ReefNet: A Large-Scale Dataset and Benchmark for Fine-Grained Coral Reef Recognition

Abdulwahab Felemban[1*], Yahia Battach[1*], Faizan Farooq Khan[1], Yuqian Fu[1], Xuhui Liu[1], Yesmeen M. Khattab[1], Yousef A. Radwan[1], Xiang Li[1], Fabio Marchese[1], Sara Beery[2], Burton H. Jones[1], Francesca Benzoni[1], and Mohamed Elhoseiny[1]

[1] King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[2] Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Abstract. Coral reefs are rapidly declining under anthropogenic pressures (e.g., climate change), creating an urgent need for scalable and automated monitoring. Progress in data-driven coral analysis, however, is constrained by the scarcity of large-scale datasets with fine-grained labels that are taxonomically consistent across sites and studies. To address this gap, we introduce ReefNet, a large-scale public coral reef image dataset with point-level annotations mapped to the World Register of Marine Species (WoRMS) taxonomy. ReefNet aggregates imagery from 76 curated CoralNet sources and an additional reef site from Al-Wajh (Red Sea), totaling approximately 925K genus-level hard coral annotations. Through expert-driven verification and targeted filtering, we derive a high-confidence benchmark subset with 92% expert agreement over 39 hard-coral label classes, enabling reliable evaluation under realistic label noise and strong class imbalance. Beyond dataset construction, we establish a comprehensive benchmark spanning *zero-shot*, *cross-domain few-shot adaptation*, *within-source* evaluation, and *cross-source* transfer to the Al-Wajh dataset. Experiments with state-of-the-art vision–language models (VLMs), multimodal large language models (MLLMs), and vision-only backbones reveal substantial degradation in zero-shot and extremely few-shot regimes, while adaptation with in-domain supervision yields large gains yet still leaves a persistent gap under cross-source shift and on long-tail genera. These results highlight fundamental challenges in applying general-purpose multimodal models to biodiversity monitoring and underscore the importance of large-scale, taxonomically grounded, high-quality datasets. ReefNet serves as both a benchmark and a training resource for advancing fine-grained coral reef understanding. Data, code, and models will be released upon acceptance.

1 Introduction

Coral reefs are among the most biodiverse ecosystems on Earth, providing immense ecological and economic value [6, 50]. However, these vibrant habitats are

* Equal contribution

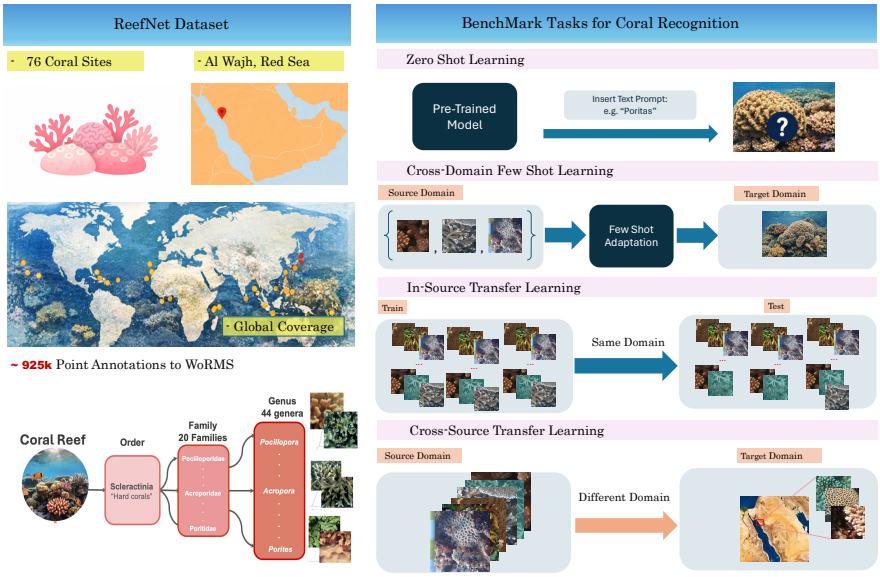


Fig. 1: ReefNet: A large-scale benchmark for fine-grained coral reef recognition. Left: ReefNet unifies coral reef imagery from 76 CoralNet sources and a newly collected Red Sea dataset, resulting in ∼925K genus-level coral annotations aligned with the WoRMS taxonomy. Right: We introduce a comprehensive evaluation benchmark spanning zero-shot learning, cross-domain few-shot adaptation, within-source evaluation, and cross-source evaluation, enabling a systematic study of fine-grained recognition and cross-reef generalization.

increasingly threatened by anthropogenic pressures, including climate change, overfishing, pollution, and ocean acidification [10, 15, 30]. As these stressors intensify, coral cover and overall reef resilience continue to decline, underscoring the urgent need for effective monitoring, conservation, and restoration efforts [10, 13]. A key component of reef conservation is habitat mapping and coverage analysis, which typically relies on expert taxonomists to annotate *in situ* underwater images [7, 28]. However, this manual annotation process is labor-intensive and difficult to scale, limiting the spatial and temporal coverage of reef monitoring programs.

Recent advances in deep learning (DL) have accelerated automation across many vision applications [21, 22, 37, 38, 42, 46, 47, 68, 70, 71]. Yet progress in coral reef recognition remains hindered by two major challenges. First, the availability of large-scale, DL-ready coral datasets with consistent taxonomy and fine-grained annotations is still limited. Second, domain shift across reef sites, caused by variations in imaging conditions, local species assemblages, and annotation conventions, often leads to substantial performance degradation when models are deployed in new locations [9, 14]. These challenges highlight the need for standardized datasets and evaluation benchmarks that enable systematic investigation of model generalization across reef environments.

Regarding source data, CoralNet [7, 14], one of the most widely used coral annotation platforms, addresses part of this challenge by allowing users to fine-tune classifiers on individual "sources." However, these source-specific models often perform poorly on unseen reef sites because of differences in taxonomy, imaging protocols, and ecological composition [65]. ReefCloud.ai [1] further improves label standardization but remains inaccessible to the broader research community. As a result, the lack of large-scale, publicly available datasets with standardized taxonomy and diverse reef coverage continues to limit the development and evaluation of modern machine-learning methods.

To address these limitations, we introduce **ReefNet**, a large-scale public coral reef image dataset with point-level annotations aligned with the taxonomy of the *World Register of Marine Species* (WoRMS) [11]. As illustrated on the left side of Figure 1, ReefNet unifies imagery from 76 carefully curated CoralNet sources and further includes 1.3K newly collected underwater images from Al-Wajh Lagoon in the Red Sea, containing 4,624 expert annotations of hard coral genera from an understudied biogeographic region. In total, ReefNet contains approximately 925K genus-level annotations of scleractinian corals (Table 1), making it one of the largest publicly available coral reef datasets for machine-learning research. To ensure annotation reliability, 8,962 annotations across multiple sources were manually verified by marine biologists, producing a high-confidence repository of coral labels spanning diverse reef ecosystems. The standardized alignment to WoRMS further enables consistent labeling across heterogeneous data sources, facilitating large-scale machine-learning research and cross-dataset comparisons. Beyond point annotations, ReefNet further incorporates textual descriptions for each coral genus, derived from digitized coral taxonomy references [60, 61, 63]. These descriptions enable language-grounded classification and support emerging research on vision-language models (VLMs) and multimodal large language models (MLLMs) for fine-grained coral reef understanding.

Beyond dataset construction, ReefNet establishes a comprehensive benchmark for fine-grained coral reef recognition. As illustrated on the right side of Figure 1, we evaluate models under multiple realistic settings, including zero-shot learning, cross-domain few-shot adaptation, within-source evaluation, and cross-source evaluation, enabling systematic analysis of model generalization across reef environments. Extensive experiments reveal that state-of-the-art models struggle with rare classes and morphologically similar coral taxa, highlighting persistent challenges related to cross-domain generalization, cross-source transfer, and fine-grained visual discrimination. By releasing the ReefNet dataset, benchmark protocols, and code, we aim to provide a standardized platform for advancing deep learning research in coral reef recognition. Moreover, the clear performance gains observed under both within-source and cross-source evaluation demonstrate the value of ReefNet not only as a benchmark dataset but also as a large-scale training resource for ecological computer vision, with coral reef understanding serving as a representative real-world application.

The main contributions of this work are summarized as follows:

- **ReefNet dataset.** We introduce ReefNet, a large-scale coral reef dataset with WoRMS-aligned standardized point annotations, aggregated from 76 CoralNet sources together with newly collected imagery from Al-Wajh Lagoon in the Red Sea. In total, ReefNet contains approximately 925K genus-level annotations of hard corals, making it one of the largest publicly available coral reef datasets for machine learning research.
- **A comprehensive benchmark for coral reef recognition.** We establish a benchmark for fine-grained coral recognition under multiple realistic settings, including zero-shot learning, cross-domain few-shot adaptation, within-source evaluation, and cross-source evaluation, enabling systematic evaluation of model generalization across reef environments.
- **Systematic evaluation and insights for ecological computer vision.** We conduct extensive experiments with fine-tuned, few-shot, and zero-shot models, revealing key challenges in long-tailed coral distributions, cross-domain generalization, and fine-grained visual discrimination, providing insights for future research in ecological computer vision.

2 Related Work

Large-Scale Biodiversity and Marine Datasets. Several large-scale datasets have recently emerged to facilitate computer vision in biodiversity classification. For instance, TreeOfLife-10M [54] consolidates over 10 million images of terrestrial and marine organisms from iNat2021 [29], BIOSCAN-1M [23], and the Encyclopedia of Life (eol.org). However, underwater taxa remain underrepresented in such general repositories due to the logistical challenges of *in situ* marine data acquisition. To bridge this gap in marine ecosystems, BenthicNet [36] aggregates seafloor imagery from multiple surveys, including the Seaview Survey Photo-quadrat [24], Reef Life Survey [17], and others [20]. BenthicNet constitutes a major step forward, supporting the WoRMS [11] taxonomy with 887,533 annotations, of which 287K are hard coral (*Scleractinia*) annotations (Table 1). Nonetheless, it primarily supports broad-scale benthic habitat labeling using the CATAMI classification scheme [3], rather than the fine-grained taxonomic granularity required for specialized coral ecological studies. In contrast, ReefNet is explicitly designed to provide the fine-grained taxonomic granularity required for specialized coral ecological studies.

Coral-specific Datasets. Existing coral-specific datasets typically face a stark trade-off between taxonomic resolution, geographic diversity, and task specificity. Datasets tailored for dense prediction, such as CoralVOS and CoralSCOP [72, 73], offer valuable pixel-level masks but categorize corals only at the coarse order level (*Scleractinia*), limiting their utility for fine-grained ecological studies. Similarly, the recent Coralscapes dataset [48] extends benthic classification to a sizeable Red Sea dataset with over 170K polygon annotations, but focuses on broader habitat cover types rather than genus- or species-level identification. In contrast, MosaicsUCSD [18] and Eilat [2, 8] offer highly detailed genus- or species-level

Table 1: Comparison With Existing Coral Classification Datasets. *Image counts reflect availability as of May 13, 2025. Hard coral/genus-level annotations are not specified for CoralNet, since its labels are not standardized. †MosaicsUCSD includes 16 annotated orthomosaics, each from ∼1.5K images. ‡CoralSCOP [72] contains 330,144 binary coral/non-coral masks (not limited to hard corals). CoralVOS [73] likewise provides only binary coral masks, so a hard coral count is N/A.

Dataset	Geographic coverage	Annotation type	Lowest taxonomic level	Mapped to WoRMS	Number of images	Number of annotations	
						Hard corals	Genus-level annotations
CoralNet [7]	**World**	Sparse points	**Species**	No	4,524,792*	N/A	No
CoralSCOP [72]	**World**	**Masks**	Order	No	41,297	330,144‡	No
CoralVOS [73]	17 sites (South China Sea)	**Masks**	Order	No	60,456	N/A	No
MosaicsUCSD [18]	Palmyra	**Masks**	**Species**	No	16†	44,008	**Yes**
Eilat [45]	Eilat	Sparse points	Genus	No	212	∼12,000	**Yes**
BenthicNet [36]	**World**	Sparse points / image label	**Species**	**Yes**	**11,408,887**	287,181	**Yes**
Coralscapes [48]	Red Sea (5 countries)	**Masks**	N/A	No	2,075	N/A	No
ReefNet (ours)	**World**	Sparse points	Genus	**Yes**	181,223	**924,626**	**Yes**

annotations but are geographically constrained to a single reef site. As a result, the localized nature heavily restricts the ability of models to generalize across different regions and imaging conditions. To overcome this geographic bias, ReefNet aggregates imagery across a diverse, global set of reef sites, ensuring broader spatial generalization.

Standardizing Coral Reef Recognition. At a broader scale, CoralNet [7] hosts vast volumes of benthic imagery with both human- and machine-generated labels, providing a rich resource for reef analysis. However, its user-generated labels are heavily source-dependent and lack a unified taxonomic standard like WoRMS, making cross-source learning and systematic evaluation exceedingly difficult. To address this critical inconsistency, ReefNet meticulously aligns approximately 925K expert-verified annotations to the universally recognized WoRMS taxonomy, ensuring strict labeling consistency. Consequently, ReefNet provides one of the most comprehensive benchmarks available, supporting a broad spectrum of computer vision tasks from habitat classification to cross-domain, fine-grained genus-level analysis. We provide further clarification of our advantages over CoralNet in the Appendix.

3 ReefNet Dataset

ReefNet is constructed through a multi-stage, expert-guided curation pipeline applied to publicly available benthic imagery and annotations from CoralNet, as shown in Figure 2. Starting from a large collection of public sources, we perform systematic source filtering, taxonomy alignment, semantic label consolidation, and expert verification to build a high-quality coral reef dataset. This section describes the dataset construction process, including data collection, label standardization,

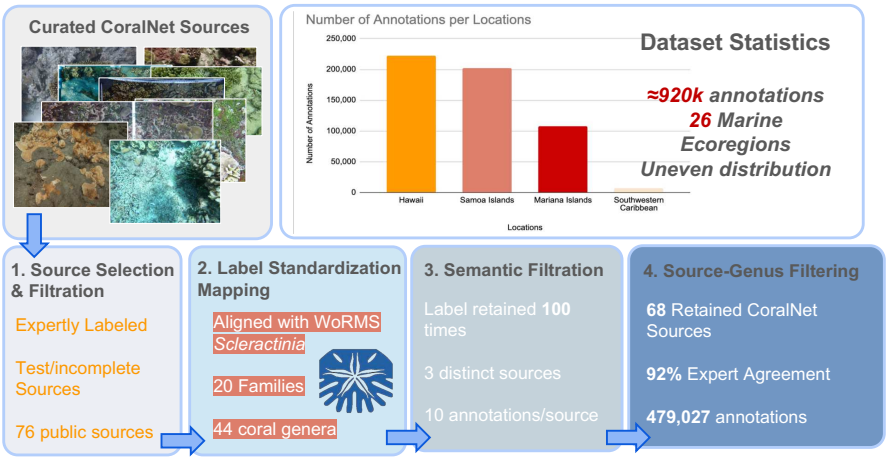


Fig. 2: ReefNet data construction pipeline. We apply a multi-stage curation process including source filtering, WoRMS taxonomy alignment, semantic label consolidation, and expert verification to construct ReefNet. The pipeline yields a high-quality coral reef dataset with ∼925K point annotations across 39 hard-coral label classes. We further augment the dataset with a newly collected Red Sea dataset from Al-Wajh Lagoon. The resulting ReefNet dataset supports multiple benchmark settings for fine-grained coral recognition.

and quality control procedures that ensure the taxonomic reliability and ecological validity of ReefNet.

3.1 Data Collection

We started with 1,366 publicly available CoralNet sources and applied a series of semantic, ecological, and technical filters to identify high-quality reef datasets with taxonomically meaningful annotations. This involved removing test or incomplete sources, selecting only expert-verified labels, and retaining sources with sufficient image and annotation counts. We then applied domain-specific criteria, such as the presence of reef-building corals and shallow, *in situ* imagery, yielding 76 public sources with approximately ∼**920K** genus-level coral annotations. The full source selection pipeline is detailed in the appendix.

Al-Wajh Lagoon Data. We additionally contribute a new dataset from the Al-Wajh Lagoon in the Red Sea (25.6°N, 36.8°E), comprising 1.3K high-resolution *in situ* images and 4,624 expert annotations of hard corals. This dataset serves as the test set for the cross-source benchmark, enabling evaluation of fine-grained classification and cross-region generalization. Data collection details are provided in the appendix.

3.2 Taxonomy Mapping and Label Standardization

To ensure taxonomic traceability and biological consistency, we mapped annotation labels to the World Register of Marine Species (WoRMS [11]). This step was applied, where relevant, to annotations from the 76 curated CoralNet sources and the Al-Wajh Lagoon dataset. Hard coral labels were manually aligned with canonical scientific names and corresponding AphiaIDs[3]. This mapping enabled consistent aggregation of biological entities across sources and ensured compatibility with other biodiversity databases. The initial mapping revealed ∼920K annotations spanning multiple taxonomic levels, which we then consolidated into a unified hard-coral label space. The final retained label set comprises 38 genera plus one family-level class (*Fungiidae*), for a total of 39 hard-coral labels spanning 20 families. However, the initial label space exhibited inconsistencies in naming, taxonomic granularity, and semantic intent, necessitating further filtering and restructuring.

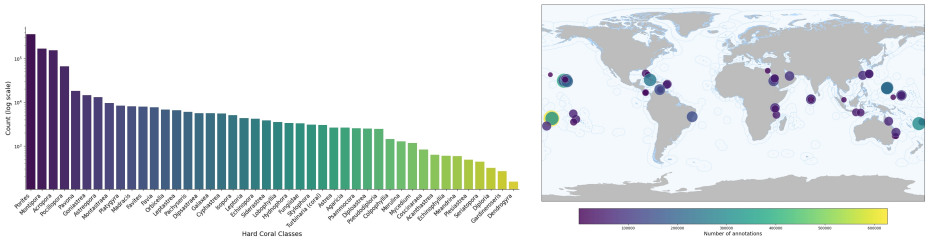
Label Filtering and Semantic Consolidation. To construct a benchmark dataset suitable for training AI models, we applied a biologically informed filtering strategy targeting syntactic variation (e.g., "Staghorn coral" vs. *Acropora cervicornis*) and taxonomic inconsistency (e.g., mixing species-, genus-, and family-level labels). This step was motivated by the inherent difficulty of fine-grained coral identification from imagery alone [14, 36], which often results in community datasets with inconsistent or imprecise taxonomic labeling.

Labels were retained only if they referred to hard corals at the genus level, except for *Fungiidae*, which was kept at the family level because semantic similarity among its genera makes genus-level verification difficult. Retained labels also had to meet the following criteria: appear at least 100 times in the dataset, be present in at least three distinct CoralNet sources with a minimum of 10 annotations per source, exhibit consistent and distinguishable visual patterns, and be taxonomically valid according to WoRMS. After filtering and consolidation, the ReefNet dataset used in our experiments includes 39 unique labels, grouped primarily at the genus level. This curated, taxonomy-aware structure enables hierarchical subsetting, scalable annotation, and custom label aggregation for ecological and machine-learning tasks.

3.3 Dataset Statistics

Annotation Distribution. The ReefNet dataset comprises **334,162** images, of which **181,223** contain hard coral annotations, totaling **924,626** point annotations. These annotations span 26 marine ecoregions across the tropical belt *sensu* [51] (see the appendix for details). The distribution is uneven: the Hawaiian Ecoregion alone contributes over 221K annotations, followed by the Samoa Islands (202K) and the Mariana Islands (108K). In contrast, regions such as the Southwestern Caribbean, Floridian, and Eastern Brazil are sparsely represented,

[3] An AphiaID is a unique identifier from the WoRMS database, linking each label to its corresponding entry in the hierarchical taxonomy. As coral taxonomy evolves, AphiaIDs allow users to verify the current validity of labels.



(a) Hard Coral Taxa **(b)** Geographic Distribution

Fig. 3: Distributions of Dataset Annotations. (a) Log-scale Distribution of Annotations Across Hard Coral Taxa. The plot includes 38 genera and one family-level class (Fungiidae). (b) Geographic Distribution of the Annotations. Sources located within two degrees of Latitude and Longitude are grouped into a single point. One source did not contain any location data and is displayed in Antarctica. Marine Ecoregions of the World are shown in light blue [51].

each with fewer than 8K samples. This geographic imbalance reflects broader disparities in coral monitoring efforts and may influence model generalization to underrepresented areas. Among all annotations, 921,351 are assigned to one of 39 labels, capturing a wide range of coral diversity. As shown in Fig. 3a, the most frequent genera, *Porites*, *Montipora*, *Acropora*, and *Pocillopora*, dominate shallow reefs globally, while several labels appear in fewer than 1K annotations, illustrating the dataset's imbalanced distribution.

Covariate Diversity. In addition to the biological breadth of ReefNet, it exhibits substantial covariate diversity across environmental conditions, camera systems, and imaging protocols (Fig. 3b), leading to significant variations in resolution, lighting, and water quality across the data. For specific target deployments, it may be advantageous to train only on images from similar environmental conditions or acquisition parameters. We provide detailed metadata on geographic context and camera setups for each CoralNet source in the Appendix , along with the metadata that will be publicly released, thereby enabling the exploration of optimal subsets.

3.4 Annotation Quality Control

To assess annotation reliability, we conducted an expert review on a stratified subset of 8,962 hard coral samples. Each review involved evaluating images tied to specific *source–genus* pairs, with labels categorized as *Correct*, *Incorrect*, *Low Quality Image*, or *Hard to Decide*. A micro-averaged expert agreement percentage was calculated as the proportion of *Correct* labels among all reviewed samples.

To support this process, we built a custom web-based application that enabled structured review of annotations across sources. Further details on the interface and the full verification pipeline are provided in the Appendix. Based on this expert feedback, which showed an overall agreement rate of 73%, we performed

Table 2: Dataset Split Statistics. ReefNet provides train/val/test splits covering 39 coral classes (38 in validation). Al-Wajh is used only for cross-source testing and contains 4,624 annotations from 13 classes overlapping with the ReefNet label set.

Split	ReefNet			Al-Wajh		
	Annotations	Classes	# Sources	Annotations	Classes	# Sources
Train	445,985	39	68	–	–	–
Val	9,999	38	63	–	–	–
Test	23,043	39	66	4624	13	1

targeted quality control, removing low-confidence genera, sources, or genus–source pairs. Guided by insights from the verified subset, this process affected a total of 920,017 hard coral annotations in the full dataset. These verification and filtering steps ensure that downstream models are trained on biologically reliable annotations and support ecologically valid interpretation of classification outcomes. Further details on the quality-control procedure are provided in the Appendix.

Source and Genus Filtering. We first excluded any source where fewer than 50% of reviewed samples were deemed correct. Similarly, coral genera with less than 50% expert agreement across all verified sources were removed. This step retained 70 sources, with a post-filter expert agreement of 78%. After filtering, 860,463 annotations remained across 39 unique hard-coral labels.

Source–Genus Filtering. We then applied a stricter filter at the source–genus level, retaining only pairs with at least 70% expert agreement in their verified samples. This resulted in a high-confidence subset with **92%** expert agreement on the representative verification set. After this step, **479,027** annotations were retained from 68 sources while preserving all 39 label classes. A summary of the filtering process and benchmark splits is presented in Table 2.

4 ReefNet Benchmarks

4.1 Dataset Splits

We evaluate all models on the ReefNet benchmark using the high-confidence subset obtained through the expert-driven quality control described in Sec. 3.4. To prevent data leakage, we partition each data source at the *image level* into training, validation, and test sets. Specifically, we employ a Multilabel Stratified Shuffle Split [12, 49] to maintain consistent class distributions across all splits. The resulting benchmark encompasses 39 hard-coral label classes (with 38 present in the validation set because one class is extremely rare).

We refer to evaluation on the 39-class ReefNet test set as the *within-source* setting. To assess generalization to unseen acquisition conditions and locations, we additionally evaluate in a *cross-source* setting on the curated Al-Wajh dataset, restricting evaluation to the 13 genera shared between ReefNet and Al-Wajh.

Since the within-source and cross-source evaluations involve different label spaces (39 vs. 13 genera), we report results separately and avoid direct comparison of absolute metric values across the two settings. Table 2 summarizes the resulting splits.

4.2 Evaluation Protocol

We next define the evaluation settings used throughout the paper. ReefNet is partitioned at the image level within each source into train/val/test splits, and all annotations extracted from a given image belong to the same split.
Within-source (ReefNet). We evaluate on the ReefNet test split (39 genera) to assess performance when training and testing data originate from the same source.
Cross-source (Al-Wajh). To measure generalization, we evaluate on the Al-Wajh test set, restricting both predictions and reported metrics to the 13 genera overlapping with ReefNet.
Few-shot support sets. For few-shot adaptation, we fine-tune on nested support sets of $k \in \{1, 5, 10, 20\}$ examples per class, sampled directly from the ReefNet training split.
Training and selection. All adaptation and full fine-tuning are conducted on the ReefNet *training* split, with model selection performed on the *validation* split. The ReefNet *test* split and the Al-Wajh dataset are reserved strictly for final evaluation.

5 Experiments

In this section, we comprehensively evaluate a diverse suite of VLMs, MLLMs, and vision-only models on the ReefNet dataset. To thoroughly assess their capabilities in hard coral genus recognition, our analysis spans zero-shot, cross-domain few-shot, and full fine-tuning paradigms.

5.1 Experimental Setup

Evaluated Models. We evaluate three families of models on ReefNet: (i) Vision–Language Models (VLMs) [31, 44, 53, 69], (ii) Multimodal Large Language Models (MLLMs) [4, 32, 33, 56, 64, 67], and (iii) vision-only models (e.g., CNNs and Vision Transformers) trained through standard supervised fine-tuning.
Evaluation settings and metric. Models are investigated under three evaluation settings: *zero-shot*, *cross-domain few-shot*, and *full fine-tuning*. To rigorously evaluate performance under the severe class imbalance of ReefNet (Fig. 3a), we adopt Macro Recall (balanced accuracy) [43] as the primary evaluation metric. Compared to standard accuracy, Macro Recall assigns equal weight to all classes, preventing the overall score from being disproportionately dominated by frequent genera. Furthermore, to assess model generalization across diverse reef environments, we report both within-source and cross-source performance across all

Table 3: Zero-shot macro recall (%) for hard-coral genus recognition. We report performance on both the within-source ReefNet benchmark and the cross-source Al-Wajh test set. Best and second-best results are **bolded** and underlined, respectively.

Category	Model	ReefNet Test	Al-Wajh Test
MLLMs	InternVL-3.5 [64]	4.0	9.8
	LLaVA-NeXT [33]	2.6	8.3
	LLaVA-OneVision [32]	3.3	9.3
	MiniCPM-4.5 [67]	3.6	9.5
	Qwen3-VL [4]	5.1	15.2
VLMs	BioCLIP2 [25]	**23.5**	**43.2**
	BioCLIP [52]	<u>11.3</u>	<u>32.2</u>
	CLIP [44]	5.2	11.2
	SigLIP [69]	9.1	18.0
	OpenCLIP [31]	4.6	13.4

tasks. For a more granular analysis, the Appendix provides detailed per-class recall, precision, and F1 scores for the top-performing models.

5.2 Zero-Shot Evaluation

The zero-shot results for the hard-coral genus recognition task, as summarized in Tab. 3, reveal a substantial performance gap between general-purpose Multimodal Large Language Models (MLLMs) and domain-specialized Vision–Language Models (VLMs). Among the MLLMs, performance is consistently low on the ReefNet within-source test, with macro recall ranging from 2.6% (LLaVA-NeXT) to 5.1% (Qwen3-VL). While these models improve on the Al-Wajh cross-source test, the best MLLM reaches only 15.2% macro recall (Qwen3-VL), which remains far below specialized VLMs such as BioCLIP2 (43.2%). These results suggest that current MLLMs, when used zero-shot, are not yet reliable for mapping subtle morphological coral cues to the correct genus labels.

In contrast, CLIP-style models demonstrate stronger performance, particularly those with domain-relevant pretraining. While standard CLIP and OpenCLIP yield modest results (5.2% and 4.6%, respectively), SigLIP improves to 9.1%, and BioCLIP reaches 11.3%. Most notably, BioCLIP2 achieves the best zero-shot performance (23.5% on the ReefNet test and 43.2% on the Al-Wajh test), outperforming all other models by a large margin. This improvement underscores the importance of biological pretraining: BioCLIP2 is trained on the TreeOfLife-200M dataset, which provides broad biological supervision [26]. Collectively, these findings suggest that zero-shot hard-coral genus recognition is primarily limited by insufficient domain-specific knowledge rather than by the general capacity of visual representations, highlighting the need for domain-aware multimodal adaptation.

Table 4: Few-Shot Hard Coral Classification Macro Recall (%). Models are fine-tuned with N examples per class from ReefNet. Results are reported for the within-source ReefNet test set and the cross-source Al-Wajh test set.

Model	Within-Source Test (ReefNet)				Al-Wajh Test			
	1-shot	5-shot	10-shot	20-shot	1-shot	5-shot	10-shot	20-shot
Qwen3-VL [4]	5.56	9.15	17.24	27.22	10.93	13.77	25.87	31.27
MiniCPM-V [67]	2.85	7.32	14.95	21.95	7.68	8.52	12.77	20.88
InternVL-3.5 [64]	2.00	2.87	10.44	18.34	10.31	9.01	12.80	12.87
CLIP [44]	8.10	23.02	38.41	45.71	23.16	29.38	47.64	57.30
OpenCLIP [31]	9.70	21.45	39.61	45.95	18.40	23.27	45.84	46.51
BioCLIP2 [25]	30.54	35.59	44.68	58.79	42.56	38.68	60.02	67.83

5.3 Cross-Domain Few-Shot Evaluation

In the *cross-domain* few-shot setting, we study few-shot adaptation of models whose *pretraining* is largely out-of-domain (i.e., not tailored to hard-coral taxonomy or reef imagery). We evaluate the top three zero-shot MLLMs and VLMs after fine-tuning with $k \in \{1, 5, 10, 20\}$ labeled examples per class sampled from the ReefNet training split. The shot configurations are nested, meaning smaller-shot sets are strict subsets of larger ones.

Overall, performance generally improves as k increases, and gains are observed both on the ReefNet within-source test set and on the cross-source Al-Wajh test set over the 13 overlapping genera. However, Tab. 4 reveals several clear trends. (i) CLIP-style VLMs consistently outperform MLLMs across all shot regimes: at 20-shot, BioCLIP2 reaches 58.79% on ReefNet and 67.83% on Al-Wajh, while the best MLLM (Qwen3-VL) attains only 27.22% and 31.27%, respectively. (ii) Standard CLIP adapts particularly well with limited supervision, improving from 8.10% to 45.71% on ReefNet and from 23.16% to 57.30% on Al-Wajh when moving from 1-shot to 20-shot. (iii) BioCLIP2 is strong even in the lowest-shot regime (30.54% on ReefNet and 42.56% on Al-Wajh at 1-shot) and remains the best overall at higher shots. (iv) We observe minor non-monotonicity in a few cases (e.g., BioCLIP2 on Al-Wajh from 42.56% at 1-shot to 38.68% at 5-shot), suggesting some sensitivity to the sampled support set. (v) While MLLMs benefit from additional supervision, they remain substantially behind VLMs even at 20-shot, indicating that current MLLMs are less effective at exploiting small labeled sets for fine-grained hard-coral genus classification.

5.4 Full Fine-tuning Evaluation

Table 5 summarizes full fine-tuning results on the within-source ReefNet test set and the cross-source Al-Wajh test set. Despite strong overall performance, errors are concentrated in rare classes and visually similar genera. In particular, many hard-coral genera share subtle morphological traits (e.g., meandroid vs.

Table 5: Hard Coral Classification Macro Recall (%). Models are fine-tuned on the high-quality ReefNet split (92% expert agreement). Results are reported for the within-source ReefNet test set and the cross-source Al-Wajh test set.

Category	Model	ReefNet Test	Al-Wajh Test
Vision	ViT-B/16 [16]	**79.84**	54.84
	EfficientNet-B3 [55]	71.29	52.58
	DeiT-B/16 [57]	76.22	51.52
	ConvNeXt-B [35]	74.70	50.47
	Swin-B [34]	77.12	50.21
	BEiT-B/16 [5]	66.66	48.23
	ResNet-50 [27]	45.55	36.93
MLLMs (LoRA)	Qwen3-VL [4]	72.99	59.80
	MiniCPM-V [67]	63.36	48.09
	InternVL-3.5 [64]	53.44	32.97
VLMs	CLIP [44]	76.34	58.51
	OpenCLIP [31]	76.78	57.48
	BioCLIP2 [25]	79.23	58.28

polygonal corallite structures) that are difficult to distinguish in standard survey imagery, suggesting that further gains may require improved architectures and/or additional curated training data for hard cases.

On the within-source ReefNet benchmark, ViT-B/16 achieves the highest macro recall among vision-only models (79.84%), with modern transformer backbones generally outperforming ResNet-50. Among VLMs, BioCLIP2 performs competitively (79.23%), close to ViT-B/16, while CLIP and OpenCLIP also achieve strong results (76.34% and 76.78%). For MLLMs, LoRA fine-tuning yields substantial improvements over zero-shot performance; Qwen3-VL is the strongest MLLM, reaching 72.99% on ReefNet.

On the cross-source Al-Wajh test set, Qwen3-VL attains the best overall macro recall (59.80%), slightly exceeding CLIP-style VLMs (58.51%–58.28%) and all vision-only models (up to 54.84%). Since within-source and cross-source evaluations are computed over different label spaces (39 vs. 13 genera), their absolute values are not directly comparable; nevertheless, the relative ordering within each setting highlights consistent trends across model families.

5.5 Additional Analysis

Fig. 4 breaks down macro recall by class-frequency groups: Head (4 genera; 78.4% of training samples), Body (23 genera; 20.4%), and Tail (12 genera; 1.2%). In the Head group, zero-shot performance for strong models such as BioCLIP2 and Qwen3-VL is competitive with cross-domain few-shot adaptation, suggesting that frequent genera benefit from robust pretrained priors and are easier to recognize even without task-specific supervision.

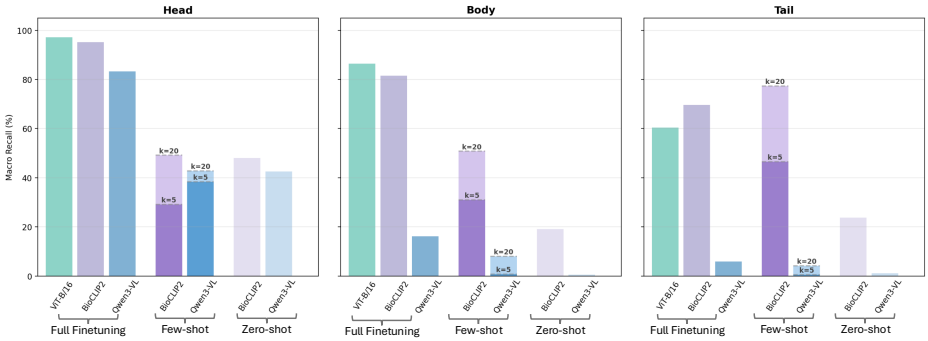


Fig. 4: Head, Body, and Tail Macro Recall. Per-group macro recall for the best-performing fully fine-tuned, cross-domain few-shot, and zero-shot models. Classes are split into head (4 classes, 78.4% of training samples), body (23 classes, 20.4%), and tail (12 classes, 1.2%) based on training-set frequency. Cross-domain few-shot bars show stacked performance at k=5 (darker shade) and the additional gain up to k=20 (lighter shade).

As class frequency decreases from Body to Tail, the gap between zero-shot and cross-domain few-shot performance widens substantially. In particular, Qwen3-VL exhibits very low zero-shot performance on Body and Tail, while few-shot adaptation—especially at k=20—provides a clear boost. This behavior is consistent with many mid- and low-frequency genera being weakly represented in large-scale pretraining corpora, requiring in-domain examples to learn fine-grained morphological distinctions.

Despite these gains, Tail (and to a lesser extent Body) performance remains lower than Head across all training paradigms, including zero-shot, few-shot, and full fine-tuning, highlighting extreme class imbalance as a major bottleneck. Notably, increasing the number of support examples from k=5 to k=20 yields a particularly large relative improvement for BioCLIP2 in the Tail regime, suggesting that additional labeled data is disproportionately valuable for rare genera compared to common ones, where pretrained priors are already strong.

6 Conclusion

We introduced **ReefNet**, a globally curated benchmark of *hard-coral genus* annotations mapped to the WoRMS taxonomy, spanning diverse biogeographic regions and acquisition conditions. Guided by expert verification, we construct a high-confidence benchmark subset with **92%** expert agreement over **39** hard-coral genera, enabling reliable and standardized evaluation across sources. Our experiments highlight three key findings. First, domain-specialized CLIP-style VLMs substantially outperform general-purpose MLLMs in the zero-shot setting, underscoring the importance of biological pretraining for fine-grained coral recognition. Second, cross-domain few-shot adaptation yields large gains for

CLIP-style models and transfers to the cross-source Al-Wajh evaluation, yet a notable gap remains between within-source and cross-source performance, reflecting the difficulty of generalization under domain shift. Third, class imbalance remains a major bottleneck: performance is strong on frequent genera but drops markedly for rare genera, even under full fine-tuning, motivating future work on long-tail learning and targeted data curation. By releasing ReefNet, along with our benchmark code and trained models, we aim to support the development of scalable and domain-adaptive tools for automated coral reef monitoring and conservation. Additional dataset details, verification protocols, and extended analyses are provided in the supplementary material.

References

1. AIMS: Reefcloud (2024). https://doi.org/10.25845/g5gk-ty57, https://reefcloud.ai
2. Alonso, I., Yuval, M., Eyal, G., Treibitz, T., Murillo, A.C.: Coralseg: Learning coral segmentation from sparse annotations. Journal of Field Robotics **36**, 1456–1477 (12 2019). https://doi.org/10.1002/rob.21915, https://onlinelibrary.wiley.com/doi/10.1002/rob.21915
3. Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schönberg, C.H.L., Stuart-Smith, R., Barrett, N., Edgar, G., Colquhoun, J., Tran, M., Jordan, A., Rees, T., Gowlett-Holmes, K.: A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: The catami classification scheme. PLOS ONE **10**(10), e0141039 (2015). https://doi.org/10.1371/journal.pone.0141039, https://doi.org/10.1371/journal.pone.0141039
4. Bai, S., Cai, Y., Chen, R., Chen, K., Chen, X., Cheng, Z., Deng, L., Ding, W., Gao, C., Ge, C., Ge, W., Guo, Z., Huang, Q., Huang, J., Huang, F., Hui, B., Jiang, S., Li, Z., Li, M., Li, M., Li, K., Lin, Z., Lin, J., Liu, X., Liu, J., Liu, C., Liu, Y., Liu, D., Liu, S., Lu, D., Luo, R., Lv, C., Men, R., Meng, L., Ren, X., Ren, X., Song, S., Sun, Y., Tang, J., Tu, J., Wan, J., Wang, P., Wang, P., Wang, Q., Wang, Y., Xie, T., Xu, Y., Xu, H., Xu, J., Yang, Z., Yang, M., Yang, J., Yang, A., Yu, B., Zhang, F., Zhang, H., Zhang, X., Zheng, B., Zhong, H., Zhou, J., Zhou, F., Zhou, J., Zhu, Y., Zhu, K.: Qwen3-vl technical report (2025), https://arxiv.org/abs/2511.21631
5. Bao, H., Dong, L., Piao, S., Wei, F.: BEit: BERT pre-training of image transformers. In: International Conference on Learning Representations (2022). https://doi.org/10.48550/arXiv.2106.08254, https://openreview.net/forum?id=p-BhZSz59o4
6. Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., Silliman, B.R.: The value of estuarine and coastal ecosystem services. Ecological Monographs **81**, 169–193 (5 2011). https://doi.org/10.1890/10-1510.1
7. Beijbom, O., Edmunds, P.J., Roelfsema, C., Smith, J., Kline, D.I., Neal, B.P., Dunlap, M.J., Moriarty, V., Fan, T.Y., Tan, C.J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B.G., Kriegman, D.: Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. PLOS ONE **10**, e0130312 (7 2015). https://doi.org/10.1371/journal.pone.0130312, https://dx.plos.org/10.1371/journal.pone.0130312
8. Beijbom, O., Treibitz, T., Kline, D.I., Eyal, G., Khen, A., Neal, B., Loya, Y., Mitchell, B.G., Kriegman, D.: Improving automated annotation of benthic survey images using wide-band fluorescence. Scientific Reports **6**, 23166 (3 2016). https://doi.org/10.1038/srep23166, https://www.nature.com/articles/srep23166
9. Belcher, B.T., Bower, E.H., Burford, B., Celis, M.R., Fahimipour, A.K., Guevara, I.L., Katija, K., Khokhar, Z., Manjunath, A., Nelson, S., et al.: Demystifying image-based machine learning: a practical guide to automated analysis of field imagery using modern machine learning tools. Frontiers in Marine Science **10** (2023). https://doi.org/10.3389/fmars.2023.1157370
10. Bellwood, D.R., Hughes, T.P., Folke, C., Nyström, M.: Confronting the coral reef crisis. Nature **429**(6994), 827–833 (2004). https://doi.org/10.1038/nature02691, https://doi.org/10.1038/nature02691
11. Board, W.E.: Worms: World register of marine species (2024). https://doi.org/10.14284/170, https://www.marinespecies.org
12. Brady, T.: Iterative stratification. https://github.com/trent-b/iterative-stratification (2017), accessed: 2025-05-16

13. Brandl, S.J., Rasher, D.B., Côté, I.M., Casey, J.M., Darling, E.S., Lefcheck, J.S., Duffy, J.E.: Coral reef ecosystem functioning: eight core processes and the role of biodiversity. Frontiers in Ecology and the Environment **17**, 445–454 (10 2019). https://doi.org/10.1002/fee.2088, https://onlinelibrary.wiley.com/doi/10.1002/fee.2088
14. Chen, Q., Beijbom, O., Chan, S., Bouwmeester, J., Kriegman, D.: A new deep learning engine for coralnet. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). vol. 2021-October, pp. 3686–3695. IEEE (10 2021). https://doi.org/10.1109/ICCVW54120.2021.00412, https://ieeexplore.ieee.org/document/9607450/
15. Cooley, S., Schoeman, D., Bopp, L., Boyd, P., Donner, S., Ghebrehiwet, D.Y., Ito, S.I., Kiessling, W., Martinetto, P., Ojea, E., Racault, M.F., Rost, B., Skern-Mauritzen, M.: Oceans and coastal ecosystems and their services. In: Climate Change 2022 – Impacts, Adaptation and Vulnerability, pp. 379–550. Cambridge University Press, Cambridge and New York (6 2023). https://doi.org/10.1017/9781009325844.005, https://www.cambridge.org/core/product/identifier/9781009325844%23c3/type/book_part
16. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021). https://doi.org/10.48550/arXiv.2010.11929, https://openreview.net/forum?id=YicbFdNTTy
17. Edgar, G.J., Cooper, A., Baker, S.C., Barker, W., Barrett, N.S., Becerro, M.A., Bates, A.E., Brock, D., Ceccarelli, D.M., Clausius, E., Davey, M., Davis, T.R., Day, P.B., Green, A., Griffiths, S.R., Hicks, J., Hinojosa, I.A., Jones, B.K., Kininmonth, S., Larkin, M.F., Lazzari, N., Lefcheck, J.S., Ling, S.D., Mooney, P., Oh, E., Pérez-Matus, A., Pocklington, J.B., Riera, R., Sanabria-Fernandez, J.A., Seroussi, Y., Shaw, I., Shields, D., Shields, J., Smith, M., Soler, G.A., Stuart-Smith, J., Turnbull, J., Stuart-Smith, R.D.: Establishing the ecological basis for conservation of shallow marine life using reef life survey. Biological Conservation **252**, 108855 (12 2020). https://doi.org/10.1016/j.biocon.2020.108855, https://linkinghub.elsevier.com/retrieve/pii/S0006320720309137
18. Edwards, C.B., Eynaud, Y., Williams, G.J., Pedersen, N.E., Zgliczynski, B.J., Gleason, A.C., Smith, J.E., Sandin, S.A.: Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. Coral Reefs **36**, 1291–1305 (12 2017). https://doi.org/10.1007/s00338-017-1624-3, http://link.springer.com/10.1007/s00338-017-1624-3
19. Ehrenberg, J., Winston, M., Oliver, T., Couch, C.: Development of a semi-automated coral bleaching classifier in coralnet: A summary of standard operating procedures and report of results. NOAA Technical Memorandum NMFS PIFSC 133, NOAA Pacific Islands Fisheries Science Center (2022). https://doi.org/10.25923/d0re-9y93, https://repository.library.noaa.gov/view/noaa/47285
20. Friedman, A.: Squidle+, a tool for managing, exploring & annotating images, video & large-scale mosaics (2020), https://squidle.org/
21. Fu, Y., Wang, R., Ren, B., Sun, G., Gong, B., Fu, Y., Paudel, D.P., Huang, X., Van Gool, L.: Objectrelator: Enabling cross-view object relation understanding across ego-centric and exo-centric perspectives. In: ICCV (2025)
22. Fu, Y., Wang, Y., Pan, Y., Huai, L., Qiu, X., Shangguan, Z., Liu, T., Fu, Y., Van Gool, L., Jiang, X.: Cross-domain few-shot object detection via enhanced open-set object detector. In: ECCV (2024)

23. Gharaee, Z., Gong, Z., Pellegrino, N., Zarubiieva, I., Haurum, J.B., Lowe, S.C., McKeown, J.T.A., Ho, C.C.Y., McLeod, J., Wei, Y.Y.C., Agda, J., Ratnasingham, S., Steinke, D., Chang, A.X., Taylor, G.W., Fieguth, P.: A step towards worldwide biodiversity assessment: The bioscan-1m insect dataset. Advances in Neural Information Processing Systems **36** (2024), http://arxiv.org/abs/2307.10455
24. González-Rivero, M., Rodriguez-Ramirez, A., Beijbom, O., Dalton, P., Kennedy, E.V., Neal, B.P., Vercelloni, J., Bongaerts, P., Ganase, A., Bryant, D.E., Brown, K., Kim, C., Radice, V.Z., Lopez-Marcano, S., Dove, S., Bailhache, C., Beyer, H.L., Hoegh-Guldberg, O.: Seaview survey photo-quadrat and image classification dataset (2019). https://doi.org/10.14264/uql.2019.9303, https://espace.library.uq.edu.au/view/UQ:734799
25. Gu, J., Stevens, S., Campolongo, E.G., Thompson, M.J., Zhang, N., Wu, J., Kopanev, A., Mai, Z., White, A.E., Balhoff, J., Dahdul, W., Rubenstein, D., Lapp, H., Berger-Wolf, T., Chao, W.L., Su, Y.: BioCLIP 2: Emergent properties from scaling hierarchical contrastive learning. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025), https://openreview.net/forum?id=yPC9zmkQgG
26. Gu, J., Stevens, S., Campolongo, E.G., Thompson, M.J., Zhang, N., Wu, J., Kopanev, A., Mai, Z., White, A.E., Balhoff, J., Dahdul, W.M., Rubenstein, D., Lapp, H., Berger-Wolf, T., Chao, W.L., Su, Y.: TreeOfLife-200M (revision a8f38b4) (2025). https://doi.org/10.57967/hf/6786, https://huggingface.co/datasets/imageomics/TreeOfLife-200M
27. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016), https://doi.org/10.48550/arXiv.1512.03385
28. Hill, J., Wilkinson, C.: Methods for ecological monitoring of coral reefs. Tech. rep., Australian Institute of Marine Science (2004)
29. Horn, G.V., Cole, E., Beery, S., Wilber, K., Belongie, S., Aodha, O.M.: Benchmarking representation learning for natural world image collections. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (3 2021), http://arxiv.org/abs/2103.16483
30. Hughes, T.P., Barnes, M.L., Bellwood, D.R., Cinner, J.E., Cumming, G.S., Jackson, J.B.C., Kleypas, J., van de Leemput, I.A., Lough, J.M., Morrison, T.H., Palumbi, S.R., van Nes, E.H., Scheffer, M.: Coral reefs in the anthropocene. Nature **546**, 82–90 (6 2017). https://doi.org/10.1038/nature22901, https://www.nature.com/articles/nature22901
31. Ilharco, G., Wortsman, M., Wightman, R., Gordon, C., Carlini, N., Taori, R., Dave, A., Shankar, V., Namkoong, H., Miller, J., Hajishirzi, H., Farhadi, A., Schmidt, L.: Openclip (Jul 2021). https://doi.org/10.5281/zenodo.5143773, https://doi.org/10.5281/zenodo.5143773
32. Li, B., Zhang, Y., Guo, D., Zhang, R., Li, F., Zhang, H., Zhang, K., Li, Y., Liu, Z., Li, C.: Llava-onevision: Easy visual task transfer (2024), https://arxiv.org/abs/2408.03326
33. Li, F., Zhang, R., Zhang, H., Zhang, Y., Li, B., Li, W., Ma, Z., Li, C.: Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models (2024), https://arxiv.org/abs/2407.07895
34. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows . In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9992–10002. IEEE Computer Society, Los Alamitos, CA, USA (Oct 2021). https://doi.org/10.1109/ICCV48922.2021.00986, https://doi.ieeecomputersociety.org/10.1109/ICCV48922.2021.00986

35. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11976–11986 (2022), https://doi.org/10.48550/arXiv.2201.03545
36. Lowe, S.C., Misiuk, B., Xu, I., Abdulazizov, S., Baroi, A.R., Bastos, A.C., Best, M., Ferrini, V., Friedman, A., Hart, D., Hoegh-Guldberg, O., Ierodiaconou, D., Mackin-McLaughlin, J., Markey, K., Menandro, P.S., Monk, J., Nemani, S., O'Brien, J., Oh, E., Reshitnyk, L.Y., Robert, K., Roelfsema, C.M., Sameoto, J.A., Schimel, A.C.G., Thomson, J.A., Wilson, B.R., Wong, M.C., Brown, C.J., Trappenberg, T.: Benthicnet: A global compilation of seafloor images for deep learning applications (5 2024), http://arxiv.org/abs/2405.05241
37. Miao, Y., Engelmann, F., Vysotska, O., Tombari, F., Pollefeys, M., Baráth, D.B.: Scenegraphloc: Cross-modal coarse visual localization on 3d scene graphs. In: ECCV (2024)
38. Miao, Y., Zaech, J.N., Wang, X., Despinoy, F., Paudel, D.P., Gool, L.V.: Langhops: Language grounded hierarchical open-vocabulary part segmentation. In: NeurIPS (2026)
39. NOAA Pacific Islands Fisheries Science Center, Ecosystem Sciences Division: National Coral Reef Monitoring Program: Benthic cover derived from analysis of images collected during stratified random surveys (StRS) across American Samoa (2018). https://doi.org/10.7289/v52v2dfw, https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:NCRMP-StRS-Images-AmSam
40. NOAA Pacific Islands Fisheries Science Center, Ecosystem Sciences Division: National Coral Reef Monitoring Program: Benthic cover derived from analysis of images collected during stratified random surveys (StRS) across the Mariana Archipelago (2018). https://doi.org/10.7289/v57m0673, https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:NCRMP-StRS-Images-Marianas
41. NOAA Pacific Islands Fisheries Science Center, Ecosystem Sciences Division: National Coral Reef Monitoring Program: Benthic cover derived from analysis of images collected during stratified random surveys (StRS) of the Hawaiian Archipelago (2018). https://doi.org/10.7289/v5js9nr4, https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:NCRMP-StRS-Images-HI
42. Pan, J., Liu, Y., Fu, Y., Ma, M., Li, J., Paudel, D.P., Van Gool, L., Huang, X.: Locate anything on earth: Advancing open-vocabulary object detection for remote sensing community. In: AAAI (2025)
43. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
44. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021), https://arxiv.org/abs/2103.00020
45. Raphael, A., Dubinsky, Z., Iluz, D., Papathomas, M.: Deep neural network recognition of shallow water corals in the gulf of eilat (aqaba). Scientific Reports **10**(1), 12959 (2020). https://doi.org/10.1038/s41598-020-69201-w, https://www.nature.com/articles/s41598-020-69201-w

46. Ren, B., Li, Y., Zheng, X., Fu, Y., Paudel, D.P., Yang, M.H., Van Gool, L., Sebe, N.: Manifold-aware representation learning for degradation-agnostic image restoration. arXiv preprint arXiv:2505.18679 (2025)
47. Ren, B., Liu, Y., Song, Y., Bi, W., Cucchiara, R., Sebe, N., Wang, W.: Masked jigsaw puzzle: A versatile position embedding for vision transformers. In: CVPR (2023)
48. Sauder, J., Domazetoski, V., Banc-Prandi, G., Perna, G., Meibom, A., Tuia, D.: The coralscapes dataset: Semantic scene understanding in coral reefs (2025), https://arxiv.org/abs/2503.20000
49. Sechidis, K., Tsoumakas, G., Vlahavas, I.: On the stratification of multi-label data. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science, vol. 6913, pp. 145–158. Springer, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23808-6_10
50. Spalding, M., Burke, L., Wood, S.A., Ashpole, J., Hutchison, J., zu Ermgassen, P.: Mapping the global value and distribution of coral reef tourism. Marine Policy **82**, 104–113 (8 2017). https://doi.org/10.1016/j.marpol.2017.05.014, https://linkinghub.elsevier.com/retrieve/pii/S0308597X17300635
51. Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., Robertson, J.: Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. BioScience **57**(7), 573–583 (07 2007). https://doi.org/10.1641/B570707, https://doi.org/10.1641/B570707
52. Stevens, S., Wu, J., Thompson, M.J., Campolongo, E.G., Song, C.H., Carlyn, D.E., Dong, L., Dahdul, W.M., Stewart, C., Berger-Wolf, T., Chao, W.L., Su, Y.: Bioclip (nov 2023). https://doi.org/10.57967/hf/1511
53. Stevens, S., Wu, J., Thompson, M.J., Campolongo, E.G., Song, C.H., Carlyn, D.E., Dong, L., Dahdul, W.M., Stewart, C., Berger-Wolf, T., Chao, W.L., Su, Y.: Bioclip: A vision foundation model for the tree of life (11 2023), http://arxiv.org/abs/2311.18803
54. Stevens, S., Wu, J., Thompson, M.J., Campolongo, E.G., Song, C.H., Carlyn, D.E., Dong, L., Dahdul, W.M., Stewart, C., Berger-Wolf, T., Chao, W.L., Su, Y.: Treeoflife-10m (2023). https://doi.org/10.57967/hf/1972, https://huggingface.co/datasets/imageomics/TreeOfLife-10M
55. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114. PMLR (2019). https://doi.org/10.48550/arXiv.1905.11946, https://doi.org/10.48550/arXiv.1905.11946
56. Team, G., Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., Perrin, S., Matejovicova, T., Ramé, A., Rivière, M., Rouillard, L., Mesnard, T., Cideron, G., bastien Grill, J., Ramos, S., Yvinec, E., Casbon, M., Pot, E., Penchev, I., Liu, G., Visin, F., Kenealy, K., Beyer, L., Zhai, X., Tsitsulin, A., Busa-Fekete, R., Feng, A., Sachdeva, N., Coleman, B., Gao, Y., Mustafa, B., Barr, I., Parisotto, E., Tian, D., Eyal, M., Cherry, C., Peter, J.T., Sinopalnikov, D., Bhupatiraju, S., Agarwal, R., Kazemi, M., Malkin, D., Kumar, R., Vilar, D., Brusilovsky, I., Luo, J., Steiner, A., Friesen, A., Sharma, A., Sharma, A., Gilady, A.M., Goedeckemeyer, A., Saade, A., Feng, A., Kolesnikov, A., Bendebury, A., Abdagic, A., Vadi, A., György, A., Pinto, A.S., Das, A., Bapna, A., Miech, A., Yang, A., Paterson, A., Shenoy, A., Chakrabarti, A., Piot, B., Wu, B., Shahriari, B., Petrini, B., Chen, C., Lan, C.L., Choquette-Choo, C.A., Carey, C., Brick, C., Deutsch, D., Eisenbud, D., Cattle,

D., Cheng, D., Paparas, D., Sreepathihalli, D.S., Reid, D., Tran, D., Zelle, D., Noland, E., Huizenga, E., Kharitonov, E., Liu, F., Amirkhanyan, G., Cameron, G., Hashemi, H., Klimczak-Plucińska, H., Singh, H., Mehta, H., Lehri, H.T., Hazimeh, H., Ballantyne, I., Szpektor, I., Nardini, I., Pouget-Abadie, J., Chan, J., Stanton, J., Wieting, J., Lai, J., Orbay, J., Fernandez, J., Newlan, J., yeong Ji, J., Singh, J., Black, K., Yu, K., Hui, K., Vodrahalli, K., Greff, K., Qiu, L., Valentine, M., Coelho, M., Ritter, M., Hoffman, M., Watson, M., Chaturvedi, M., Moynihan, M., Ma, M., Babar, N., Noy, N., Byrd, N., Roy, N., Momchev, N., Chauhan, N., Sachdeva, N., Bunyan, O., Botarda, P., Caron, P., Rubenstein, P.K., Culliton, P., Schmid, P., Sessa, P.G., Xu, P., Stanczyk, P., Tafti, P., Shivanna, R., Wu, R., Pan, R., Rokni, R., Willoughby, R., Vallu, R., Mullins, R., Jerome, S., Smoot, S., Girgin, S., Iqbal, S., Reddy, S., Sheth, S., Põder, S., Bhatnagar, S., Panyam, S.R., Eiger, S., Zhang, S., Liu, T., Yacovone, T., Liechty, T., Kalra, U., Evci, U., Misra, V., Roseberry, V., Feinberg, V., Kolesnikov, V., Han, W., Kwon, W., Chen, X., Chow, Y., Zhu, Y., Wei, Z., Egyed, Z., Cotruta, V., Giang, M., Kirk, P., Rao, A., Black, K., Babar, N., Lo, J., Moreira, E., Martins, L.G., Sanseviero, O., Gonzalez, L., Gleicher, Z., Warkentin, T., Mirrokni, V., Senter, E., Collins, E., Barral, J., Ghahramani, Z., Hadsell, R., Matias, Y., Sculley, D., Petrov, S., Fiedel, N., Shazeer, N., Vinyals, O., Dean, J., Hassabis, D., Kavukcuoglu, K., Farabet, C., Buchatskaya, E., Alayrac, J.B., Anil, R., Dmitry, Lepikhin, Borgeaud, S., Bachem, O., Joulin, A., Andreev, A., Hardin, C., Dadashi, R., Hussenot, L.: Gemma 3 technical report (2025), https://arxiv.org/abs/2503.19786

57. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jegou, H.: Training data-efficient image transformers &; distillation through attention. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 10347–10357. PMLR (18–24 Jul 2021), https://proceedings.mlr.press/v139/touvron21a.html
58. Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeirotis, P., Perona, P., Belongie, S.: Building a Bird Recognition App and Large Scale Dataset with Citizen Scientists: The Fine Print in Fine-Grained Dataset Collection. In: CVPR (2015)
59. Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S.: The iNaturalist Species Classification and Detection Dataset. In: CVPR. pp. 8769–8778 (2018)
60. Veron, J.: Corals of the World. Volume 1. Australian Institute of Marine Science and CRR Qld Pty Ltd, Townsville, Australia (2000)
61. Veron, J.: Corals of the World. Volume 3. Australian Institute of Marine Science and CRR Qld Pty Ltd, Townsville, Australia (2000)
62. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The Caltech-UCSD Birds-200-2011 Dataset. In: Computation & Neural Systems Technical Report (2011)
63. Wallace, C.C.: Staghorn Corals of the World: A Revision of the Coral Genus "Acropora" (Scleractinia; Astrocoeniina; Acroporidae) Worldwide, with Emphasis on Morphology, Phylogeny and Biogeography. CSIRO Publishing, Collingwood, VIC, Australia (1999)
64. Wang, W., Gao, Z., Gu, L., Pu, H., Cui, L., Wei, X., etc., Z.L.: Internvl3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency (2025), https://arxiv.org/abs/2508.18265

65. Williams, I.D., Couch, C.S., Beijbom, O., Oliver, T.A., Vargas-Angel, B., Schumacher, B.D., Brainard, R.E.: Leveraging automated image analysis tools to transform our capacity to assess status and trends of coral reefs. Frontiers in Marine Science **6** (4 2019). https://doi.org/10.3389/fmars.2019.00222, https://www.frontiersin.org/article/10.3389/fmars.2019.00222/full
66. Wu, J., Hu, X., Guo, J., Liu, J., Zhao, Q., Chen, Z., Ji, X., et al.: IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition. In: CVPR Workshops (2019)
67. Yao, Y., Yu, T., Zhang, A., Wang, C., Cui, J., Zhu, H., Cai, T., Li, H., Zhao, W., He, Z., et al.: Minicpm-v: A gpt-4v level mllm on your phone. arXiv preprint arXiv:2408.01800 (2024)
68. Yu, H., Zhang, D., Xie, P., Zhang, T.: Point-based radiance fields for controllable human motion synthesis. arXiv preprint arXiv:2310.03375 (2023)
69. Zhai, X., Mustafa, B., Kolesnikov, A., Beyer, L.: Sigmoid loss for language image pre-training (2023), https://arxiv.org/abs/2303.15343
70. Zheng, X., Liu, Y., Lu, Y., Hua, T., Pan, T., Zhang, W., Tao, D., Wang, L.: Deep learning for event-based vision: A comprehensive survey and benchmarks. arXiv preprint arXiv:2302.08890 (2023)
71. Zheng, X., Zhu, J., Liu, Y., Cao, Z., Fu, C., Wang, L.: Both style and distortion matter: Dual-path unsupervised domain adaptation for panoramic semantic segmentation. In: CVPR (2023)
72. Zheng, Z., Liang, H., Hua, B.S., Wong, Y.H., Ang, P., Pui, A., Chui, Y., Yeung, S.K.: Coralscop: Segment any coral image on this planet. In: IEEE/CVF CIEEE/CVF Conference on Computer Vision and Pattern Recognitiononference on Computer Vision and Pattern Recognition (2024), https://coralscop.hkustvgd.com
73. Ziqiang, Z., Yaofeng, X., Haixin, L., Zhibin, Y., Yeung, S.K.: Coralvos: Dataset and benchmark for coral video segmentation (2023). https://doi.org/10.48550/arXiv.2310.01946

This appendix is organized as follows. Appendix 7 positions ReefNet relative to CoralNet, detailing the limitations of CoralNet for ML use and how ReefNet addresses them. Appendix 8.1 describes the full ReefNet data collection methodology, including the multi-stage source filtering pipeline. Appendix 8.2 details the annotation quality control process and the custom verification platform. Appendix 8.3 provides further details on the centralized re-verification protocol, reviewer background, and sampling strategy. Appendix 8.4 reports covariate diversity across ecoregions, image white balance, and resolution statistics. Appendix 8.5 describes the Al-Wajh Lagoon dataset used as the cross-source benchmark test set. Appendix 9 provides full training details, including vision model architectures, MLLM LoRA fine-tuning setup, and zero-shot prompting protocols. Appendix 10 presents additional quantitative results, including per-class recall, overall accuracy, and precision/F1 scores. Finally, we provide additional qualitative examples, limitations of the dataset, contributor attribution, and information on data and code availability.

7 ReefNet in Relation to CoralNet

Despite CoralNet's [7] wide adoption as a platform for coral annotations and dataset publication, it presents several challenges that limit its usability for

machine-learning practitioners. This stands in contrast to other biological domains, where the community has built well-established machine-learning-ready datasets (e.g., CUB-200-2011, NABirds, iNaturalist, IP102) [58,59,62,66]. ReefNet addresses these shortcomings, making it a more reliable resource for scientific exploration and algorithmic development. The key limitations of CoralNet are:

- **Lack of a unified label set across sources.** CoralNet aggregates data with heterogeneous and often incompatible label sets, making large-scale integration infeasible.
- **Absence of taxonomically verified hard coral labels.** Many sources use outdated, ambiguous, or generic labels instead of scientifically recognized names (e.g., those in the World Register of Marine Species).
- **No standardized quality control.** Without systematic quality checks, it is difficult to identify reliable samples, which is critical for both training and evaluation.
- **No large-scale benchmark framework.** CoralNet does not provide a standardized evaluation setting for consistent model comparison.

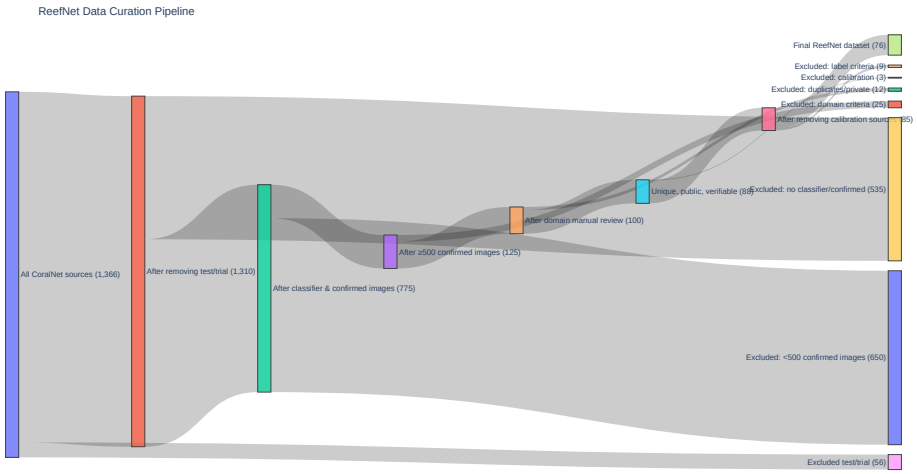


Fig. 5: ReefNet Curation Pipeline. The diagram illustrates the sequential filtering and exclusion steps applied to 1,366 CoralNet image sources to arrive at the final ReefNet dataset of 76 sources. Sources were excluded due to an insufficient number of human-verified annotations, ecological or privacy concerns, and calibration-related issues.

To address these challenges, we propose ReefNet with the following key contributions:

- **World Register of Marine Species (WoRMS) [11]-based unified taxonomic labeling:** ReefNet adopts a unified labeling scheme grounded in WoRMS to integrate heterogeneous sources. This harmonization also enabled the inclusion

of a Red Sea–focused dataset. We anticipate that this taxonomy will encourage future work to adopt scientifically sound annotation practices.
- **Rigorous re-verification:** CoralNet sources were annotated by distinct groups with variable expertise, leading to inconsistencies in label provenance. While acceptable within individual sources, these inconsistencies hinder multi-source integration. ReefNet mitigates this issue by introducing a centralized verification step: a dedicated review team, composed of individuals with complementary backgrounds, re-evaluated all included sources. This ensures a consistent, cross-source quality standard unattainable through direct dataset merging.
- **Standardized benchmarks:** ReefNet establishes reproducible benchmarks with clearly defined evaluation settings, including: i) *Within-source evaluation*, which assesses performance when training and testing data originate from the same set of CoralNet sources and provides a controlled measure of classification accuracy across 39 hard-coral label classes; ii) *Cross-source evaluation*, which measures generalization to a held-out dataset (Al-Wajh Lagoon, Red Sea) with different acquisition conditions, geographic region, and reef assemblage, restricted to the 13 genera shared with ReefNet; iii) *Zero-shot evaluation*, which tests the ability of VLMs and MLLMs to recognize coral genera without any task-specific training; and iv) *Cross-domain few-shot evaluation*, which probes how efficiently models adapt to the ReefNet label space when given only $k \in \{1, 5, 10, 20\}$ labeled examples per class.

8 Additional Information on the Dataset

8.1 ReefNet Data Collection Methodology

The **ReefNet** dataset was constructed through a multi-stage curation pipeline applied to publicly available sources hosted on *CoralNet*. Figure 5 summarizes this pipeline. Beginning with all **1,366 publicly listed CoralNet sources**, we applied a series of semantic, ecological, and technical filters to isolate high-quality, taxonomically relevant reef imagery and annotations. These steps ensured that retained sources featured dense, consistent, and biologically meaningful annotation data.

We first excluded sources with names containing keywords such as *"test"* or *"trial"*—commonly created by users experimenting with the platform—reducing the pool to **1,310 sources**. An exception was made for *CoralNet Assistance Test*, which was retained after manual inspection confirmed its annotation reliability and ecological relevance.

Next, we removed sources lacking a trained classifier or any *confirmed* (i.e., human-annotated) images, yielding **775 sources**. Applying a minimum threshold of **500 confirmed images** further reduced the set to **125 sources**. These were manually reviewed against two domain-specific criteria: (i) the presence of annotations for **Scleractinian** (reef-building hard coral) taxa, and (ii) use of **in situ** imagery from shallow, tropical reef environments, resulting in **100 sources**.

To ensure uniqueness, we eliminated duplicate annotations based on filename, label identity, and point-level annotation coordinates—occasionally leading to

the exclusion of entire sources. We also removed any sources that had become private after our data crawl, yielding **88 public, verifiable sources**. Following consultation with data owners, we excluded three calibration sources used for training novice annotators, resulting in **85 sources**.

A final filtering stage retained only labels that met all of the following criteria: (i) ecologically relevant (i.e., hard coral taxa), (ii) appeared at least 100 times, (iii) were present in a minimum of three distinct sources with at least 10 annotations per source, (iv) exhibited consistent, visually distinguishable patterns, and (v) were taxonomically valid per WoRMS. After filtering and consolidation, the final **ReefNet** dataset consisted of **76 sources** and approximately **925,000 hard coral annotations**.

In parallel, we standardized associated metadata for each source, including geographic location, contributor affiliation, and source history (see Table 7). The complete metadata will be released as a CSV file.

CoralNet Point Sampling Strategies Randomized point sampling is a long-established ecological standard and has also been widely adopted in terrestrial ecology (e.g., canopy cover, bird populations, vegetation structure). While ReefNet models treat each annotation point as an independent patch-level sample, the underlying point annotations originate from CoralNet-selected sources, which employ automated sampling methods across sites with:

- **Simple random:** 51 sources
- **Stratified random:** 23 sources
- **Uniform grid:** 2 sources

ReefNet includes a total of 924,626 point annotations across 181,223 images, with the following per-image statistics:

- **Mean:** 40 points per image
- **Median:** 22 points per image
- **Range:** 25–180 points per image

8.2 ReefNet Data Quality Control

To ensure taxonomic reliability and consistency across ReefNet annotations, we implemented a multi-stage expert verification and filtering process. This process was supported by a custom web-based application developed specifically for large-scale, structured review of coral genus annotations by marine biologists. Expert feedback collected through this platform was used to assess annotation quality, identify systematic labeling errors, and curate high-confidence subsets for benchmark construction.

Manual Verification Tool. We developed a custom web-based platform to support expert review of coral genus annotations (Figure 6). Marine biologists used the tool to verify stratified subsets of annotations across sources and taxa. Each annotation was labeled as *Correct*, *Incorrect*, *Low Image Quality*, or *Hard to Decide*, based on a zoomable image patch together with the full-image context.

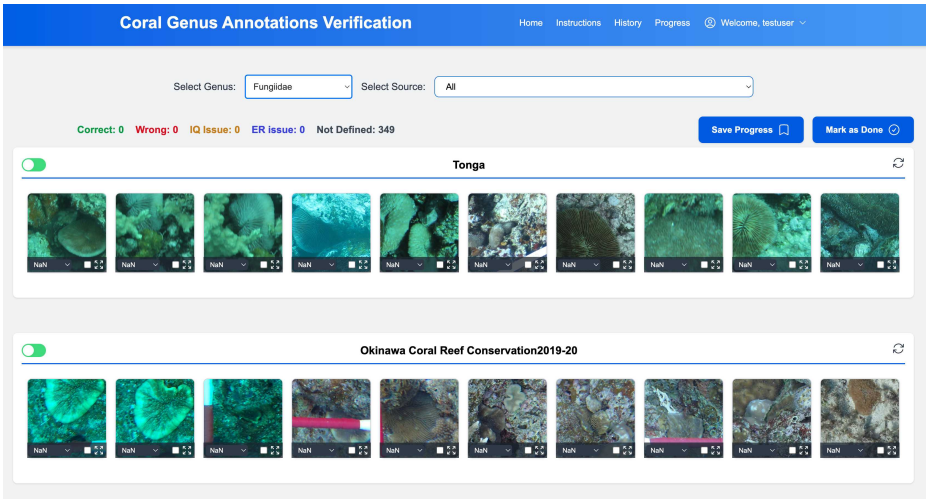


Fig. 6: Screenshot of the ReefNet manual verification platform. Experts assess genus-level annotations by reviewing image patches alongside full-image context and assign structured labels through a dropdown interface. The tool supports efficient, large-scale quality control of coral annotations.

The interface supports genus- and source-specific filtering, progress tracking, and pseudonymized user sessions. The system is lightweight, scalable, and integrated into the ReefNet curation pipeline. All verification actions are logged with metadata to support auditability and downstream filtering. Expert feedback informed exclusion criteria and helped construct the high-confidence benchmark subset, which achieved 92% agreement. These labels were also used to refine taxonomic definitions and guide model retraining.

8.3 More Details on the Centralized Re-Verification Protocol

Our reviewers are not inherently superior to the original CoralNet annotators. However, our audit revealed systematic issues, including taxonomic inconsistencies across sources, frequent mislabeling (e.g., genus-level errors in *Acropora*), and the absence of standardized label conventions. Moreover, because each CoralNet source was annotated by a distinct group with varying taxonomic expertise, labels that may be acceptable within individual sources can introduce substantial noise and bias when merged across datasets. ReefNet addresses this issue by applying centralized expert verification across all sources, thereby establishing a consistent cross-source quality baseline that CoralNet alone cannot provide.

Expert Reviewers' Background. The verification team included one highly experienced coral taxonomist, three senior coral ecologists, and six PhD-level specialists in coral systematics and reef monitoring. Full reviewer details will be provided in the acknowledgments in the camera-ready release.

Review Protocol. We implemented a stratified random sampling procedure covering 8,962 patches (10 per genus per source). Each reviewer independently annotated the selected patches and flagged uncertain samples. Unresolved cases were either excluded or explicitly marked as low confidence. This protocol increased expert agreement from 73% to 92% by retaining only source–class pairs that received high-confidence consensus across reviewers.

Goals of Expert Filtering. The objective of our verification process is not to override CoralNet annotations, but to augment them with standardized, reproducible curation. Specifically, the process (i) standardizes labels through WoRMS AphiaIDs, (ii) resolves cross-source inconsistencies, and (iii) produces a high-confidence dataset suitable for benchmarking machine-learning models in ecological research.

Scalability Considerations. Although the verification tool enables efficient, structured feedback from expert reviewers, its scalability remains constrained by the availability of qualified taxonomists. As ReefNet grows or is extended to more diverse regions, broader human-in-the-loop verification workflows will be needed to support higher annotation throughput. Future directions include active learning to prioritize ambiguous or low-confidence cases, hierarchical review pipelines with tiered expertise, and exploration of consensus-based crowdsourcing for preliminary filtering stages. Addressing these limitations will be critical for sustaining high-quality annotation pipelines in large-scale ecological monitoring.

8.4 Additional Details on Covariate Diversity

This section highlights three key aspects of metadata in ReefNet that are essential for both machine learning and biological applications: i) geographic location (Figure. 7), ii) image white balance, and iii) image resolution.

Ecoregions Distribution. To assess the geographic and ecological diversity of ReefNet, we grouped all annotations by marine ecoregion using the Marine Ecoregions of the World (MEOW) classification system. Figure 7 summarizes the distribution of sources, images, and annotations across 26 unique ecoregions.

The dataset spans a wide range of coral reef habitats, including the Red Sea, Caribbean, Central Indo-Pacific, and Central Pacific. Notably, several biodiversity hotspots are heavily represented: the Hawaiian Islands (18 sources, 221,419 annotations), Samoa Islands (2 sources, 201,685 annotations), and Mariana Islands (6 sources, 107,553 annotations). Other regions, such as the East Caroline Islands and Fiji, also contribute substantial volumes of data, enriching the taxonomic and ecological breadth of the dataset.

The number of image sources per ecoregion varies significantly, reflecting uneven global monitoring efforts. While some ecoregions are represented by multiple contributors and thousands of images, others—such as Tweed–Moreton or the Southwestern Caribbean—appear undersampled. This variation is important for evaluating model robustness and generalization across biogeographically distinct environments.

All counts are shown on a log scale to accommodate differences spanning several orders of magnitude. Overall, this distribution highlights ReefNet's po-

tential to support geographically robust coral classification and cross-region generalization studies.

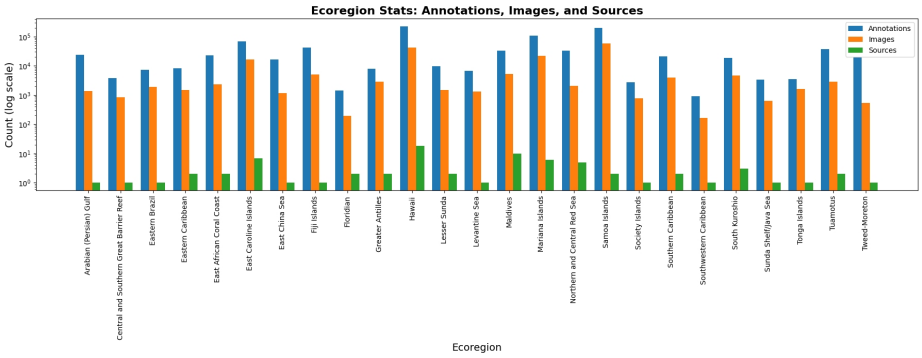


Fig. 7: Distribution of annotations, images, and sources across marine ecoregions. Each bar group represents one of 26 marine ecoregions covered in the ReefNet dataset, as defined by the MEOW classification. The plot shows the number of annotations, distinct images, and image sources per ecoregion (log scale). Geographic coverage is notably high in regions like Hawaii, Samoa, and the Mariana Islands, but more limited in some Atlantic and Western Indian Ocean regions.

Image White Balance. Variations in water clarity, depth, light penetration, and camera configuration can lead to strong color casts in underwater imagery. We therefore document the average intensities of the red, green, and blue color channels across CoralNet sources (Figure 14). Sources with non-overlapping peaks in their RGB distributions generally exhibit strong color casts due to poor white balancing. The variety of color casts captured in ReefNet can improve model robustness; however, certain applications may benefit from using imagery whose lighting conditions more closely match the target deployment setting.

Image Resolution. Image resolution in ReefNet ranges from 0.2 to 27 megapixels, with a median resolution of 12 megapixels. This diversity enhances the generalizability of AI models trained on ReefNet data. As with white balance, however, users may prefer to extract images with a more consistent resolution for specific tasks. Most sources in ReefNet exhibit relatively uniform resolution distributions as a result of standard camera systems, whereas a few sources show substantial variation, generally because benthic quadrats were manually cropped from the original images (Figure 15).

To facilitate the creation of customized sub-datasets, source-level metadata—including geographic location, RGB color profiles, and resolution statistics—will be released with the ReefNet dataset on Hugging Face.

8.5 Al-Wajh Lagoon Dataset

Dataset Overview. As part of a larger environmental monitoring effort, nearly 300 coral reefs were surveyed in the Al-Wajh lagoon (25.6°N, 36.8°E) between March and September 2021. Surveyed reefs were strategically selected using a stratified random design to ensure representation of the region's diverse reef habitats, including Reef Walls, Reef Crests, Reef Slopes, Patch Reefs, and Algal Reefs. The Al-Wajh Test dataset is composed of a subset of this imagery collected from the outer reefs of the lagoon. Imagery in the Al-Wajh Test dataset was collected using purpose-built photoquadrats, ensuring images were taken from a consistent distance from the seafloor. The camera was Canon PowerShot G9 X Mark II which shoots images of 20.2 MP. This represents the upper end of the resolutions available in the ReefNet dataset (Figure 15). A total of **1,376 images** were manually annotated on the CoralNet platform. Per image, 10 points were generated following a stratified random design of 2 rows and five columns, generating a total of **4,624** annotations of hard corals. The geographic coverage and annotation locations of this benchmark subset are shown in Figure 8.

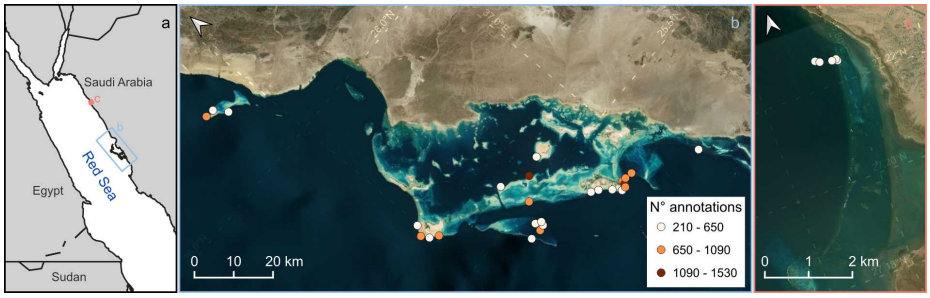


Fig. 8: Map showing the origin of the imagery in the Al-Wajh test dataset. Panel a shows the location of panel b and c in the Red Sea. Panel b shows the Al-Wajh lagoon with the specific locations of the annotations in the test set. Panel c shows the location of a limited set of images taken outside the Al-Wajh lagoon. Map contains Natural Earth data and Bing satellite imagery.

Annotation Distribution. The Al-Wajh Lagoon dataset covers 13 coral genera that overlap with those used to train the different models. Figure 9 shows the annotation distribution, which exhibits a bias toward more abundant classes such as *Porites*, *Acropora*, *Pocillopora*, and *Montipora*.

9 Experimental details

9.1 Training Setup

General Settings. Unless otherwise specified, all vision-only models and VLMs were trained for **20 epochs** using the AdamW optimizer with a base learning

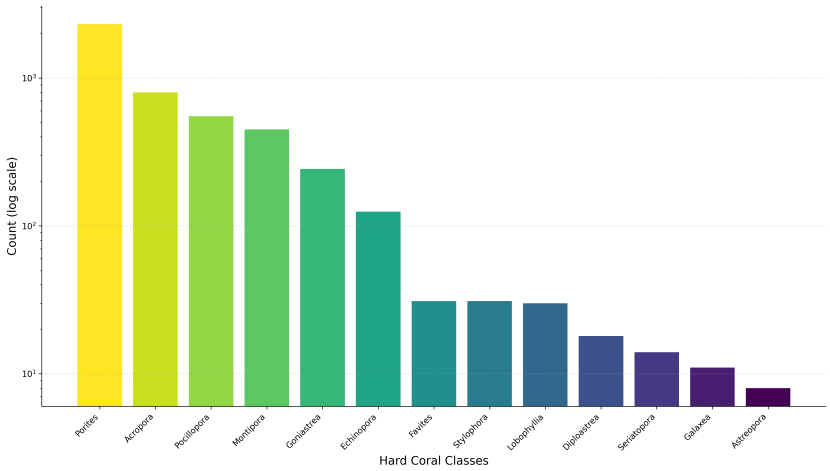


Fig. 9: Log-scale Distribution of Annotations Across 13 Hard Coral Genera in the Al-Wajh Lagoon Dataset. The distribution shows a bias toward some classes, comparable to the distribution in ReefNet.

rate of 2×10^{-5} and a batch size of 16 per GPU across 3 A100 GPUs. Inputs were normalized using ImageNet statistics: mean $[0.485, 0.456, 0.406]$ and standard deviation $[0.229, 0.224, 0.225]$. No data augmentation was applied during training. Training was conducted with automatic mixed precision (BF16) and cosine learning rate scheduling with 5 warmup epochs. Each training sample consists of a 1024×1024 crop extracted on-the-fly from the original full-resolution CoralNet image, centered on each point annotation. A held-out validation set was used to monitor performance and select the best-performing model checkpoint. No validation samples were included in training, ensuring a clean separation between training and evaluation.

Architectures. We evaluated seven vision-only models: ViT-B/16, DeiT-B/16, Swin-B, ResNet-50, ConvNeXt-B, EfficientNet-B3, and BEiT-B/16. All models used pretrained weights from ImageNet-21K or ImageNet-1K (as publicly available via `timm`). All models are fine-tuned and evaluated at the 1024×1024 input resolution described above. We also evaluated five VLMs: CLIP [44], OpenCLIP [31], SigLIP [69], BioCLIP [53], and BioCLIP2 [25]. Among these, CLIP, OpenCLIP, and BioCLIP2 were additionally fine-tuned under full and few-shot settings; SigLIP and BioCLIP were evaluated zero-shot only. For MLLMs, five models were evaluated zero-shot: Qwen3-VL-8B [4], InternVL3.5-8B [64], MiniCPM-V-4.5 [67], LLaVA-NeXT [33], and LLaVA-OneVision [32]. Among these, Qwen3-VL, MiniCPM-V, and InternVL-3.5 were additionally fine-tuned using LoRA.

VLM Fine-Tuning Setup. CLIP, OpenCLIP, and BioCLIP2 were fine-tuned for 20 epochs using AdamW with learning rate 2×10^{-5}, cosine scheduling, and no augmentation. For each genus, we constructed a taxonomy string by concatenating all hierarchical levels from kingdom to genus based on WoRMS (e.g., `"Animalia;`

`Cnidaria;Hexacorallia;Scleractinia;Acroporidae;Acropora"`), and used these strings as text queries. BioCLIP2 was initialized from pretrained weights on TreeOfLife-200M [26]. Training used local and global contrastive loss objectives as in the original BioCLIP setup.

9.2 MLLM Fine-Tuning Setup

All MLLMs were fine-tuned using LoRA via the LLaMA-Factory framework. We used LoRA rank 8 applied to all linear layers (`lora_target: all`), training for **2 epochs** with AdamW (learning rate 5×10^{-5}, cosine schedule, warmup ratio 0.03). Training used BF16 precision with a per-device batch size of 8 and no gradient accumulation. Images were resized to a maximum of 1024×1024 pixels. Flash Attention 2 was enabled where supported. The fine-tuning prompt was: `"What is the genus in the center of the image? Answer with one word."` with the genus label as the target response. Models were trained on 3 A100 GPUs. The three MLLMs fine-tuned were: Qwen3-VL-8B [4], InternVL3.5-8B [64], and MiniCPM-V-4.5 [67].

9.3 Zero-Shot Prompting Protocol

Vision–Language Models (VLMs): For all VLM-based models (CLIP, OpenCLIP, SigLIP, BioCLIP, and BioCLIP2), we use a simple zero-shot text template: "A photo of a {genus}."

For each of the 39 candidate coral genera, we instantiate the template with the genus name and compute the image–text similarity score. The predicted genus corresponds to the text prompt with the highest similarity score.

Multimodal Large Language Models (MLLMs): For MLLMs (Qwen3-VL, InternVL3.5, MiniCPM-V, LLaVA-NeXT, and LLaVA-OneVision), we use the following prompt structure.

> You are an expert marine biologist specializing in hard coral genera recognition.
> Task:
> You are given:
> 1. An image of a coral colony.
> 2. A list of 39 possible coral genera.
> Your job is to identify which genus the coral in the image belongs to.
> Rules:
> - You must select exactly ONE genus.
> - The answer MUST be one of the provided genera.
> - Do NOT provide any explanation.
> - Do NOT describe the coral.
> - Output ONLY the genus name.
> Genus List:
> genus_list

The placeholder `{genus_list}` contains the 39 candidate genus names, which are explicitly included in the prompt.

Output Parsing: We parse the model output using string matching against the provided list of genus names. If the model output contains additional tokens, we extract the substring corresponding to the closest matching genus name from the candidate list.

10 Additional Results

10.1 Per-Class Analysis

Figures 10 and 11 show the per-class recall and normalized confusion matrix for ViT-B/16 on the ReefNet within-source test set (39 genera, 23,043 samples; macro recall 79.8%).

Per-class recall. The four dominant genera—*Acropora*, *Montipora*, *Pocillopora*, and *Porites*—are correctly classified in over 95% of cases, reflecting their abundance in training (collectively 78.4% of samples). Performance degrades progressively for body and tail classes. The three lowest-recall genera are *Favia* (0%, n=3 test samples), *Plesiastrea* (17%, n=29), and *Gardineroseris* (31%, n=26); for *Favia* the test set contains only 3 samples, making any estimate unreliable.

Confusion patterns. The most frequent misclassifications occur among the three most abundant genera: *Acropora*→*Pocillopora* (1.3%), *Pocillopora*→*Acropora* (2.4%), and *Montipora*→*Acropora* (1.4%). Beyond these abundance-driven confusions, several ecologically interpretable patterns emerge. Genera with compact, polygonal corallites—*Acanthastrea*, *Dipsastraea*, and *Favites*—are mutually confused: *Acanthastrea*→*Dipsastraea* (12.0%), *Dipsastraea*→*Favites* (5.9%), and *Favites*→*Dipsastraea* (6.9%). *Platygyra* is confused with *Favites* in 11.7% of cases, reflecting their similar polygonal skeletal patterns. *Astreopora* is misclassified as *Montipora* in 5.8% of cases, a known difficulty arising from their inconspicuous corallite structures in patch-scale images. *Echinopora* is confused with *Montipora* (5.6%) and *Porites* (2.4%), genera sharing comparable surface textures and encrusting growth forms. Among the lowest-recall genera, *Plesiastrea* (n=29) is predominantly misclassified as *Hydnophora* (55.2%) and *Cyphastrea* (20.7%), while *Gardineroseris* (n=26) is most often confused with *Favites* (34.6%) and *Porites* (23.1%).

10.2 Overall Accuracy, Precision, and F1 Score

Table 6 reports macro precision, macro F1, and overall accuracy (micro recall) for all models evaluated on the ReefNet within-source test set (39 genera, 23,043 samples), complementing the macro recall reported in the main paper. Macro recall is included in the table for reference. A note on precision and F1: because some models—particularly in the zero-shot setting—never predict certain classes, macro-averaged precision and F1 may be artificially deflated for those models (per-class precision is defined as 0 for classes with no predicted positives). This

does not affect macro recall, which depends only on predicted vs. ground-truth labels within each class. As an extreme example, LLaVA-NeXT in zero-shot mode collapses to predicting a single genus (*Favia*) for virtually all inputs, yielding near-zero precision and F1.

11 Additional Qualitative Examples

This section presents qualitative examples illustrating model predictions and annotations from the ReefNet benchmarks. Figures 12 and 13 highlight predictions from the within-source ReefNet test set and the cross-source Al-Wajh test set, respectively, offering insights into model behavior across diverse hard-coral categories.

In Figure 12, predictions from a ViT model demonstrate its ability to classify a range of coral genera such as *Porites*, *Acropora*, and *Montipora*. Ground-truth (GT) labels are displayed alongside model predictions and confidence scores, with correct classifications shown in green and misclassifications in red. For instance, while the model achieves high precision on *Acropora* and *Porites*, it occasionally misclassifies *Porites* as *Montipora*, reflecting the difficulty of distinguishing morphologically similar coral types.

Figure 13 presents examples from the Al-Wajh dataset, which poses additional region-specific challenges. This dataset includes coral genera such as *Stylophora* and *Favia*. Notably, consistent confusion between *Porites* and *Montipora* underscores the difficulty of separating genera with overlapping morphological features. Nevertheless, the model demonstrates high confidence in classifying visually distinctive classes like *Goniastrea*.

These qualitative examples emphasize the inherent complexity of coral reef imagery, driven by factors such as morphological similarity, environmental variability (e.g., water clarity and lighting conditions), the presence of multiple benthic organisms in the same patch, and fuzzy class boundaries.

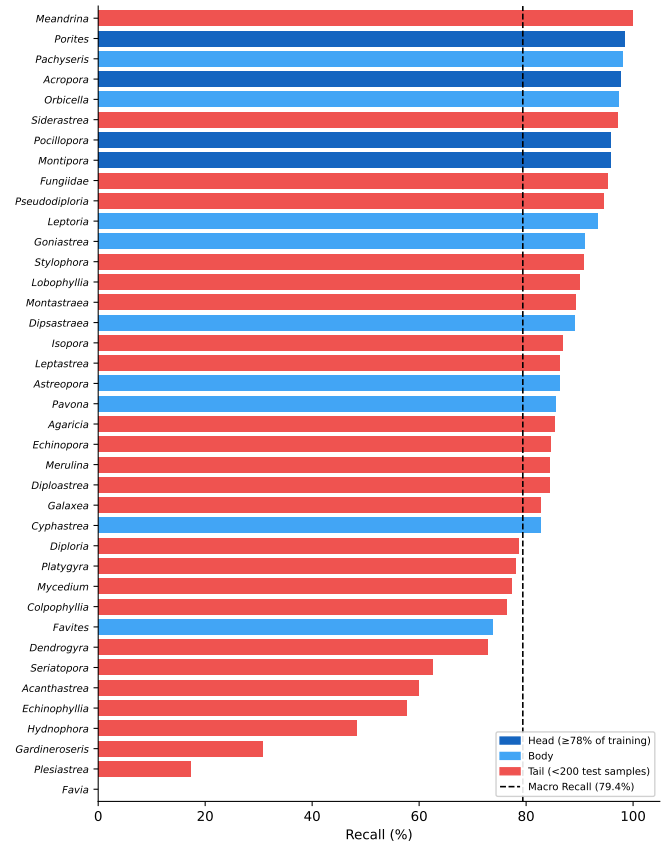


Fig. 10: Per-class recall of ViT-B/16 on the ReefNet within-source test set. Classes are sorted by recall (ascending). Dark blue bars indicate the four head genera (collectively 78.4% of training samples); medium blue bars are body genera; red bars are tail genera with fewer than 200 test samples. The dashed line shows the macro recall (79.8%).

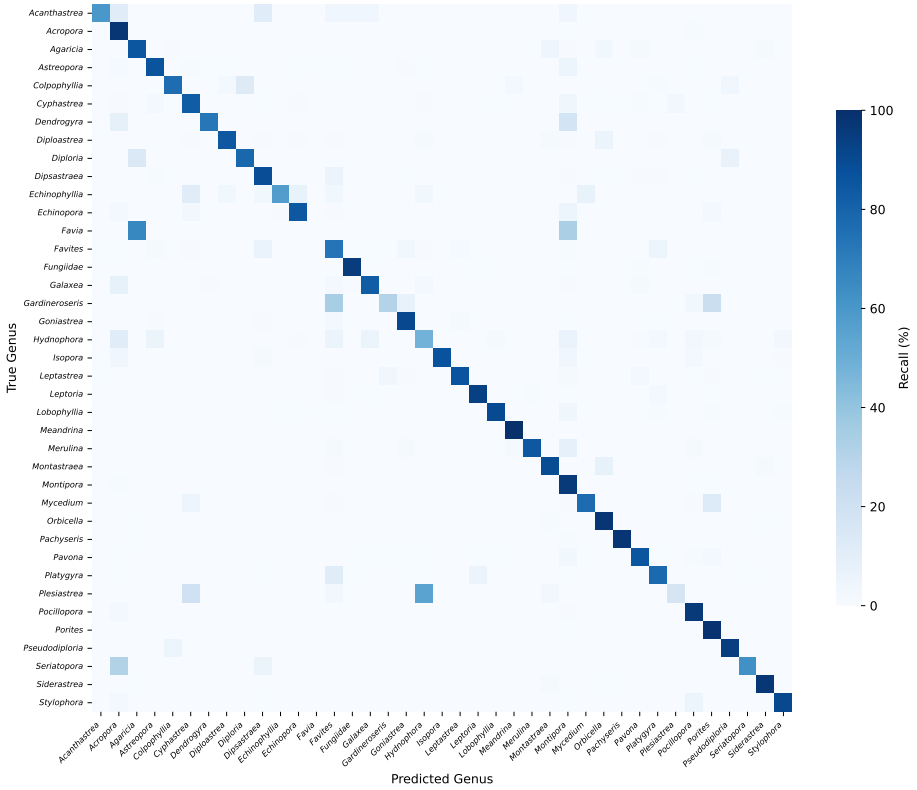


Fig. 11: Normalized confusion matrix for ViT-B/16 on the ReefNet within-source test set. Each row is normalized by the true class count (row sum = 100%), so diagonal entries equal per-class recall. Notable off-diagonal clusters include the abundant genera (*Acropora*, *Pocillopora*, *Montipora*), the compact-corallite group (*Acanthastrea*, *Dipsastraea*, *Favites*), and pairs such as *Platygyra*↔*Favites* and *Astreopora*→*Montipora*.

Table 6: Additional metrics on the ReefNet within-source test set (39 genera, 23,043 samples). Macro recall—the primary metric reported in the main paper—is included for reference. Precision and F1 may be affected by classes absent from model predictions; a caveat noted in Section 10.2. Bold indicates the best result within each model group.

Model	Macro Recall	Macro Precision	Macro F1	Accuracy
Vision Models				
ResNet-50	45.6	55.9	48.3	80.7
EfficientNet-B3	71.3	75.7	72.4	90.2
ViT-B/16	**79.8**	**83.0**	**80.7**	**94.6**
DeiT-B/16	76.2	80.8	77.6	93.7
BEiT-B/16	66.7	73.7	71.0	89.6
Swin-B	77.1	81.9	79.7	93.3
ConvNeXt-B	74.7	80.6	76.6	92.0
VLMs — Zero-Shot				
CLIP-L/14	5.2	3.8	3.4	24.4
OpenCLIP-L/14	4.6	3.9	3.7	29.9
SigLIP	9.0	12.4	5.1	24.1
BioCLIP	11.6	9.6	7.4	19.8
BioCLIP2	**23.5**	**21.6**	**17.1**	**41.2**
MLLMs — Zero-Shot				
LLaVA-NeXT	2.6	0.0	0.0	0.0
LLaVA-OneVision	3.3	2.1	1.1	5.4
MiniCPM-V	3.6	3.3	2.8	22.3
Qwen3-VL	**5.1**	**3.8**	**4.1**	**35.4**
InternVL3	4.0	3.5	2.4	15.8
VLMs — Fine-Tuned				
CLIP-L/14	76.3	65.0	67.6	89.3
OpenCLIP-L/14	76.8	72.9	73.8	91.8
BioCLIP2	**79.2**	**73.8**	**75.5**	**92.2**
MLLMs — Fine-Tuned				
MiniCPM-V	63.4	74.9	65.3	86.6
InternVL3	53.4	64.7	55.5	81.2
Qwen3-VL	**73.0**	**82.2**	**75.5**	**92.1**

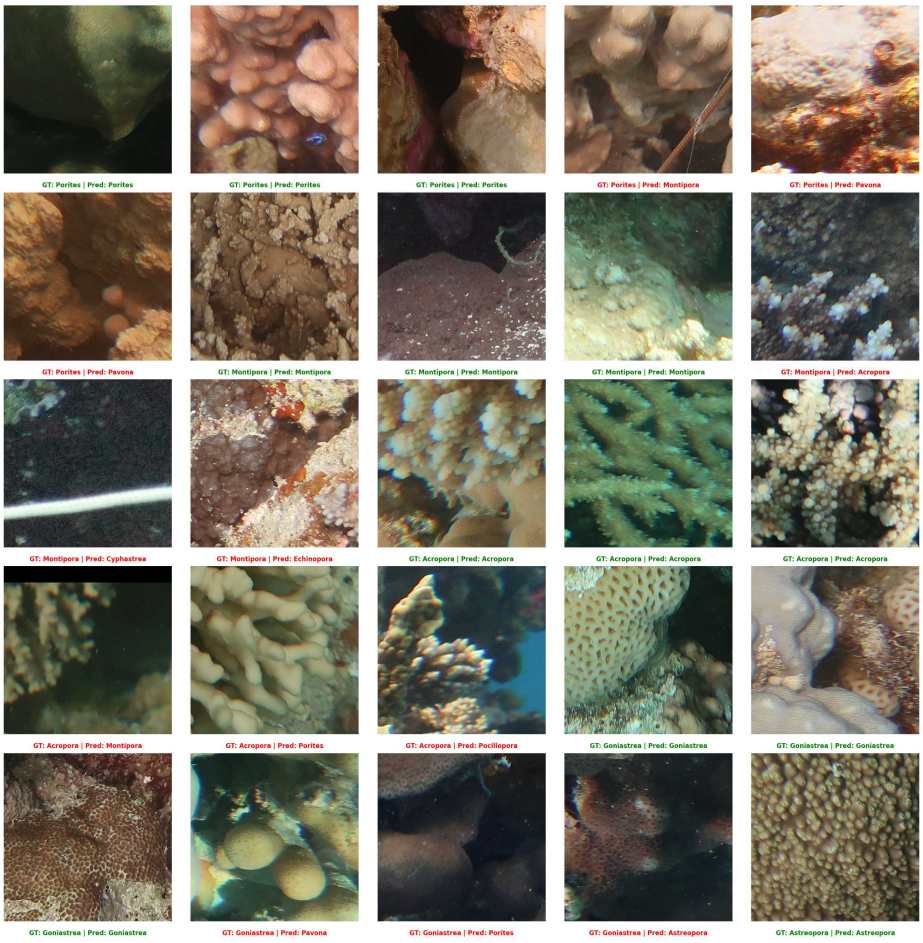


Fig. 12: Qualitative examples from the ReefNet within-source test set. The model shown is a ViT. GT: Ground truth label; Pred: Model prediction.

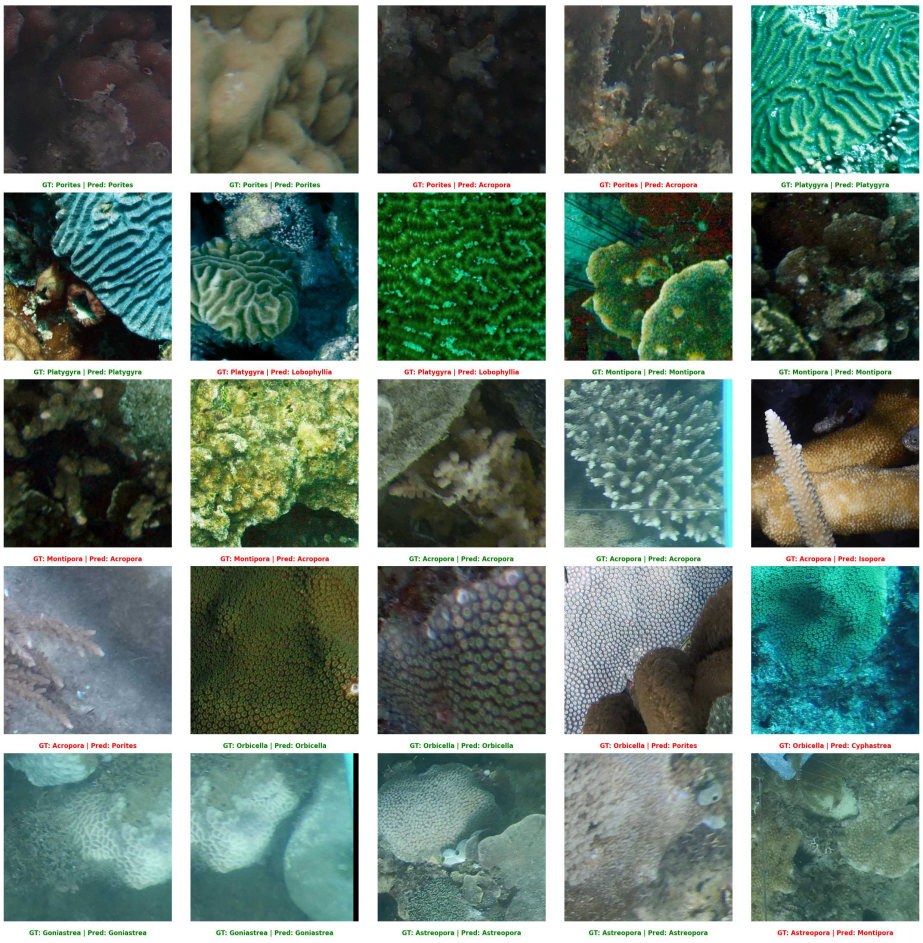


Fig. 13: Qualitative examples from the Al-Wajh dataset. The model shown is a ViT. GT: Ground truth label; Pred: Model prediction.

12 Limitations

While the ReefNet dataset marks a significant advancement in providing standardized, taxonomically fine-grained annotations for coral reef imagery, several limitations should be considered when training models or developing applications based on this dataset:

- **Dynamic Taxonomy:** Coral taxonomy is an evolving scientific field, and taxonomic classifications are subject to revision over time. While the dataset reflects the most current understanding available at the time of compilation, future changes in taxonomy may render some annotations outdated. Users are encouraged to leverage the provided AphiaIDs to verify the latest accepted taxonomy through the World Register of Marine Species (WoRMS) [11].
- **Patch-Based Annotations:** ReefNet annotations are exclusively patch-based, which may not fully capture the spatial variability and ecological context of coral reefs (e.g., Figure 12). Patch-based models, while valuable, may be less accurate than models leveraging more comprehensive semantic segmentation. Although recent large-scale segmentation datasets, such as CoralVOS and CoralSCOP [72, 73], offer valuable insights, they lack fine-grained taxonomic data. ReefNet complements these datasets by providing extensive taxonomically detailed annotations, filling a critical gap for future integration with segmentation-based approaches.

13 Contributor Attribution and Ethical Data Use.

To ensure transparency, traceability, and proper credit to original data providers, we compiled a comprehensive list of all 85 CoralNet sources that passed the pre-final curation stages of ReefNet, along with their corresponding contributors and institutional affiliations. This attribution acknowledges the global community of researchers and practitioners who have made their data publicly available through CoralNet, often as part of long-term ecological monitoring efforts. The information was curated through a combination of CoralNet metadata, institutional websites, and direct outreach to contributors. In line with principles of responsible data stewardship, we have preserved all original attribution metadata and encourage future users of ReefNet to do the same. Note that while the attribution covers all 85 sources that passed pre-final filtering, the final ReefNet dataset consists of 76 sources after the ecological and taxonomic quality filters described in Section 8.1; the classification benchmark in the main paper further restricts to the 68 high-confidence sources that met expert-agreement criteria.

14 Resources Availability

The taxonomically mapped annotations, ReefNet metadata, trained models, and the Al-Wajh Lagoon image collection will be made publicly available on the Hugging Face platform. For the corresponding CoralNet imagery, we will provide

a list of source image URLs and a script to download the data directly from CoralNet, thereby leaving control over the images with the original owners. The code will be released on GitHub.

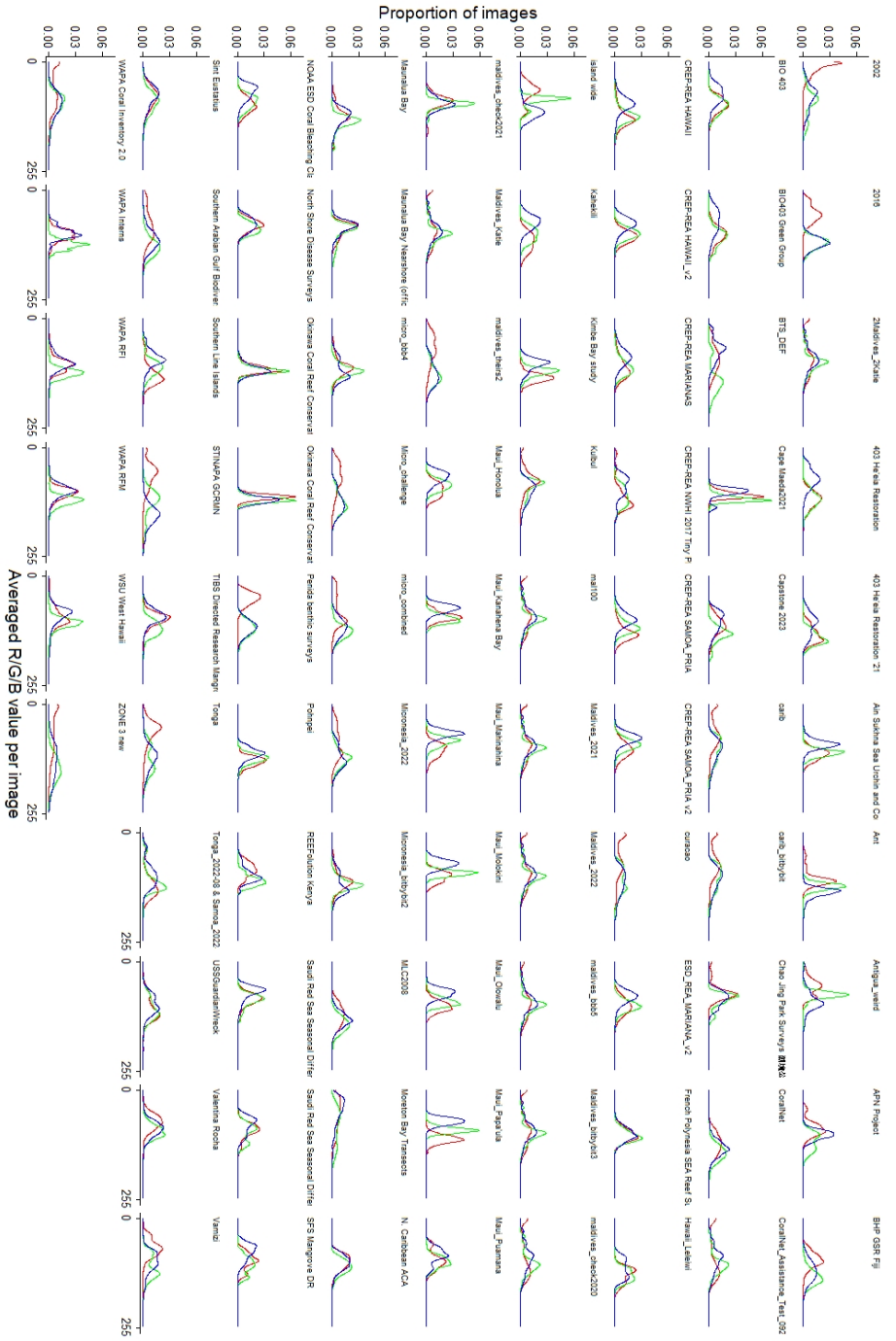


Fig. 14: Density plots per CoralNet source of the average Red, Green, and Blue values of each image.

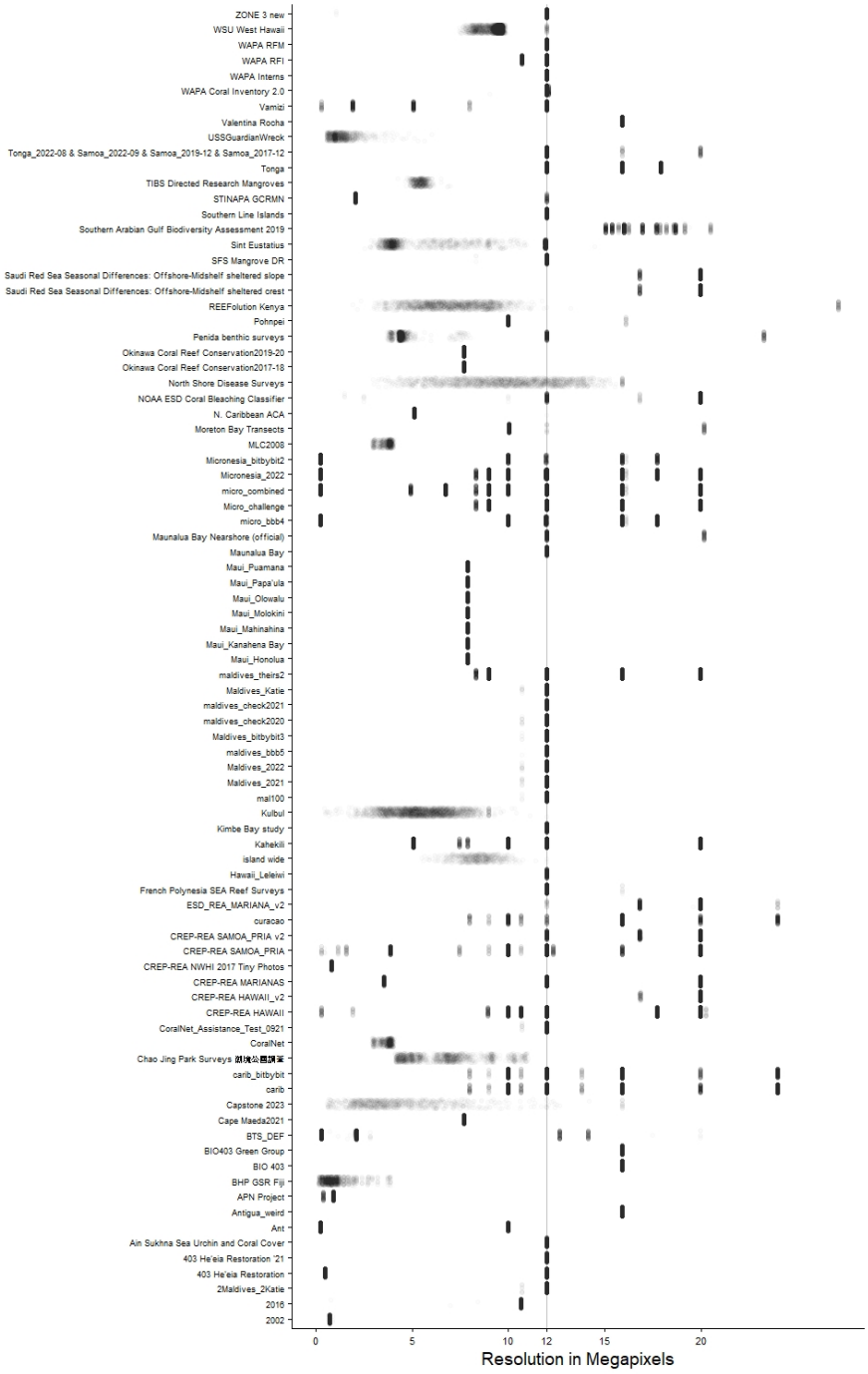


Fig. 15: Scatter plot showing the resolution of each annotated image per source.

Table 7: Contributors and affiliations for the 85 CoralNet sources included in the ReefNet dataset. This list is based on publicly available information from the CoralNet platform, supplemented by our in-depth online research into each source and outreach to their original contributors.

CoralNet source	Listed contributors	Listed affiliations
2002	Zuhairah Dindar	University of New South Wales
2016	Zuhairah Dindar	University of New South Wales
2Maldives_2Katie	Katie Lubarsky, Hugh Runyan	University of California San Diego
403 He'eia Restoration	Cynthia Hunter, Dorian Brunzelle, Kaitlyn Jacobs, Alana Minato, Cody Powers, Connor Antonis, Devon Stapleton, Dylan Rich, Jeany Robledo, Jacob Nygaard, Jordan Pounds-Crihfield, Jacquelyn Simpson, Kevin Christensen, Kaitlin Hooper, Madeline Payne, Renee Wold, Shayna Arakaki, Zachary Clark	Hawaii Institute of Marine Biology, University of Hawaii at Manoa
403 He'eia Restoration '21	Cynthia Hunter, Dorian Brunzelle, Kaitlyn Jacobs, Keisha Bahr, Brooklyn Bennett, Marisa Bhao-Intr, Brittany Kernodle, Corey Ling, Dan Zhuo, Desiree Shaw, Johann Vollrath, Kenzie Vierra, Lena Marinkovich, Lauryn Pisciotto, Mellisa Gajardo, Madeleine Perez, Sophia Hanscom, Samantha Thomas, Toranosuke Degawa, Zada Boyce-Quentin	Hawaii Institute of Marine Biology, University of Hawaii at Manoa
Ain Sukhna Sea Urchin and Coral Cover	Omar Attum	Indiana University Southeast
Ant	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
Antigua_weird	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
APN Project	Roslizawati Ab Lah, Kirsten Benkendorff, Zoe White	University of Malaysia Terengganu, Southern Cross University
BHP GSR Fiji	John Stratford, Jason Lynch	University College London, Zoological Society London

Continued on the next page

CoralNet source	Listed contributors	Listed affiliations
BIO 403	Morgan Guadagnoli, Ana Velasquez, Eliza Beckwith, Haley Weis, Isabella Davila, Jamie Mazurski, Rachel Bagnas, Terra Stevens	University of Hawaii at Manoa
BIO403 Green Group	Morgan Guadagnoli, Ana Velasquez, Eliza Beckwith, Haley Weis, Terra Stevens	University of Hawaii at Manoa
BTS_DEF	Henrique dos Santos, Igor Cruz, Joao Ferreira, Ian Vinicio	Universidade Federal da Bahia
Cape Maeda2021	Tomofumi Nagata	Okinawa Environment Science Center
Capstone 2023	Emily Ogawa, David Hyrenbach, Kristina Bechthold, Leslie Rosa, Ivy Haxo	Hawaii Pacific University
carib	Hugh Runyan, Nathaniel Hanna Holloway	Scripps Institution of Oceanography, University of California San Diego
carib_bitbybit	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
Chao Jing Park Surveys	Emma Chen, Shinya Shikina, Kaixiang Yang, Chu Yu Ling, Joey Hsia, Yu-Chund Chuan, Tzu-Cheng Lin, Cheng Yin Chu, Wen Teng Huang, Yuenyi Leung	National Taiwan Ocean University
CoralNet	Dong Li	Zhejiang University
CoralNet_Assistance_Hugh_2021	Hugh Runyan, Ceiba Becker, Esmaralda Alcantar, Nicole Pedersen	Scripps Institution of Oceanography, University of California San Diego
CREP-REA HAWAII [41]	Annette DesRochers, Andrew Gray, Brett Schumacher, Bernardo Vargas Angel, Courtney Couch, Ivor Williams, Jonathan Charendoff, Morgan Winston Pomeroy, Paula Misa, Tom Oliver, Troy Kanemura, Ari Halperin, Chelsie Counsell, Colt Davis, Isabelle Basden, Jon Ehrenberg, Kerry Reardon, Mia Lamirand, Roseanna Lee	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
CREP-REA MARIANAS [40]	Annette DesRochers, Andrew Gray, Bernardo Vargas Angel, Courtney Couch, Jonathan Charendoff, Morgan Winston Pomeroy, Paula Misa, Tom Oliver, Troy Kanemura, Kaylyn McCoy, Kevin Lino, Marie Ferguson, Mia Lamirand, Nalani Kito-Ho, Winter Jimenez, Brett Schumacher	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center

Continued on the next page

CoralNet source	Listed contributors	Listed affiliations
CREP-REA NWHI 2017 Tiny Photos [41]	Andrew Gray, Bernardo Vargas Angel, Courtney Couch, Jon Ehrenberg	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
CREP-REA SAMOA/PRIA [39]	Annette DesRochers, Andrew Gray, Brett Schumacher, Bernardo Vargas Angel, Courtney Couch, Ivor Williams, Jonathan Charendoff, Morgan Winston Pomeroy, Paula Misa, Tom Oliver, Troy Kanemura, Isabelle Basden, Mia Lamirand, Andrew Shantz, Brittany Huntington, Corinne Amir, Hatsue Bailey, Marie Ferguson, Mollie Asbury, Nalani Kito-Ho, Winter Jiminez, Helen Ford, Nicole Kamalu	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
curacao	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
ESD_REA HAWAII_v2 [41]	Annette DesRochers, Bernardo Vargas Angel, Courtney Couch, Jonathan Charendoff, Morgan Winston Pomeroy, Tom Oliver, Isabelle Basden, Jon Ehrenberg, Hatsue Bailey, John Morris, Mia Larimand, Paula Misa	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
ESD_REA_MARIANA_v2 [40]	Andrew Gray, Bernardo Vargas Angel, Courtney Couch, Jonathan Charendoff, Kaylyn McCoy, Tom Oliver, Ari Halperin, Mia Lamirand, Hatsue Bailey, Nicolas Osborn, Jon Ehrenberg	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
ESD_REA_SAMOA_PRIA_v2 [39]	Andrew Gray, Bernardo Vargas Angel, Courtney Couch, Jonathan Charendoff, Morgan Winston Pomeroy, Paula Misa, Ari Halperin, Isabelle Basden, Mia Lamirand, Hatsue Bailey, Nicolas Osborn, Tom Oliver	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
French Polynesia SEA Reef Surveys	Elliott Bates	International Master of Science in Marine Biological Resourses, Sea Education Association
Hawaii_Leleiwi	Russel Sparks, Devon Aguiar	Department of Land and Natural Resources - Aquatics

Continued on the next page

CoralNet source	Listed contributors	Listed affiliations
island wide	Daniela Escontrela, Elena Turner	University of Hawaii
Kahekili	Bernardo Vargas Angel, Ivor Williams, Andrew Gray,Tye Kindinger, Mia Lamirand,	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
Kimbe Bay study	Alice Williams, Kitty Watts	University of Bristol
Kulbul	Ben Murphy, Caitlin Younis, Hannah Kish, Azri Saparwan, Justin Bovery-Spencer, Tarquin Singleton	GBR Biology
mal100	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
Maldives_2021	Katie Lubarsky, Hugh Runyan, Anupama Sethuraman, Ceiba Becker, Jamie Pettengell	Scripps Institution of Oceanography, University of California San Diego
Maldives_2022	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
Maldives_bitbybit3	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
maldives_bbb5	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
maldives_check2020	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
maldives_check2021	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
Maldives_Katie	Hugh Runyan, Katie Lubarsky	Scripps Institution of Oceanography, University of California San Diego
maldives_theirs2	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego

Continued on the next page

CoralNet source	Listed contributors	Listed affiliations
Maui_Honolua	Russel Sparks	Department of Land and Natural Resources - Aquatics
Maui_Kanahena Bay	Russel Sparks	Department of Land and Natural Resources - Aquatics
Maui_Mahinahina	Russel Sparks	Department of Land and Natural Resources - Aquatics
Maui_Molokini	Russel Sparks	Department of Land and Natural Resources - Aquatics
Maui_Olowalu	Russel Sparks, Tatiana Martinez	Department of Land and Natural Resources - Aquatics
Maui_Papa'ula	Russel Sparks, Tatiana Martinez	Department of Land and Natural Resources - Aquatics
Maui_Puamana	Russel Sparks	Department of Land and Natural Resources - Aquatics
Maunalua Bay	Paula Moehlenkamp	Univeristy of Hawaii
Maunalua Bay Nearshore (official)	Pamela Weiant, Alexandria Barkman	Malama Maunalua
micro_bbb4	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
Micro_challenge	Hugh Runyan, Katie Lubarsky	Scripps Institution of Oceanography, University of California San Diego
micro_combined	Hugh Runyan, Katie Lubarsky, Chris Sullivan, Ahmyia Cacapit, Charles Hambley, Isa Bersamin, Sarah Romero	Scripps Institution of Oceanography, University of California San Diego
Micronesia_2022	Hugh Runyan, Katie Lubarsky, Chris Sullivan	Scripps Institution of Oceanography, University of California San Diego
Micronesia_bitbybit2	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
MLC2008	Dong Li	Zhejiang University
Moreton Bay Transects	Joshua Wirth, Gal Eyal	University of Queensland

Continued on the next page

CoralNet source	Listed contributors	Listed affiliations
N. Caribbean ACA	Alexandra Ordonez Alvarez, Brianna Bambic, Myles Phillips, Bernadette Charpentier	National Geographic, Queensland University, Wildlife Conservation Society, University of Ottowa
NOAA ESD Coral Bleaching Classifier [19]	Courtney Couch, Jonathan Charendoff, Morgan Winston Pomeroy, Tom Oliver, Jon Ehrenberg	National Oceanographic and Atmospheric Administration, Pacific Islands Fisheries Science Center
North Shore Disease Surveys	Julianna Renzi, Maddie Cunningham	University of California Santa Barbara
Okinawa Coral Reef Conservation2017-18	Tomofumi Nagata, Eiji Yamakawa	Okinawa Environment Science Center
Okinawa Coral Reef Conservation2019-20	Tomofumi Nagata	Okinawa Environment Science Center
Penida benthic surveys	Pascal Sebastian, Rinaldi Gotama	Indo Ocean Project
Pohnpei	Hugh Runyan	Scripps Institution of Oceanography, University of California San Diego
REEFolution Kenya	Ewout G. Knoester, Anniek Vos, Bulisa Masiga, Jowan van Lente, Luc Visser, Mercy Zawadi Katana, Omar F. Yusuf	Wageningen University and Research, REEFolution Trust
Saudi Red Sea Seasonal Differences: Offshore-Midshelf sheltered crest	Clara Nuber, Matt Tietbohl, Karla Gonzalez	King Abdullah University of Science and Technology
Saudi Red Sea Seasonal Differences: Offshore-Midshelf sheltered slope	Clara Nuber, Matt Tietbohl, Karla Gonzalez	King Abdullah University of Science and Technology
SFS Mangrove DR	Max Vierling, Samantha Krausse, Toni Trinh	School for Field Studies
Sint Eustatius	Myrsini Lymperaki	Univeristy of Amsterdam
Southern Line Islands	Nicole Pedersen, Samantha Clements	Scripps Institution of Oceanography, University of California San Diego
STINAPA GCRMN	Caren Eckrich, Tessa Haanskorf, Angelica Verschragen	STIchting NAtionale PArken Bonaire, Wageningen University and Research, University of Amsterdam

Continued on the next page

CoralNet source	Listed contributors	Listed affiliations
TIBS Directed Research Mangroves	Emma Greenberg, Jenna Shea, Rachel Schneider	School for Field Studies
Tonga	Patrick Smallhorn-West, Lucy Southworth	James Cook University
Tonga_2022-08 & Samoa_2022-09 & Samoa_2019-12 & Samoa_2017-12 [39]	Chris Sullivan, Katie Lubarsky, Gloria Mariño-Briceño, Hannah Gower, Kylie Yogi, Phi Lang	Scripps Institution of Oceanography, University of California San Diego
USSGuardianWreck	Catherine Kim, Ben Neal, Dominic Bryant	Univeristy of Queensland
Valentina Rocha	Valentina Rocha, Emily Esplandiu	University of Miami
Vamizi	Marques da Silva Isabel, Erwan Sol, Felix Domadoma	Center for Research and Environmental Conservation - Lurio University
WAPA Coral Inventory 2.0	David Burdick, Colin Lock, Melissa Vaccarino	National Park Service
WAPA Interns	Ashton Williams, Marisa Agarwal, Andrew O'Connor, Christina Kilkeary, Emma Vaughn, Erin Mullins, Katherine Tangney, Malvika Shrimali, Michelle Diminuco, Motusaga Vaeoso, Natalie Scott, Nicholas Burgos, Philippe Astier, Ryan Stanley, Sarah Yokota, Serena Butler, Ashley Swafford, Xavier Quinata	National Park Service
WAPA RFI	Anneke Padmos, Julia Padilla, Ronja Steinbach, Tim Clark, Terence Dela Cruz, Eliza Frances Manglona	National Park Service
WAPA RFM	Anneke Padmos, Julia Padilla, Ronja Steinbach, Tim Clark, Terence Dela Cruz	National Park Service
WSU West Hawaii	Brian Tissot, Molly Bogeberg	Washington State University
ZONE 3 new	Jamila Hassan, Sulemani Mohamed	Wildlife Conservation Society