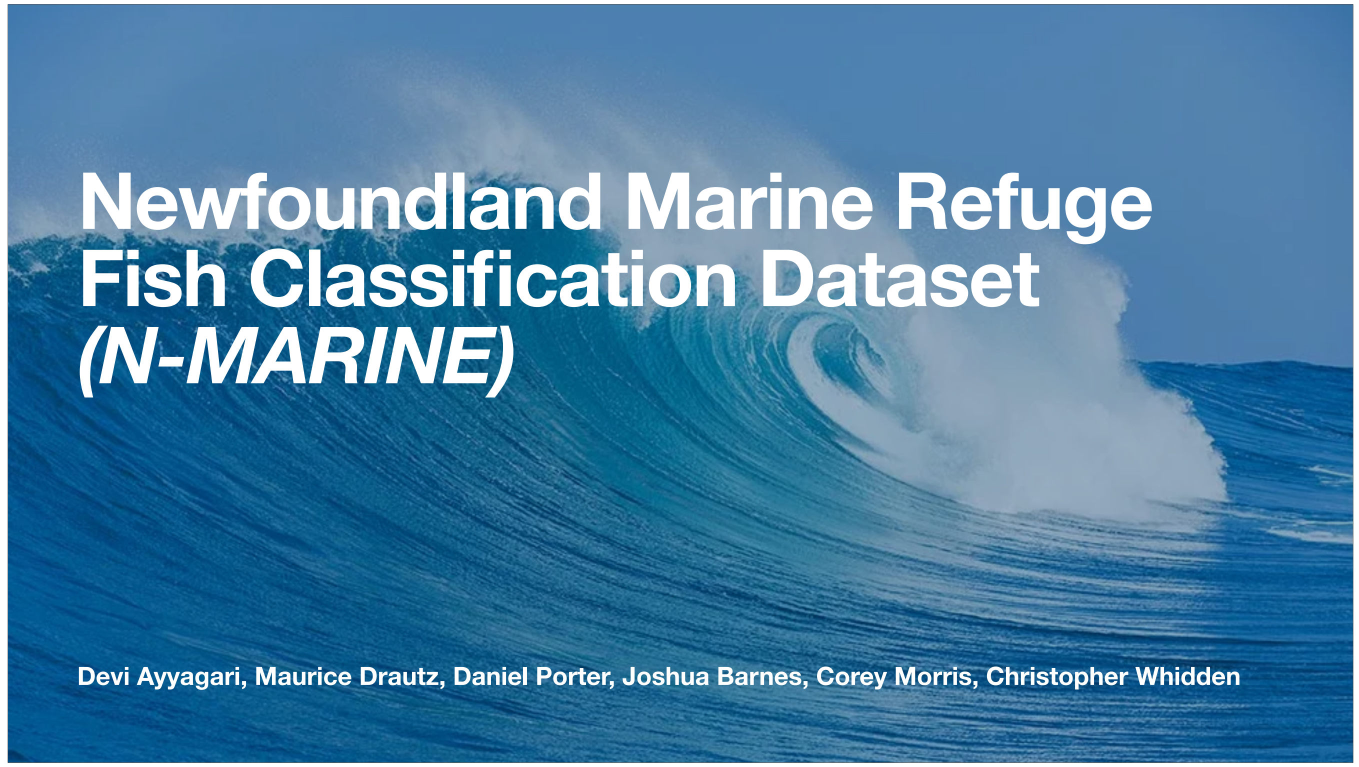
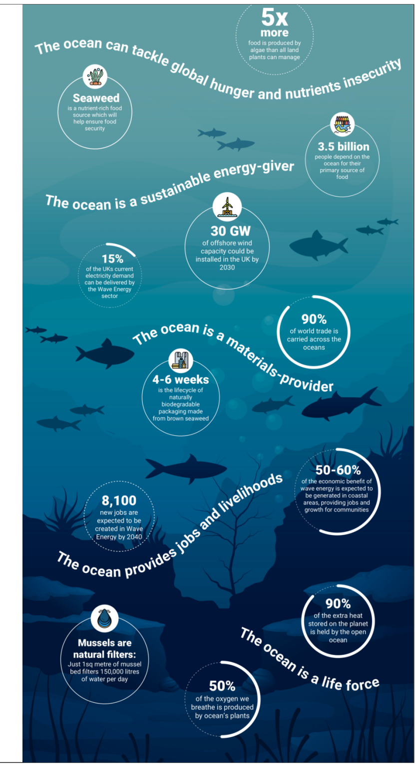


“Hello everyone, I’m Devi Ayyagari, a PhD student at Dalhousie University. Today I’m introducing N-MARINE—a North Atlantic underwater image dataset for fish detection and classification. I’ll walk through the dataset, the baseline we’re releasing for benchmarking, and—most importantly—why datasets like this are essential for accelerating climate-aware marine monitoring. I’ll also briefly highlight the gaps between ecologists and computer scientists, and how we can enable better collaboration by being intentional about the datasets and the baselines we release.”

The Need for Scalable and Non-Invasive Monitoring tools

- **Oceans are vital:** Oceans cover **71%** of Earth, act as a major **carbon sink**, and sustain the livelihoods of **millions** in coastal communities.
- **Rapid warming** is reshaping oceans: shifting species ranges, **population collapses** (e.g., snow crab), and increasing **policy tensions**.
- **Oceans are changing:** With warmer waters, **>80% of marine life** is moving and changing **breeding/feeding** patterns.
- **Immediate need**: scalable, non-invasive tools to **monitor marine ecosystem health**.

Pic courtesy: https://simplybluegroup.com/news/oceans-for-good/



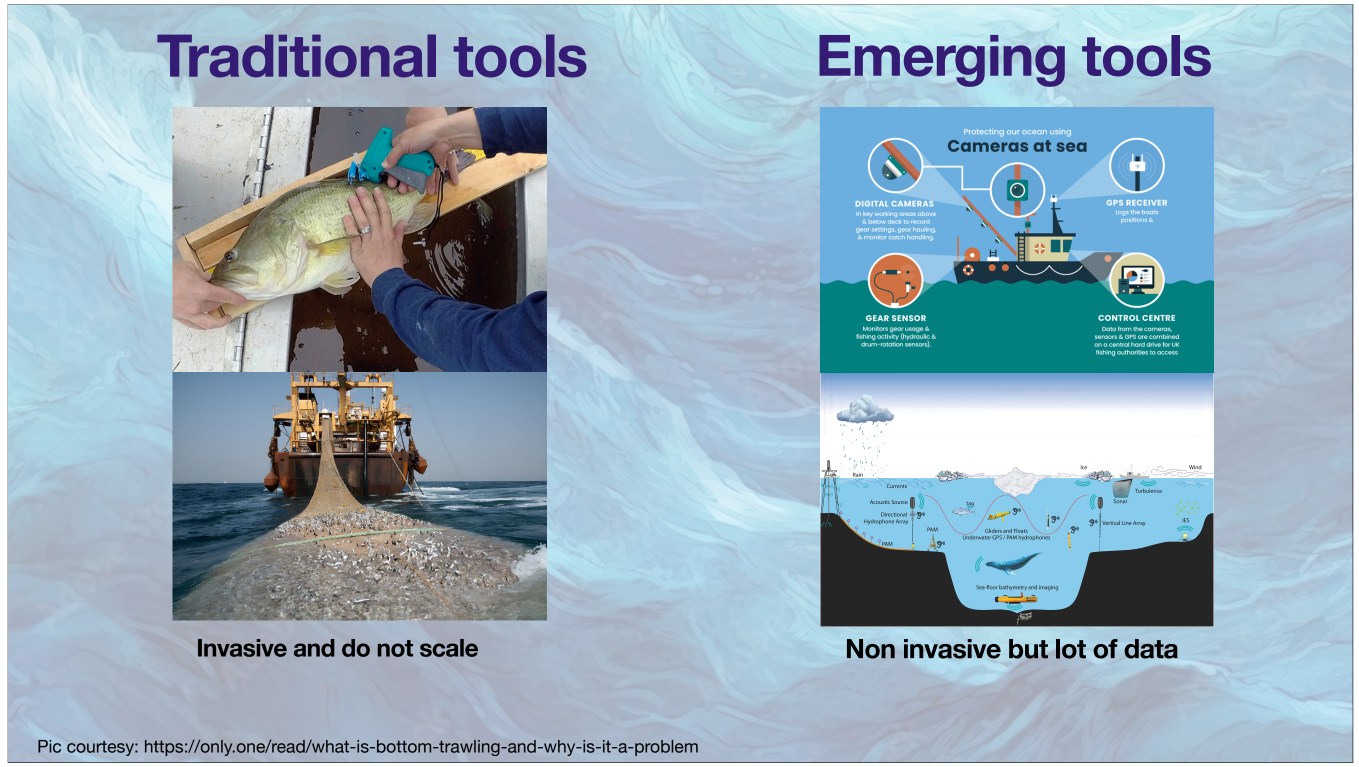
Oceans cover ~71% of Earth and are a major carbon sink and support millions of coastal livelihoods.

In 2025, the ocean remains exceptionally warm: ocean heat content is at record highs (set in 2024), the ocean continues to absorb ~90% of excess planetary heat.

This rapid warming is reshaping ecosystems.
With warmer waters, >80% of marine life is moving poleward and deeper, altering breeding/feeding patterns and management lines.

For example, an estimated 10 billion snow crab—about 90% of the stock—disappeared off Alaska between 2018 and 2021, putting heavy pressure on a fishery worth over $150 million.

We need systematic, scalable, non-invasive monitoring to track ecosystem health and catch these shifts early—after all, we cannot manage what we do not measure.



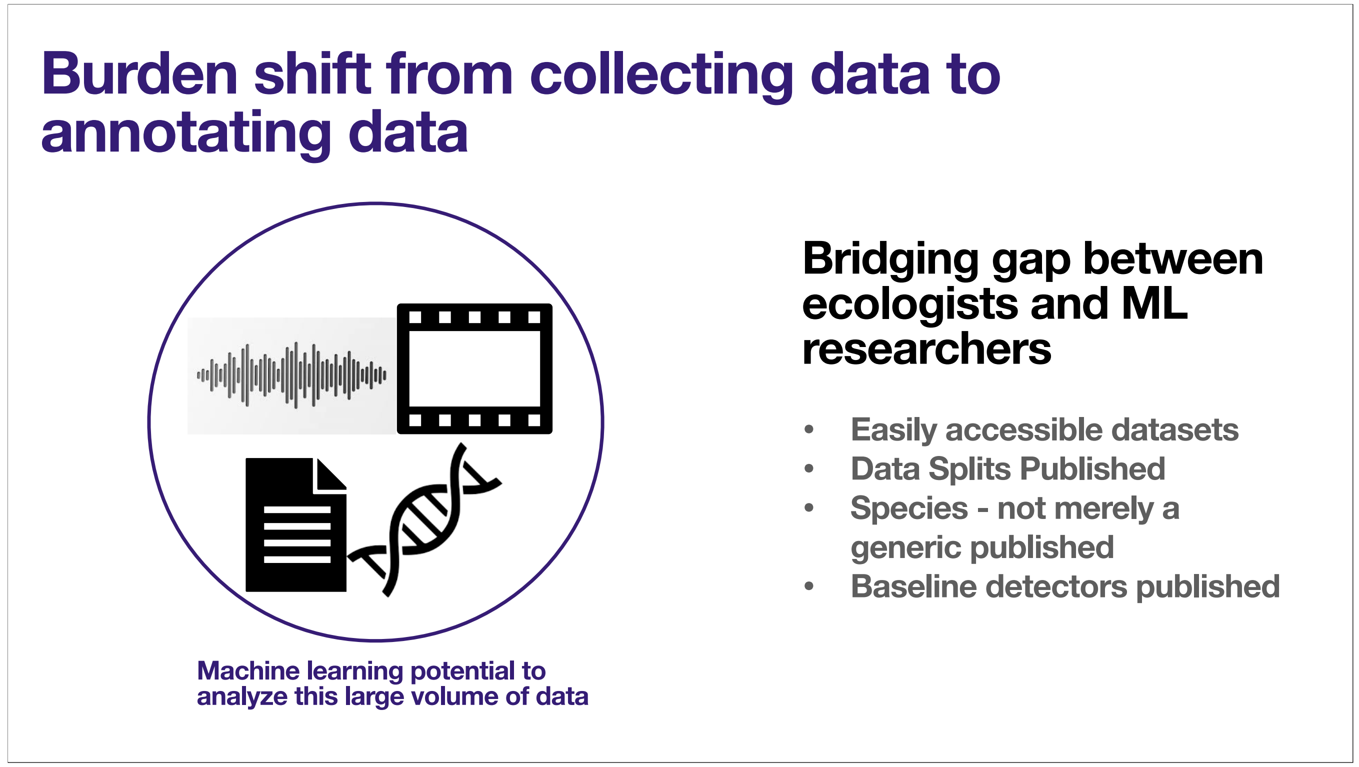
Traditionally, we've monitored marine populations with tags, trawl/field surveys, and on-deck visual observations. These are invasive, labor-intensive, and don't scale.

Newer, non-invasive tools are changing that:

Acoustic sensors – uses sound to detect and size schools and map habitats.
Underwater cameras – fixed or baited systems that capture species, behaviors, and habitat.
eDNA – detects species from genetic material in water samples.

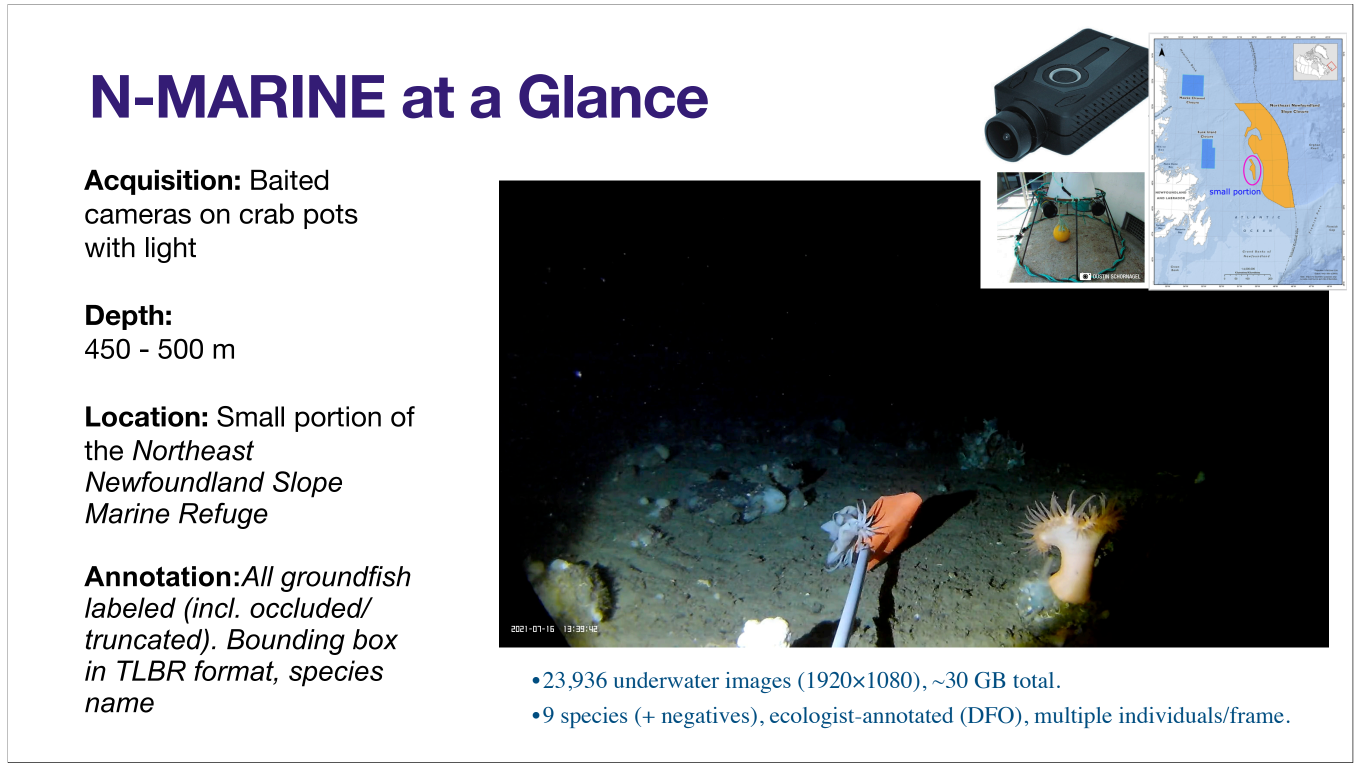
However, these sensors generate massive, multi-format data streams—far more than humans can review.

That's where machine learning can help.



Machine learning can triage massive sensor streams, but to analyze behavior, count individuals, and estimate populations, it needs large volumes of high-quality annotations.
As a result, the bottleneck has shifted from collecting data to annotating it—work that demands ecologists with specialized training.
Despite unprecedented data, we're still far from turning it into actionable insight.
To bridge ecology and computer science and truly accelerate monitoring, we need data systems and pipelines that process the raw data and focus scarce expert time where it matters most—interpreting model results and deriving actionable insights such as predicting species shifts and informing management strategies.
We think that Public, open datasets—from diverse regions, ecosystems, and camera systems—with exhaustive annotations, strong baselines, and species labels down to the lowest taxonomic level—that form the foundation for developing transferrable generalizable models are a practical way to do that.

We surveyed existing public image datasets and evaluated how they can be used to train machine learning models with small, locally annotated datasets, published in ICES this year. The core gap we found is that public data is concentrated in reefs, fish markets, and on-deck footage, while deep-underwater datasets are severely underrepresented. As a result, models trained under these conditions struggle to adapt to region-specific datasets acquired in deep water. This underscores the need for region-specific benchmarks.

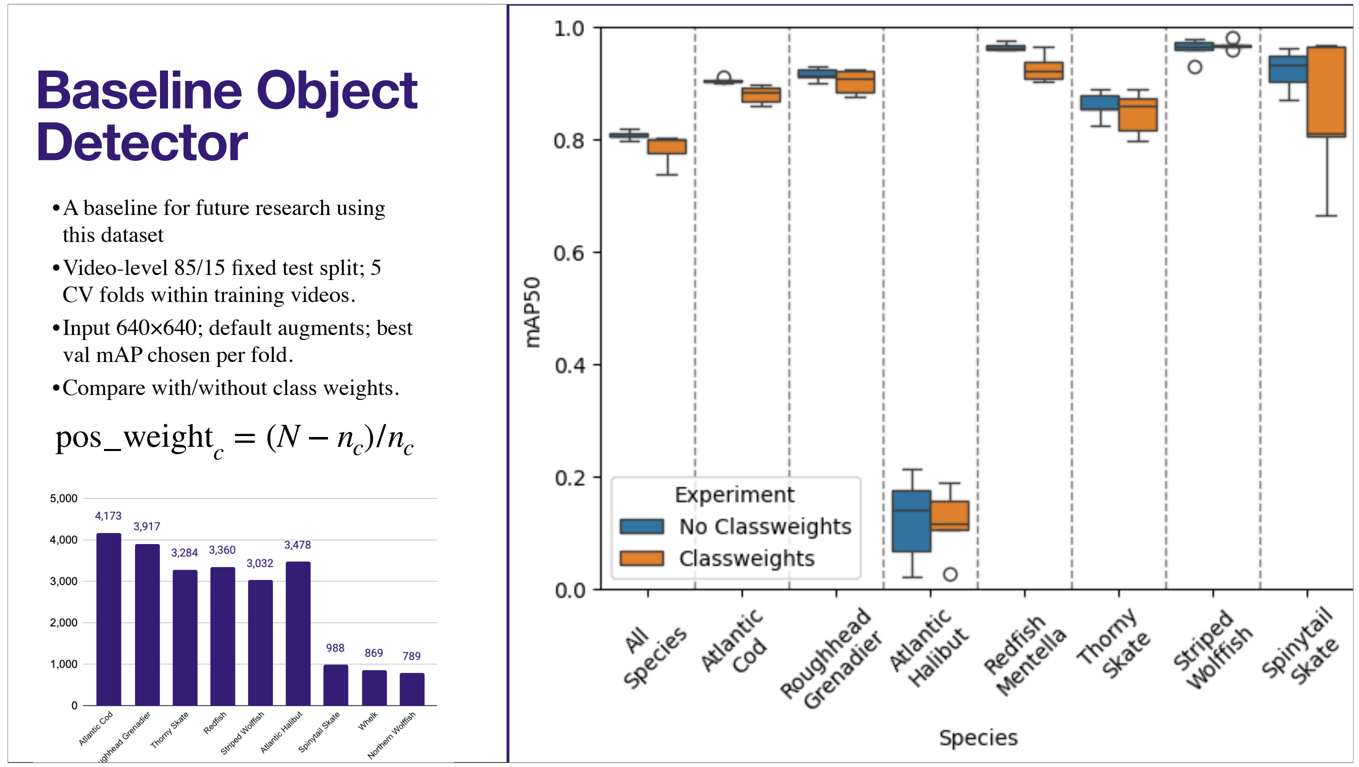


To that end, we publish N-MARINE - a deep underwater image dataset acquired in low light conditions in North Atlantic Ocean. The image here is a representative sample of images in the dataset.

The dataset was acquired using baited underwater cameras mounted on crab pots with light attached to it at depths of roughly 450–500 m. The location is a small portion of the Northeast Newfoundland Slope Marine Refuge. An expert ecologist at DFO labeled all 9 species of groundfish in the dataset, including occluded or truncated individuals.

Contains 23,936 images—about 30 GB. Boxes are stored in TLBR format with the species name.

This dataset was originally acquired to study the impact of seismic activity in the small portion of the marine refuge.
Alongside the data, we train a YOLOv7 baseline to detect the 9 different ground fish species in the datasets dn publish the baselines as well. Datasets published on open Canada portal
Baselines on hugging face



We trained a **YOLOv7** detector on N-MARINE using a **fixed 85/15 video-level split** to prevent train–test leakage. There is class imbalance in the dataset some species (e.g., Atlantic Cod) are common, while others (e.g., Northern Wolffish) are rare.
Note: **Whelk** and **Northern Wolffish** do not appear in the **test** split.

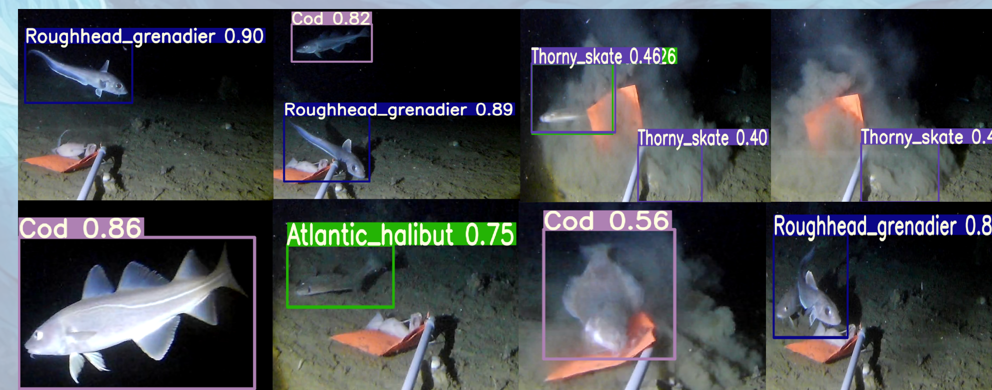
We have two Key observations

Firstly, Spinytail skate (the rarest class in the test set) outperforms Atlantic halibut. We hypothesize that this is because spinytail skate comes from a **crisp video**, while halibut often **kicks up sediment**, creating turbidity that suppresses detection. We'll show examples next.

Secondly, A model trained with **inverse-frequency class weights** - to offset the bias of larger sample size; performed **slightly worse** than the unweighted baseline. We are curious if exploring other loss functions like**focal loss**and **targeted augmentations** for rare classes would improve tail performance without sacrificing head-class precision.

Challenges for the AI community

- **Missed detections > misclassifications:** Poor visibility (turbidity, occlusion) suppresses recall more than it harms class ID.
- **Class-specific fragility:** *Halibut* performance collapses with a single turbid video (median mAP@0.5 ≈ 0.141).
- **Over-prediction of abundant classes:** Bias toward frequent species inflates false positives in busy scenes.



Example detections: columns 1–2 show correct detections; columns 3–4 show typical errors

Here are few challenges with using our baseline detector to detect ground fish in the dataset.
The left two columns, shows examples where the model performs well—clear water, clean silhouettes, and the model boxes the fish correctly.
Last 2 columns, the model struggles: the model misses detections when fish are partially occluded—for instance, one fish behind another simply disappears to the model.
The problem gets worse in turbid footage effecting performance of species like Atlantic halibut.

We invite the ML community to tackle these challenges: design pipelines that are visibility-aware, use temporal tracking to recover through occlusions, and treat rare classes explicitly so these models can perform well in real, messy ocean data.

Discussion

- **Strong domain sensitivity:** Small changes in lighting, habitat, or species mix cause site-to-site drift; models that work in one deployment often need adaptation/new labels elsewhere.
- **Active learning pipelines:** combine the data growth with self-/active-learning steps to focus labeling efforts
- **Collaboration:** Invite community to contribute new sites/species; align with self/active learning.

Using N-MARINE

- **Train & stress-test across datasets:** Build detectors/classifiers by **combining multiple marine datasets** to improve robustness to new acquisition setups and shifting species distributions; evaluate under turbidity, low light, and occlusion.
- **Cross-dataset transfer:** Run **pretrain → finetune** experiments (e.g., OzFish → N-MARINE) to assess generalization and **augment smaller datasets**.
- **Applied workflows:** Use as a starting point for **abundance estimation**, **behavior analysis**, and **semi-automated annotation**.
- **Teaching & upskilling:** Provide an educational resource that **bridges ecology and computer vision** and trains the next generation.

Dataset

Baseline object detector

Two takeaways. First, models are very domain-sensitive—small shifts in lighting, habitat, or species mix cause performance to drift, so a model that works at one site often needs adaptation or new labels at the next. This emphasizes the need active-learning pipelines: let the system surface the hardest, most uncertain frames so ecologists spend time where it matters in correcting model suggestions. And we hope to see more open datasets from diverse regions with diverse habitat and species distributions enabling us to develop generalizable models.

Second, on using N-MARINE. We anticipate that researchers use it to stress-test models under turbidity, low light, and occlusion; do cross-dataset transfer—pretrain on N-MARINE to boost performance on small local datasets. It could also be a starting point for downstream tasks like abundance estimation, studying fish behavior, and semi-automated annotation, and most importantly, we think it would be a valuable teaching resource that bridges ecology and computer vision. If you want to dive in, the left QR is the dataset; the right QR is the baseline detector.

Thanks for listening. I am excited to see what the marine and computer science community builds with this dataset! We hope that N-MARINE accelerates non-invasive, scalable monitoring in the North Atlantic and sparks collaboration on robustness under turbidity, transfer learning, and open-set recognition.
I'm happy to connect about data access, training details, or integration into monitoring pipelines.

I want to acknowledge and thank my advisor Dr. Whidden, Maurice - undergrad visiting student in our lab, DFO and the ecologists for the dataset and the annotations, DRAC for compute resources and NSERC and OFI for the funding for the development of this work!