

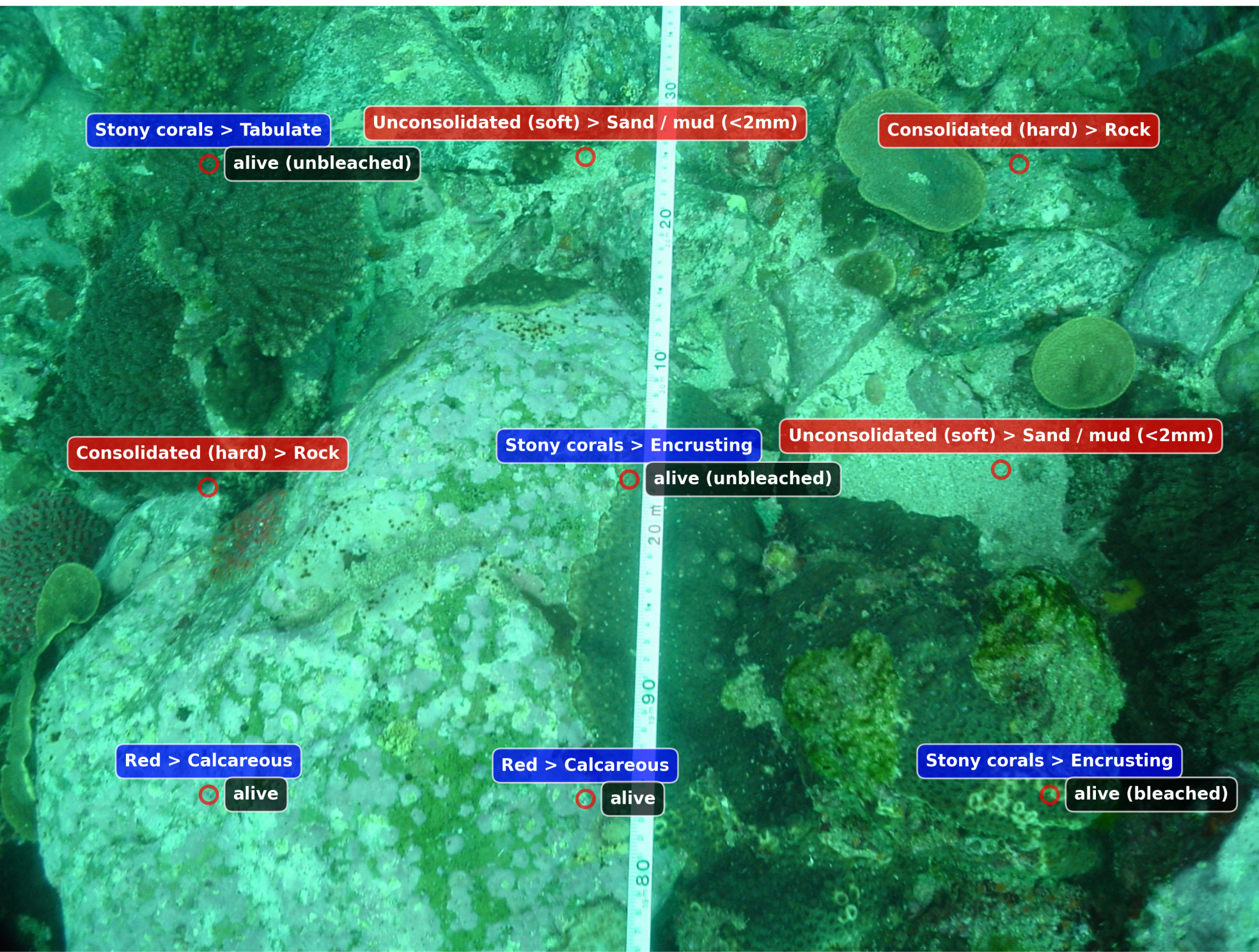
BenthicNet: A global compilation of seafloor images for deep learning applications



Scott C. Lowe*, Benjamin Misiuk*, Isaac Xu*, Shakhboz Abdulazizov, Amit .R. Baroi, Vicki Ferrini, Ariell Friedman, Deborah R. Hart, Daniel Ierodiaconou, Jacquomo Monk, John M. O’Brien, Elizabeth Oh, Luba Y. Reshitnyk, Chris M. Roelfsema, Jessica A. Sameoto, Alexandre C.G. Schimel, Craig J. Brown†, Thomas Trappenberg†

Key stats

Total collection sites: 24k
Total images: 11.4M
Labelled images: 190k
Total annotations: 3.1M
Annotation types: Substrate, Morphology, Taxonomy, ...
Metadata: Latitude/Longitude, Datetime, Depth



Deep learning applications

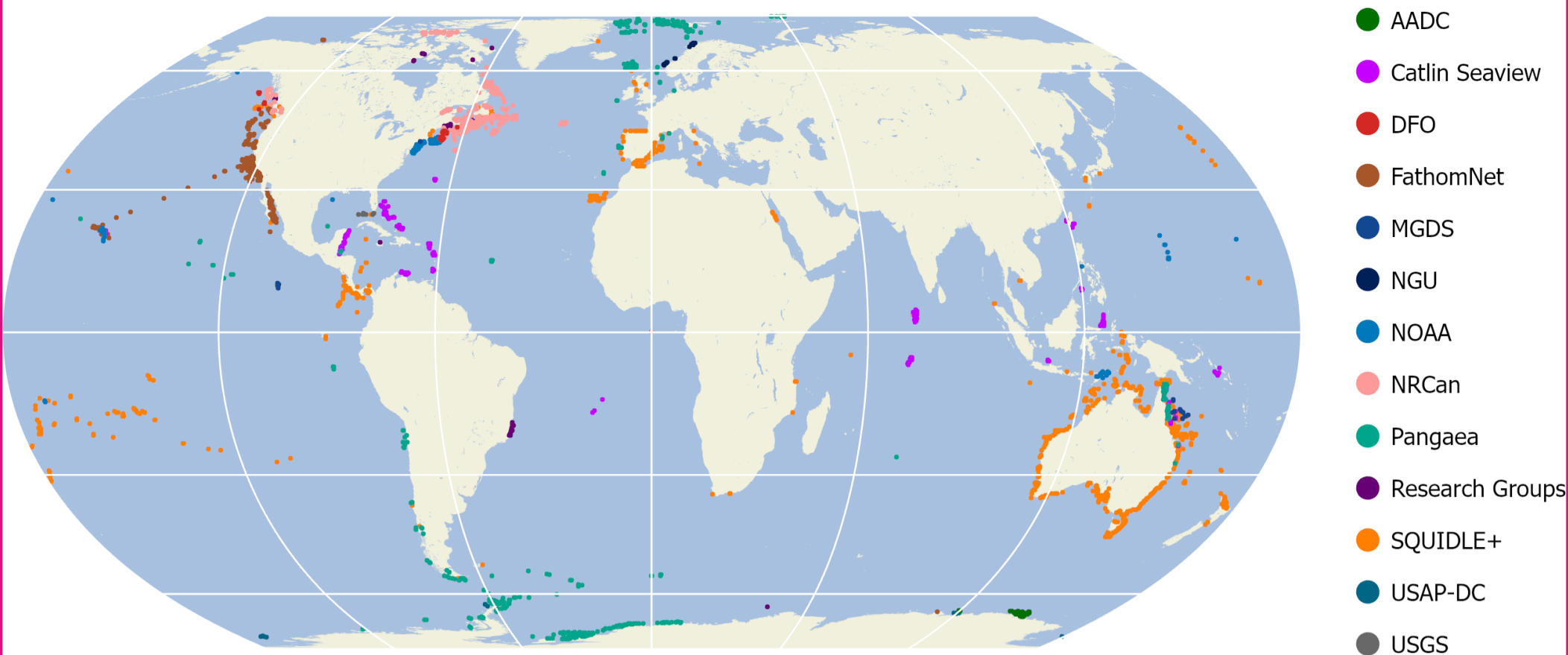
Feature → Task
Large unlabelled data → Self/semi-supervised learning
Full image labels → Image classification
Point annotations → Sparse semantic segmentation
Hierarchical labels → Hierarchical modelling
Domain shift/diversity → Robustness
Spatiotemporal location → Geospatial modelling

Motivation

Benthic habitats — seafloor environments where living things make their home — are **under threat** from fishing, ocean acidification, etc.
Photography is a key component of **monitoring** benthic habitats, and can be performed at **scale** with underwater drones. For continuous monitoring to keep up with data collection, we need **automated ML** analysis.

Spatial distribution

Images span the globe, mostly focused on the continental shelves along the coasts.

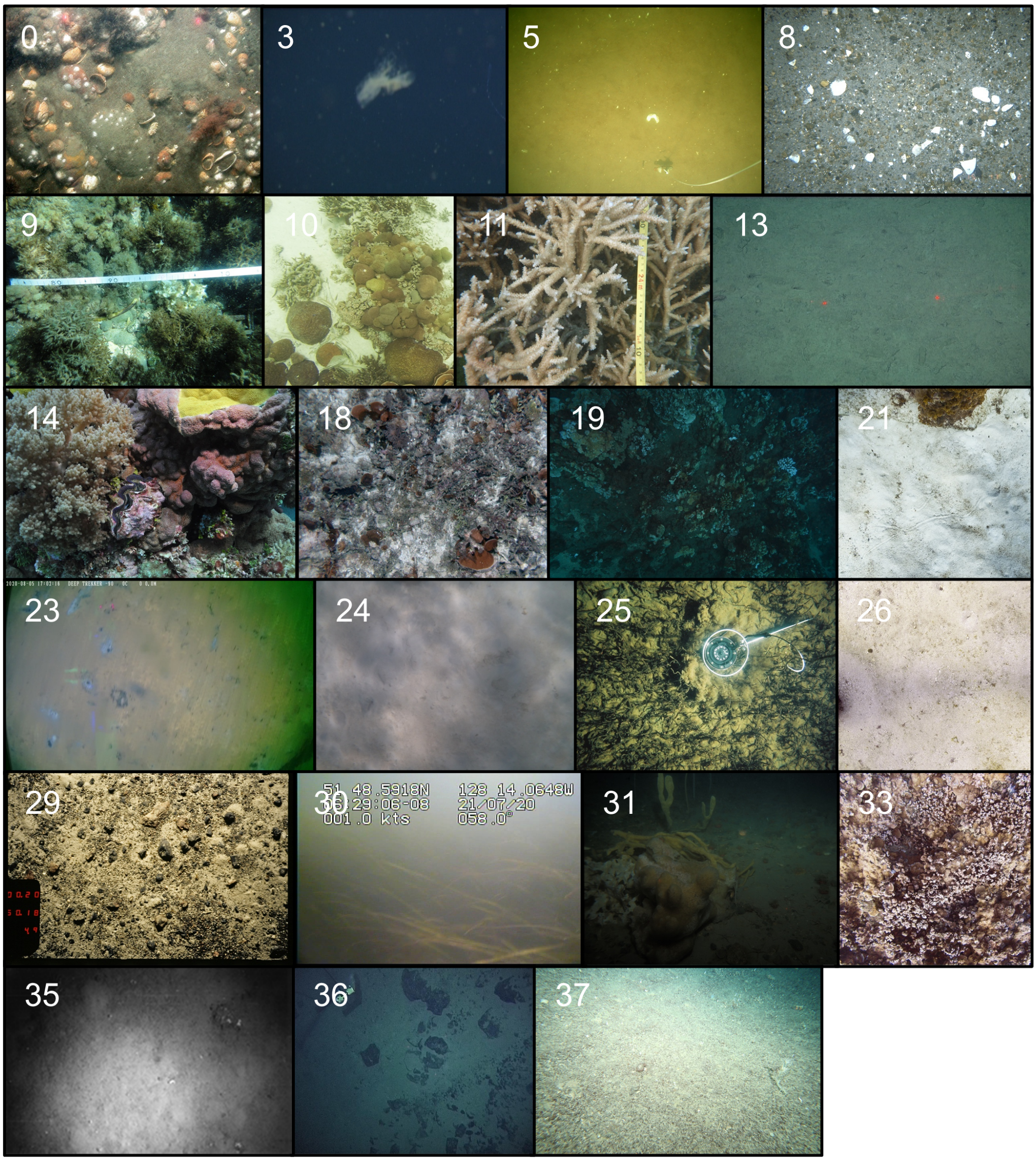


Dataset curation

Sources: Collated photographs from 2k datasets across a variety of public repositories and private holdings.
Public: SQUIDLE+, Catlin SV, PANGAEA, USGS, NOAA, FathomNet, NRCAN.
Private: NGU, DFO, MUN, Dalhousie, Hakai, 4D Oceans, EAC, UFES.
Spatial range: Global dataset, spanning all world oceans.
Temporal range: Most images collected 2005–2022, but some older: oldest are from 1960s.
Label standardization: Diverse annotations converted to CATAMI (substrate, bedforms, biotic morphology), WoRMS (biotic taxonomy), & qualifiers i.e. bleached, dead
Partitioning: Multi-labelled train/test set split with spatial blocking.

Table 1: Label standardization examples.

Original	CATAMI		WoRMS
	Substrate	Biota	Taxonomy
Mud and tube worms	Substrate <ul style="list-style-type: none">Unconsolidated (soft)Sand / mud (<1mm)Mud / silt (<64um)	Worms <ul style="list-style-type: none">PolychaetesTube worms	Annelida <ul style="list-style-type: none">Polychaeta
Hard Coral: Non hermatypic: Free living (Fungia etc)	–	Cnidaria <ul style="list-style-type: none">Corals<ul style="list-style-type: none">Stony coralsSolitaryFree living	Cnidaria <ul style="list-style-type: none">Hexacorallia<ul style="list-style-type: none">Scleractinia
Pocillopora sp.	–	Cnidaria <ul style="list-style-type: none">Corals<ul style="list-style-type: none">Stony corals	Cnidaria <ul style="list-style-type: none">Hexacorallia<ul style="list-style-type: none">Scleractinia<ul style="list-style-type: none">PocilloporidaePocillopora



Acknowledgments

Funding: Ocean Frontiers Institute (OFI).
Data contributions: Thanks to Reef Life Survey divers; Australia’s IMOS, ACFR, and Reef Builder programs; the Geological Survey of Norway; and more (see paper/dataset for a full list).

