
Newfoundland Marine Refuge Fish Classification Dataset (*N-MARINE*)

Devi Ayyagari
Dalhousie University
Halifax, NS 15213
devi.ayyagari@dal.ca

Maurice Drautz
Dalhousie University
Halifax, NS 15213
ny538703@dal.ca

Daniel Porter *
St. John's, Canada
porterdr5@gmail.com

Joshua Barnes
National Research Council Canada
St. John's, Newfoundland joshua.barnes@nrc-cnrc.gc.ca

Corey Morris
Fisheries and Oceans Canada
St. John's, Canada
corey.morris@dfo-mpo.gc.ca

Christopher Whidden
Dalhousie University
Halifax, NS 15213
cwhidden@dal.ca

Abstract

Scaling marine ecosystem monitoring is increasingly urgent as ocean warming accelerates ecological change. Current techniques like trawling are invasive, harmful to habitats, and infrequent, leading to missed shifts in species distributions such as the snow crab collapse in the Bering Sea after the 2018–2019 marine heatwave and the spread of invasive European green crab in Atlantic Canada. Machine learning offers a pathway to scalable, automated monitoring, but progress in regions such as the North Atlantic has been constrained by the scarcity of annotated data. We present N-MARINE, the first open-source North Atlantic underwater image dataset for groundfish detection and classification. It contains 23,936 ecologist-annotated ocean-floor camera images across nine species, with bounding boxes, standardized splits, and YOLOv7 baselines (best mAP@0.5 = 0.808). N-MARINE provides a foundation to build upon and advance research on generalizable visual representations and transfer learning for regional underwater ecosystems.

1 Introduction

The oceans cover 71% of Earth's surface, function as a major carbon sink, and support the welfare and livelihoods of millions in coastal communities [1, 2]. In 2018, about 38% of the global population lived within 100 km of the coast [3]. The ocean also sustains substantial economic activity: international ocean-economy trade amounts to trillions of U.S. dollars, and the export value of fish and fishery products alone reached \$156 billion in 2021 [4]. The ocean is warming rapidly; in 2024, global sea-surface temperatures and ocean heat content reached record highs, intensifying ongoing climate trends [5]. Consistent with warming, many marine species are shifting their geographic ranges, with leading edges generally moving poleward and/or deeper [6]. Recent events underscore the ecological and socioeconomic stakes—for example, the abrupt collapse of eastern Bering Sea snow crab to historical lows by 2021 [7]. Projections indicate that—regardless of emissions scenario—at least 37% and 54% of straddling fish stocks will shift between Exclusive Economic Zones (EEZs) and

*Work conducted while employed at Fisheries and Oceans Canada (DFO); now retired.

Species	Train inst.	Train vids	Test inst.	Test vids	Total inst.
<i>Gadus morhua</i> (Atlantic Cod)	3,448	98	725	23	4,173
<i>Macrourus berglax</i> (Roughhead Grenadier)	3,308	42	609	8	3,917
<i>Amblyraja radiata</i> (Thorny Skate)	2,701	27	583	6	3,284
<i>Sebastes mentella</i> (Redfish Mentella)	3,186	27	174	3	3,360
<i>Anarhichas lupus</i> (Striped Wolffish)	2,658	14	374	3	3,032
<i>Hippoglossus hippoglossus</i> (Atlantic Halibut)	3,329	13	149	1	3,478
<i>Bathyraja spinicauda</i> (Spinytail Skate)	930	4	58	2	988
<i>Buccinum undatum</i> (Whelk)	869	1	0	0	869
<i>Anarhichas denticulatus</i> (Northern Wolffish)	789	1	0	0	789
Negative (background) images	2,558	151	442	26	3,000
No. of sampled videos	157		28		185

Table 1: Class distribution of object instances and sampled videos across training and test splits.

the high seas by 2030 and 2050, respectively [8]. These trends heighten the urgency of systematic, scalable monitoring of marine ecosystems [9].

Traditionally, marine ecosystem monitoring relies on deck-based visual surveys, trawls, gill nets, and animal tagging—methods that can be lethal for some organisms and, in the case of bottom trawls, disturb benthic habitats [10]. Over the past decade, non-invasive, sensor-based approaches have expanded: active acoustics (scientific echosounders) to estimate fish abundance/biomass; passive acoustics (soundscape recordings) [11] to monitor vocal taxa and habitat use; environmental DNA (eDNA) [12] to detect species—including rare/cryptic and invasive taxa—and characterize taxonomic composition; and optical platforms (BRUVS and towed or stationary cameras) to acquire underwater video and still images [13, 14]. These sensors now generate vast, diverse datasets across habitats and time, yet existing workflows cannot scale quickly enough to measure and interpret rapid ecological change. Advances in computing power have made complex machine-learning analyses widely accessible, offering the potential to automate detection, classification, and abundance estimation [15, 16], but progress is constrained by the scarcity of well-annotated training and benchmark datasets [17].

Existing open datasets for underwater detection and classification—e.g., FathomNet [18], DeepFish [19], and OzFish [13]—primarily cover other regions, habitats, and species; to our knowledge, none focus on North Atlantic waters. N-MARINE addresses this gap by providing ecologist-annotated imagery of groundfish from a region that is uniquely important: it is currently the only province in Canada with an active offshore oil and gas industry [20]. It is designed not only as a standalone dataset but also to complement existing corpora—supporting transfer learning, enabling systematic evaluation of how models trained in one region generalize to others, and facilitating integration into larger training pipelines. By releasing images, labels, standardized splits, and baseline models, our goal is to lower adoption barriers and provide a resource that others can build upon, advancing scalable marine ecosystem monitoring and the development of more generalizable visual representations.²

2 Dataset description

N-MARINE is a marine-taxa detection and classification dataset collected in 2021 using underwater video cameras mounted on crab pots equipped with lights and bait, deployed at depths of 400–500 m within the Small portion of the Northeast Newfoundland Slope Closure. Further details on survey design and acquisition procedures are provided in [21]. The dataset contains 23,936 images (1920 × 1080 px; 1 MB per image; 30 GB total), each annotated with bounding boxes and species labels. Nine marine species—annotated at the species rank to facilitate future hierarchical learning—are documented, and images may contain multiple individuals per frame.

N-MARINE is a curated subset of 185 videos drawn from a larger corpus of 221 videos originally collected to study the impacts of seismic oil and gas exploration surveys. Each video is approximately five minutes long, recorded at 30 fps. From each video, 1,500 frames were annotated, but only a subset of these annotated frames is included in the open-source release, resulting in a total of 23,936 images in the dataset. All groundfish, including partially occluded and truncated individuals, were labeled by a Fisheries and Oceans Canada (DFO) ecologist in DIVE [22].

²Models are uploaded on https://huggingface.co/WhiddenLab/N-MARINE_baseline_classifiers, the dataset is available at <https://open.canada.ca/data/en/dataset/2ae46860-f82a-4127-bb1f-b02e36ef6a70>, and supplementary training material is uploaded on https://github.com/Pentaerythritetranitrat/N-MARINE_dataset_supplementary.

YOLOv7 Model	Precision	Recall	mAP@0.5	mAP@[0.5:0.95]
With class weights	0.783 ± 0.0223	0.757 ± 0.030	0.784 ± 0.025	0.464 ± 0.020
Without class weights	0.807 ± 0.036	0.764 ± 0.014	0.808 ± 0.007	0.494 ± 0.008

Table 2: Baseline YOLOv7 performance on N-MARINE (mean $\pm$ std. across cross-validation folds).

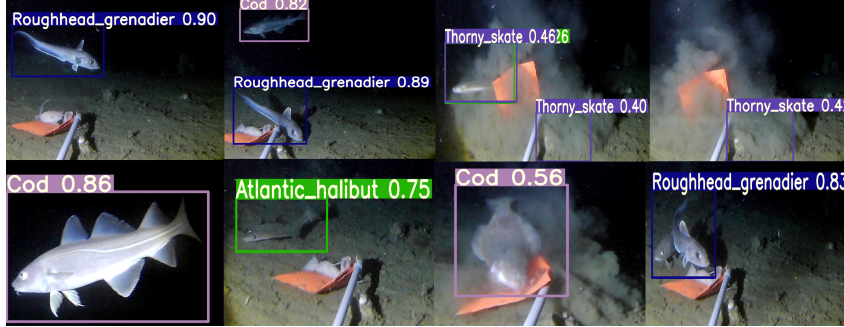


Figure 1: Example detections from the best YOLOv7 baseline on N-MARINE: columns 1–2 show correct detections; columns 3–4 show typical errors (misses and misclassifications)

We organize images by source video with a video identifier and frame index in both filenames and label files, allowing identification of the original video for each image. Up to 3,000 positive images were randomly sampled per species, with all available images used when fewer than 3,000 examples were available. Additionally, 3,000 negative (background) images were randomly selected from the 185 videos. We provide annotations in YOLO-style per-video CSV files, containing both absolute pixel coordinates (top-left and bottom-right) for visualization and normalized coordinates for training.

3 Baseline object detector

We trained a baseline object detector using the YOLOv7[23] framework to characterize model performance under fully supervised training conditions on the N-MARINE dataset and to perform detection and classification of nine marine species in the dataset. To train this detector, the dataset was first split into training and test sets at the *video* level to prevent leakage between splits. Videos were randomly assigned in an 85/15 ratio while preserving class balance where possible. In rare cases where a species appeared in only one video—specifically, *Buccinum undatum*(Whelk) and *Anarhichas denticulatus*(Northern Wolffish)—if that video was initially assigned to the test set, it was reassigned to the training set so the model would encounter that species during learning. Per-species counts by split are reported in Table 1.

The test split was fixed at 15% of videos, and within the remaining training videos, five stratified, video-level folds were defined for statistical analysis of the baseline. In each fold, the training portion was further partitioned into an 85/15 train/validation split (still at the video level to avoid leakage). Final results are reported on the fixed test set using the model with the highest validation mAP in its corresponding fold (the training curves are attached in the Appendix), and the species distributions and exact file lists for all folds and splits will be published in the repository. We observed notable class imbalance in the dataset and investigated whether applying class weights could improve performance. For class c with n_c positives out of N training examples, we set $\text{pos_weight}_c = (N - n_c)/n_c$. Training details for the object detector are in the Appendix. We include the configuration files and command lines for full reproducibility.

4 Discussion

The YOLOv7 baseline without class weighting achieved the highest overall performance, with $\text{mAP}@0.5 = 0.808 \pm 0.0067$ and $\text{mAP}@[0.5:0.95] = 0.494 \pm 0.0076$, alongside a precision of 0.8066 ± 0.0360 and recall of 0.7640 ± 0.0140 (Table 2). Introducing inverse-frequency class weights slightly improved mAP for Spinytail skate (increasing the median $\text{mAP}@0.5$ to 0.61 from 0.52) - a rare species but reduced aggregate mAP to $\text{mAP}@0.5 = 0.784 \pm 0.0245$ and $\text{mAP}@[0.5:0.95] =$

0.4644 ± 0.0203 . This suggests that, rather than relying on naive re-weighting, model training could benefit from more targeted imbalance-handling strategies such as focal loss, targeted data augmentation for rare classes, class-aware sampling, per-class confidence thresholds, or increasing the diversity of training data for under-represented species.

Following the aggregate results, we examined per-class performance to assess how species representation influenced detection. Class-wise results (Figure 2) show that Atlantic halibut (median $\text{mAP}@0.5 \approx 0.141$) performs much worse than all other classes. The confusion matrix for the model with the best test mAP (see in Appendix) shows an extremely high background false-negative rate for halibut (0.99), indicating frequent misses. These omissions are likely driven by visibility limitations—poor contrast, partial occlusion, turbidity—with all halibut test examples coming from a single video where sediment clouds obscured body outlines. Figure 1 illustrates common failure modes—partial occlusions, turbidity, overlapping individuals—seen across classes. Overall, poor performance is more strongly linked to missed detections under difficult visibility than to inter-class confusion (as can be observed in the confusion matrix in the Appendix), and abundant species risk over-prediction in background regions. Improving rare-class performance will require novel algorithms and expanded data collection to ensure adequate representation in both training and evaluation splits.

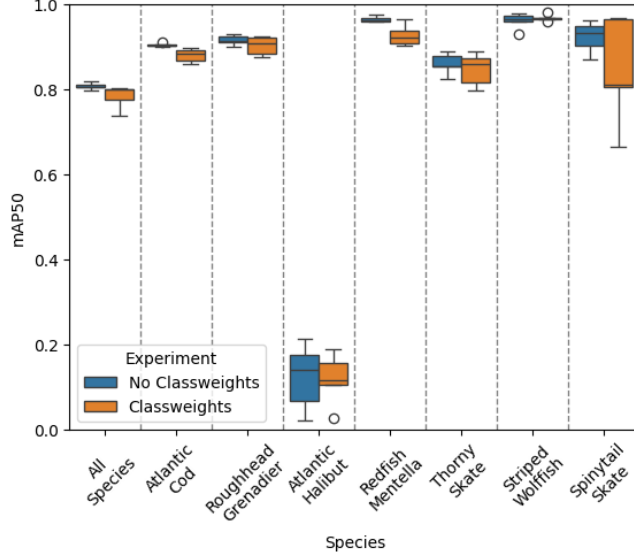


Figure 2: Per-class $\text{mAP}@0.5$ on the fixed test split across five folds for YOLOv7 baselines

These findings underscore a broader challenge: detector performance is highly sensitive to acquisition conditions. Small changes in camera placement, lighting, turbidity, habitat background, or local species composition can yield error patterns similar to those observed for halibut. Consequently, models that work well in one deployment often require adaptation—or new annotations—before being applied elsewhere. In practice, the bottleneck has shifted from data collection and frame-by-frame counting to obtaining sufficiently high-quality labels for each new dataset. Large observatory networks such as Ocean Networks Canada record thousands of hours of seafloor video, yet much of it remains underutilized due to limited annotation capacity. N-MARINE provides a reproducible baseline, but sustainable monitoring will require workflows that combine selective expert labeling, domain adaptation, and active learning to handle the diversity of real-world conditions.

This need is particularly acute in the Northwest Atlantic, including the Northeast Newfoundland Slope where these data were collected. This shelf system has warmed rapidly in recent decades and is among the fastest-warming ocean regions globally [24, 25, 26], with documented impacts on marine habitats and fisheries. Shifts in thermal habitats are already driving distributional changes in key species such as gadids and Atlantic Halibut [27, 28], underscoring the urgency for consistent, high-resolution biological monitoring [29]. While DFO has begun standardizing environmental tracking through initiatives like the Newfoundland and Labrador Climate Index [24], complementary species-level image datasets remain scarce. More broadly, marine ecosystems worldwide face climate-driven shifts in distribution, timing, and abundance, yet many regions still lack sustained, high-resolution biodiversity records. To address this gap, we release the curated N-MARINE dataset, with the full set of 221 annotated videos available upon request for specialized research needs. By making this data accessible, we aim to close critical knowledge gaps, foster collaboration between computer vision engineers and ecologists, and support policy development to mitigate the impacts of a rapidly changing ocean.

Acknowledgments and Disclosure of Funding

Use unnumbered first-level headings for the acknowledgments. All acknowledgments go at the end of the paper before the list of references. Moreover, you are required to declare funding (financial activities supporting the submitted work) and competing interests (related financial activities outside the submitted work). More information about this disclosure can be found at: <https://neurips.cc/Conferences/2020/PaperInformation/FundingDisclosure>.

Do **not** include this section in the anonymized submission, only in the final paper. You can use the ack environment provided in the style file to automatically hide this section in the anonymized submission.

References

- [1] Martin Visbeck. Ocean science research is key for a sustainable future. *Nature Communications*, 9(1):690, February 2018. Publisher: Nature Publishing Group.
- [2] Pierre Friedlingstein, Michael O’Sullivan, Matthew W. Jones, Robbie M. Andrew, Dorothee C. E. Bakker, Judith Hauck, Peter Landschützer, Corinne Le Quéré, Ingrid T. Luijkx, Glen P. Peters, et al. Global carbon budget 2023. *Earth System Science Data*, 15:5301–5369, 2023.
- [3] Arthur G. Cosby, Viswadeep Lebakula, C. N. Smith, David W. Wanik, K. Bergene, A. N. Rose, David A. Swanson, and David E. Bloom. Accelerating growth of human coastal populations at the global and continent levels: 2000–2018. *Scientific Reports*, 14:22489, 2024.
- [4] Zhaoyuan Shi and Desheng Xue. Measuring the international ocean economy trade: Method and application. *Marine Policy*, 175:106636, 2025.
- [5] Lijing Cheng, John Abraham, Kevin E. Trenberth, James Reagan, Huai-Min Zhang, Andrea Storto, Karina Von Schuckmann, Yuying Pan, Yujing Zhu, Michael E. Mann, Jiang Zhu, Fan Wang, Fujiang Yu, Ricardo Locarnini, John Fasullo, Boyin Huang, Garrett Graham, Xungang Yin, Viktor Gouretski, Fei Zheng, Yuanlong Li, Bin Zhang, Liying Wan, Xingrong Chen, Dakui Wang, Licheng Feng, Xiangzhou Song, Yulong Liu, Franco Reseghetti, Simona Simoncelli, Gengxin Chen, Rongwang Zhang, Alexey Mishonov, Zhetao Tan, Wangxu Wei, Huifeng Yuan, Guancheng Li, Qiuping Ren, Lijuan Cao, Yayang Lu, Juan Du, Kewei Lyu, Albertus Sulaiman, Michael Mayer, Huizan Wang, Zhanhong Ma, Senliang Bao, Henqian Yan, Zenghong Liu, Chunxue Yang, Xu Liu, Zeke Hausfather, Tanguy Szekely, and Flora Gues. Record High Temperatures in the Ocean in 2024. *Advances in Atmospheric Sciences*, 42(6):1092–1109, June 2025.
- [6] Malin L Pinsky, Rebecca L Selden, and Zoë J Kitchel. Climate-driven shifts in marine species ranges: Scaling from organisms to communities. *Annual review of marine science*, 12:153—179, January 2020.
- [7] Cody S. Suzwalski, Kerim Aydin, Erin J. Fedewa, Brian Garber-Yonts, and Michael A. Litzow. The collapse of eastern bering sea snow crab. *Science*, 382(6668):306–310, 2023.
- [8] Juliano Palacios-Abrantes, William W. L. Cheung, M. A. Oyinlola, Gabriel Reygondeau, U. Rashid Sumaila, et al. Climate change drives shifts in straddling fish stocks in the world’s ocean. *Science Advances*, 11(31):eadq5976, 2025.
- [9] Nathaniel L. Bindoff, William W. L. Cheung, James G. Kairo, Javier Arístegui, Valeria A. Guinder, Robert Hallberg, Nathalie Hilmi, Nianzhi Jiao, Md Saiful Karim, Lisa A. Levin, Sean O’Donoghue, Sara R. Purca Cuicapusa, Baruch Rinkevich, Toshio Suga, Alessandro Tagliabue, and Phillip Williamson. Changing ocean, marine ecosystems, and dependent communities. In Hans-Otto Pörtner, Debra C. Roberts, Valérie Masson-Delmotte, Panmao Zhai, Melinda Tignor, Elvira S. Poloczanska, Katja Mintenbeck, Andrés Alegria, Maïke Nicolai, Andrew Okem, Jan Petzold, Bard Rama, and Nora M. Weyer, editors, *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change*, pages 447–588. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.
- [10] Natalya D Gallo, Noelle M Bowlin, Andrew R Thompson, Erin V Satterthwaite, Briana Brady, and Brice X Semmens. Fisheries surveys are essential ocean observing programs in a time of global change: a synthesis of oceanographic and ecological data from us west coast fisheries surveys. *Frontiers in Marine Science*, 9:757124, 2022.
- [11] Anas Yassir, Said Jai Andaloussi, Ouail Ouchetto, Kamal Mamza, and Mansour Serghini. Acoustic fish species identification using deep learning and machine learning algorithms: A systematic review. *Fisheries Research*, 266:106790, 2023.

- [12] Fan Xiong, Lu Shu, Honghui Zeng, Xiaoni Gan, Shunping He, and Zuogang Peng. Methodology for fish biodiversity monitoring with environmental dna metabarcoding: The primers, databases and bioinformatic pipelines. *Water Biology and Security*, 1(1):100007, 2022.
- [13] Australian Institute of Marine Science (AIMS), University of Western Australia (UWA), and Curtin University. Ozfish dataset - machine learning dataset for baited remote underwater video stations. Available at: <https://doi.org/10.25845/5e28f062c5097>, 2019. Accessed February 28, 2025.
- [14] Alzayat Saleh, Marcus Sheaves, and Mostafa Rahimi Azghadi. Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish and Fisheries*, 23(4):977–999, 2022.
- [15] Daniel Marrable, Kathryn Barker, Sawitchaya Tippaya, Mathew Wyatt, Scott Bainbridge, Marcus Stowar, and Jason Larke. Accelerating species recognition and labelling of fish from underwater video with machine-assisted deep learning. *Frontiers in Marine Science*, 9:944582, 2022.
- [16] Hasna Maudi Lathifah, Ledy Novamizanti, and Syamsul Rizal. Fast and accurate fish classification from underwater video using you only look once. *IOP Conference Series: Materials Science and Engineering*, 982(1):012003, dec 2020.
- [17] Devi Ayyagari, Talukder Wasi Alavi, Navlika Singh, Joshua Barnes, Corey Morris, and Christopher Whidden. Dataset selection is critical for effective pre-training of fish detection models for underwater video. *ICES Journal of Marine Science*, 82(4):fsaf039, 04 2025.
- [18] Kakani Katija, Eric Orenstein, Brian Schlining, Lonny Lundsten, Kevin Barnard, Giovanna Sainz, Oceane Boulais, Megan Cromwell, Erin Butler, Benjamin Woodward, and Katherine L. C. Bell. Fathomnet: A global image database for enabling artificial intelligence in the ocean. *Scientific Reports*, 12(1):15914, 2022.
- [19] Nahuel Garcia-d’Urso, Alejandro Galan-Cuenca, Paula Pérez-Sánchez, Pau Climent-Pérez, Andres Fuster-Guillo, Jorge Azorin-Lopez, Marcelo Saval-Calvo, Juan Eduardo Guillén-Nieto, and Gabriel Soler-Capdepón. The deepfish computer vision dataset for fish instance segmentation, classification, and size estimation. *Scientific Data*, 9(1):287, 2022.
- [20] Canada Energy Regulator. Market snapshot: Atlantic canada offshore oil and natural gas production. <https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/market-snapshots/2017/market-snapshot-25-years-atlantic-canada-offshore-oil-natural-gas-production.html>, 2017. Accessed: 2025-08-18.
- [21] Corey Morris, Joshua Barnes, Dustin Schornagel, Christopher Whidden, and Phillippe Lamontagne. Measuring effects of seismic surveying on groundfish resources off the coast of newfoundland, canada. *Journal of Ocean Technology*, 16(3):57–63.
- [22] Matthew Dawkins, Linus Sherrill, Keith Fieldhouse, Anthony Hoogs, Benjamin Richards, David Zhang, Lakshman Prasad, Kresimir Williams, Nathan Lauffenburger, and Gaoang Wang. An open-source platform for underwater image and video analytics. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 898–906, 2017.
- [23] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7464–7475, 2023.
- [24] Frédéric Cyr and Peter S. Galbraith. A climate index for the newfoundland and labrador shelf. *Earth System Science Data*, 13:1807–1828, 2021.
- [25] Andrew J. Pershing, Michael A. Alexander, Carlos M. Hernandez, Lisa A. Kerr, Arnaud Le Bris, Katherine E. Mills, Janet A. Nye, Nicholas R. Record, Hillary A. Scannell, James D. Scott, Graham D. Sherwood, and Andrew C. Thomas. Slow adaptation in the face of rapid warming leads to collapse of the gulf of maine cod fishery. *Science*, 350(6262):809–812, 2015.
- [26] Dan Seidov, Andrey Mishonov, and Robert Parsons. Recent warming and decadal variability of gulf of maine and slope water. *Limnology and Oceanography*, 2021.
- [27] David Côté, Matthew Roberts, Matthew J. S. Windle, Linda Van Guelpen, Paul Regular, and et al. Forecasted shifts in thermal habitat for cod species in the northwest atlantic and eastern canadian arctic. *Frontiers in Marine Science*, 8:764072, 2021.

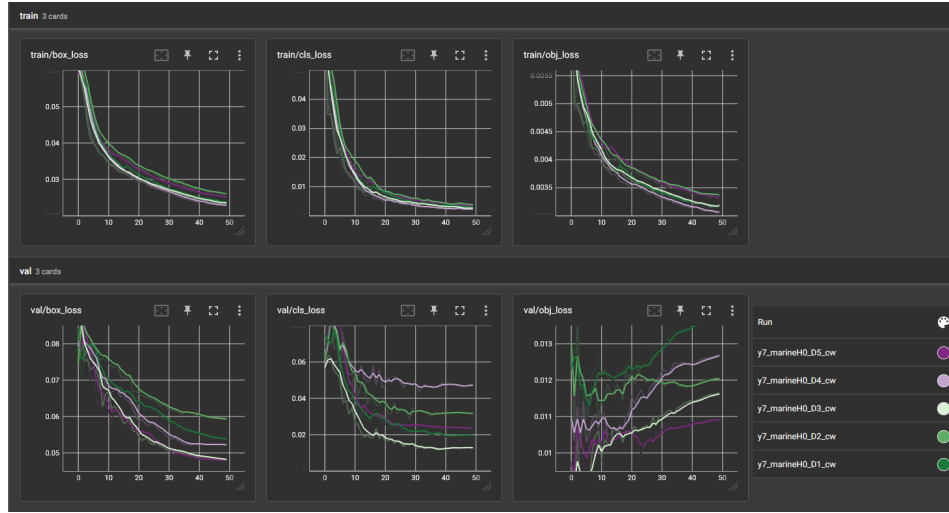
- [28] Nancy L. Shackell, Jonathan A. D. Fisher, Cornelia E. den Heyer, Daniel R. Hennen, Andrew C. Seitz, Arnault Le Bris, Dominique Robert, Michael E. Kersula, Steven X. Cadrin, Richard S. McBride, Christopher H. McGuire, Tony Kess, Krista T. Ransier, Chang Liu, Andrew Czich, and Kenneth T. Frank. Spatial ecology of atlantic halibut across the northwest atlantic: A recovering species in an era of climate change. *Reviews in Fisheries Science & Aquaculture*, 30(3):281–305, 2022.
- [29] Fisheries and Oceans Canada. Canada’s oceans now: Atlantic ecosystems, 2022. <https://www.dfo-mpo.gc.ca/oceans/publications/soto-rceo/2022/report-rapport-eng.html>, 2024. Synthesis of recent observations (2018–2021) in Atlantic Canada; web report published 2024-03-05.

5 Appendix

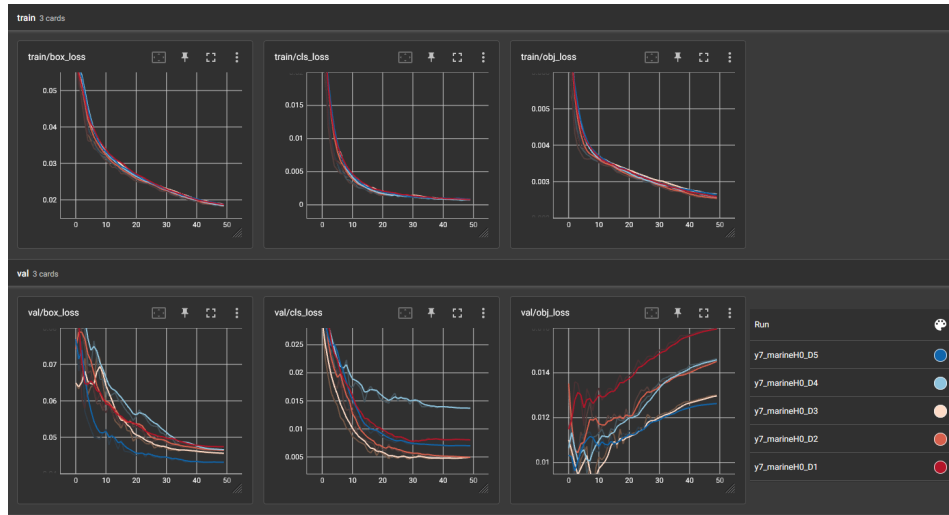
5.1 Baseline Training Details

Bounding box annotations were initially provided in TLBR format (absolute pixel coordinates for the top-left and bottom-right corners). These were converted to YOLO format, with per-video CSVs split into individual per-image text files containing normalized coordinates and class labels. Each YOLO-compatible annotation file was stored in the same directory as its corresponding image, following YOLO training requirements.

All baseline models were trained using the official YOLOv7 implementation (commit a207844b1ce82d204ab36d87d496728d3d2348e7) with the authors' *default* optimization, augmentation, and inference settings, unless otherwise specified. The initial learning rate (lr0) was reduced from 0.01 to 0.001. Models were initialized with MS COCO pre-trained weights, modifying the detection head to output the nine N-MARINE classes. Training was run for 50 epochs with a batch size of 32, letterboxed inputs of 640×640 pixels, and an NVIDIA A100 SXM4 40GB GPU. Inference was performed with a confidence threshold of 0.001, NMS IoU threshold of 0.65, and no test-time augmentation (TTA). For experiments using class weighting, only the scalar classification positive-weight parameter (cls_pw) was adjusted; all other configuration parameters remained at default values.



(a) With class weights.



(b) Without class weights.

Figure 3: Training and validation loss curves (box loss, classification loss, and objectness loss) for five cross-validation splits of the YOLOv7 model. Smoothed curves represent per-epoch loss values, showing consistent decreases in training loss and variation in validation performance across splits.

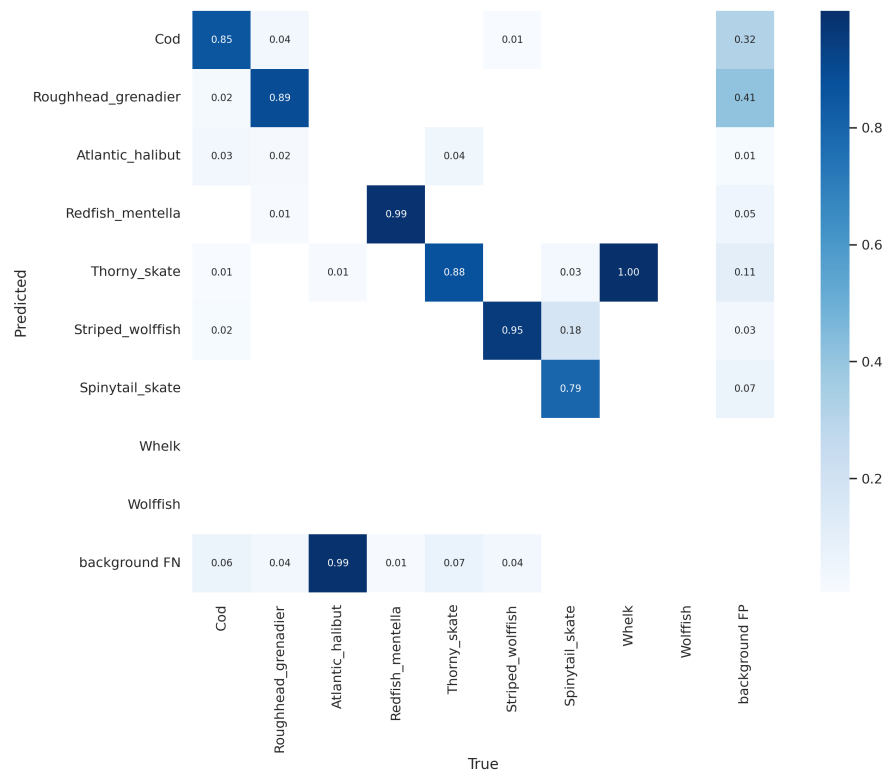


Figure 4: Confusion matrix for the model from the split with the highest test set mAP, trained without class weights. Values are normalized per true class using the test set.