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

Anonymous Author(s)
Affiliation
Address
email

Abstract

Advances in underwater imaging enable collection of extensive seafloor image datasets necessary for monitoring important benthic ecosystems. The ability to collect seafloor imagery has outpaced our capacity to analyze it, hindering mobilization of this crucial environmental information. Machine learning approaches provide opportunities to increase the efficiency with which seafloor imagery is analyzed, yet large and consistent datasets to support development of such approaches are scarce. Here we present BenthicNet: a global compilation of seafloor imagery designed to support the training and evaluation of large-scale image recognition models. An initial set of over 11.4 million images was collected and curated to represent a diversity of seafloor environments using a representative subset of 1.3 million images. These are accompanied by 3.1 million annotations translated to the CATAMI scheme, which span 190 000 of the images. A large deep learning model was trained on this compilation and preliminary results suggest it has utility for automating large and small-scale image analysis tasks.

1 Introduction

Underwater imagery, including both still photographs and video, is among the most common forms of data used to inform *benthic habitat mapping*. Benthic habitat maps describe both biotic and abiotic elements of the seafloor [8, 47], useful for marine management goals such as monitoring species and habitats of interest, informing policy decisions, and guiding sustainable ocean resource use [27, 3]. Seabed imagery has great utility for characterizing benthic environments for several reasons: it is non-invasive and minimally destructive, it may be collected remotely, it may be analyzed for multiple purposes (e.g. biology, geology), and it is more efficient to collect and store than physical samples (e.g. grabs, cores, preserved specimen).

Benthic imagery is increasingly collected using automated and remote underwater vehicles (AUVs, ROVs), which have the potential to generate larger volumes of data than previous methods—faster even than it can be analyzed [74, 24, 66]. The manual classification, annotation, and labelling of seabed imagery therefore acts as a bottleneck in the habitat mapping workflow [6] — a challenge which automation with machine learning (ML) could address.

Successfully training large-scale deep learning models from scratch requires large volumes of data. However, a deep neural network that has previously been trained on one task can be repurposed for a new task through transfer learning, provided the new task uses similar input stimuli to that used when training the original network. For vision models, transfer learning is typically performed by reusing models pretrained on the ImageNet dataset [64], comprised of photographs of terrestrial and anthropocentric objects. However, since this data does not represent subaqueous environments, the shift in domain may limit the capacity for transfer learning to benthic habitats [57, 42]. This motivates the need for a large-scale dataset of global seabed imagery.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Cumulatively, adequate volumes of benthic image data currently exist to support the development of large ML models, but they are spread globally among various research groups, government data portals, and open data repositories. Hence there is a need to compile and curate datasets for the development of large-scale image recognition models. Moreover, unlike terrestrial and anthropocentric images, there is no objective label for many seabed habitats, biological communities, substrate types, or organisms. Numerous different classification schemes are used to label benthic features [1, 50, 44], with small studies often coming up with their own bespoke classification scheme. Because no single vocabulary is universally applied to describe these features, we currently lack large sets of consistently labelled images that are necessary for training deep learning models for benthic environments. We note an outstanding need to develop standardized protocols for the translation of common marine image labelling schemes.

For benthic data, as is the case for most domains, unlabelled data is much easier to obtain at scale than annotated data. Fortunately, self-supervised learning (SSL) techniques have been developed which can harness unlabelled data for the initial pretraining stage of the neural network. In SSL, a pretext task is constructed automatically from the input data itself [4, 13, 28, 14, 26, 15, 17, 11, 29], enabling training of large-scale models on unlabelled imagery which can be more easily collected at scale. After pretraining with SSL, models have already learnt to see and understand stimuli and can be fine-tuned for specific downstream tasks without needing large volumes of annotated data.

The whole-image labels that typify benthic habitat image datasets may differ from other forms of marine image labelling that focus on locating specific objects, semantic labelling, bounding boxes, and masking. These forms of labelling are well suited to applications focusing on single taxa, pelagic biota, and object detection or tracking, and efforts to establish extensive image datasets for those applications are also underway. FathomNet [7, 36], contains over 100 000 images and over 200 000 localization labels with bounding boxes focused generally on marine biota, while Orenstein et al. [53] present a dataset of 3.4 million plankton images that have been labelled and used to train deep learning models. Hong et al. [30] have established the TrashCan dataset, containing over 7 000 images of marine debris with corresponding bounding box and segmentation masks. Other comparable image datasets include WildFish for classifying fish species [83], the OUC-vision large-scale underwater image database for underwater salient object detection [34], and the Brackish dataset [56] for detecting fish, crabs and starfish in brackish waters. Multiple datasets have been established to support the automated annotation of coral imagery including Moorea Labeled Corals [5], the Gulf of Eilat dataset [60], and notably, CoralNet [6, 12]. In addition to serving as a data repository where users can upload and share underwater image data and point labels, CoralNet provides a web interface to facilitate labelling and development of image recognition models. Several other comparable data portals and software packages enable the labelling and centralization of marine image data in this way (e.g. FathomNet, 36; SQUIDLE+; BIIGLE, 39; VIAME).

Here we describe BenthicNet: a global compilation of seafloor images that is designed to support development of automated image processing tools for benthic habitat data. With this compilation, we strive to obtain thematic diversity by (i) compiling benthic habitat images from locations around the world, and (ii) representing habitats from a broad range of marine environments. The compiled dataset is assessed for these qualities. Additionally, we aim to achieve diversity of non-thematic image characteristics (e.g. image quality, lighting, perspective) by obtaining data from a range of acquisition platforms and camera configurations. Diversity in the data is important for ensuring as much benthic imagery as possible is within the domain of our dataset. The dataset is presented in three parts: a diverse collection of over 11 million seafloor images from around the world, provided without labels (BenthicNet-11M); a rarefied subset of 1.3 million images, selected to maintain diversity in the imagery while reducing redundancy and volume (BenthicNet-1M); and a collection of 188 688 labelled images bearing 3.1 million annotations (BenthicNet-Labelled). We provide a large SSL model pretrained on BenthicNet-1M, and demonstrate its application using examples from BenthicNet-Labelled. The compilation and SSL model are made openly available to foster further development and assessment of benthic image automation tools.

2 Data curation

In order to achieve a diverse collection of benthic habitat images for training deep learning models, data spanning a range of environments and geographies was obtained from a variety of sources. These initially included project partners and research contacts, which were leveraged to establish additional

Table 1: **Summary of BenthicNet data sources** including the number of images in BenthicNet-11M (Full collection), BenthicNet-1M (Subsampled), and BenthicNet-L (Labelled). Further details on the individual datasets are provided within the BenthicNet metadata.

				# Samples		
Source	Region	# Datasets	# Sites	Full collection	Subsampled	Labelled
Online Repository/Collection						
AADC	Antarctic	2	86	2 056	2 024	203
Catlin Seaview	Global	22	861	1 082 452	283 674	11 346
FathomNet	W. USA	8	3 381	68 908	58 196	0
MGDS	Global	6	32	15 023	6 154	0
NOAA (via OneStop)	USA	18	526	73 019	40 714	4 543
NRCan	Canada	78	1 804	23 855	18 851	3 595
PANGAEA	Global	1 191	1 196	764 924	236 968	40 204
SQUIDLE+	Global	691	14 187	9 166 472	608 576	85 387
USAP-DC	Antarctic	5	27	4 144	2 886	0
USGS	USA	5	38	104 155	7 035	0
Individual Contributions						
4D Oceans	E. Canada	2	274	3 008	2 715	3 000
DFO (BIO)	E. Canada	6	381	7 773	5 981	7 762
DFO (IOS)	W. Canada	7	9	16 247	1 993	10 106
EAC	E. Canada	1	7	1 220	1 015	886
Hakai Institute	W. Canada	2	45	4 735	3 609	1 697
HAL	Jamaica	1	1	505	505	0
LaboGeo/UFES	E. Brazil	1	359	359	287	359
MUN	Arctic	4	135	10 691	6 403	10 687
NGU	Norway	4	580	50 290	50 275	0
NOAA (NEFSC)	N.E. USA	1	2	2 240	2 065	2 240
SEAM	E. Canada	3	284	6 811	5 170	6 673
Total	Global	2 058	24 215	11 408 887	1 345 096	188 688

data partnerships with individuals, academic and not-for-profit research groups, and government organizations. The largest data volumes were obtained from several academic, government, and third-party public data repositories. The acquisition of labelled data was prioritized in all cases, but extensive high quality unlabelled data collections were also included where feasible.

2.1 Data compilation and quality control

The formats and varieties of data were diverse, including collections of images with spreadsheet metadata, images with metadata contained in file names, GIS files containing images from which metadata was extracted, lists of URL image links, and raw video with text file annotations. Datasets not formatted as a single folder of images or list of URL links with CSV metadata were re-formatted. Metadata contained in image file names was parsed and used to construct a metadata CSV file where necessary. Image data within GIS files was extracted using ArcGIS Pro and the ArcPy Python package, along with other metadata contained within the files. Geographic coordinates were converted to decimal degrees using the WGS 84 datum. Video files were subsampled by extracting still frames according to their metadata using FFmpeg. After formatting, all datasets were subjected to quality control checks for missing entries, duplicates, label consistency, image quality, and matches between images and metadata. Quality control of image labels was performed by sampling the metadata and comparing labels to corresponding images for each dataset. Datasets where notable label inconsistencies were detected were rejected. Data columns were renamed to match a standardized format for the BenthicNet dataset. Acquisition methodology for each source is described in detail in Appx. A. Statistics for the various data sources are summarized in Tab. 1.

2.2 Data management

In total, 11 408 887 images were collected from the sources described above (see §2.1). Of all the images acquired, 188 688 included labels corresponding to visible benthic elements. Labels enable training and validation for supervised modelling tasks, such as localized species or substrate

Table 2: **Examples of original image labels translated to hierarchical labels** according to CATAMI v1.4 and WoRMS. Some original labels indicated both substrate and biota, while others indicated only one of these. For biota, some original labels provided more morphological detail and others more taxonomic; as much detail was retained as possible in both the CATAMI Biota and WoRMS taxonomic translations, respectively.

	CATAMI		WoRMS	
Original	Substrate	Biota	AphiaID	Taxonomy
Mud and tube worms	Substrate ↳ Unconsolidated (soft) ↳ Sand / mud (<1mm) ↳ Mud / silt (<64um)	Worms ↳ Polychaetes ↳ Tube worms	883	Annelida ↳ Polychaeta
Hard Coral: Non hermatypic: Free living (Fungia etc)	–	Cnidaria ↳ Corals ↳ Stony corals ↳ Solitary ↳ Free living	1363	Cnidaria ↳ Hexacorallia ↳ Scleractinia
Pocillopora sp.	–	Cnidaria ↳ Corals ↳ Stony corals	206938	Cnidaria ↳ Hexacorallia ↳ Scleractinia ↳ Pocilloporidae ↳ *Pocillopora*

identifications [33, 58], or bottom type classification [18]. There are several ways though that unlabelled data may still be utilized using unsupervised [77], semi-supervised [2, 54, 73, 78], and self-supervised [31, 4, 13, 28, 82, 26, 29] approaches. To facilitate a range of potential applications, we consider the dataset in two ways hereafter: the full set of images without their labels (BenthicNet-11M) and the set of labelled images (BenthicNet-Labelled).

2.2.1 Labelled data

Label translation. In order to increase the utility of the compiled data, and to facilitate validation of models trained on it, image labels from all datasets were translated to the CATAMI classification scheme (v1.4) [1], which spans both substrate and biota categories. Biota labels were additionally mapped to the World Registry of Marine Species (WoRMS) taxonomy [76].

Images were originally labelled according to a range of different established and bespoke schemes, but a large number of these (e.g. from SQUIDLE+) were readily available as CATAMI labels. CATAMI is a flexible framework that offers several advantages to accommodate translation and integration of a broad range of other labelled data. First, CATAMI supports labels for multiple classes of benthic features, including "branches" for both biota and physical elements such as substrate, bedforms, and relief. This enables the translation of a range of labelled datasets that were initially collected for a variety of different purposes. Second, labels within these branches are hierarchical. This means that objects may be labelled at different or even multiple levels of detail depending on the quality of the data, the confidence of the analyst, or the requirements of a particular application. This is critical for the translation of the multi-source data compiled here, which were initially analyzed at a range of thematic (e.g. taxonomic) resolutions for different purposes. Finally, CATAMI implements labels that are designed to be visually recognizable from image data. At a coarse level, these may distinguish broad groups or phyla of biota, but at finer levels, where identifying individual genera or species may become difficult using image data alone, morphological labels may be applied. These describe the size, shape, colour, and growth form of an organism, which may be recognizable even though the taxonomy is ambiguous.

We translated all image labels to the CATAMI (e.g. Tab. 2). All unique labels were extracted for each labelled image dataset in turn. Each unique label was translated to its closest CATAMI equivalent(s), maintaining the hierarchical level of the original data as closely as possible. In some cases, additional information within the metadata such as comments or auxiliary labels were used to complete the translation. Some annotations were provided in schemes that extend versions of CATAMI, such as the Australian Morphospecies Catalogue, which provides more precise morphological detail for the shape of sponges. Where this was done systematically and with more than 10 samples, we extended our scheme to match this increased level of morphological detail. Some annotations included man-made objects, such as trash or cables, which fell outside the scope of the CATAMI scheme, but may have value toward monitoring the anthropogenic impact on benthic habitats. Thus we also added an additional Anthropogenic branch to the hierarchy to cater to these annotations. We also include fields for CATAMI modifiers that indicate additional information such as whether organisms are bleached or dead, or their colours.

Some datasets provided taxonomic labels of biota at a high level of detail (genus or species level). To retain this information, taxonomic biota labels were additionally assigned an AphiaID from the World Registry of Marine Species (WoRMS). Where detailed taxonomic labels could not be determined, remaining biota annotations (e.g. morphological descriptions from CATAMI) were also mapped onto the WoRMS taxonomy at the highest level of specificity possible (typically phylum, class, or order).

In total, there were 188 688 labelled images, 3 091 158 individual CATAMI labels, and 1 131 391 WoRMS taxonomic labels for the BenthicNet-Labelled compilation.

Partitioning. To enable consistent validation and benchmarking between models using the BenthicNet dataset compilation, we provide train and test partitions of the labelled data. Test data were selected according to a partially spatial and class-label stratified procedure in order to ensure representation of a broad range of labels, and to reduce the degree of similarity between test and training partitions caused by spatial autocorrelation [37, 48]. The challenge in partitioning the dataset stems from the multi-label nature and imbalanced proportions of labels in BenthicNet. Firstly, imbalance necessitates careful assignment of rarer labels in the dataset. Additionally, a single image may have any combination of labels across multiple branches of the CATAMI hierarchy. If an image is assigned to test or training partitions due to a particular label, we must consider how the assignment affects other labels on the same image, some of which may be rarer.

Our partitioning process was as follows. We selected the target number of annotations per label to place in the test set as the smaller of 15% of the number of samples for the most frequently occurring label and 35% of the samples for the median label. We added images to train or test partitions one at a time, selecting the next label to add based on the following factors, in order of priority.

1. Ensure at least two samples for each label can be placed in the train partition.
2. Ensure at least 50% of the samples for each label can be placed in the train partition without using samples within 50 m of a test sample.
3. Ensure at least 15% of the samples for each label can be placed in the test partition without using samples within 50 m of a train sample.
4. Ensure no more than 35% of the samples for each label would be placed in the test partition.
5. Prioritize the CATAMI label with the fewest remaining unallocated images.

After determining the next label to add to a partition, we selected an image bearing that annotation to add to the partition as follows:

1. Of unallocated images bearing the label, if any are within 50 m of an image already in the target partition, randomly select an image from the closest 10% of those images.
2. Otherwise, randomly select an image bearing the label which is not within 50 m of an image already allocated to the other partition. Images violating this were used if needed to satisfy minimum populations described above (50% in train, 15% in test).

In practice, our partitions grow spatially outwards from initial seed locations, with new locations seeded to represent new label classes. Any remaining samples were allocated to the test partition if within 50 m of an image already in test, and to train otherwise. Effectively, test data was selected to prioritize representation of CATAMI labels, and then to minimize spatial overlap with the training data, as much as possible. This algorithm was run in parallel on 37 subsets of the data, each corresponding to a different ecological marine unit [65] (EMU; see Appx. C). 142 767 images (75.66%) were assigned to the training partition and 45 921 (24.34%) to test. The code for obtaining training and test partitions of the labelled data is provided at the BenthicNet code repository.

2.2.2 Unlabelled data

All images in BenthicNet may be used for applications which do not require labels, thus for the "unlabelled" slice of the data we considered all 11 408 887 images. We refer to the full collection as BenthicNet-11M. These images were not distributed uniformly in space; some datasets had a low sampling intensity, e.g. a few images per recording site taken by divers; while others were densely

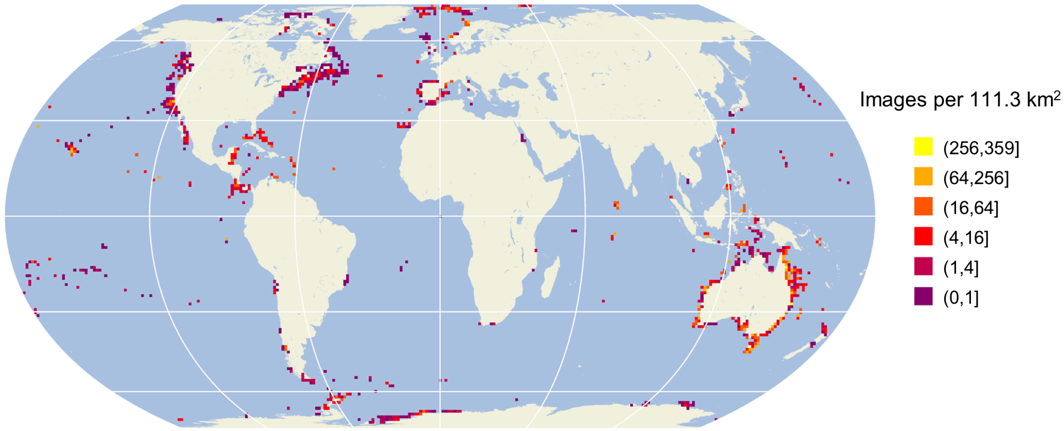

Figure 1: **Distribution of BenthicNet-1M images** projected to Equal Earth.

sampled, e.g. AUV video. To reduce the redundancy of the densely sampled data (thereby also reducing data volume and imbalance) the data was subsampled spatially, as described in Appx. B. We refer to the subsampled dataset as BenthicNet-1M.

3 Data Records

All BenthicNet data, metadata, and models described here are available from [REDACTED]. These include (i) a CSV file with an entry for each image in the subsampled compilation, BenthicNet-1M, conforming to the convention presented in Tab. 5; (ii) a single CSV file with an entry for each label of each image of the labelled compilation, BenthicNet-Labelled, conforming to the format presented in Tab. 3; (iii) a tarred directory containing each image in BenthicNet-1M and BenthicNet-Labelled, resized and compressed to JPEG format; (iv) a version of the entire image compilation tarred at the individual subdataset level; and (v) the ResNet-50 model weights resulting from SSL pretraining on BenthicNet-1M as described in §4. We additionally include CSV files listing the counts of each individual CATAMI label present, and a list of WoRMS taxonomical labels present. The metadata and models are available for use without restriction under the Creative Commons Attribution 4.0 License (CC-BY-4.0). Most images are available under CC-BY-4.0, except where the original licenses of subdatasets indicate limitations to derivative or commercial uses. Individual licenses for all subdatasets are retained and available along with the metadata within the repository.

3.1 Data formats

All unlabelled image metadata were standardized to a common format (Tab. 5). The datetime field was completed to the highest level of precision possible. Times were converted to UTC where timezones were indicated, and assumed to be UTC otherwise; it is not possible to guarantee all times are in UTC. Missing datetime and coordinate information was imputed where reasonably possible, e.g., by assigning the geographic mean centre of the acquisition site where coordinates were missing for some images. Labelled images were additionally assigned metadata describing the original and translated CATAMI labels, and WoRMS AphiaIDs (Tab. 3). Metadata indicating the pixel location of image labels were retained where provided.

4 Technical Validation

The subsampled BenthicNet dataset contains images from locations around the world (Fig. 1). Several regions are densely sampled: the Australian coast, Iberian Peninsula, Norwegian and Greenland Seas, North-Eastern and Western Canadian and U.S. continental shelves, and some Antarctic coast. Others are under-sampled, such as the Indian Ocean and the South Atlantic including the eastern coast of South America and west coast of Africa. Images collected in the open oceans are more dispersed than those at the continental shelves.

Table 3: **Format for compiled BenthicNet-Labelled image metadata.** Coverage is the fraction of images that have at least one such metadata entry.

Column	Contents	Data-type	Units	Coverage
url	URL address for this image	String		100.00%
source	Data provider/repository	String		100.00%
dataset	Name of dataset	String		100.00%
site	Image location name	String		100.00%
image	Image filename	String		100.00%
latitude	Latitude (WGS 84)	Float	Decimal degree	100.00%
longitude	Longitude (WGS 84)	Float	Decimal degree	100.00%
datetime	Acquisition date and time (UTC)	String	YYYY-MM-DD HH:mm:ss	100.00%
partition	Train/test split allocation	String		100.00%
annotation_column	Relative x location of labelled pixel	Float	Fraction of image width	53.11%
annotation_row	Relative y location of labelled pixel	Float	Fraction of image height	53.11%
original_label	Original image label	String		82.10%
catami_biota	CATAMI biota label	String		75.59%
catami_substrate	CATAMI substrate label	String		70.30%
catami_bedforms	CATAMI bedform label	String		6.87%
catami_relief	CATAMI relief label	String		2.46%
catami_qualifiers	CATAMI label qualifier	String		10.82%
colour_qualifier	Label colour qualifier	String		6.52%
bleached	Whether biota is bleached	Float	Values 0 or 1	13.44%
dead	Whether biota is deceased	Float	Values 0 or 1	25.76%
aphia_id	WoRMS taxon AphiaID label	Integer		60.73%
gebco_bathymetry	Depth interpolated from GEBCO2022	Float	Metres	100.00%
emu	Nearest ecological marine unit	Integer		99.98%

4.1 Environmental heterogeneity

Images in the compiled datasets were acquired between 1965–2021 from depths ranging from < 1 m to over 5 500 m (Fig. 2). We analyzed the representativeness of image samples across the broader global oceans by comparing the sampled frequency to the area of each ocean basin (Fig. 3), demonstrating a similar distribution. We extended this analysis in more detail by considering the distribution of ecological marine units (EMUs) captured by the dataset (see Appx. C). The nearest EMU to each image is provided as a metadata field for both the BenthicNet-1M and BenthicNet-Labelled datasets; depths from the GEBCO2022 grid are also provided for each image (assigned using bilinear interpolation; Tables 3 and 5).

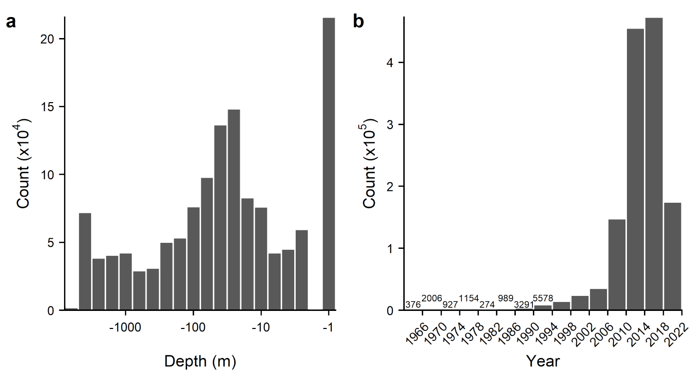

Figure 2: **Distribution of BenthicNet-1M images** according to (a) log-scale **depth** data retrieved from the GEBCO2022 grid [23] and (b) **year of acquisition**.

4.2 Model training

We ran experiments to explore the utility of pretraining with SSL on the unlabelled BenthicNet-1M dataset for automating downstream benthic image labelling tasks. We identified two labelling tasks from within our data. **BenthicNet-Substrate-d2.** We considered all images in BenthicNet-Labelled with a whole-frame CATAMI substrate label to at least depth 2. We used the depth-2 labels,

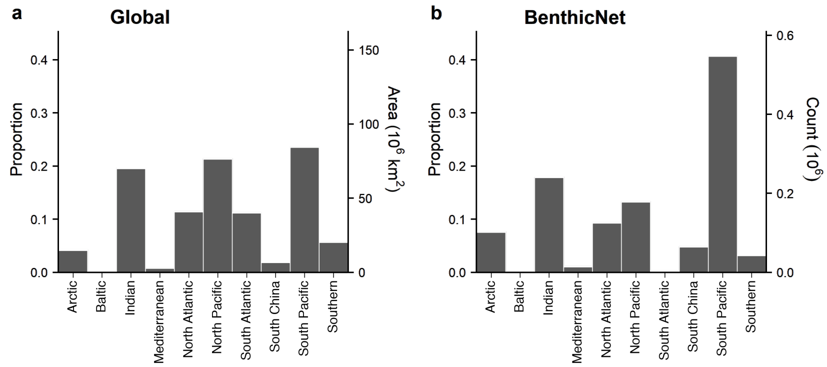

Figure 3: **Distribution of BenthicNet-1M images according to global ocean/sea basins** [65]. (a) Proportion and area of ocean basins. (b) Proportion of BenthicNet image samples from each ocean basin.

comprised of 5 classes "Sand/mud", "Pebble/gravel", "Cobbles", "Boulders", and "Rock". This subset, "BenthicNet-Substrate-d2", comprised of 43 430 training and 13 719 test images. This size was made possible by our large scale data collation and label mapping pipeline, but we anticipate that a pretrained model may be used on small datasets using a bespoke labelling scheme for the area. We thus also considered a classification task using the original labels of an input dataset: **German Bank 2010**, provided by DFO [9], which had whole-image "benthoscape" that describe recognizable combinations of substrate, bedforms, and biology visible in 3 181 images, collected off the southwest coast of Nova Scotia, Canada. The labels are (1) "reef", where boulders or bedrock with frequent epifauna comprise more than 50% of images; (2) "glacial till", consisting of mixed sediments (cobble, gravel, sand); (3) "silt/mud" with frequent evidence of infaunal bioturbation; (4) "silt with bedforms"; and (5) "sand with bedforms", which commonly included sand dollars (*Echinarachnius parma*). 2 681 images were used for training and 500 for testing. Both tasks have one-hot targets, which simplifies model training and evaluation.

We initially considered four recent instance-learning SSL methods: SimSiam [15], Bootstrap Your Own Latent (BYOL) [26], Momentum Contrast (MoCo-v2) [16], and Barlow Twins (BT) [82]. Using a ResNet-50 trained on BenthicNet-1M with these methods, we found they performed similarly, with BT performing consistently well at downstream classification tasks. Our analyses here thus show BT as a representative SSL method. Using BT, we trained a ResNet-50 model on the BenthicNet-1M dataset for 100, 200, and 400 epochs with the LARS optimizer [79]. The hyperparameters were as for Zbontar et al. [82], except the learning rate which was set to 2×10^{-3} and annealed using one-cycle with 10 epochs of warm-up [68]. Models were trained using four Nvidia A100 GPUs, with total batch size 512. We compare against transfer learning from a publicly available ResNet-50 model[1] pretrained with cross-entropy on ImageNet-1k (600 epochs), provided by torchvision [55].

Using one of the pretrained ResNet-50 backbone encoders, we performed a linear probe by training a linear classifier head while keeping the encoder frozen. We trained the classifier head for 100 epochs with one-cycle LR schedule, maximum LR 3×10^{-5}, 10 ep. warm-up epochs. We performed full fine-tuning with an unfrozen encoder and classifier head initialized from the linear probe. We trained the network end-to-end with one tenth the learning rate used for the linear probe for 300 epochs. The transfer-learning models were compared to models trained from scratch for 100 or 400 epochs, using the one-cycle schedule with peak learning rate 3×10^{-5}.

As shown in Tab. 4, pretraining with BT on BenthicNet-1M consistently outperformed training from scratch on the labelled data—the FT model outperformed training from scratch by +4.3% on BenthicNet-Substrate-d2 and +22.9% on German Bank 2010, and the linear probe outperformed training from scratch 100 ep. The BenthicNet-1M pretrained model outperformed the ImageNet-1k model across the linear probes, indicating its superior alignment to the downstream tasks of benthic habitat mapping. The smaller dataset, German Bank 2010, was much more challenging, with lower performance from all models and very poor linear probe performance from the ImageNet-1k model (37.6%). With fine-tuning, the ImageNet-1k pretrained model was able to perform well on the larger BenthicNet-Substrate-d2 evaluation task, demonstrating the utility of the scale of the labelled dataset.

[1] `torchvision.models.ResNet50_Weights.IMAGENET1K_V2` [recipe]

Table 4: **Micro-accuracy and macro F1-score (%)** on two downstream test datasets when training from scratch (No pretraining), or using a pretrained encoder either with linear probe (frozen weights; ❄) or full fine-tuning (🔥). Mean (± std. err.) over 3 random seeds (same pretrained backbones over seeds). Bold: **best** performance.

Pretraining				BenthicNet-Substrate-d2		German Bank 2010	
Dataset	Loss	FT	Epochs	Accuracy ↑	F1-score ↑	Accuracy ↑	F1-score ↑
No pretraining		🔥	100	81.0 ± 0.6	55.3 ± 1.5	53.4 ± 2.4	43.0 ± 3.8
ImageNet-1k	Cross-entropy	❄	100	81.8 ± 0.1	56.6 ± 0.3	37.6 ± 5.2	30.0 ± 2.9
BenthicNet-1M	Barlow Twins	❄	100	**83.6** ± 0.1	**57.7** ± 0.3	**55.9** ± 2.4	**43.2** ± 6.0
No pretraining		🔥	400	83.8 ± 0.1	61.8 ± 0.3	54.1 ± 3.2	46.7 ± 3.2
ImageNet-1k	Cross-entropy	🔥	100+300	**88.3** ± 0.1	**69.5** ± 0.3	65.9 ± 4.0	59.2 ± 4.2
BenthicNet-1M	Barlow Twins	🔥	100+300	88.1 ± 0.1	68.5 ± 0.2	**77.0** ± 0.7	**72.3** ± 0.8

The ImageNet-1k model was even able to outperform our BenthicNet-1M pretrained model by a small margin (0.2% acc.) on this task. However, on the smaller and more challenging German Bank 2010, the BenthicNet-1M pretraining was greatly superior (+21.1%), demonstrating the utility of a task-aligned pretrained model for transferring to small data regimes.

The confusion matrices (Appx. D) show the FT models have similar biases on BenthicNet-Substrate-d2, confusing the same classes as each other (Fig. 8; Cobbles→Boulders; Pebble/gravel→Sand/mud; etc.). On German Bank 2010 (Fig. 9), the BenthicNet-1M pretrained model was able to greatly increase the recall of silt/mud, silt with bedforms, and glacial till, almost entirely removing ImageNet-1k's most confused class pair, glacial till ↔ sand with bedforms.

5 Usage Notes

Labels translated to the CATAMI scheme were sourced from a wide variety of scientific studies with the express intent of supporting the training and validation of large image recognition models. Jointly, these labels should be analyzed with care, particularly if utilized for other purposes. Some datasets included whole image labels indicating the presence of a single benthic feature (e.g. organism, substrate), while others supplied single labels indicating multiple features, or multiple labels for different features within an image. One result of such diversity is variation in the completeness of labels from different datasets—some, for example, focus on a the presence of single species, or only focus on the most conspicuous or abundant substrate types. For some datasets, it is thus reasonable to expect a larger proportion of false negative labels if the data is treated in a presence/absence manner. In other words, many benthic features are likely visible in the images, which have not been labelled. We operate under the assumption, though, that labels within a dataset were assigned consistently. If performing analyses at the dataset level using the compilation presented here, it is important to investigate the specifics of the dataset(s) in question.

Similarly, the diversity of labelling methodologies has resulted in a number of different schema by which original labels were translated to CATAMI equivalents. For example, some labels indicating the percent cover of organisms or substrate types in an image were converted to binary presence/absence. Additionally, auxiliary information such as annotator notes were used in some cases to obtain a CATAMI label. Efforts were made to indicate the original data label as closely as possible, but it was not always possible to include all information used to translate an original label to CATAMI. Therefore, original labels provided in our metadata may not contain all available information for each image, and the source datasets should be referenced as the authoritative source in all cases.

The examples provided here focus on the physical environment, but there are abundant opportunities to explore use of the biological labels. Through use of the SSL pretrained encoder, we anticipate training and deployment of hierarchical morphological and biological identification models is possible. A challenging component of this task is the imbalance of biota labels. Methods such as over-sampling, weighting, and data augmentation may be necessary to achieve effective large-scale supervised models in the biota hierarchy, both to address the label imbalance and distributional shift from the labelled subset to the full range of ocean imagery. These applications will be explored in coming work.

References

[1] Franziska Althaus, Nicole Hill, Renata Ferrari, Luke Edwards, Rachel Przeslawski, Christine H. L. Schönberg, Rick Stuart-Smith, Neville Barrett, Graham Edgar, Jamie Colquhoun, Maggie Tran, Alan Jordan, Tony Rees, and Karen Gowlett-Holmes. A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: The CATAMI classification scheme. *PLOS ONE*, 10(10):1–18, 10 2015. doi:10.1371/journal.pone.0141039. URL `https://doi.org/10.1371/journal.pone.0141039`.

[2] Riccardo Arosio, Brandon Hobley, Andrew J. Wheeler, Fabio Sacchetti, Luis A. Conti, Thomas Furey, and Aaron Lim. Fully convolutional neural networks applied to large-scale marine morphology mapping. *Frontiers in Marine Science*, 10:1228867, July 2023. ISSN 2296-7745. doi:10.3389/fmars.2023.1228867. URL `https://www.frontiersin.org/articles/10.3389/fmars.2023.1228867/full`.

[3] Elaine K. Baker and Peter T. Harris. Habitat mapping and marine management. In Peter T. Harris and Elaine K. Baker (eds.), *Seafloor Geomorphology as Benthic Habitat*, pp. 17–33. Elsevier, 2nd edition, 2020. ISBN 978-0-12-814960-7. doi:10.1016/B978-0-12-814960-7.00002-6. URL `https://linkinghub.elsevier.com/retrieve/pii/B9780128149607000026`.

[4] Randall Balestriero, Mark Ibrahim, Vlad Sobal, Ari Morcos, Shashank Shekhar, Tom Goldstein, Florian Bordes, Adrien Bardes, Gregoire Mialon, Yuandong Tian, Avi Schwarzschild, Andrew Gordon Wilson, Jonas Geiping, Quentin Garrido, Pierre Fernandez, Amir Bar, Hamed Pirsiavash, Yann LeCun, and Micah Goldblum. A cookbook of self-supervised learning. *arXiv preprint*, arXiv:2304.12210, 2023. doi:10.48550/arxiv.2304.12210.

[5] Oscar Beijbom, Peter J Edmunds, David I Kline, B Greg Mitchell, and David Kriegman. Automated annotation of coral reef survey images. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 1170–1177. IEEE, 2012. doi:10.1109/CVPR.2012.6247798.

[6] Oscar Beijbom, Peter J. Edmunds, Chris Roelfsema, Jennifer Smith, David I. Kline, Benjamin P. Neal, Matthew J. Dunlap, Vincent Moriarty, Tung-Yung Fan, Chih-Jui Tan, Stephen Chan, Tali Treibitz, Anthony Gamst, B. Greg Mitchell, and David Kriegman. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. *PLOS ONE*, 10(7):1–22, 07 2015. doi:10.1371/journal.pone.0130312. URL `https://doi.org/10.1371/journal.pone.0130312`.

[7] Océane Boulais, Ben Woodward, Brian Schlining, Lonny Lundsten, Kevin Barnard, Katy Croff Bell, and Kakani Katija. FathomNet: An underwater image training database for ocean exploration and discovery. *arXiv preprint*, arXiv:2007.00114, 2020. doi:10.48550/arxiv.2007.00114.

[8] Craig J. Brown, Stephen J. Smith, Peter Lawton, and John T. Anderson. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, 92(3):502–520, May 2011. ISSN 02727714. doi:10.1016/j.ecss.2011.02.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S0272771411000485`.

[9] Craig J Brown, Jessica A Sameoto, and Stephen J Smith. Multiple methods, maps, and management applications: Purpose made seafloor maps in support of ocean management. *Journal of Sea Research*, 72:1–13, 2012. doi:10.1016/j.seares.2012.04.009.

[10] Craig J Brown, Jonathan Beaudoin, Mike Brissette, and Vicki Gazzola. Multispectral multibeam echo sounder backscatter as a tool for improved seafloor characterization. *Geosciences*, 9(3): 126, 2019. doi:10.3390/geosciences9030126.

[11] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9630–9640, 2021. doi:10.1109/ICCV48922.2021.00951.

[12] Qimin Chen, Oscar Beijbom, Stephen Chan, Jessica Bouwmeester, and David Kriegman. A new deep learning engine for CoralNet. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 3686–3695, 2021. doi:10.1109/ICCVW54120.2021.00412.

[13] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1597–1607. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/chen20j.html.

[14] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 22243–22255. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/fcbc95ccdd551da181207c0c1400c655-Paper.pdf.

[15] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15750–15758, June 2021. doi:10.1109/CVPR46437.2021.01549.

[16] Xinlei Chen, Haoqi Fan, Ross B. Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arxiv preprint*, arXiv:2003.04297, 2020. doi:10.48550/arXiv.2003.04297. URL https://arxiv.org/abs/2003.04297.

[17] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9640–9649, October 2021. doi:10.1109/ICCV48922.2021.00950.

[18] Andre Diegues, Jose Pinto, Pedro Ribeiro, Roberto Frias, and do Campo Alegre. Automatic Habitat Mapping using Convolutional Neural Networks. In *2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV)*, pp. 1–6, Porto, Portugal, November 2018. IEEE. ISBN 978-1-72810-253-5. doi:10.1109/AUV.2018.8729787. URL https://ieeexplore.ieee.org/document/8729787/.

[19] Graham J. Edgar and Rick D. Stuart-Smith. Systematic global assessment of reef fish communities by the Reef Life Survey program. *Scientific Data*, 1(1):140007, May 2014. ISSN 2052-4463. doi:10.1038/sdata.2014.7. URL https://doi.org/10.1038/sdata.2014.7.

[20] Graham J. Edgar, Antonia Cooper, Susan C. Baker, William Barker, Neville S. Barrett, Mikel A. Becerro, Amanda E. Bates, Danny Brock, Daniela M. Ceccarelli, Ella Clausius, Marlene Davey, Tom R. Davis, Paul B. Day, Andrew Green, Samuel R. Griffiths, Jamie Hicks, Iván A. Hinojosa, Ben K. Jones, Stuart Kininmonth, Meryl F. Larkin, Natali Lazzari, Jonathan S. Lefcheck, Scott D. Ling, Peter Mooney, Elizabeth Oh, Alejandro Pérez-Matus, Jacqueline B. Pocklington, Rodrigo Riera, Jose A. Sanabria-Fernandez, Yanir Seroussi, Ian Shaw, Derek Shields, Joe Shields, Margo Smith, German A. Soler, Jemina Stuart-Smith, John Turnbull, and Rick D. Stuart-Smith. Establishing the ecological basis for conservation of shallow marine life using Reef Life Survey. *Biological Conservation*, 252:108855, December 2020. ISSN 00063207. doi:10.1016/j.biocon.2020.108855.

[21] HE Garcia, RA Locarnini, TP Boyer, JI Antonov, OK Baranova, MM Zweng, JR Reagan, and DR Johnson. World Ocean Atlas 2013, Volume 3: Dissolved oxygen, apparent oxygen utilization, and oxygen saturation. In S Levitus (ed.), *NOAA Atlas NESDIS 75*. United States Department of Commerce, 2014. Mishonov, A. technical ed.

[22] HE Garcia, RA Locarnini, TP Boyer, JI Antonov, OK Baranova, MM Zweng, JR Reagan, and DR Johnson. World Ocean Atlas 2013, Volume 4: Dissolved inorganic nutrients (phosphate, nitrate, silicate). In S Levitus (ed.), *NOAA Atlas NESDIS 76*. United States Department of Commerce, 2014. Mishonov, A. technical ed.

[23] GEBCO Compilation Group. GEBCO_2022 Grid, 2022.

[24] Manuel González-Rivero, Pim Bongaerts, Oscar Beijbom, Oscar Pizarro, Ariell Friedman, Alberto Rodriguez-Ramirez, Ben Upcroft, Dan Laffoley, David Kline, Christophe Bailhache, et al. The Catlin Seaview Survey–kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 24(S2):184–198, 2014. doi:https://doi.org/10.1002/aqc.2505.

[25] Manuel González-Rivero, Oscar Beijbom, Alberto Rodriguez-Ramirez, Tadzio Holtrop, Yeray González-Marrero, Anjani Ganase, Chris Roelfsema, Stuart Phinn, and Ove Hoegh-Guldberg. Scaling up ecological measurements of coral reefs using semi-automated field image collection and analysis. *Remote Sensing*, 8(1):30, January 2016. ISSN 2072-4292. doi:10.3390/rs8010030. URL `http://www.mdpi.com/2072-4292/8/1/30`.

[26] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, koray kavukcuoglu, Remi Munos, and Michal Valko. Bootstrap your own latent - a new approach to self-supervised learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 21271–21284. Curran Associates, Inc., 2020. URL `https://papers.nips.cc/paper/2020/file/f3ada80d5c4ee70142b17b8192b2958e-Paper.pdf`.

[27] Peter T. Harris and Elaine K. Baker. Why map benthic habitats? In Peter T. Harris and Elaine K. Baker (eds.), *Seafloor Geomorphology as Benthic Habitat*, pp. 3–15. Elsevier, 2nd edition, 2020. ISBN 978-0-12-814960-7. doi:10.1016/B978-0-12-814960-7.00001-4. URL `https://linkinghub.elsevier.com/retrieve/pii/B9780128149607000014`.

[28] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. doi:10.1109/CVPR42600.2020.00975.

[29] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15979–15988, June 2022. doi:10.1109/CVPR52688.2022.01553.

[30] Jungseok Hong, Michael Fulton, and Junaed Sattar. TrashCan: A semantically-segmented dataset towards visual detection of marine debris. *arXiv preprint*, arXiv:2007.08097, 2020. doi:10.48550/arxiv.2007.08097. URL `https://arxiv.org/abs/2007.08097`.

[31] Mengxiao Huang, Yang Wang, Weiqi Zou, and Yang Cao. Fast adaptive self-supervised underwater image enhancement. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 3371–3375, 2022. doi:10.1109/ICIP46576.2022.9897298.

[32] Robert Huber, Egor Gordeev, Markus Stocker, Aarthi Balamurugan, and Uwe Schindler. pangaeapy - a Python module to access and analyse PANGAEA data, September 2020. URL `https://doi.org/10.5281/zenodo.4013941`.

[33] Chris Jackett, Franziska Althaus, Kylie Maguire, Moshiur Farazi, Ben Scoulding, Candice Untiedt, Tim Ryan, Peter Shanks, Pamela Brodie, and Alan Williams. A benthic substrate classification method for seabed images using deep learning: Application to management of deep-sea coral reefs. *Journal of Applied Ecology*, 60(7):1254–1273, 2023. doi:10.1111/1365-2664.14408.

[34] Muwei Jian, Qiang Qi, Junyu Dong, Yinlong Yin, Wenyin Zhang, and Kin-Man Lam. The OUC-vision large-scale underwater image database. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1297–1302, 2017. doi:10.1109/ICME.2017.8019324.

[35] A Jordan and P Hedge. Marine biodiversity hub impact report. synopsis of research impacts 2007–2020. Technical report, Department of Agriculture, Water and the Environment. Canberra, Australia, 2020. URL `https://www.nespmarine.edu.au/system/files/Marine_Biodiversity_Hub_Impact_Report_FINAL_DEC2020.pdf`.

[36] 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, September 2022. ISSN 2045-2322. doi:10.1038/s41598-022-19939-2. URL `https://www.nature.com/articles/s41598-022-19939-2`.

[37] Matthew S Kendall, Olaf P Jensen, Clark Alexander, Don Field, Greg McFall, Reed Bohne, and Mark E Monaco. Benthic mapping using sonar, video transects, and an innovative approach to accuracy assessment: a characterization of bottom features in the Georgia Bight. *Journal of Coastal Research*, 2005(216):1154–1165, 2005. doi:10.2112/03-0101R.1.

[38] Myriam Lacharité, Craig J Brown, and Vicki Gazzola. Multisource multibeam backscatter data: Developing a strategy for the production of benthic habitat maps using semi-automated seafloor classification methods. *Marine Geophysical Research*, 39(1):307–322, 2018. doi:10.1007/s11001-017-9331-6.

[39] Daniel Langenkämper, Martin Zurowietz, Timm Schoening, and Tim W. Nattkemper. BIIGLE 2.0 - Browsing and annotating large marine image collections. *Frontiers in Marine Science*, 4, 2017. ISSN 2296-7745. doi:10.3389/fmars.2017.00083. URL `https://www.frontiersin.org/article/10.3389/fmars.2017.00083`.

[40] Barbara H Lidz and David G Zawada. Possible return of *Acropora cervicornis* at Pulaski Shoal, Dry Tortugas National Park, Florida. *Journal of Coastal Research*, 29(2):256–271, 2013. doi:10.2112/JCOASTRES-D-12-00078.1.

[41] RA Locarnini, AV Mishonov, JI Antonov, TP Boyer, HE Garcia, OK Baranova, Zweng MM, CR Paver, JR Reagan, Johnson DR, M Hamilton, and D Seidov. World Ocean Atlas 2013, Volume 1: Temperature. In S Levitus (ed.), *NOAA Atlas NESDIS 73*. United States Department of Commerce, 2013. Mishonov, A. technical ed.

[42] Scott C. Lowe, Joakim Bruslund Haurum, Sageev Oore, Thomas B. Moeslund, and Graham W. Taylor. An empirical study into clustering of unseen datasets with self-supervised encoders. *arXiv preprint arXiv:2406.02465*, 2024. doi:10.48550/arxiv.2406.02465.

[43] J. Mackin-McLaughlin, S. Nemani, B. Misiuk, A. Templeton, P. Gagnon, Edinger, and K. Robert. Spatial distribution of benthic flora and fauna of coastal placentia bay, an ecologically and biologically significant area of the island of Newfoundland, Atlantic Canada. *Frontiers in Environmental Science*, 10:999483, 2022. doi:10.3389/fenvs.2022.999483.

[44] Marine and Coastal Spatial Data Subcommittee. Coastal and marine ecological classification standard (CMECS). Technical Report FGDC-STD-018-2012, United States Federal Geographic Data Committee, 2012. URL `https://repository.library.noaa.gov/view/noaa/27552`.

[45] Pedro S. Menandro, Ana Carolina Lavagnino, Fernanda V. Vieira, Geandré C. Boni, Tarcila Franco, and Alex C. Bastos. The role of benthic habitat mapping for science and managers: A multi-design approach in the Southeast Brazilian Shelf after a major man-induced disaster. *Frontiers in Marine Science*, 9:1004083, October 2022. ISSN 2296-7745. doi:10.3389/fmars.2022.1004083. URL `https://www.frontiersin.org/articles/10.3389/fmars.2022.1004083/full`.

[46] Pedro S. Menandro, Benjamin Misiuk, Craig J. Brown, and Alex C. Bastos. Multispectral multibeam backscatter response of heterogeneous rhodolith beds. *Scientific Reports*, 13(1): 20220, November 2023. ISSN 2045-2322. doi:10.1038/s41598-023-46240-7. URL `https://www.nature.com/articles/s41598-023-46240-7`.

[47] Benjamin Misiuk and Craig J. Brown. Benthic habitat mapping: A review of three decades of mapping biological patterns on the seafloor. *EarthArXiv*, September 2023. doi:10.31223/X5DD4S.

[48] Benjamin Misiuk, Trevor Bell, Alec Aitken, Craig J Brown, and Evan N Edinger. Mapping Arctic clam abundance using multiple datasets, models, and a spatially explicit accuracy assessment. *ICES Journal of Marine Science*, 76(7):2349–2361, 2019. ISSN 1054-3139. doi:10.1093/icesjms/fsz099.

[49] Benjamin Misiuk, Markus Diesing, Alec Aitken, Craig J Brown, Evan N Edinger, and Trevor Bell. A spatially explicit comparison of quantitative and categorical modelling approaches for mapping seabed sediments using random forest. *Geosciences*, 9(6):254, 2019. doi:10.3390/geosciences9060254.

[50] Richard Mount, Phillippa Bricher, and Jenny Newton. National Intertidal/Subtidal Benthic (NISB) habitat classification scheme. Technical report, Australian Greenhouse Office, National Land & Water Resources Audit, School of Geography and Environmental Studies, University of Tasmania, October 2007. URL https://ozcoasts.org.au/wp-content/uploads/2018/05/pn21267.pdf.

[51] Shreya Nemani, David Cote, Benjamin Misiuk, Evan Edinger, Julia Mackin-McLaughlin, Adam Templeton, John Shaw, and Katleen Robert. A multi-scale feature selection approach for predicting benthic assemblages. *Estuarine, Coastal and Shelf Science*, pp. 108053, 2022. ISSN 02727714. doi:10.1016/j.ecss.2022.108053. URL https://linkinghub.elsevier.com/retrieve/pii/S0272771422003110.

[52] John O'Brien, Melisa Wong, and Betty Roethlisberger. Benthic imagery from nearshore drop camera surveys along Eastern Shore (Nova Scotia, Canada) to characterize shallow subtidal habitats. Mendeley Data, 2022. Available at https://data.mendeley.com/datasets/xfby2gf6kp/1.

[53] Eric C Orenstein, Oscar Beijbom, Emily E Peacock, and Heidi M Sosik. WHOI-Plankton- A large scale fine grained visual recognition benchmark dataset for plankton classification. *arXiv preprint*, arXiv:1510.00745, 2015. doi:10.48550/arxiv.1510.00745.

[54] Yassine Ouali, Céline Hudelot, and Myriam Tami. An overview of deep semi-supervised learning. *arXiv preprint*, arXiv:2006.05278, 2020. doi:10.48550/arxiv.2006.05278.

[55] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[56] Malte Pedersen, Joakim Bruslund Haurum, Rikke Gade, and Thomas B Moeslund. Detection of marine animals in a new underwater dataset with varying visibility. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 18–26, 2019. URL https://openaccess.thecvf.com/content_CVPRW_2019/html/AAMVEM/Pedersen_Detection_of_Marine_Animals_in_a_New_Underwater_Dataset_with_CVPRW_2019_paper.html.

[57] Malte Pedersen, Niels Madsen, and {Thomas B.} Moeslund. No machine learning without data: Critical factors to consider when collecting video data in marine environments. *The Journal of Ocean Technology*, 16(3):21–30, October 2021. ISSN 1718-3200. Funding Information: Funded by the Independent Research Fund Denmark under case number 9131-00128B.

[58] N Piechaud, C Hunt, Pf Culverhouse, Nl Foster, and Kl Howell. Automated identification of benthic epifauna with computer vision. *Marine Ecology Progress Series*, 615:15–30, 2019. doi:10.3354/meps12925. URL https://www.int-res.com/abstracts/meps/v615/p15-30/.

[59] A Post, P E O'Brien, L K Armand, and A Carroll. Seafloor image annotations from the Sabrina upper slope, East Antarctica, Ver. 1. Australian Antarctic Data Centre Dataset, 2020.

[60] Alina Raphael, Zvy Dubinsky, David Iluz, Jennifer IC Benichou, and Nathan S Netanyahu. Deep neural network recognition of shallow water corals in the Gulf of Eilat (Aqaba). *Scientific reports*, 10(1):1–11, 2020. doi:10.1038/s41598-020-69201-w.

[61] Chris M. Roelfsema, Eva M. Kovacs, Juan Carlos Ortiz, David P. Callaghan, Karlo Hock, Mathieu Mongin, Kasper Johansen, Peter J. Mumby, Magnus Wettle, Mike Ronan, Petra Lundgren, Emma V. Kennedy, and Stuart R. Phinn. Habitat maps to enhance monitoring and management of the Great Barrier Reef. *Coral Reefs*, 39:1039–1054, August 2020. doi:10.1007/s00338-020-01929-3.

[62] Christiaan M Roelfsema, Eva M Kovacs, and Stuart R Phinn. Georeferenced photographs of benthic photoquadrats acquired along 160 transects distributed over 23 reefs in the Cairns to Cooktown region of the Great Barrier Reef, January and April/May, 2017, 2017. URL https://doi.org/10.1594/PANGAEA.877578.

[63] Christiaan M Roelfsema, Eva M Kovacs, Douglas Stetner, and Stuart R Phinn. Georeferenced benthic photoquadrats captured annually from 2002-2017, distributed over Heron Reef flat and slope areas, 2018. URL https://doi.org/10.1594/PANGAEA.894801.

[64] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, April 2015. doi:10.1007/s11263-015-0816-y.

[65] Roger G. Sayre, Dawn J. Wright, Sean P. Breyer, Kevin A. Butler, Keith Van Graafeiland, Mark J. Costello, Peter T. Harris, Kathleen L. Goodin, John M. Guinotte, Zeenatul Basher, Maria T. Kavanaugh, Patrick N. Halpin, Mark E. Monaco, Noel Cressie, Peter Aniello, Charles E. Frye, and Drew Stephens. A three-dimensional mapping of the ocean based on environmental data. *Oceanography*, 30, March 2017. doi:10.5670/oceanog.2017.116.

[66] Timm Schoening, Jonas Osterloff, and Tim W Nattkemper. RecoMIA—Recommendations for marine image annotation: Lessons learned and future directions. *Frontiers in Marine Science*, 3: 59, 2016. doi:10.3389/fmars.2016.00059.

[67] J Smith. High resolution still photographs of the seafloor across the Mertz Glacier Region, Ver. 1. Australian Antarctic Data Centre Dataset, 2017.

[68] Leslie N. Smith. A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. *arXiv preprint*, arXiv:1803.09820, 2018. doi:10.48550/arxiv.1803.09820. URL http://arxiv.org/abs/1803.09820.

[69] BJ Todd, AA Boyce, CB Chapman, VE Kostylev, WA Rainey, Spencer PL, and MS Uyesugi. Expedition report 2000-047: CCGS *Hudson*, southern Scotian Shelf. Technical Report Open File 3911, Geological Survey of Canada, 2001. URL https://doi.org/10.4095/212966.

[70] BJ Todd, KW Asprey, AS Atkinson, R Blasco, S. Fromm, PR Girouard, VE Kostylev, O Longva, T Lynds, WA Rainey, PL Spencer, MS Uyesugi, and PC Valentine. Expedition report CCGS *Hudson* 2002-026: Gulf of Maine. Technical Report Open File 1468, Geological Survey of Canada, 2003. URL https://doi.org/10.4095/214143.

[71] BJ Todd, WA Rainey, PR Girouard, PC Valentine, CB Chapman, GD Middleton, AA Boyce, MS Uyesugi, VE Kostylev, ST Rumbolt, SA Fromm, SE Hynes, and PL Spencer. Expedition report CCGS *Hudson* 2003-054: German Bank, Gulf of Maine. Technical Report Open File 4728, Geological Survey of Canada, 2005. URL https://doi.org/10.4095/216676.

[72] M John Tremblay, Stephen J Smith, Brian J Todd, Pierre M Clement, and David L McKeown. Associations of lobsters (*Homarus americanus*) off southwestern Nova Scotia with bottom type from images and geophysical maps. *ICES Journal of Marine Science*, 66(9):2060–2067, 2009. doi:10.1093/icesjms/fsp178.

[73] Jesper E. van Engelen and Holger H. Hoos. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440, Feb 2020. ISSN 1573-0565. doi:10.1007/s10994-019-05855-6. URL https://doi.org/10.1007/s10994-019-05855-6.

[74] Stefan B Williams, Oscar Pizarro, Michael Jakuba, and Neville Barrett. AUV benthic habitat mapping in south eastern Tasmania. In *Field and Service Robotics*, pp. 275–284. Springer, 2010. doi:10.1007/978-3-642-13408-1_25.

[75] Brittany R Wilson, Craig J Brown, Jessica A Sameoto, Myriam Lacharité, Anna M Redden, and Vicki Gazzola. Mapping seafloor habitats in the Bay of Fundy to assess megafaunal assemblages associated with *Modiolus modiolus* beds. *Estuarine, Coastal and Shelf Science*, 252:107294, 2021. ISSN 0272-7714. doi:10.1016/j.ecss.2021.107294.

[76] WoRMS Editorial Board. World register of marine species (WoRMS). https://www.marinespecies.org, 2024. Accessed: 2023-04-03.

[77] Takaki Yamada, Adam Prügel-Bennett, and Blair Thornton. Learning features from georeferenced seafloor imagery with location guided autoencoders. *Journal of Field Robotics*, 38(1): 52–67, 2021. doi:10.1002/rob.21961.

[78] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):8934–8954, 2023. doi:10.1109/TKDE.2022.3220219.

[79] Yang You, Igor Gitman, and Boris Ginsburg. Large batch training of convolutional networks. *arXiv preprint*, arXiv:1708.03888, 2017. doi:10.48550/arXiv.1708.03888. URL http://arxiv.org/abs/1708.03888.

[80] Roman N. Zajac, Lauren M. Stefaniak, Ivar Babb, Christian W. Conroy, Shannon Penna, Deena Chadi, and Peter J. Auster. Chapter 10 - An integrated seafloor habitat map to inform marine spatial planning and management: a case study from Long Island Sound (Northwest Atlantic). In Peter T. Harris and Elaine Baker (eds.), *Seafloor Geomorphology as Benthic Habitat*, pp. 199–217. Elsevier, 2nd edition, 2020. ISBN 978-0-12-814960-7. doi:10.1016/B978-0-12-814960-7.00010-5. URL https://www.sciencedirect.com/science/article/pii/B9780128149607000105.

[81] David G Zawada, Rob Ruzicka, and Michael A Colella. A comparison between boat-based and diver-based methods for quantifying coral bleaching. *Journal of Experimental Marine Biology and Ecology*, 467:39–44, 2015. doi:10.1016/j.jembe.2015.02.017.

[82] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stephane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12310–12320. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/zbontar21a.html.

[83] Peiqin Zhuang, Yali Wang, and Yu Qiao. Wildfish: A large benchmark for fish recognition in the wild. In *Proceedings of the 26th ACM International Conference on Multimedia*, MM ’18, pp. 1301–1309, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356657. doi:10.1145/3240508.3240616. URL https://doi.org/10.1145/3240508.3240616.

[84] MM Zweng, JR Reagan, JI Antonov, RA Locarnini, AV Mishonov, TP Boyer, HE Garcia, OK Baranova, DR Johnson, D Seidov, and MM Biddle. World Ocean Atlas 2013, Volume 2: Salinity. In S Levitus (ed.), *NOAA Atlas NESDIS 74*. United States Department of Commerce, 2013. Mishonov, A. technical ed.

Appendices

A Data collation—additional details

As described in the main text (§2.1), we collated imagery from a variety of sources and converted them to a standardized format (Tab. 5). In this section, we describe our protocol for each source. The distribution of sources around the world is indicated in Fig. 4.

Table 5: **Format for compiled BenthicNet-1M unlabelled image metadata.**

Column	Contents	Data-type	Units	Coverage
url	URL address for this image	String		100.00%
source	Data provider/repository	String		100.00%
dataset	Name of dataset	String		100.00%
site	Image location name	String		100.00%
image	Image filename	String		100.00%
latitude	Latitude (WGS 84)	Float	Decimal degree	99.63%
longitude	Longitude (WGS 84)	Float	Decimal degree	99.63%
datetime	Acquisition date and time (UTC)	String	YYYY-MM-DD HH:mm:ss	99.85%
gebco_bathymetry	Depth interpolated from GEBCO2022	Float	Metres	99.63%
emu	Nearest ecological marine unit	Integer		99.63%

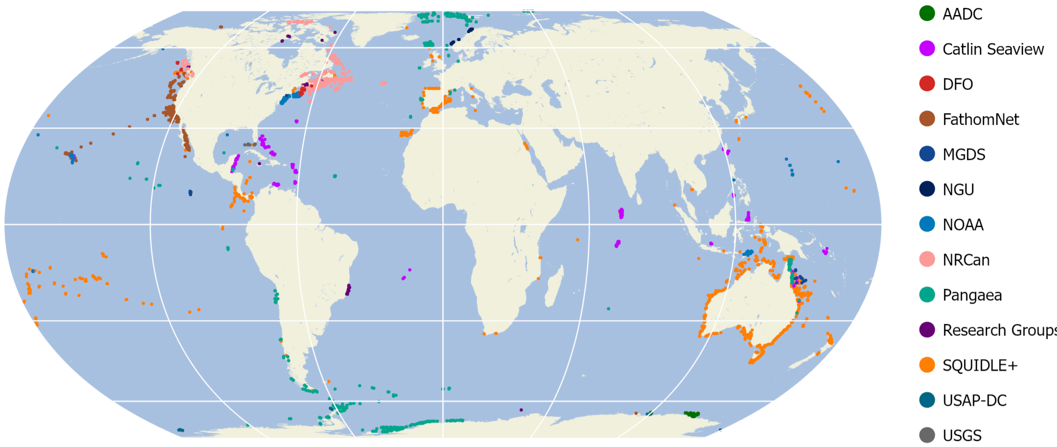

Figure 4: **Distribution of images** projected to Equal Earth, according to data source.

A.1 Individual contributions

A number of datasets were contributed by individual project partners; several of these were from eastern Canada. The Seascape Ecology and Mapping (SEAM) Lab at Dalhousie University provided three datasets for the BenthicNet compilation from this region. Still images were provided ($n = 2281$) that were extracted from passive drop down video drifts conducted in the Bay of Fundy at 281 sites between 2017–2019 using a 4k camera system [75]. Whole-image labels were supplied according to site-specific "benthoscapes" interpreted by the image analyst, which are recognizable combinations of dominant substrate type and biological characteristics [8, 47]. All megafauna were additionally identified to the highest possible taxonomic resolution for each image. A dataset of high definition benthic photographs ($n = 4064$) was also provided from surveys conducted between 2009–2014 at the St Anns Bank marine protected area [38], which included whole-image benthoscape labels defined for the site. Finally, the SEAM lab contributed photographs of the seabed ($n = 62$) used for the 2017 R2Sonic Multispectral Challenge in the Bedford Basin, Nova Scotia [10], which included broad whole-image substrate descriptions and, occasionally, biological observations. The 4D Oceans lab at the Fisheries and Marine Institute of Memorial University of Newfoundland provided still images ($n = 3000$) extracted from underwater video, as part of the project "Coastal Habitat Mapping of Placentia Bay" conducted off the coast of Newfoundland, which included whole-image substrate-derived bottom class labels [43, 51]. The Ecology Action Centre (EAC) provided 1 220 images collected by citizen scientists via Go Pro-mounted kayak between 2019–2021 at shallow eelgrass sites in Nova Scotia. These included whole-image labels for the presence or absence of eelgrass (*Zostera marina*).

Several datasets collected by researchers at Memorial University of Newfoundland (MUN) were also contributed from northern Canada. These included 895 images collected for a benthic mapping project in Frobisher Bay, Nunavut, between 2015–2016 [49]; 1059 images from Wager Bay, Nunavut, collected in collaboration with Parks Canada as part of the Ukkusiksalik National Park Marine Baseline Data Collection Project; 541 images from Chesterfield Inlet, Nunavut, collected for a local benthic habitat mapping project conducted in coordination with the Government of Nunavut, and University of Manitoba; and 8 443 images from the area around Qikiqtarjuaq, Nunavut, which were obtained as part of a mapping campaign to monitor a locally harvested soft-shell clam population [48]. These datasets were each accompanied by site-specific whole-image labels describing the dominant substrate types visible in each image.

Several image datasets were provided by the Hakai Institute from western Canada. A total of 8 787 images were obtained from nearshore benthic surveys conducted between 2017–2020 from sites on the central coast of British Columbia and sites within Pacific Rim National Park Reserve (PRNPR). This data was comprised of still images from ROV deployments and GIS-annotated drop camera videos collected primarily for the purposes of mapping eelgrass meadows (*Zostera marina*). Still images were extracted from videos using the methods described above (i.e. using FFmpeg). Whole-image labels were provided corresponding to the dominant visible substrate and vegetation type present in each image.

Individual datasets were also acquired from outside Canada. The Marine Geosciences Lab (LaboGeo) at Universidade Federal do Espirito Santo (UFES) provided quadrat sample images acquired by drop camera during rhodolith surveys off the east coast of Brazil between 2015–2020 [45, 46]. These were cropped to remove the quadrat frame, and 360 images were included in the BenthicNet compilation. Whole-image labels were provided that identify the presence of rhodoliths and select biogenic substrate types. A dataset of 505 images was provided by the Hierarchical Anticipatory Learning (HAL) lab at Dalhousie University, which was collected from Ocho Rios, Jamaica, in shallow water by snorkeler in 2022. Images were unlabelled, and comprised coral reef and a range of substrate types.

A.2 DFO

Fisheries and Oceans Canada (DFO) is a federal institution responsible for managing many of Canada's marine resources. DFO provided three separate contributions to the BenthicNet compilation. The Population Ecology Division at the Bedford Institute of Oceanography (BIO) contributed 645 annotated images from George's Bank, which separates the Gulf of Maine from the Northwest Atlantic. These images were collected by the Geological Survey of Canada (GSC) Atlantic for programs under Natural Resources Canada (NRCan) using the Campod digitial camera system deployed from the CCGS *Hudson* in 2000 [69] and 2002 [70]. Annotations included whole-image benthoscape labels describing the primary substrate and presence of characteristic biota. Benthic images were also contributed from a GSC survey on German Bank off the southwest coast of Nova Scotia in 2003 [71] using Campod ($n = 641$), and from DFO Ecosystems and Ocean Science Sector surveys in 2006 [72] ($n = 2044$), and 2010 [9] ($n = 3181$) using the Towcam underwater imaging platform. These images included whole-image labels describing the dominant visible substrate type, some of which additionally included detailed comments describing the proportion of cover for multiple substrate types. A separate contribution from the Habitat Ecology Section at BIO comprised 1 262 images from coastal eelgrass and macroalgae surveys along the Eastern Shore of Nova Scotia between 2019 and 2020 [52]. These images were extracted from video footage captured by a GoPro HERO7 (1080p or 2.7k resolution) deployed from a drop-down platform for passive drifts at 269 sites. Substrate labels were provided at the whole-image level according to the Coastal and Marine Ecological Classification Standard (CMECS) [44], as were labels for particular biota, including macroalgae and seagrasses. Finally, the DFO Deep-sea Ecology Program at the Institute of Ocean Sciences (IOS), British Columbia, contributed data collected during the 2018 Northeast Pacific Seamount Expedition using the ROV Hercules. Northeast Pacific Seamount Expedition Partners and Ocean Exploration Trust collected imagery at SGaan Kinghlas-Bowie, Explorer, and Dellwood Seamounts off the west coast of Canada in 2018. Video frames were extracted every 10 seconds for analysis, and 16 247 were included here. Labels were provided for some images describing the primary substrate type and also the "biotope" observed, which broadly describes the benthic community and/or habitat context (e.g. coral garden, vertical wall, sponge ground). Some images overlapped and were thus not originally labelled; in such cases, neighbouring image labels were interpolated where not initially assigned due to overlap with other images.

A.3 NRCan

Natural Resources Canada (NRCan) is a federal organization responsible for managing and researching a range of natural resources at the national scale. NRCan makes data freely available via the Canada Open Government Portal. The NRCan/GSC Seabed Photo Collection was acquired for this project, which includes 20 260 images recorded from 1 804 camera stations across 78 expeditions distributed throughout the waters surrounding Canada. These photographs were collected between 1965 and 2015 using a range of equipment; photographs taken before 1978 were in greyscale, and after 1978 in colour. Photographs before 2000 were collected using film and after 2004 were digital, with both used in the interim. 3 767 of the photographs were annotated with verbose descriptions of either geological features, biological contents, or both. These descriptions were parsed in order to apply whole-image substrate and biota labels (see §2.2). The full list of expeditions associated with this dataset was obtained along with URLs of corresponding metadata CSV files in GeoDataBase format from the NRCan FTP server. The GeoDataBase file was processed with geopandas, and CSV files were downloaded for each expedition location (URLs were manually corrected for expedition `82FOGO-ISLE`, for which the CSV files were available at URLs containing the string `82FOGO_ISLE` instead). These CSV files, containing URLs for individual images from the expeditions, were merged together. The year of acquisition was inferred from the expedition name, and columns were renamed to match the standardized dataset format. Sample images were inspected from each expedition to verify their appropriateness. All images from expedition 71014 consisted of collages formed of 2–6 individual photographs, and were excluded.

A.4 NGU

The Geological Survey of Norway (NGU) is responsible for national geological mapping and research, including marine applications. NGU contributed 50 290 images to this project, which were extracted from 581 underwater video transects acquired during six cruises. These were carried out between 2010 and 2017 in coastal areas and fjords of Norway (Astafjorden, Frohavet, Søre Sunmøre, Sogn og Fjordane, Ofoten, Tysfjorden, and Tjeldsundet), as part of several "Marine Base Maps" projects. The videos were acquired using a camera rig towed near the seafloor (0 m to 200 m depth) from the NGU research vessel *Seisma*. The 2010 cruises (codes 1002 and 1007) used a 720x480 digital video camera, while all the other cruises (codes 1408, 1508, 1511, and 1706) used a higher-resolution GoPro HERO3+. The images were obtained by extracting one video frame every 10 seconds of video footage.

A.5 MGDS

The Marine Geoscience Data System (MGDS) is a data repository that offers public access to a curated collection of marine geophysical data products and complementary data related to understanding the formation and evolution of the seafloor and sub-seafloor. MGDS provides tools and services for the discovery and download of data collected throughout the global oceans produced primarily by researchers funded by the U.S. National Science Foundation. Six datasets were obtained from MGDS, in collaboration with the Lamont-Doherty Earth Observatory at Columbia University. Four of these were collected from the Long Island Sound Estuary in 2012 and 2013 using the United States Geological Survey (USGS) Seabed Observation and Sampling System (SEABOSS), Integrated Seafloor Imagery System camera sled, and the Kraken2 ROV [80]. One dataset was obtained from the East Pacific Rise Spreading Center during the 2011 Atlantis expedition, using an Insite Scorpio Digital Camera mounted on the ROV Jason II. The final dataset was acquired by the Schmidt Ocean Institute (SOI) during the 2020 R/V *Falkor* expedition FK200429 off the northeast coast of Australia. Here, the ROV SuBastian was mobilized and images were obtained using a Subsea Systems and Inc. Z70 Digital Camera. All datasets from MGDS were manually reviewed and filtered to remove surface images (e.g. on the research vessel) and duplicates.

A.6 NOAA

The U.S. National Oceanic and Atmospheric Administration (NOAA) is a federal science institution that conducts extensive marine research. NOAA hosts diverse collections of environmental data that are made available to the public. Benthic images were sourced from the NOAA data repository for addition to the BenthicNet dataset. Candidate data were identified using the NOAA OneStop portal, using the search strings "benthic", "habitat", "image", "camera", and "photograph". Datasets returned not containing image files were rejected. The remainder were reviewed manually, and datasets were additionally rejected that did not meet quality or content standards. Reasons for rejection included substantial proportions of non-benthic images (e.g. above-water, pelagic, individual animals, air photos), partial or full scene obstruction by non-benthic objects (e.g. equipment, ROV/AUV parts), highly inconsistent image content or quality, and incoherent dataset or metadata formatting (e.g. unorganized collections of various types of data, metadata not readable via script). Datasets were also excluded that did not meet the metadata requirements of this project—namely, those lacking metadata entirely, or lacking geographic locations for images. Where the latter occurred, efforts were made to estimate image locations using available information; for example, by assigning general study site coordinates to images, or by assigning the mean geographic centre of other images at the study site. Datasets that were otherwise suitable for inclusion were generally not rejected due to poor image quality or low resolution alone. All datasets were subjected to the quality control checks listed previously before downloading for inclusion in the BenthicNet collection, and columns were renamed to match the standardized dataset format. Several datasets included labels associated with the National Coral Reef Monitoring Program (NCRMP) describing the benthic cover, which primarily comprised coral taxa and substrate labels applied to both whole-images and points. These labels were retained.

Additional data was contributed by the NOAA Northeast Fisheries Science Center (NEFSC). These included benthic images from Georges Bank, the Mid-Atlantic Bight, and off the coast of Cape Cod ($n = 2240$). Image surveys were conducted in 2015 using the NOAA HabCam benthic imaging platform. Whole-image labels were provided indicating the primary and secondary substrate types, and also the presence of certain taxa (mussels, *Didemnum* tunicates, bryozoans).

A.7 USGS

The United States Geological Survey (USGS) is a federal organization that conducts earth science research and provides public geoscience information and data. A series of unlabelled benthic image datasets were retrieved from the USGS Science Data Catalogue. Several of these were initially discovered from review of the scientific literature [40, 81], and the remainder were discovered by querying the repository using the search strings "benthic", "habitat", "image", "camera", and "photograph". Candidate datasets were screened using the same methodology as outlined above for data retrieved from the NOAA repository. Datasets were rejected that did not contain images, contained non-benthic images, were largely obstructed by non-benthic objects, or were formatted incoherently. Where precise image locations were not provided, estimates were obtained using the mean centre of the study site bounding box coordinates. All candidate datasets were subjected to the quality control checks listed previously and columns were renamed to match the standardized dataset format.

A.8 USAP-DC

The U.S. Antarctic Program Data Center (USAP-DC) is funded by the U.S. National Science Foundation and is a domain repository for U.S. Antarctic Research data from all disciplines. Five unlabelled datasets were obtained from USAP-DC. These were discovered from the USAP-DC website using the search strings "benthic", "habitat", "image", "camera", and "photograph". Datasets were screened using the methodology described for the NOAA and USGS repositories. Additionally, some images that did not depict the seabed (e.g. pictures on the boat deck) were manually omitted. The mean centre of the study site bounding boxes were used to estimate image locations where precise positioning was not provided. These were checked for quality using the methodology described previously and columns were renamed to match the standardized dataset format.

A.9 AADC

The Australian Antarctic Data Centre (AADC) is a long-term repository for Australia's Antarctic data. This data is freely and openly available for scientific use. Two datasets were obtained for this project from the AADC data portal. Seafloor images ($n = 203$) from the Sabrina slope, East Antarctica, were collected in 2017 over four transects during survey "IN2017_V01" using the Australian CSIRO Marine National Facility's Deep Tow Camera [59], and were downloaded along with associated metadata from AADC. These included whole-image labels indicating the substrate type coverage and the presence of biota; the former were retained here. Additionally, Geoscience Australia and the Australian Antarctic Division collected underwater photographs in 2011 at 97 sites in the Mertz Glacier region of Antarctica [67], and 1 853 images were acquired for this project. Images and metadata from both datasets were checked for quality and formatted for standardization with the BenthicNet compilation.

A.10 SQUIDLE+

SQUIDLE+ is an online tool for managing, exploring, and annotating images and video of the seafloor. It also serves as a global repository, containing standardized records for images collected by different groups around the world. SQUIDLE+ is a living product that is updated continuously with new images and labels. A snapshot of the images available on SQUIDLE+ was acquired on April 13, 2023. The SQUIDLE+ web API was used to download the records for every image on SQUIDLE+, totalling 9 166 472 at that time. The paginated download was joined together and merged into a single CSV file, and columns were renamed to match our standardized format for the compilation.

Several of the large individual SQUIDLE+ datasets in this collection additionally included publicly accessible image annotations. These included Australia's Integrated Marine Observing System (IMOS), which distributes oceanographic data from a consortium of Australian institutions that is freely and openly available to the scientific community. This data included a large number of images collected by the IMOS AUV Facility, notably, using Sirius and Nimbus AUVs. IMOS images available from SQUIDLE+ were cross-referenced with data entries from the Australian Ocean Data Network (AODN) portal for this project. Labelled images were also provided by the Reef Life Survey (RLS) [19, 20], which is a global citizen science program that trains SCUBA divers to conduct underwater visual surveys of shallow reef biodiversity in temperate and tropical reef habitats, typically between 2 m – 20 m depth. Divers capture approximately 20 images per survey using an underwater camera positioned approximately 50 cm from the substrate, and images vary in resolution and quality due to camera configuration and environmental conditions. The Schmidt Ocean Institute (SOI) is a non-profit foundation established to advance global oceanographic research that hosts a large labelled image collection on SQUIDLE+. Deployed from the SOI R/V *Falkor*, the ROV *SuBastian* has collected high resolution images from waters around the world, including the deep ocean. All oceanographic data collected by the SOI are made openly available for research purposes. The National Environmental Science Program (NESP) Marine Biodiversity Hub [35] has also provided a large labelled image dataset. This project aims to provide foundational science for conservation in Australian and provides data openly in support of marine research. Each of the above datasets included sub-image point labels identifying underlying physical or biological elements according to the CATAMI scheme [1]. Finally, the image dataset presented by Yamada et al. [77] collected via AUV from the Southern Hydrate Ridge was downloaded from a separate SQUIDLE domain, SOI SQUIDLE+, along with point annotations describing substrate or biotic elements according to a site-specific scheme.

A.11 FathomNet

FathomNet [36] is an open-source underwater image database with global scope operated by the Monterey Bay Aquarium Research Institute (MBARI). FathomNet is soliciting contributions from around the world to develop a large open-source database of images that may be used to develop artificial intelligence algorithms, with a focus on identifying marine species. Like SQUIDLE+, FathomNet is a living product that is updated continuously. We used the FathomNet Python API to download a snapshot of the images available on FathomNet as of April 6, 2023. The code for this API call is provided in Appx. F. At the time of downloading, these images were primarily acquired from Pacific Waters around California, Western USA. Records were partitioned into "sites" based on the directory structure in the URL. Where not available in the record itself, timestamps were extracted from image names, where possible. Columns were renamed to match our standardized format. Many of the images were annotated with bounding boxes around animals and other concepts appearing in the images. However, annotations were available only under a No Derivatives license (CC BY-ND 4.0), which prohibited conversion to other schemes and formats. All FathomNet annotations were thus discarded.

A.12 PANGAEA

PANGAEA is an open access repository aimed at archiving, publishing and distributing georeferenced data from earth system research, hosting 678 projects and 408 811 datasets from various fields at the time of writing. We searched and retrieved benthic image datasets from PANGAEA with a combination of API calls and web-scraping, then pruned the resulting datasets and reformatted them. The pangaeapy Python package [32] was used to interface with the PANGAEA library. Using the `PanQuery` API, PANGAEA was searched for 20 queries with various combinations of benthic environment related keywords to find photographs of the seafloor (see Appx. E for complete list). The `PanDataSet` API was used to retrieve the metadata for the dataset IDs identified in these searches. Some IDs corresponded to dataset series, which list multiple child datasets. In these cases, all child datasets were retrieved. Some datasets were available in tabular format, and were downloaded directly. Other datasets were paginated, with images hosted on webpages on PANGAEA; these could not be downloaded with the API and were scraped with a custom webscraper using the `BeautifulSoup4` and `request` libraries.

All datasets returned by this search as of January 1, 2024 were downloaded and results were filtered as follows. (1) Datasets that did not possess a column containing the word "url" or "image" that was populated by hyperlinks to files in an image format (TIFF, JPEG, PNG, BMP, CR2) were removed to enable automation of the data acquisition process. It was not possible to verify any ZIP file would contain images without downloading it, and was impractical to automatically associate metadata with the images within a ZIP file of unknown structure. Datasets with images only available to download as a ZIP file were thus discarded. (2) False positives from the search (datasets comprising imagery not of the seafloor) were filtered out by removing datasets with titles containing undesired keywords appearing in a manually curated blacklist (e.g. "aquarium", "meteorological observations", "sea ice conditions", "do not use"). (3) URLs for images consisting of maps, other dataset summary figures, and inappropriate photo subjects were filtered out by removing data hosted on PANGAEA subdomains dedicated to subjects such as maps, projects, publications, sea ice, and satellite imagery. (4) Images were removed where the URL contained text indicating the subject matter was otherwise inappropriate (e.g. "dredgephotos", "grabsample", "core", "aquarium", "divemap"). Finally, the columns in the CSV files were renamed to our standardized format.

Several of the datasets obtained from PANGAEA included thematic labels corresponding to benthic images. Many of these were labels of specific biota identified to the highest possible taxonomic resolution, some of which included estimates of percentage cover of each organism in the image. Several of the latter datasets comprised experimental growth plates harbouring the labelled biota. Some datasets additionally included labels for trash and anthropogenic debris. All labels were dropped where datasets indicated usage of machine-assisted annotation instead of manual annotation. Finally, additional point labels were obtained for datasets from the Great Barrier Reef Marine Park, eastern Australia, collected for habitat mapping purposes by the University of Queensland Remote Sensing Research Centre. These datasets comprised quadrat images collected via snorkel and diving from over 100 reefs throughout the Great Barrier Reef Marine Park [62, 61, 63]. Points were labelled according to a custom scheme used for these projects at the Great Barrier Reef that describe biotic and abiotic elements found within the reef. Additional labels were also provided indicating the biotic functional group, and a simplified classification scheme applicable to a global context.

A.13 XL Catlin Seaview Survey

The XL Catlin Seaview Survey was a large-scale project undertaken between 2012–2018 to document and study the status of coral reefs globally using underwater imagery. Surveys focused on shallow reefs typically around 10 m depth and comprised linear transects ranging between 1.6 km to 2 km in length. Downward-facing seabed images of approximately 1 m^2 were acquired using Canon 5D MII cameras mounted on a self-propelled diver-operated platform called the "SVII" [24, 25]. Data from the project is made openly available for further scientific research. For this project, 1 082 452 images from 860 surveys organized into 22 regional datasets were downloaded from the University of Queensland data repository. Tabular data providing image metadata was also acquired in CSV format, including image point labels identifying biotic and abiotic elements using the global scheme applied above for the Great Barrier Reef mapping projects. The metadata were renamed and formatted to match the standardized BenthicNet compilation.

B Spatial subsampling

The aim of the subsampling procedure was to obtain a manageable unlabelled data volume without reducing the breadth of benthic environments represented. Many datasets indicated which images were collected at the same recording station, or the same camera deployment/transect. We collectively refer to this location annotation as a "site". To maximize spatial and thematic diversity of images, subsampling was performed separately for each unique site in the unlabelled datatset.

In order to subsample the data spatially, we first determined a desirable number of images that should be drawn from a given site based on the data density. The base target number of images sought at each site was set to 250, meaning that the subsampling procedure would not reduce the number of images below this number. Not all component datasets indicated whether images were collected at the same site, despite containing images from multiple distinct locations that would meet our "site" criteria. To address this, we automatically detected the number of "pseudo-sites" within an annotated site, or within a dataset originally lacking any site labels. Pseudo-sites were determined as clusters of samples at least 1 000 m from each other. The target number of samples was scaled up by the number of pseudo-sites within a labelled site. Some (pseudo-)sites additionally had gaps between them of several hundred metres, which we refer to as "subsites". The target number of samples for a site was increased by 50 for each subsite within it separated by at least 100 m.

After determining the target number of images to draw from each site in the unlabelled dataset, the data was subsampled spatially. Sites with fewer than 40 samples per pseudo-site were not subsampled. At sites with more than 40 images, images were subsampled with a target separation distance of $\Delta = 1.25$ m according to the following procedure.

1. Add the first image in the dataset.
2. Continue through the list of images in the dataset (sorted in collection order; i.e. chronologically) until finding the first image at least $\Delta = 1.25$ m from the last image added to the dataset.
3. Add either this image or the previous image in the list, whichever is closest to being a distance $\Delta = 1.25$ m from the last image added to the dataset.
4. From the list of remaining images to consider, remove all images collected within $\Delta/2 = 0.625$ m of this image.
5. Return to Step 2; repeat until reaching the end of the dataset.
6. Add the last image in the dataset if it was at least $\Delta/2 = 0.625$ m from all other images.

Sites lacking precise coordinate information for each image could not be subsampled spatially. In these cases, sites were subsampled by keeping every n-th image (ordered chronologically) at the site to achieve the desired number.

Many sites still had more images than their target number of samples after this initial spatial subsampling, so this process was repeated with larger separation distances until the target subsample size was achieved at each site, or a maximum downsampling separation distance of 20 m was reached. Separation distances were scaled up by factors of 2, 3, 4, 6, 8, 10, 12, 14, or 16 compared to the base subsampling of 1.25 m target separation to achieve the desired subsample size (i.e. $\Delta = 2.5$ m, 3.75 m, ..., 20 m). The subsampling distance selected (and hence subsampled set of images at that site) was the largest distance that did not reduce the total number of images below the target for the site (250+), determined as described above. The subsampling procedure selected 1 345 096 images (11.8% of the total) to be included in the subsampled BenthicNet dataset (Fig. 1), which we refer to as BenthicNet-1M.

C Ecological Marine Units

Sayre et al. [65] introduced a three-dimensional partitioning of the global oceans into statistical clusters based on a 57-year climatology of physiochemical oceanographic measurements [41, 84, 21, 22]. These 37 "ecological marine units" (EMUs) represent a concise and objective summary of global marine environments at 0.25° horizontal resolution. The bottom-layer EMUs were extracted to assess the distribution of BenthicNet image samples across global benthic environmental regions. Each image was assigned the nearest bottom-layer (i.e. seafloor) EMU in space to compare the sampled frequency of each environment to the proportion of area covered by each EMU (Figs. 5 and 6).

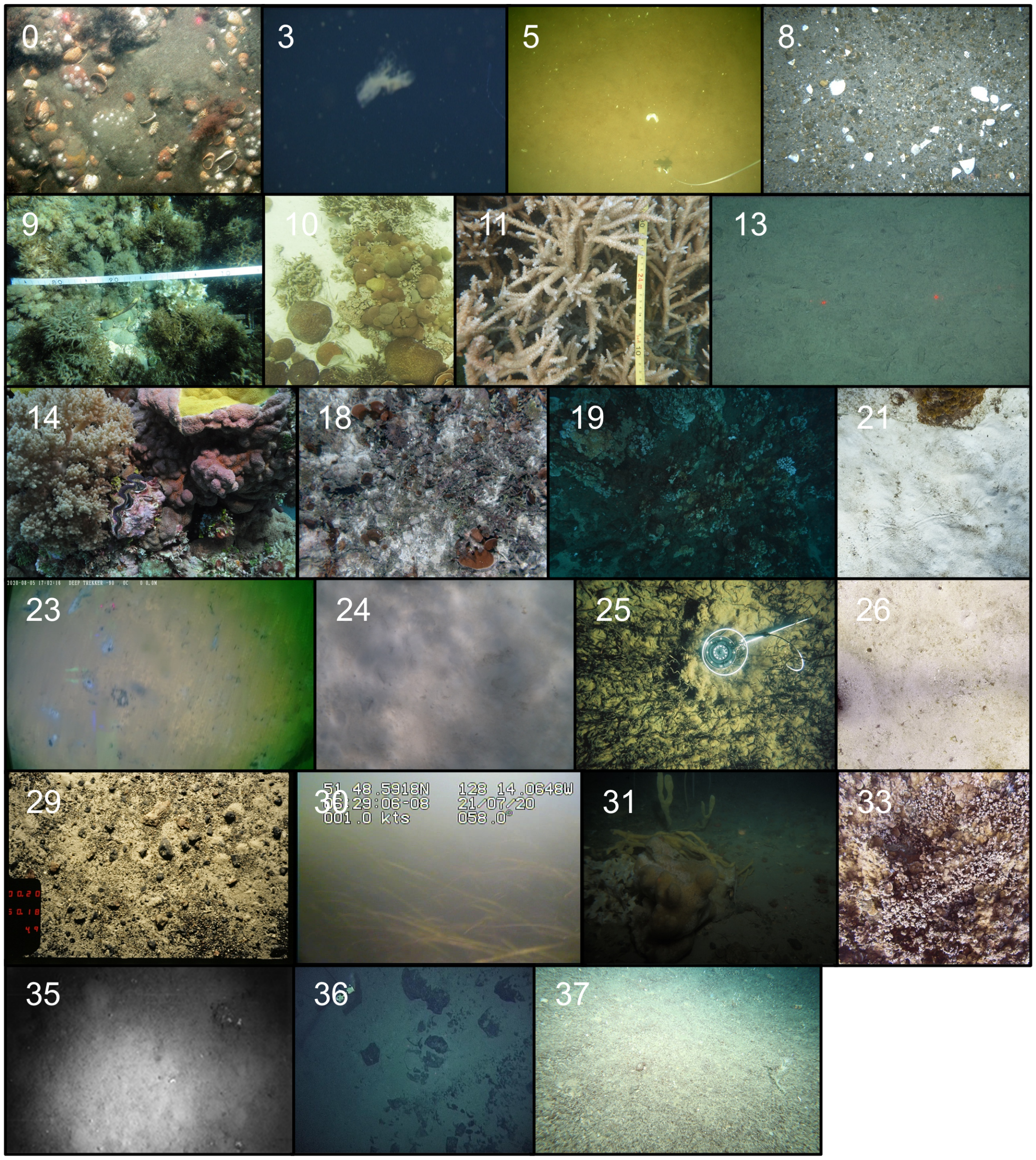

Figure 5: **Examples of BenthicNet images** from each sampled ecological marine unit (EMU), indicated by a number in white overlaid in the top-left of each image. See Sayre et al. [65] for a full description of the EMU classes.

C.1 Unlabelled data exploratory analysis

Generally, images were distributed more evenly across the bottom-layer EMUs than would be expected from a random sample, while the distribution across the major ocean basins more closely

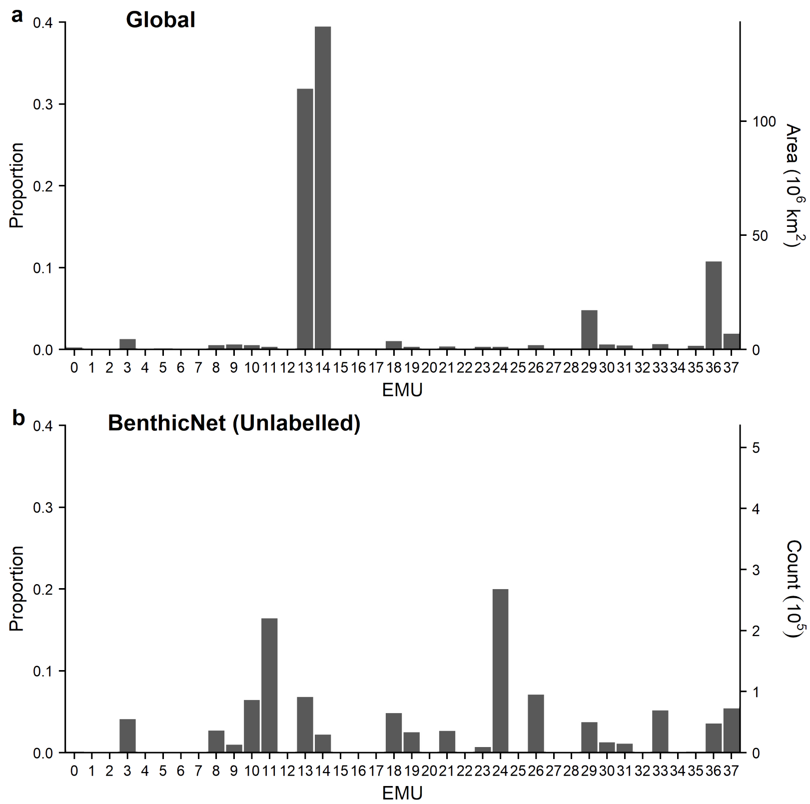

Figure 6: **Distribution of BenthicNet-1M images according to bottom layer ecological marine units** (EMUs). (a) Proportion and area of global oceans classified into each EMU. (b) Proportion of BenthicNet image samples from each EMU. See Sayre et al. [65] for a full description of the EMU classes.

matched expectation. The majority of the global seafloor (82.4%) is classified into EMUs 14 (deep, very cold, normal salinity, moderate oxygen, high nitrate, low phosphate, high silicate), 13 (deep, very cold, normal salinity, low oxygen, high nitrate, medium phosphate, high silicate), and 36 (deep, very cold, normal salinity, moderate oxygen, medium nitrate, low phosphate, low silicate) [65], comprising most of the Pacific, Indian, and polar oceans. These environments are not over-represented in the BenthicNet dataset, with no single EMU accounting for $> 20.6\%$. The three most common EMUs sampled (47.6%) were 24 (shallow, warm, normal salinity, moderate oxygen, low nitrate, low phosphate, low silicate), 11 (shallow, cool, normal salinity, moderate oxygen, low nitrate, low phosphate, low silicate), and 13 (deep, very cold, normal salinity, low oxygen, high nitrate, medium phosphate, high silicate), representing continental shelves in the equatorial regions, the shallow sub-tropics, and the deep Pacific and Indian oceans. The distribution of images across ocean basins was generally proportionate to the expectation given the area of each ocean, but notable exceptions include an apparent under-representation of the South Atlantic, and over-representation of the South Pacific.

C.2 Labelled data exploratory analysis

The BenthicNet-Labelled data spans an environmental extent similar to that of the BenthicNet-1M data. Two of the EMUs that were abundantly sampled with unlabelled imagery were also prominently represented in the labelled dataset; EMUs 11 (shallow, cool, normal salinity, moderate oxygen, low nitrate, low phosphate, low silicate) and 24 (shallow, warm, normal salinity, moderate oxygen, low nitrate, low phosphate, low silicate) comprised a near-majority (49.82%) of the labelled dataset (Fig. 7). These two environments are broadly distributed in space [65], and here primarily represent datasets from Australia, Tasmania, and Central America.

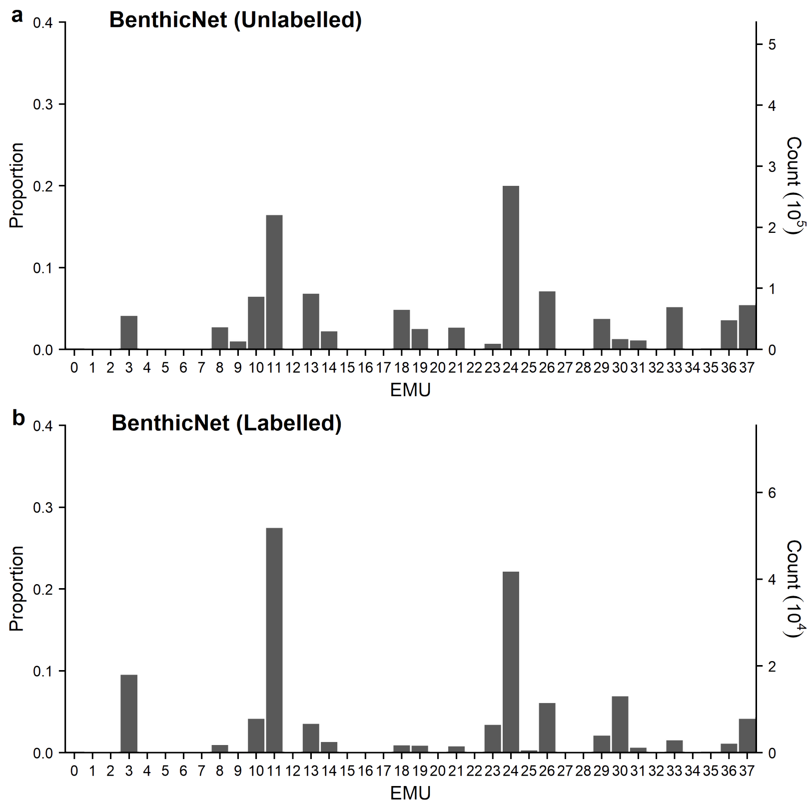

Figure 7: **Distribution of BenthicNet images according to bottom layer ecological marine units** (EMUs) for (a) unlabelled and (b) labelled datasets. See Sayre et al. [65] for a full description of the EMU classes.

D Confusion matrices

In this section, we provide confusion matrices for the fine-tuned models described in §4.2.

An important observation is that for both supervised classification tasks, and both transfer models, the best-predicted classes tended to be those that are most distinct, while the intermediate classes were subject to confusion. For example, "cobble" was the most difficult label to predict in the BenthicNet-Substrate-d2 dataset, and indeed, it can be difficult even for a human to differentiate cobbles from pebbles or boulders in underwater imagery. These substrate class boundaries are defined arbitrarily at a particular length scale (2 mm and 64 mm) that may only be determined through accurate measurement or image scaling; there is substantial possibility of incorrect or subjective human labels for such data. Additionally, the imbalanced priors for the classes may also play a role in predictive success. Sand and mud labels dominate both data subsets—it is not surprising that the models have a tendency to predict sand for other classes, and to perform strongly on sand.

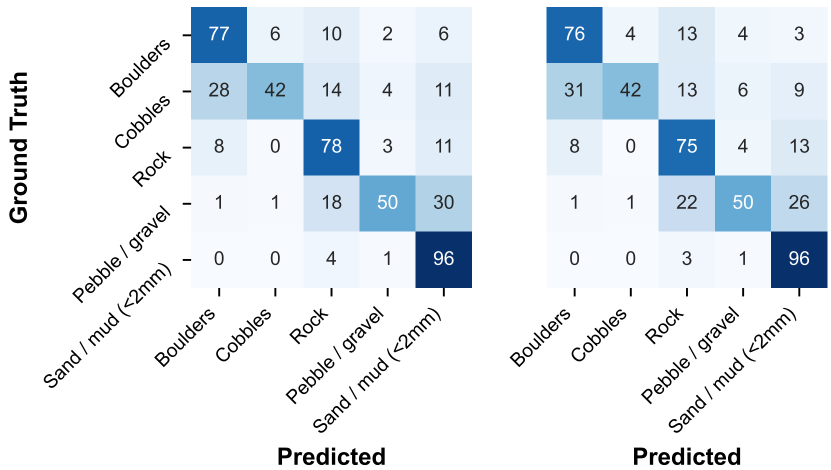

Figure 8: **Confusion matrix (% of ground truth) for CATAMI Substrate predictions on BenthicNet-Substrate-d2 test data.** (Left) Model pretrained with cross-entropy on ImageNet-1k, fine-tuned on BenthicNet-Substrate-d2. (Right) Model pretrained with Barlow Twins on BenthicNet-1M, fine-tuned on BenthicNet-Substrate-d2.

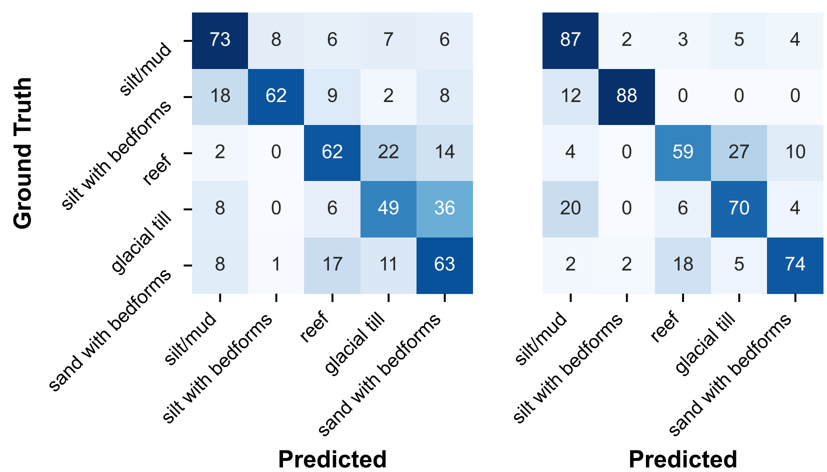

Figure 9: **Confusion matrix (% of ground truth) for the German Bank 2010 test data.** (Left) Model pretrained with cross-entropy on ImageNet-1k, fine-tuned on German Bank 2010. (Right) Model pretrained with Barlow Twins on BenthicNet-1M, fine-tuned on German Bank 2010.

E PANGAEA Search

To thoroughly search PANGAEA for seafloor imagery, we used 20 search terms with a range of synonyms for the content of interest. The PANGAEA search API is comprehensive and allows terms be combined with AND or OR operators, and negative search terms to be used. However, we could not merge all our synonyms together into a single, large query because the number of results which can be returned by one query is limited to 500 records.

The search terms used were as follows:

```
(seabed OR "sea bed" OR "sea-bed") (image OR imagery OR photo OR photograph
    OR "photo-transect" OR photoquad* OR photo-quad* OR jpg OR jpeg OR png OR tif
    OR tiff)
(seafloor OR "sea floor" OR "sea-floor") (image OR imagery OR photo OR photograph
    OR "photo-transect" OR photoquad* OR photo-quad* OR jpg OR jpeg OR png OR tif
    OR tiff)
("ocean floor" OR "ocean-floor") (image OR imagery OR photo OR photograph
    OR "photo-transect" OR photoquad* OR photo-quad* OR jpg OR jpeg OR png OR tif
```

```
    OR tiff)
underwater (habitat* OR substrate OR sediment) (image OR imagery OR photo
    OR photograph OR "photo-transect" OR photoquad* OR photo-quad* OR jpg OR jpeg
    OR png OR tif OR tiff)
benthic (image OR imagery OR photo OR photograph OR "photo-transect" OR photoquad*
    OR photo-quad* OR jpg OR jpeg OR png OR tif OR tiff)
(benthos or benthoz) (image OR imagery OR photo OR photograph OR "photo-transect"
    OR photoquad* OR photo-quad* OR jpg OR jpeg OR png OR tif OR tiff)
(coral OR reef OR seagrass OR "sea grass") (image OR imagery OR photo OR photograph
    OR "photo-transect" OR photoquad* OR photo-quad* OR jpg OR jpeg OR png OR tif
    OR tiff)
(auv OR rov OR uuv OR "underwater vehicle") (image OR imagery OR photo OR photograph
    OR "photo-transect" OR photoquad* OR photo-quad* OR jpg OR jpeg OR png OR tif
    OR tiff)
benthoscape habitat* image
benthoscape habitat* imagery
benthoscape habitat* photo
benthoscape habitat* photograph
benthoscape habitat* ("photo-transect" OR photoquad* OR photo-quad*)
benthoscape habitat* (jpg OR jpeg OR png OR tif OR tiff)
benthoscape image
benthoscape imagery
benthoscape photo
benthoscape photograph
benthoscape ("photo-transect" OR photoquad* OR photo-quad*)
benthoscape (jpg OR jpeg OR png OR tif OR tiff)
```

Each search term was prefixed with a set of negative search terms to remove false positives, given as follows

```
-microscop? -"Meteorological observations" -topsoil -soil -sky
    -"wind vector" -"wind stress" -"vertical profile" -"vertical distribution"
```

The full code for our PANGAEA search is publicly available at [REDACTED].

F FathomNet Python API Code

We retrieved the full set of images on FathomNet by using the FathomNet API from the fathonnet-py Python package as follows.

```python
import fathomnet.api.images
import pandas as pd

keys = ["url", "uuid", "timestamp", "latitude", "longitude"]

records = []
for submitter in fathomnet.api.images.find_distinct_submitter():
    for image in fathomnet.api.images.find_by_contributors_email(submitter):
        records.append({k: getattr(image, k) for k in keys})

df = pd.DataFrame.from_records(records)
df.drop_duplicates(subset="url", inplace=True)
```