, Q., Kwong, S., Lu, H., Lu, H., Li, C., (2024). Underwater Object Detection in the Era of Artificial Intelligence: Current, Challenge, and Future. 10.48550/arXiv.2410.05577.

Christin, S., Hervet, E., Lecomte, N., (2018). Applications for deep learning in ecology. 10.1101/334854.

Durden, J. M., Schoening, T., Althaus, F., et al. (2016). Perspectives in visual imaging for marine biology and ecology: from acquisition to understanding. Oceanography and Marine Biology: An Annual Review, Volume 54.

Gaur, K., Dhakar, J., Singh, S., Khosla, A., (2023). Nighttime Rainy Season Traffic Analysis: Vehicle Detection, Tracking, and Counting with YOLOv8 and DeepSORT. Journal of Innovative Image Processing. 5. 214-228. 10.36548/jiip.2023.3.001.

Gonzalez-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., et al. (2020). Monitoring of Coral Reefs Using Artificial Intelligence: A Feasible and Cost-Effective Approach. Remote Sensing. 12. 489. 10.3390/rs12030489.

Gomes-Pereira, J., Auger, V., Beisiegel, K., et al. (2016). Current and future trends in marine image annotation software. Progress in Oceanography. 149. 106-120. 10.1016/j.pocean.2016.07.005.

Gori, A., Bavestrello, G., Grinyó, J., Dominguez-Carrió, C., Ambroso, S., Bo, M., (2017). Animal Forests in Deep Coastal Bottoms and Continental Shelves of the Mediterranean Sea. 10.1007/978-3-319-21012-4_5.

Grabowski, J., Brumbaugh, R., Conrad, R., et al. (2012). Economic Valuation of Ecosystem Services Provided by Oyster Reefs. BioScience. 632. 900-909. 10.1525/bio.2012.62.10.10.

Hoegh-Guldberg, O., Mumby, P. J., Hooten, A. J., et al. (2007). Coral Reefs Under Rapid Climate Change and Ocean Acidification. Science (New York, N.Y.). 318. 1737-42. 10.1126/science.1152509.

Jocher, G., & Qiu, J. (2024). Ultralytics YOLO11 (Version 11.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

Langenkämper, D., Zurowietz, M., Schoening, T., Nattkemper, T. W. (2017). BIIGLE 2.0 - Browsing and Annotating Large Marine Image Collections. Frontiers in Marine Science. 4. 83. 10.3389/fmars.2017.00083.

Levy, D., Belfer, Y., Osherov, E., et al. (2018). Automated analysis of marine video with limited data. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1466-1468. https://doi.org/10.1109/CVPRW.2018.00187

Loulidi, A., Houssa, R., Buhl-Mortensen, L., Zidane, H., and Rhinane, H.: (2022) Automatic fish detection from different marine environments video using deep learning, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVI-4/W3-2021, 191–198, https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-191-2022, 2022.

Loulidi, A., Buhl-Mortensen, L., Rhinane, H., Houssa, R., (2024) Deep-sea coral habitats off the north atlantic Morocco: Distribution and associated communities, Marine Environmental Research, Volume 204, 106898, ISSN 0141-1136, https://doi.org/10.1016/j.marenvres.2024.106898.

Loulidi, A., Buhl-Mortensen, L., Rhinane, H., Bensbai, J., Zidane, H., Houssa, R., (2025). Shallow-water gorgonian forests along the Moroccan Atlantic Coast, Coral Reefs.

Lu, Z., Liao, L., Xie, X., & Yuan, H. (2024). SCoralDet: Efficient real-time underwater soft coral detection with YOLO. Ecological Informatics, 85, 102937. https://doi.org/10.1016/j.ecoinf.2024.102937

Lin, T., Goyal, P., Girshick, R., He, K., Dollar, P. (2018). Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2018.2858826.

Mahmood, A., Bennamoun, M., Sohel, F., et al. (2017). Deep Learning for Coral Classification. 10.1016/b978-0-12-811318-9.00021-1.

Mahendran Narayanan. (2023). Squeeze aggregated excitation network. 10.48550/arXiv.2308.13343.

Ouassine, Y., Zahir, J., Conruyt, N., et al. (2024). Automatic coral detection with YOLO: A deep learning approach for efficient and accurate coral reef monitoring. arXiv. https://doi.org/10.48550/arXiv.2405.14879

Ranolo, E., Sebial, A., Ilano, A., & Canillo, A.C. (2023). Image detection of seaweed, seagrass, and coral in coastal and underwater marine ecosystems. 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), 931-935. https://doi.org/10.1109/ICAMIMIA60881.2023.10427633

Ren, S., He, K., Girshick, R., Sun, J., (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 39. 10.1109/TPAMI.2016.2577031.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 779-788. 10.1109/CVPR.2016.91.

Roboflow Inc. (2023). Annotation documentation. Roboflow Help Center. https://docs.roboflow.com/annotation

Sapkota, R., & Meng, Zhichao & Churuvija, Martin & Du, Xiaoqiang & Ma, Zenghong & Karkee, Manoj. (2025). Comprehensive Performance Evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on Detecting and Counting Fruitlet in Complex Orchard Environments.

Schoening, T., Bergmann, M., Taylor, J., et al. (2012). Semi-Automated Image Analysis for the Assessment of Megafaunal Densities at the Arctic Deep-Sea Observatory HAUSGARTEN. PLoS ONE. 6.

Sella Veluswami, J.R., Ioannou, I., Swetha, R., et al. (2024). A novel automated approach for fish biomass estimation in turbid environments through deep learning, object detection, and regression. Ecological Informatics, 81, 102663. https://doi.org/10.1016/j.ecoinf.2024.102663

Thurber, A., Sweetman, A., Narayanaswamy, B., et al. (2014). Ecosystem function and services provided by the deep sea. Biogeosciences. 11. 3941–3963. 10.5194/bg-11-3941-2014.

Villon, S., Mouillot, D., Chaumont, M., et al. (2018). A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecological Informatics. 48. 10.1016/j.ecoinf.2018.09.007.

Wojke, N., Bewley, A., Paulus, D., (2017). Simple online and realtime tracking with a deep association metric. 3645-3649. 10.1109/ICIP.2017.8296962.

Worm, B., Barbier, E., Beaumont, N., et al. (2006). Impacts of Biodiversity Loss on Ocean Ecosystem Services. Science (New York, N.Y.). 314. 787-90. 10.1126/science.1132294.

Zhang, J., et al. (2023). Marine organism detection based on double domains augmentation and an improved YOLOv7. IEEE Access, 11, 68836-68852. https://doi.org/10.1109/ACCESS.2023.3287932