Front Matter

Full Title

Digitizing the coral reef: machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats

Short title

Digitizing the coral reef with spectral imaging

Authors

Daniel Schürholz[1,2], Arjun Chennu[2]

1. Microsensor Group, Max Plank Institute for Marine Microbiology, dschuerh@mpi-bremen.de
2. Data Science and Technology, Leibniz Centre for Tropical Marine Research, arjun.chennu@leibniz-zmt.de

Abstract

Existing coral reef survey methods struggle to accurately capture the taxonomic detail within the complex spatial structure of benthic communities. We propose a workflow to leverage underwater hyperspectral transects and two machine learning algorithms to produce dense habitat maps of 1150 m^2 of reefs across the Curaçao coastline. All 500+ million pixels were assigned to 43 labels at taxonomic family, genus or species level for corals, algae, sponges, or to substrate labels such as sediment, turf algae and cyanobacterial mats. With low annotation effort (2% pixels) and no external data, this workflow enables accurate (Fbeta 87%) survey-scale mapping, with unprecedented thematic and spatial detail. Our assessments of the composition and configuration of 23 transect communities showed high consistency. The dense habitat maps reveal the inadequacies of point sampling methods to accurately describe reef benthic communities. Digitizing the reef habitat structure enables validation and novel analysis of pattern and scale in coral reef ecology.

Introduction

Under rapidly changing environmental conditions, the need for accurate and speedy ecological assessment of marine and freshwater ecosystems has greatly increased. This is particularly pressing for coral reefs, which are the most biologically diverse marine ecosystems on the planet, but have suffered significant deterioration in recent years due to a variety of stressors (1, 2). The continuing deterioration of coral reef health worldwide will endanger many of the ecosystems services that these reefs provide to coastal populations and other associated systems (3). Effective protection and restoration measures are required to alleviate the anthropogenic pressure exerted on the reefs. To be effective, these measures must be guided by frequent and accurate monitoring of coral reef health. However, the long-term degradation of reefs confounds an inter-generational understanding of baseline reef health that informs reef restoration and management interventions (4), thus highlighting the need for objective assessments of reef health through monitoring.

Modern reef monitoring efforts focus on the creation of benthic habitat maps, as they capture the spatial distributions of species, habitat features and their environmental context (5, 6). Such information captured over a long time series forms the basis of scientific evaluation of the ecosystem's evolution, and underpins the decisions for management, conservation and restoration (7, 8). The spatial, temporal and thematic scales of ecosystem mapping have a critical influence on the viability of specific scientific analyses (9), such as elucidating functional drivers, detecting community phase shifts or signalling deterioration of habitats. Capturing the structure and function of benthic communities in a coral reef helps reef ecologists to understand the sensitivity of the reef structure to drivers of change (10, 11). Due to the large spatial range (coastlines) and regular (daily-to-weekly) coverage by satellite imagers, habitat maps created from remote sensing efforts enable the analysis of regional and global phenomena affecting coral reefs (12–14). However, studying the local interactions between the biotic constituents of a reef with the surrounding abiotic milieu through satellite-derived maps has remained constrained due to insufficient spatial and thematic resolution (9). Such analyses require reef inventories in the form of benthic habitat maps with deep taxonomic and spatial detail, as often done in terrestrial studies (15, 16). These detailed inventories, typically derived from in-situ surveys, have been carried out in less than a fraction of coral reef regions worldwide (17). Thus, the divergence in spatial scales (range and resolution), temporal scales (repeatability of a study) and thematic scales (taxonomic/geomorphic) between remote sensing and in-situ studies needs to be reduced to be able to produce coherent comparisons between the benthic habitat maps produced by each method.

The latest improvements in Earth observation technologies have enhanced the spectral and spatial resolutions of remotely sensed images (0.5–10 m per pixel) that are used in global reef monitoring efforts. As a consequence, better information products can be extracted, such as geomorphological zonation of reefs (18), coral bleaching events maps (12, 19), bathymetric maps (20, 21), reef probability maps (22) and benthic habitat maps (18, 23). These habitat maps have informed the implementation of several marine protected areas (6, 9). The long-term success of implemented conservation efforts is underpinned by the accuracy of maps used to design and track the protected area (9). Measuring the accuracy of satellite-derived maps is a difficult task, impeded by the lack of in-situ validation datasets and the lack of error estimation in those few in-situ datasets that do exist (17, 24). While remote sensing offers a viable approach for large scale analyses of reefs (14), current satellite sensors lack spatial resolution to represent small organisms (<0.5m) and the spectral resolution to potentially differentiate organisms to a deep taxonomic description (25, 26).

Recent meta-analysis has revealed the problems in our current understanding of the drivers of change in coral reef ecosystems (17). The study shows that habitat metrics provided by many surveys do not report any uncertainty values with their results. This, together with the fact that only 0.01% – 0.1% of the world's reefs have been surveyed, compounds the uncertainty in scaling up to ecosystem-level effects. Not only should in-situ reef surveys be prioritized in all biogeographic regions, but they should also phase out sparse descriptions and instead adopt dense and detailed descriptions of the reef structure. It is also recommended to communicate objective estimates of uncertainty or error to improve the accuracy of projections from the survey scale to the ecosystem scale. Therefore, it is important to develop capabilities for dense and detailed mapping within the survey scale that is efficient and objective. Such in-situ maps can provide state-of-the-art validation for current and upcoming satellite missions with improved imaging capabilities.

In-situ surveys are the key to alleviate this lack of validation through carefully designed 'ground-truthing' through underwater/proximal sensing and mapping of shallow coastal ecosystems (27). The lack of conformity in the set of labels (labelspace hereafter) used between satellite and in-situ maps/studies is typically a confounding factor for a meaningful validation of habitat maps (28). The labelspace of global maps usually include broad functional groups in the reef (coral, algae, sediment, etc.), as well as some status indicators (dead, alive, bleached, etc.) or morphological descriptions (branching, massive, weedy, etc.) of colonies (29). This multi-faceted and easy-to-interpret view of the reef structure is useful for coastal management (6). In contrast, in-situ studies enable thematic scales that identify organisms down to genus or species level, as well as different substrate-associated communities such as cyanobacterial mats and turf algae (30). More so, this detailed view includes surrounding abiotic parameters such as substrate type (sand, rock, rubble) and pollutants or artefacts in the reef. Nonetheless, capturing habitat structure with a detailed labelspace is typically limited by the availability of expert time or by logistical constraints. For this reason the reef structure is severely undersampled – both spatially and thematically – in a majority of reef studies. Classic benthic surveying techniques – such as line intersects or point count quadrat sampling – typically pick a low number of points in most surveying protocols (31, 32). While a thematically detailed analysis is possible with these approaches, they are typically sparse in terms of spatial resolution. Thus validation of satellite-derived habitat maps with these techniques is limited to comparing coverage percentages for large coastal regions, neglecting spatial distributions and locations of benthic components (17).

Newer in-situ monitoring techniques are providing ways to densely describe benthic reef habitats. Recent studies have used photogrammetry to stitch a large number of images to produce 3D models, orthomosaics and digital elevation models of the reef benthos (33–35). The acquisition of large number of images has been enabled by the advancements in instrumentation (36–38) and robotic platforms, both aerial (39, 40) and underwater (41–43). Improvements in machine learning (ML) and computer vision tools, especially the application of artificial neural networks, have contributed to better accuracy and throughput of efforts in benthic classification (36, 44, 45) or semantic segmentation (46–48). Semantic segmentation is the task of assigning each pixel in an image with categorical labels, which meaningfully delineate objects and hence digitizes the structure of the scene. State-of-the-art semantic segmentation studies of coral reef sites provide high accuracy of both classification and contouring of biotic targets (such as coral colonies or sponges) and abiotic targets (such as sediment) (48). A recent study used photogrammetry, sparse label augmentation and semantic segmentation networks to produce dense benthic maps with taxonomic detail (up to 35 labels) (46). While the authors report a mean pixel accuracy of 49.93% when augmenting from 300 sparse pixels,

note that their label augmentation procedure often neglects smaller specimens and rarer classes. Another study leverages the images that create each point in a 3D point cloud to classify 10 labels directly into the 3D model (47). While promising, the accuracy in thematically detailed dense maps must be improved for operational deployment for full survey analysis.

Here we demonstrate how to produce accurate, dense and detailed maps of coral reefs from underwater surveys (Fig 1). Dense means that all regions/pixels (biotic and abiotic) in each transect are assigned a habitat label, resulting in full semantic segmentation of the transect without any catch-all 'background' labels. Detailed refers to the thematic detail that is captured by the labels, such as either being taxonomic (species, genus, etc.) or functional (corals, sponges, etc.). We leverage machine learning of underwater hyperspectral transects captured over multiple weeks and locations along the coast of the Caribbean island of Curaçao (38, 49). Our workflow description (Fig 1) considers all the steps from the field survey to the classification of 500+ million pixels to the validation of aggregate habitat descriptions derived from the detailed habitat maps. We show how dense maps can be produced with flexible thematic scales, either at a functional ("reefgroups") or taxonomic ("detailed") labelspace, with clear assessments of methodological accuracy. By implementing two independent ML methods (neural networks and object-based ensemble classification) in parallel, we show that the machine-produced habitat maps are consistent in terms of habitat descriptors, and can be generated with workflows of different complexities. These ML methods can be used to map and analyse transect data at the survey scale, without the need to source training data from external datasets. We provide an assessment of the consistency between the habitat structure of the benthic community as described by the maps produced with each ML method. Finally, we exploit the dense habitat maps of island-wide transects to simulate different sampling efforts and reveal the inadequacies of point sampling methods to accurately describe reef benthic communities.

Results

Annotated survey data

Underwater hyperspectral transects were acquired with the HyperDiver device in an island-wide survey of coral reefs (38). The detailed description of the acquisition and processing is available in the data descriptor (49) and in the *Methods*. From the total 147 transects a subset of 31 were annotated (Fig 1). The annotated data consisted of 23 "learning transects" (373+ million pixels), that were used to train and test classifiers in the machine learning steps of our workflow, and 8 separate "validation transects" (150+ million pixels) that were used to assess the performance of our workflow on unseen data outside of transects in the learning transects.

Following a targeted low-effort annotation protocol (49), the annotation of 2089 small polygonal regions – with a combined 8.2 million pixels – resulted in the "detailed" labelspace of 43 labels which contained both taxonomic labels (genus, species, etc.) and substrate labels (sediment, turf algae, etc.) (Fig 2). The taxonomic labels were remapped to their functional groups, such as *Coral*, *Macroalgae*, *Sponge*, etc., and the substrate labels were retained, which resulted in 11 classes in the "reefgroups" labelspace. Therefore, the same set of reference annotations (polygon regions) labeled by a human expert were represented as two different descriptions of the reef benthic community. Loosely speaking, the detailed labelspace describes the habitat in the perspective of a reef ecologist, while the reefgroups labelspace describes the habitat and community from the perspective of a reef manager. Overall, it was possible to represent each annotated transect with two types of signal (radiance or reflectance) and labelspaces (detailed or reefgroups) for machine learning towards automated classification.

The reference annotations were created with a bias towards reducing human effort (49) rather than providing uniform coverage of samples across the survey data. This resulted in a ML dataset with a relatively high degree of label imbalance, both when considered as a set of polygons or as a set of pixels across the annotated transects (Fig S1, S2). The degree of imbalance meant that, for example, the top 5 classes (*Sediment*, *Turf algae*, *D. strigosa*, *Dictyota*, *S. siderea*) had 789 polygons and 3.16 million pixels, while the bottom 5 rarest classes (*A. cauliformis*, *B. asbestinum*, *D. stokesii*, *Zoanthid*, *L. variegata*) had only 14 polygons and 18464 pixels.

Machine learning for benthic classification

Our workflow successfully employed two different ML methods (see *Methods*) to classify the transect images into habitat maps (Fig 1). The "segmented" method which used ensemble classifiers on irregular image segments generated through unsupervised segmentation, whereas the "patched" method used convolutional neural networks on regular image patches to produce predictions. Both methods were set up to take as input an image segment/patch from either of the signal types and produce the same output: an array of probability values for each label in the labelspace. The predictive performance of the trained classifiers was tested on a set of image annotations, which was spatially disjoint (no shared pixels) from the annotations used for training, as recommended in recent reviews (50). This testing set comprised of 15496 image segments for the segmented method and 50000 pixels for the patched method. The ML setup allowed us to combine the low-complexity segmented method (ensemble classifiers) and the high-complexity

patched method (deep learning) as interchangeable ML components in the workflow for scalable reef mapping.

For the detailed labelspace, the segmented+reflectance combination had the best predictive performance with an Fbeta score of 84.5% (Fig 3A; Table 1). The classifier had 80% to 96% reliable recall for a majority of the 43 labels (see diagonal of Fig 3A), which had sufficient data support. Some labels with low data support showed excellent recall (*A. archeri*, *A. crassa*, *B. asbestinum* and *D. stokesii*), while others showed significant errors (*Zoanthid*, *L. variegata* and *D. anchorata*). Despite having to distinguish between 43 labels, the segmented+reflectance classifier had a high Cohen's kappa coefficient indicating performance which is 83.5% of a perfect classifier (Fig 3A; Table 1). For the same labelspace, the patched+radiance combination (Fig 3B; larger version in Fig S3) performed with a 9% lesser Fbeta score (76.7%). Most classes had a recall value between 70% and 90%, with some rare classes, such as *A. cauliformis* and *B. asbestinum*, having significant errors.

For the reefgroups labelspace, the best predictive performance was 87.5% in the segmented+reflectance combination (Fig 3C) followed closely by the patched+radiance combination with 86.9% (Fig 3D). The latter showed high recall values for all 11 classes, even reaching 96% correctness for the *Coral* class – which consists of 19 different coral genera and species (Fig 2). The highest confusion occurred between *Turf algae* and *Sediment* classes. For both ML methods classifiers had a Cohen's kappa score of 83% towards a perfect classifier (Fig 3; Table 1). Overall, both ML methods showed Fbeta scores between 72% to 87% with better performances on the reefgroups labelspace than on the detailed labelspace (Fig 3; Fig S3-S7; Table 1).

We studied how the classifier performance depended on the quantity and quality of annotated spectral pixels. The quantity of annotated data was measured as unique pixels used during training. Classifier performance was measured by training on various quantities of data but keeping the amount of computing effort constant (see *Methods*). The classifier performance in the segmented method improved significantly with greater quantities of training data (Fig 4A). The greatest improvement was an Fbeta from 76% with 1.47 million pixels to 86% with 5.89 million pixels seen in the detailed labelspace using reflectance spectra. The performance also improved with greater quantities of radiance spectra, but overall the segmented method performed better on reflectance rather than radiance spectra. In contrast, the patched method showed no improvement with larger quantities of training data (Fig 4B). Instead, the performance in the detailed labelspace showed a 1% deterioration when the computing effort was distributed over all the available data (5.93 million pixels). This seemingly unexpected result of poorer performance with more data can be understood by considering the lower number of training iterations over the larger dataset in order to maintain the same computing effort.

The impact of data quality, or spectral resolution, on predictive performance was assessed by varying the number of spectral bands in the training data. We subsampled 10, 25, 50 and 100 uniformly distributed spectral bands out of the original 200 bands. In both the segmented and patched methods (Fig 4C), the predictive performance showed a strong 5% to 15% improvement when the number of spectral bands was increased from 10 to 25 with diminishing improvements when using 50, 100 or 200 spectral bands. Overall, the availability of greater spectral resolution, even when the bands were chosen without special consideration, had a large effect on the performance of the ML methods.

Smooth habitat maps

To generate habitat maps, each segment or patch of a transect image was passed to the trained classifiers (of the corresponding spectral type) to predict the probability array for the labelspace (Fig 1 "Habitat mapping"). The probability map obtained directly from the classifier's prediction (Fig 5) showed generally noisy spatial distribution, with many areas of low confidence. This effect of low and noisy confidence was larger in the detailed than the reefgroups labelspace. Habitat maps produced by taking the label with the maximum probability at each location exhibited correspondingly noisy distribution of labels. Processing the predicted probability map with dense conditional random fields (DCRF) utilized the global context of the transect map to produce a more uniform and high confidence map of probabilities (Fig 5C,E), in which the only remaining low confidence regions corresponded to the border pixels between adjacent targets. The habitat maps produced from the processed probability maps better delineated the benthic scene in the natural view with smooth and contiguous regions (Fig 5A,B).

Beyond the visual cohesiveness, the smoothed habitat maps were quantitatively more consistent than the raw habitat maps in all combinations of ML methods and labelspaces for transects (Fig 5C,F for radiance spectra and Fig S8A,B for reflectance spectra). The map consistency was measured as the average of the label accuracy in each of the annotation regions in all of the annotated transects, i.e. regions used for both training and testing the ML methods. The consistency of the habitat maps in the regions of the validation transects were lower than in the learning transects (Fig 5C-F): consistency for the segmented method dropped from 94% to 43% and from 95% to 56% with the detailed and reefgroups labelspaces (Fig 5D,F), and from 88% to 66% and from 92% to 74% for the patched method (Fig 5C,E). This indicated that the patched method (convolutional neural networks) was better than the segmented method (ensemble object classifiers) at generalizing to unseen data. Classification of transects with the reflectance signals resulted in more extreme drops in consistency up to 76% (Fig S8). Overall, the best predictive performance on data from the validation transects, which was not used in any ML step, was from the patched method.

Smooth habitat maps in both labelspaces were produced for all 31 transects. With each transect approximately 50 m $\times$ 1 m in size, this task involved assigning each of the 500+ million spectral pixels one of 43 labels (detailed) or one of 11 labels (reefgroups) independently. Montages of the habitat maps for all transects were visualized (Fig S9-S12). A selection of interesting sections of these habitat maps were visualized (Fig 6) with the natural view (rows 1 and 4), the detailed map (rows 2 and 5) and the reefgroups map (rows 3 and 6) shown together. In image A:2, the habitat map shows correctly classified and well delineated instances of three coral species (*D. strigosa*, *P. asteroides* and *M. cavernosa*), a sponge (*N. nolitangere)* as well as regions of *Sediment*, *Turf algae* and *Dictyota* macroalgae. In the corresponding reefgroups map in A:3, the same benthic targets are identified to their correct functional group and well delineated. The maps in F:5 and F:6 show another example of fine-grained segmentation of the branches of a specimen of the *Plexauridae* soft coral family. Comparing the maps in D:2 and D:3, or in H:5 and H:6, shows the number of different coral species that can be identified under the broad *Coral* functional group (in orange). Small encrusting taxa such as *Coralline algae* are visible in B:2 and E:2. The delineation of *Cyanobacterial mat* in G:2 and I:2, along with the many regions of *Sediment* and *Turf algae* represent substrate and microbial components of coral reef benthos. The habitat map sections also display some confusions: different sections of the same colony in C:2 are assigned to *Plexauridae* and *Gorgoniidae*, which are both soft coral families with similar digitate morphologies, while in C:3 this same colony is assigned between *Coral* and *Soft coral*. The incorrectly labelled regions of *N. nolitangere* sponge in E:2 (and *Sponge* in

E:3) along the image edge are also errors. These errors likely occur due to the poorly illuminated shadow regions that received predictions with low confidence, and then got reassigned to the *Sponge* class by the DCRF process due to nearby high-confidence regions. Note however that this did not occur for the *Transect tape* in the shadow regions which was predicted with high confidence.

Assessing the habitat structure

The compositional and configurational structures of the reef benthos from the 23 (learning) transects distributed across Curaçao island were assessed and compared (Fig 7). Since both ML methods independently generated habitat maps in each labelspace of each transect, the habitat maps derived from each method were used for pairwise comparisons of the transect's habitat structures. A Bland-Altman analysis of the coral coverage (Fig 7A,B) and the Shannon diversity index (Fig 7C,D) from the reefgroups labelspace maps showed a high degree of correlation and low bias between the segmented and patched methods. One outlier transect was left out of the calculations (see *Methods*). Note that each of the metrics were calculated from all points in the mapped transects. The coverage of all other reefgroups classes, derived from either mapping method, were comparable across the range of values in all the transects (Fig S13). The Bray-Curtis similarity, which compares the compositional structure between the communities in the maps of both methods, had a median value of 82% in the detailed and reefgroups labelspaces with a quartile range of 72%-89% and 75%-90%, respectively (Fig 7E). With the *Sediment* and *Turf algae* substrate classes merged together, the median value for the similarity rose to 89% for the detailed and reefgroups labelspaces, with a quartile range of 78%-92% and 79%-94%, respectively. This improvement in the similarity index indicates a large effect of the inherent definition problem of *Turf algae* on reef habitat mapping (see *Discussion*).

To measure the workflow's ability to classify into both labelspaces correctly, we created an abstracted "detailed-to-reefgroups" version of the detailed lablespace maps, that is assigning all labels to their corresponding reefgroups label. Then the reef community composition was compared between the detailed-to-reefgroups maps and the reefgroups maps. The similarity assessment across the learning transects for the patched method showed an 88% similarity median value with a quartile range of 84%-92%, while the segmented method showed an 87% median value with a quartile range of 77-91%. With the *Sediment* and *Turf algae* classes merged, the patched method showed a similarity median of 90% (quartile range 88%-97%), whereas the segmented method showed a barely improved median similarity of 88% (quartile range 77%-93%). Our proposed workflow recovered a reef community composition that was highly consistent between the taxonomic and functional group descriptions of the reef benthos.

Going beyond community composition, we also compared the consistency in the spatial configuration by calculating the Jaccard index (Fig 7F) for each label between the detailed-to-reefgroups and reefgroups maps for each transect (see *Methods*). Three labels showed better configurational similarity for the patched method, *Cyanobacterial mat* (64% vs. 40%), *Coral* (74% vs. 57%) and *Macroalgae* (51% vs. 29%). Two labels showed better configurational similarity for the segmented method, *Sediment* (83% vs. 78%) and *Turf algae* (72% vs. 67%). The Jaccard score was lower for both methods on rarer classes (Fig S14).

Evaluating the effort-vs-error trade off in reef sampling

One benefit of having dense and accurate habitat maps that capture the composition and configuration of reef benthic habitats is the opportunity to revisit the effort-vs-error relationship of sparser reef sampling techniques. A common technique is to use random sampling of a small subset of points (usually 5 to 25) on

photo-quadrats to count habitat labels and derive community descriptions of reef transects. We used our dense habitat maps to conduct simulations of quadrat-wise random and sparse sampling to estimate habitat metrics – percent coverage or Shannon index (H') – through point-count estimates (see *Methods*). The number of point samples required to achieve a relative error lesser than 5% was assessed (Fig 8). To recover the hard coral coverage in transect T1, which has low biodiversity (H' = 0.6) and low coral coverage (2.3%), 960 points per quadrat were required (Fig 8B,C). For the transect T4, that has the highest biodiversity (H' = 2.6) and coral coverage (34.2%), only 80 points per quadrat were required. Similarly, 960 points per quadrat were needed to recover the sediment coverage in transect T4, where sediment covers only 2% of the benthos (Fig 8D,E). In contrast, only 5 points per quadrat were required to capture the sediment coverage in transect T1, which has the highest sediment coverage (84.6%). The Shannon index was recovered with 10 sampling points per quadrat for transect T4 (H' = 2.6), with 40 points per quadrat for the transects T2 (H' = 1.6) and T3 (H' = 1.3), and with 160 points per quadrat for transect T1 (H' = 0.6) (Fig 8F,G). Overall, higher sampling effort is required to accurately recover the coverage of rare species or to capture the Shannon biodiversity index of scenes with low biotic coverage. To accurately estimate habitat metrics in reef transects, where the expected diversity or coverage is not previously known, at least 160 points per quadrat should be used.

Discussion

The presented reef mapping workflow was able to produce dense habitat maps with an unprecedented degree of thematic (43 labels) and spatial (~2.5 cm/pixel) detail. The rich spectral detail in the HyperDiver data was leveraged by machine learning classifiers to produce highly accurate habitat maps (87% Fbeta) with little annotation effort (2% pixels). These maps cover an area of ~1150m^2 distributed across the reefs of an island coastline. The two labelspaces used in the maps describe the reef benthic biodiversity down to genus and species level as well as abiotic and microbial components, such as turf algae and cyanobacterial mats. Our habitat maps provide a no-pixel-left-behind dense view of entire 50 m long transects, which allows us to localize and identify the visible components of the reef benthic community. Two powerful ML methods with different operational paradigms were used independently to produce highly accurate habitat maps, thus enabling an objective comparison of reef descriptions at big data scale. Our workflow provides a deep habitat structure description (diversity, coverage, composition and configuration), which demonstrated high convergence between both ML methods. Nonetheless, our detailed assessment indicates that deep learning classifiers (i.e. the patched method) are better at generalizing towards new and unseen datasets under comparable annotation and computational effort. Our dense habitat maps created from real-world survey data allowed us to analyze the effort-vs-error relationship in describing reef habitats using traditional point-count sampling methods.

Recently an increasing number of in-situ reef surveys have been conducted globally, providing better data support for coverage estimates of key functional classes (51). The increased availability of data has inspired large scale analyses of reef structure and drivers (17, 52). However, as noted by the authors, data integration remains challenging due to the heterogeneity of the in-situ data in terms of thematic, temporal and spatial scales, the different methodologies used for the surveys, the difficulty to capture cryptic biota and the lack of error estimates (25, 52, 53). Furthermore, considering only coverage information of coarse reef functional groups can hide intra-group shifts and conceal key dynamics of coral reef communities (54). Therefore, reef mapping products must be designed to be flexible in terms of ontology, while capturing as much spatial and thematic detail as possible, to serve the needs of different expert groups, such as reef ecologists or managers (23, 37, 55).

Our workflow uses a detailed thematic scale in the form of 43 benthic categories that describe biotic and abiotic components of the reef habitat. Subsequently a reefgroups labelspace, comprised of functional groups of reef biodiversity, can be abstracted from the taxonomically detailed labelspace through interconnected ontologies, similar to a hierarchical geomorphic zones developed previously (29). To assess the versatility of our workflow to address the needs of reef managers, we directly predicted into the reefgroups maps in parallel. These independent methods of creating reefgroups maps resulted in a consistent retrieval of the composition and configuration of the reef transects (Fig 7E,F). Being able to 'translate' between the two different views of the reef composition allows for efficient exchange between experts in different fields. Reef ecologists can leverage the detailed descriptions to tackle more intricate questions about the reef community and functions, while environmental managers can use the higher level descriptions for zonation analysis. Flexibility in the thematic scale should be a key consideration to digitize the structure of coral reefs across different usage contexts.

To automate the digitization of reef structure, careful consideration of workflow parameters is recommended. In our workflow, we independently utilized two ML approaches: one based on deep neural networks and the other on object-based image analysis. Continual increase in complexity and specificity of ML tasks for automation impedes a clear judgment of an ML method's ability to generalize (50). We implemented two ML methods with different operational paradigms to explore the trade-off in performance vs. complexity of automation. We optimized the hyperparameters, but not extensively, and instead maintained consistency of initial settings or computational effort or data quantity in our comparative runs (Fig 4). By testing the performance on a non-overlapping dataset, we validate the performance of the algorithms. Both segmented and patched methods showed 80% ± 5% weighted accuracies in both labelspaces (Fig 3), the segmented approach performing better with reflectance data and the patched method with radiance data. Another work targeting a similar sized labelspace (35 labels) achieves a mean pixel accuracy of 49.9% with a deep semantic segmentation network (DeepLabV3+) on sparse samples in 729 test images of coral reef orthomosaics (46). Given only 2% of annotated pixels, our workflow mapped, with impressive accuracy on 43 labels, underwater transects with high natural variability. We show that both simple and complex ML methods can produce accurate mapping of reef transects, apparently due to the spectral detail (Fig 4C).

Another methodological focus was the data requirements for the ML algorithms. We assessed the performance of both methods based on the number of unique pixels (in segments or patches) used for training (Fig 4A,B). The patched method needed less data to achieve its peak performance under the same amount of computing power and annotation effort (number of training samples). Despite the variable lighting conditions (i.e. diver shadows, red-channel attenuation, wave focusing) and methodological artefacts (i.e. diving speeds, camera angles, diving altitude) between training and validation transects, the patched method classified into both labelspaces more consistently (Fig 5C-F). Although classifier-level metrics are better for the segmented method (Fig 3), the patched method was 23% better at classifying out-of-distribution data in the detailed labelspace and 18% in the reefgroups labelspace (Fig 5C-D). Given that expert annotation is the bottleneck for reef survey analysis (27, 44), the patched method, with its better generalization capability, provides better performance-per-human-effort compared to the segmented method.

Despite the high accuracies and consistency achieved by the classifiers, pre- and post-processing routines could enhance their outputs. Recent studies describe pre-processing methods to normalize the lighting and angle variability across transects, using colour cards and levellers to correct spectra and rectify shapes in the scenes (56). Also, on a post-processing step expanding the patched method's superior performance on non-training data (i.e. validation transects), more robust decision strategies can be added, such as Monte Carlo dropout inference or ensemble models (57–59). We implemented DCRFs as a post-processing step, which have proven to enhance classification accuracy in previous works (60, 61). DCRFs are a mechanism to include local or global scene-level context of the probability map to refine the predictions at a given image region. We observed an increase in map consistency when post-processing the classifier probabilities with DCRFs, for both ML methods and both labelspaces (Fig 5). DCRFs effectively render a spatial smoothing effect on the habitat maps, removing regions with low classification confidence that are surrounded by higher confidence regions. While this has the effect of boosting both visual and quantitative consistency with the natural view of the scene, DCRFs should be carefully considered to not inadvertently remove small but desirable survey targets, such as juveniles or small colonies.

The final habitat maps from our workflow show spatial and thematic detail of the structure of the coral reef benthic community (Fig 6). Benthic targets are clearly separated into meaningful regions, which represent different substrata, different organisms of various sizes, and even survey artefacts such as reference boards. Small coral colonies and intricate shapes of branching corals, soft corals and sponges, and even transect tapes are correctly delineated and classified (see *Results*). Our workflow is able to accurately map bare sediment, turf algae and cyanobacterial mats achieving a previously missing capability in reef habitat mapping: dense mapping of the microbial components of reef substrata, while integrating them into a benthic community labelspace (Fig 6). This can be used to quantify changes in abundance of cyanobacterial mats or turf algae, usually an indicator of reef deterioration (11). Our workflow achieved a dense no-pixel-left-behind mapping of the intricate spatial structure and taxonomic diversity of coral reef communities.

The primary focus of our workflow development was to convert underwater spectral images into dense habitat maps. The target of our assessment was to go beyond comparing classifier-level metrics, and assess the final habitat maps for consistency and accuracy. We considered the goal of comparing the densely classified maps to photo-quadrats with point-count estimates (49), but the large difference in sample sizes – ~80 x 25 quadrat points vs. millions of classified pixels – made the experiment statistically unsound. Given that 98% pixels (out of 500+ million) do not have reference label annotations and the spatial structure of the transects (Fig S9-S12), it is difficult to assess the maps accuracy on a pixel level. To overcome this limitation, we exploited the fact that the two ML methods produced the same type of output, but worked fully independent of each other in terms of method, input signal and parametrization. We compared the habitat structure of the dense maps in describing the same physical transect. To achieve this we used coverage, as well as composition and configuration metrics, which are key descriptors of habitat structure (62, 63). We consider that if the statistical properties of the habitat maps are similar, our workflow will have succeeded in representing the true composition and configuration of the coral reef transects captured by the underwater surveys.

We assessed the convergence in values of hard coral coverage and Shannon diversity index from the dense habitat maps through a Bland-Altman analysis. The analysis showed that the habitat maps agree on coverage and diversity metrics, except for some small discrepancies (Fig 7A,B; Fig S13). Coverage of corals, sponges and cyanobacterial mats were overestimated by the segmented method in some transects (Fig S13 I-J). The composition of the communities between pairs of dense habitat maps (considering all pixels) was highly similar as shown by the Bray-Curtis similarity index (Fig 7E). This shows that both ML methods independently ascribe similar classes and similar number of those classes to the same transect. Furthermore, the assessment of the communities between the two labelspaces was also very similar (Fig 7E), providing confidence in the thematic flexibility of our workflow. We infer that automated mapping with ML methods of underwater hyperspectral transects can handle intra-class variability (detailed labelspace) as well as inter-group variability (reefgroups labelspace) with high accuracy.

Going beyond the composition, we also assessed the spatial configuration similarity of the benthic community in the transects described by both ML methods. We used the Jaccard index, which is often used to asses instance and semantic segmentation networks in their ability to predict the shapes of the reference annotations. The regions from the abstracted detailed-to-reefgroups maps and direct reefgroups maps yielded Jaccard scores over 60% for the dominant labels (Fig 7F). Given that these comparisons are across hundreds of millions of pixels and over 43 different labels, these results indicate high configurational

similarity between the maps. Therefore, our workflow is able to correctly localize and delineate important targets in benthic habitat maps, despite the degree of thematic detail. Nonetheless, these assessments are inter-comparisons within our workflow and a correct assessment of the configuration of the habitat structure would require dense manual annotations of the transects. The high degree of convergence in the habitat structure mapped independently by the two ML methods using two different hyperspectral signal types gives confidence in the ability of the workflow to produce accurate in-situ descriptions of coral reef habitats with unprecedented detail and analytical throughput.

There has been some debate about the amount of sampling effort necessary to accurately measure reef habitat structure. In part, this is due to the fact that marine habitat surveying is undertaken with a number of different methods and analysis tools, each with different potential sources of error (64, 65). In benthic surveys with photo-quadrats and point count analysis, the 'adequate' number of samples (points) per quadrat to accurately describe the community has been a topic of debate (31, 66–68). The goal is to find a good balance between expert effort (labelling the points in images) and reliability in the derived scene description. Different numbers (from 5 points to 99 points per quadrat) have been proposed and used in survey analyses (67, 69). Simulations of sampling synthetic habitat maps, based on normal distribution of class abundances, to recover the coral and sponge coverage estimates with less than 5% relative error, revealed that the optimal number of points per quadrat ranged between 13 points for a heterogeneous area with high coral coverage and over 600 points for a very homogeneous region with low coral coverage (31). The recommended number of points per quadrat was 80 for transects of unknown community structure, but generally depended on the true underlying diversity and dispersion of the community configuration. Studies on empirical community maps (70) show that more than 340 random points per quadrat may be necessary, depending on the distributional properties of the community.

We contribute to the debate with a reef-scale analysis based on empirical community structure derived from our benthic habitat maps (Fig 8). We simulated quadrat sampling with various degrees of sampling effort, and found that the number of point per quadrat to accurately recover the coverage of rare species exceeded the recommended values in the literature. Even with 1000 sampling points per quadrat, rare species at the transect level could not be detected within the 5% error limit (Fig 8B,D). In contrast to the coverage of individual classes, the Shannon diversity index was captured within 5% relative error with 160 sampling points per quadrat (Fig 8E). We confirm that the rarity and skewness of occurrence significantly impacts the error associated with a constant sampling effort. When dense mapping is not possible, and no prior knowledge of the biodiversity at a site exists, it is advisable to sample over 160 points per quadrat during generation of baseline data, as well as to communicate the uncertainty generated from this sampling design (17). If rare species are an important survey target, then dense habitat mapping represents the least effort towards accurate estimations of species coverage.

We presented an end-to-end workflow that produced consistent habitat maps for a large area of coral reefs. Although the results are encouraging enough to recommend the different methods applied in this work, some limitations are noteworthy. For example, the nature of the images gathered by the push-broom hyperspectral camera (without any georeferencing) meant that the resulting transects are not ideal for photogrammetric techniques, thus hindering the generation of orthomosaics and 3D models. Recent studies are investigating novel techniques to overcome this limitation and have succeeded in producing rectified orthomosaics of hyperspectral transects (71, 72). Furthermore, a recent study has shown that cryptic biota is

often neglected in 2D top-down views of the reef and that 3D models of reef structure help establish the contribution of these organisms to reef biomass (53). In our study, in order to avoid uncontrolled overlap, a one meter gap was left in between adjacent transects. This resulted in the transects providing incomplete coverage of the seafloor in the study site, a methodological occlusion of the reef benthic habitat. Nonetheless, the study sites contained 20 transects each, and represent an extremely detailed sampling of the benthic community structure due to dense mapping. With the blistering pace of development in machine learning, novel architectures and paradigms could be implemented to render better spatial-spectral classification with fewer annotations. Furthermore, image processing techniques that exploit underwater image formation models to account for the water column properties can aid standardization of data for ML (73). Regardless of the data and ML method developments used in reef surveys, we consider it important to evaluate the dense habitat maps (and not the classifiers) in terms of accuracy and completeness, as the measure of progress. To disentangle the effects of changes in ML methods and data, we urge that the original images and annotations be made publicly available so that they can be re-evaluated independently.

Our proposed workflow showcases a way to generate dense habitat maps of coral reefs with flexible thematic detail. This thematic and spatial detail in the maps enables fine-grained analyses of coral reef functions and community dynamics by coral reef ecologists. We seek to unify the perspective of ecologists, environmental managers, remote sensing and machine learning communities involved in the study of coral reefs. Particularly for ecologists and managers, it provides a consistent habitat description with adaptive thematic detail. Between remote sensing and machine learning experts, it offers a perspective on bridging the 'measurement gap' between ML classifiers and the ultimate data products, i.e. habitat maps. The consistency achieved by our mapping workflow, and the patched method in particular, is related to the richness in spectral detail and the spatial acuity of our proximal sensing vantage point of underwater surveys (38, 49). With improvements in cost and performance of underwater spectral surveying technology, it will become feasible to integrate it as a standard in-situ reef monitoring technique. The widespread use of underwater spectral surveying and automated benthic habitat mapping promises to provide the best validation data for aerial Earth observation efforts to map coral reefs globally. The integration of thematic detail into global habitat mapping promises to enable novel analyses of pattern and scale in coral reef ecology.

Materials and Methods

The data acquisition and annotation effort of the underwater hyperspectral transects are described (in full length) in the data descriptor that precedes this work (49). In the following two subsections we provide a brief summary.

Underwater hyperspectral surveying

The HyperDiver (38) is a diver-operated surveying system that contains a hyperspectral push-broom (or line-scan) camera that captures spectral images of the seafloor. Hyperspectral transects were acquired in a survey conducted along the leeward coastline of the Caribbean island of Curaçao. In total, 8 coral reef sites were selected to cover a variety of habitat types with varying environmental and anthropogenic factors, such as proximity to coastal urban settlements, tourism impact, current directions, nutrient run-offs, etc. At each site 10 to 20 transects were acquired, each transect measuring 50 m long and 1 m wide. Transects were acquired at varying depths (3 m to 9 m range) by swimming in a lawnmower pattern and leaving a 1 m gap between adjacent transects (see *Discussion*). The resulting dataset contains 147 hyperspectral transects, from which 31 transects were selected for testing the proposed workflow. Transect images, consisting of a series of spectral lines captured along the transect, were manually rectified due to lack of underwater georeferencing. Each of the 640 pixels in the line-of-view contained 12-bits of radiometric information for each of the 480 wavelengths bands in the 400-800 nm range. The spectral images were interpolated and reduced to 200 bands of 8-bit radiometric precision. From this reference format of transect spectral images, a further subset (N=[10, 25, 50, 100, 150]) of equally spaced bands were selected to test the impact of spectral detail on the ML methods (Fig 4C).

Each of the transect images was captured under natural, and hence varying, light conditions due to depth, cloud cover, water surface conditions and other survey artefacts. The average radiance signal of a grey reference board present in each transect scene was divided from the radiance signal of each pixel to derive pseudo-reflectance signals. All radiance transects were transformed to reflectance transects and both signal types bundled into datasets that were used to create habitat maps.

Benthic annotation and thematic flexibility

Human expert annotations were created to support automated classification of the transects by machine learning methods. The annotators had the natural view (RGB rendering) of the hyperspectral transect and video obtained by a GoPro camera (attached to the HyperDiver) at their disposal to visually identify targets. The annotations consist of small polygons of different sizes laid on the transects to form separate regions (Fig S1, S2). Each region was then assigned a semantic label out of 46 labels. This labelspace grew organically as new targets were being identified by the experts. If possible, biotic classes were annotated to the deepest taxonomic level possible, such as family, genus or species. Substrate classes are represented as coarser categories, such as sediment, cyanobacterial mat and turf algae. Coral rubble was initially considered as a separate class in the labelspace. However, since it was generally covered with turf algae it was removed to prevent confusions in the ML classifications. Survey materials were also included to give semantic labels to any object found in the transects, i.e. transect tape or reference board. This "detailed" labelspace, consisted of 46 benthic classes, of which 3 were removed given their very low representation in the selected transects (<2 annotated regions). The resulting labelspace had 43 final labels.

Each label in the detailed labelspace was then abstracted to a corresponding functional reef group class (Fig 2). As an example, the 19 coral species and genera were abstracted to a class called "Coral". The 11 resulting classes formed the "reefgroups" labelspace. Given that the original annotations were made in the detailed labelspace, the labels were always interchangeable between labelspaces. This flexibility allowed to run the ML setup with annotations in either labelspace, to test the ML classifiers' accuracy; as well as remapping the resulting detailed habitat maps to the reefgroups labelspaces (detailed-to-reefgroups maps) in order to assess the consistency of the final products, the habitat maps.

Machine learning for benthic mapping

Machine learning classifiers were created to predict the identity of each image region based on its spectral-spatial features. The predictions were in the form of a probability value corresponding to each label in a labelspace. The ML methods used in this work were selected for their different approach to predict patterns in the hyperspectral images (see *Discussion*). For each ML method (patched, segmented), classifier models were trained independently for each combination of signal type (radiance, reflectance) and labelspace (detailed, reefgroups). Each of the trained classifiers were tested on a spatially disjoint dataset corresponding to the training dataset.

For the patched method, a deep learning network called the spectral-spatial residual network (SSRN) (74) – developed for remote sensing hyperspectral image analysis – was used to train a classifier. This network identifies spectral and spatial features by first convolving 1D filters in a spectral branch and then convolving 2D filters in a spatial branch over patches from the hyperspectral image. We ran hyper-parameter tuning experiments for the network, and based on good performance we chose parameter values close to the values provided in the original study (74), as illustrated in Figure 1. Other parameters used for training the network were a batch size of 32, a number of epochs of 32 and an input patch size of 13 × 13 pixels. For each pixel in the image, the probability of labels is predicted for the central pixel base don the neighbouring pixels in the image patch. The image was padded with reflection of border pixels to enable selection of patches at the image edges. Cross entropy loss with weight balancing based on class abundance was used as the loss function during network training. The optimizer was a rectified Adam function with a weight decay value of 1.5 and had a β of 0.999 (75). The learning rate value was adjusted by a cyclic learning rate scheduler, oscillating from 1e-8 to 1e-3 in triangular ramps with a step of 250 batches of data samples. For each centre pixel in a patch the SSRN algorithm returned vector of unary potentials which were transformed into label probabilities with the softmax function. The label corresponding to the highest probability in the vector was used to select the corresponding label. After training, each transect was mapped by passing every image patch through the trained network to obtain the predicted label probabilities for the central pixel. The whole network training workflow was built with the PyTorch library (76).

The segmented method consisted of three sequential processing steps to obtain the the label probabilities for each image region (Fig 1). The first step was to reduce each transect image to it's 6 principal components (PC) using the scikit-learn library (77). The second step was superpixel over-segmentation of the transect using the mean PC image as input to the watershed algorithm in the scikit-image library. (78). The parameters for the watershed algorithm were a batch size of 2000 x 640 pixels, 12000 markers per batch and a compactness of 1000. This reduced the transect image into a set of irregularly shaped superpixels which were contiguous groups of pixels with similar characteristics. Descriptors for each hyperspectral superpixel were calculated along the spectral dimension: mean, standard deviation, minimum and maximum values

were calculated for each of the 200 bands. Concatenation resulted in a feature vector of 800 features for each of superpixel. These vectors were then used as input samples for the random forest (RF) ensemble classifier in the scikit-learn library. The parameters for the RF classifier used for transect mapping were 300 base estimators , 2 minimum leaves per tree, 25 as the maximum tree depth and a minimum of 3 samples per split. The function used to measure the quality of a split was the Gini inequality function and the class weights were adjusted inversely to the class abundance in the samples. The RF classifier output was the label probabilities for a superpixel, which were assigned to all the pixels within the superpixel to generate the class probability map of each transect.

The class probability maps of each transects from either ML method was smoothed by refining the probabilities with dense conditional random fields (DCRFs) (60). DCRFs were used to update the probability of each pixel based on the surrounding context, i.e. class probabilities. DCRFs interconnect every pixel in an image through a graph model, thus allowing fusing of long-range and short-range context within the image. The class probability maps were used as unary potential inputs to the DCRF, implemented with the pyDenseCRF python library to obtain the output of smoothed probability maps (Fig 5).

Each class probability map -- from either ML method and with or without smoothing -- was converted to a class map by assigning the identity of the class with the highest probability at each transect image location. The result was a categorical habitat map where every pixel was assigned to one label in the used labelspace.

Comparison of habitat structure

We assessed the habitat maps through simple coverage metrics and through aggregate metrics for diversity, composition and configuration of the habitat structure. For each transect the percentage cover (P_i) was calculated as:

$$P_i = 100 \times (C_i / N)$$

where C_i is the count of pixels of class i in the transect and N is the total pixel count in the transect.

As a diversity measure for each transect we used Shannon diversity index (H'), defined as:

$$H' = -\sum_{i=1}^{R} p_i \ln(p_i)$$

where p_i is the proportion of elements of a class i and R is the total number of classes in the labelspace. To identify biases in the calculated habitat metrics from the habitat maps of the ML methods we applied a Bland-Altman analysis (Fig 7A-D). This analysis consists of two plots that help identify agreement between two quantitative methods of measurement. The first plots the values of both methods for the specific variable – percentage cover of a class or Shannon diversity index – against each other, to identify values deviating from the one-to-one correlation line. In the second plot the differences of the paired measurements are plotted against their averages, to identify the mean of the difference and its ±1.96 standard deviation lines. The bias is read as the gap between the mean of the difference and the 0 difference line. The two methods are considered to be in agreement if 95% of the values lie within the standard deviation lines in the second plot.

To measure the compositional similarity between two classified maps we used the Bray-Curtis similarity (BCS) index (also known as quantitative Sørensen similarity index), defined as:

$$BCS = 1 - \frac{\sum_{i=1}^{R} |A_i - B_i|}{\sum_{i=1}^{R} |A_i + B_i|}$$

where A_i is the count of pixels of class i in map A and B_i is the count of pixels of class i in map B . R is the total number of classes in the labelspace. We used the "braycurtis" function from scipy python library (79). The closer the BCS value is to 1 the more similar the composition of the two compared communities.

We calculated the similarity in configuration of the communities in two habitat maps by calculating the Jaccard score for each reefgroups class. The Jaccard score is also known as the intersection over the union (IoU_i) and is defined as:

$$IoU_i = \frac{A_i \cap B_i}{A_i \cup B_i}$$

Where A_i is the count of pixels of class i in map A and B_i is the count of pixels of class i in map B . When the Jaccard score is 1, then the two maps are identical in configuration and when the score is 0 then the maps are entirely dissimilar. A high Jaccard score requires that across the transect pixels, both the identity and the location of the two maps are a match. Thus, it is a stringent measure of the similarity between two maps or sets.

Effort-vs-error of point-count sampling

We conducted virtual sampling simulations to estimate the error associated with a certain level of sampling effort in assessing the diversity and coverage of key groups through sparse point sampling of transects. For this experiment, we selected 4 transects with Shannon diversity index from low to high (H' = {0.61, 1.26, 1.6, 2.64}). Each transect was divided into 50 non-overlapping quadrats of size (640 pixels x 406-705 pixels) . Each of these quadrats was sparsely sampled with N = {5, 10, 20, 40, 80 ,160, 240, 320, 480, 640, 960} randomly selected points. This means that when N = 5 points, for example, 250 points (50 quadrats x 5 points) were selected from the habitat map of the transect. The random sampling of the quadrats was conducted 250 times for each effort level. From each set of sparse subsampled points in each transect, the coral coverage, sediment coverage and Shannon diversity index were calculated. For each metric, the relative deviation of the value obtained from the subsampled points from the value obtained from all the points in each transect was calculated. We selected 5% relative error as the limit for acceptable error, similar to a previous simulation study (31). The resulting error from changing sampling effort was compared for each of the transects containing a significantly different species diversity and coverage distribution (Fig 8).

References

1. T. P. Hughes, D. R. Bellwood, S. R. Connolly, H. V. Cornell, R. H. Karlson, Double jeopardy and global extinction risk in corals and reef fishes. *Curr. Biol. CB*. **24**, 2946–2951 (2014).

2. D. J. Hughes, R. Alderdice, C. Cooney, M. Kühl, M. Pernice, C. R. Voolstra, D. J. Suggett, Coral reef survival under accelerating ocean deoxygenation. *Nat. Clim. Change*. **10**, 296–307 (2020).

3. O. Hoegh-Guldberg, E. S. Poloczanska, W. Skirving, S. Dove, Coral Reef Ecosystems under Climate Change and Ocean Acidification. *Front. Mar. Sci.* **4** (2017) (available at https://www.frontiersin.org/article/10.3389/fmars.2017.00158).

4. M. Muldrow, E. C. M. Parsons, R. Jonas, Shifting baseline syndrome among coral reef scientists. *Humanit. Soc. Sci. Commun.* **7**, 1–8 (2020).

5. A. Guisan, R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. T. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, T. G. Martin, J. R. Rhodes, R. Maggini, S. A. Setterfield, J. Elith, M. W. Schwartz, B. A. Wintle, O. Broennimann, M. Austin, S. Ferrier, M. R. Kearney, H. P. Possingham, Y. M. Buckley, Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**, 1424–1435 (2013).

6. C. M. Roelfsema, E. M. Kovacs, J. C. Ortiz, D. P. Callaghan, K. Hock, M. Mongin, K. Johansen, P. J. Mumby, M. Wettle, M. Ronan, P. Lundgren, E. V. Kennedy, S. R. Phinn, Habitat maps to enhance monitoring and management of the Great Barrier Reef. *Coral Reefs*. **39**, 1039–1054 (2020).

7. S. A. Foo, G. P. Asner, Scaling Up Coral Reef Restoration Using Remote Sensing Technology. *Front. Mar. Sci.* **6** (2019) (available at https://www.frontiersin.org/article/10.3389/fmars.2019.00079).

8. D. O. Obura, G. Aeby, N. Amornthammarong, W. Appeltans, N. Bax, J. Bishop, R. E. Brainard, S. Chan, P. Fletcher, T. A. C. Gordon, L. Gramer, M. Gudka, J. Halas, J. Hendee, G. Hodgson, D. Huang, M. Jankulak, A. Jones, T. Kimura, J. Levy, P. Miloslavich, L. M. Chou, F. Muller-Karger, K. Osuka, M. Samoilys, S. D. Simpson, K. Tun, S. Wongbusarakum, Coral Reef Monitoring, Reef Assessment Technologies, and Ecosystem-Based Management. *Front. Mar. Sci.* **6**, 580 (2019).

9. V. Lecours, On the Use of Maps and Models in Conservation and Resource Management (Warning: Results May Vary). *Front. Mar. Sci.* **4**, 288 (2017).

10. D. M. De Bakker, E. H. Meesters, R. P. M. Bak, G. Nieuwland, F. C. Van Duyl, Long-term Shifts in Coral Communities On Shallow to Deep Reef Slopes of Curaçao and Bonaire: Are There Any Winners? *Front. Mar. Sci.* **3** (2016), doi:10.3389/fmars.2016.00247.

11. D. M. de Bakker, F. C. van Duyl, R. P. M. Bak, M. M. Nugues, G. Nieuwland, E. H. Meesters, 40 Years of benthic community change on the Caribbean reefs of Curaçao and Bonaire: the rise of slimy cyanobacterial mats. *Coral Reefs*. **36**, 355–367 (2017).

12. G. Liu, C. M. Eakin, M. Chen, A. Kumar, J. L. De La Cour, S. F. Heron, E. F. Geiger, W. J. Skirving, K. V. Tirak, A. E. Strong, Predicting Heat Stress to Inform Reef Management: NOAA Coral Reef Watch's 4-Month Coral Bleaching Outlook. *Front. Mar. Sci.* **5**, 57 (2018).

13. S. F. Heron, L. Johnston, G. Liu, E. F. Geiger, J. A. Maynard, J. L. De La Cour, S. Johnson, R. Okano, D. Benavente, T. F. R. Burgess, J. Iguel, D. I. Perez, W. J. Skirving, A. E. Strong, K. Tirak, C. M. Eakin, Validation

of Reef-Scale Thermal Stress Satellite Products for Coral Bleaching Monitoring. *Remote Sens.* **8**, 59 (2016).

14. J. D. Hedley, C. M. Roelfsema, I. Chollett, A. R. Harborne, S. F. Heron, S. Weeks, W. J. Skirving, A. E. Strong, C. M. Eakin, T. R. L. Christensen, V. Ticzon, S. Bejarano, P. J. Mumby, Remote Sensing of Coral Reefs for Monitoring and Management: A Review. *Remote Sens.* **8**, 118 (2016).

15. M. Brandt, C. J. Tucker, A. Kariryaa, K. Rasmussen, C. Abel, J. Small, J. Chave, L. V. Rasmussen, P. Hiernaux, A. A. Diouf, L. Kergoat, O. Mertz, C. Igel, F. Gieseke, J. Schöning, S. Li, K. Melocik, J. Meyer, S. Sinno, E. Romero, E. Glennie, A. Montagu, M. Dendoncker, R. Fensholt, An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature*. **587**, 78–82 (2020).

16. M. Onishi, T. Ise, Explainable identification and mapping of trees using UAV RGB image and deep learning. *Sci. Rep.* **11**, 903 (2021).

17. E. J. Hochberg, M. M. Gierach, Missing the Reef for the Corals: Unexpected Trends Between Coral Reef Condition and the Environment at the Ecosystem Scale. *Front. Mar. Sci.* **8**, 1191 (2021).

18. M. B. Lyons, C. M. Roelfsema, E. V. Kennedy, E. M. Kovacs, R. Borrego-Acevedo, K. Markey, M. Roe, D. M. Yuwono, D. L. Harris, S. R. Phinn, G. P. Asner, J. Li, D. E. Knapp, N. S. Fabina, K. Larsen, D. Traganos, N. J. Murray, Mapping the world's coral reefs using a global multiscale earth observation framework. *Remote Sens. Ecol. Conserv.* **6**, 557–568 (2020).

19. J. Li, N. S. Fabina, D. E. Knapp, G. P. Asner, The Sensitivity of Multi-spectral Satellite Sensors to Benthic Habitat Change. *Remote Sens.* **12**, 532 (2020).

20. J. Li, D. E. Knapp, M. Lyons, C. Roelfsema, S. Phinn, S. R. Schill, G. P. Asner, Automated Global Shallow Water Bathymetry Mapping Using Google Earth Engine. *Remote Sens.* **13**, 1469 (2021).

21. J. Li, D. E. Knapp, S. R. Schill, C. Roelfsema, S. Phinn, M. Silman, J. Mascaro, G. P. Asner, Adaptive bathymetry estimation for shallow coastal waters using Planet Dove satellites. *Remote Sens. Environ.* **232**, 111302 (2019).

22. J. Li, D. E. Knapp, N. S. Fabina, E. V. Kennedy, K. Larsen, M. B. Lyons, N. J. Murray, S. R. Phinn, C. M. Roelfsema, G. P. Asner, A global coral reef probability map generated using convolutional neural networks. *Coral Reefs*. **39**, 1805–1815 (2020).

23. C. M. Roelfsema, M. Lyons, N. Murray, E. M. Kovacs, E. Kennedy, K. Markey, R. Borrego-Acevedo, A. Ordoñez Alvarez, C. Say, P. Tudman, M. Roe, J. Wolff, D. Traganos, G. P. Asner, B. Bambic, B. Free, H. E. Fox, Z. Lieb, S. R. Phinn, Workflow for the Generation of Expert-Derived Training and Validation Data: A View to Global Scale Habitat Mapping. *Front. Mar. Sci.* **8**, 228 (2021).

24. S. R. Phinn, C. M. Roelfsema, P. J. Mumby, Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. *Int. J. Remote Sens.* **33**, 3768–3797 (2012).

25. E. J. Hochberg, M. J. Atkinson, S. Andréfouët, Spectral reflectance of coral reef bottom-types worldwide and implications for coral reef remote sensing. *Remote Sens. Environ.* **85**, 159–173 (2003).

26. F. E. Muller-Karger, E. Hestir, C. Ade, K. Turpie, D. A. Roberts, D. Siegel, R. J. Miller, D. Humm, N. Izenberg, M. Keller, F. Morgan, R. Frouin, A. G. Dekker, R. Gardner, J. Goodman, B. Schaeffer, B. A. Franz, N. Pahlevan, A. G. Mannino, J. A. Concha, S. G. Ackleson, K. C. Cavanaugh, A. Romanou, M. Tzortziou, E. S. Boss, R. Pavlick, A. Freeman, C. S. Rousseaux, J. Dunne, M. C. Long, E. Klein, G. A. McKinley, J. Goes, R. Letelier, M.

Kavanaugh, M. Roffer, A. Bracher, K. R. Arrigo, H. Dierssen, X. Zhang, F. W. Davis, B. Best, R. Guralnick, J. Moisan, H. M. Sosik, R. Kudela, C. B. Mouw, A. H. Barnard, S. Palacios, C. Roesler, E. G. Drakou, W. Appeltans, W. Jetz, Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. *Ecol. Appl.* **28**, 749–760 (2018).

27. C. M. Roelfsema, S. R. Phinn, in *Coral Reef Remote Sensing: A Guide for Mapping, Monitoring and Management*, J. A. Goodman, S. J. Purkis, S. R. Phinn, Eds. (Springer Netherlands, Dordrecht, 2013; https://doi.org/10.1007/978-90-481-9292-2_14), pp. 375–401.

28. G. M. Foody, Thematic Map Comparison. *Photogramm. Eng. Remote Sens.* **70**, 627–633 (2004).

29. E. V. Kennedy, C. M. Roelfsema, M. B. Lyons, E. M. Kovacs, R. Borrego-Acevedo, M. Roe, S. R. Phinn, K. Larsen, N. J. Murray, D. Yuwono, J. Wolff, P. Tudman, Reef Cover, a coral reef classification for global habitat mapping from remote sensing. *Sci. Data*. **8**, 196 (2021).

30. F. Althaus, N. Hill, R. Ferrari, L. Edwards, R. Przeslawski, C. H. L. Schönberg, R. Stuart-Smith, N. Barrett, G. Edgar, J. Colquhoun, M. Tran, A. Jordan, T. Rees, K. Gowlett-Holmes, A Standardised Vocabulary for Identifying Benthic Biota and Substrata from Underwater Imagery: The CATAMI Classification Scheme. *PLOS ONE*. **10**, e0141039 (2015).

31. E. Pante, P. Dustan, Getting to the Point: Accuracy of Point Count in Monitoring Ecosystem Change. *J. Mar. Biol.* **2012**, e802875 (2012).

32. K. E. Kohler, S. M. Gill, Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. *Comput. Geosci.* **32**, 1259–1269 (2006).

33. E. Prado, A. Rodríguez-Basalo, A. Cobo, P. Ríos, F. Sánchez, 3D Fine-scale Terrain Variables from Underwater Photogrammetry: A New Approach to Benthic Microhabitat Modeling in a Circalittoral Rocky Shelf. *Remote Sens.* **12**, 2466 (2020).

34. R. Ferrari, J. X. Leon, A. J. Davies, J. H. R. Burns, S. A. Sandin, W. F. Figueira, M. Gonzalez-Rivero, Editorial: Advances in 3D Habitat Mapping of Marine Ecosystem Ecology and Conservation. *Front. Mar. Sci.* **8** (2022) (available at https://www.frontiersin.org/article/10.3389/fmars.2021.827430).

35. C. A. Lepczyk, L. M. Wedding, G. P. Asner, S. J. Pittman, T. Goulden, M. A. Linderman, J. Gang, R. Wright, Advancing Landscape and Seascape Ecology from a 2D to a 3D Science. *BioScience*. **71**, 596–608 (2021).

36. M. González-Rivero, O. Beijbom, A. Rodriguez-Ramirez, D. E. P. Bryant, A. Ganase, Y. Gonzalez-Marrero, A. Herrera-Reveles, E. V. Kennedy, C. J. S. Kim, S. Lopez-Marcano, K. Markey, B. P. Neal, K. Osborne, C. Reyes-Nivia, E. M. Sampayo, K. Stolberg, A. Taylor, J. Vercelloni, M. Wyatt, O. Hoegh-Guldberg, Monitoring of Coral Reefs Using Artificial Intelligence: A Feasible and Cost-Effective Approach. *Remote Sens.* **12**, 489 (2020).

37. E. E. Peterson, E. Santos-Fernández, C. Chen, S. Clifford, J. Vercelloni, A. Pearse, R. Brown, B. Christensen, A. James, K. Anthony, J. Loder, M. González-Rivero, C. Roelfsema, M. J. Caley, C. Mellin, T. Bednarz, K. Mengersen, Monitoring through many eyes: Integrating disparate datasets to improve monitoring of the Great Barrier Reef. *Environ. Model. Softw.* **124**, 104557 (2020).

38. A. Chennu, P. Färber, G. De'ath, D. de Beer, K. E. Fabricius, A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Sci. Rep.* **7**, 1–12 (2017).

39. E. Casella, A. Collin, D. Harris, S. Ferse, S. Bejarano, V. Parravicini, J. L. Hench, A. Rovere, Mapping coral reefs using consumer-grade drones and structure from motion photogrammetry techniques. *Coral Reefs*. **36**, 269–275 (2017).

40. K. E. Joyce, S. Duce, S. M. Leahy, J. Leon, S. W. Maier, K. E. Joyce, S. Duce, S. M. Leahy, J. Leon, S. W. Maier, Principles and practice of acquiring drone-based image data in marine environments. *Mar. Freshw. Res.* **70**, 952–963 (2018).

41. M. González-Rivero, P. Bongaerts, O. Beijbom, O. Pizarro, A. Friedman, A. Rodriguez-Ramirez, B. Upcroft, D. Laffoley, D. Kline, C. Bailhache, R. Vevers, O. Hoegh-Guldberg, The Catlin Seaview Survey – kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **24**, 184–198 (2014).

42. R. A. Armstrong, O. Pizarro, C. Roman, in *Mesophotic Coral Ecosystems*, Y. Loya, K. A. Puglise, T. C. L. Bridge, Eds. (Springer International Publishing, Cham, 2019; https://doi.org/10.1007/978-3-319-92735-0_51), *Coral Reefs of the World*, pp. 973–988.

43. M. Johnson-Roberson, O. Pizarro, S. B. Williams, I. Mahon, Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *J. Field Robot.* **27**, 21–51 (2010).

44. O. Beijbom, P. J. Edmunds, C. Roelfsema, J. Smith, D. I. Kline, B. P. Neal, M. J. Dunlap, V. Moriarty, T.-Y. Fan, C.-J. Tan, S. Chan, T. Treibitz, A. Gamst, B. G. Mitchell, D. Kriegman, Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation. *PLOS ONE*. **10**, e0130312 (2015).

45. I. D. Williams, C. S. Couch, O. Beijbom, T. A. Oliver, B. Vargas-Angel, B. D. Schumacher, R. E. Brainard, Leveraging Automated Image Analysis Tools to Transform Our Capacity to Assess Status and Trends of Coral Reefs. *Front. Mar. Sci.* **6**, 222 (2019).

46. I. Alonso, M. Yuval, G. Eyal, T. Treibitz, A. C. Murillo, CoralSeg: Learning coral segmentation from sparse annotations. *J. Field Robot.* **36**, 1456–1477 (2019).

47. B. M. Hopkinson, A. C. King, D. P. Owen, M. Johnson-Roberson, M. H. Long, S. M. Bhandarkar, Automated classification of three-dimensional reconstructions of coral reefs using convolutional neural networks. *PLOS ONE*. **15**, e0230671 (2020).

48. G. Pavoni, M. Corsini, N. Pedersen, V. Petrovic, P. Cignoni, Challenges in the deep learning-based semantic segmentation of benthic communities from Ortho-images. *Appl. Geomat.* **13**, 131–146 (2021).

49. A. R. Rashid, A. Chennu, A Trillion Coral Reef Colors: Deeply Annotated Underwater Hyperspectral Images for Automated Classification and Habitat Mapping. *Data*. **5**, 19 (2020).

50. M. E. Paoletti, J. M. Haut, J. Plaza, A. Plaza, Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.* **158**, 279–317 (2019).

51. "The Sixth Status of Corals of the World: 2020 Report" (Global Coral Reef Monitoring Network, International Coral Reef Initiative, 2020), (available at https://gcrmn.net/2020-report/).

52. M. Reverter, S. B. Helber, S. Rohde, J. M. de Goeij, P. J. Schupp, Coral reef benthic community changes in the Anthropocene: Biogeographic heterogeneity, overlooked configurations, and methodology. *Glob. Change Biol.* **28**, 1956–1971 (2022).

53. N. A. Kornder, J. Cappelletto, B. Mueller, M. J. L. Zalm, S. J. Martinez, M. J. A. Vermeij, J. Huisman, J. M. de Goeij, Implications of 2D versus 3D surveys to measure the abundance and composition of benthic coral reef communities. *Coral Reefs*. **40**, 1137–1153 (2021).

54. M. Brito-Millán, M. J. A. Vermeij, E. A. Alcantar, S. A. Sandin, Coral reef assessments based on cover alone mask active dynamics of coral communities. *Mar. Ecol. Prog. Ser.* **630**, 55–68 (2019).

55. V. Lecours, R. Devillers, D. C. Schneider, V. L. Lucieer, C. J. Brown, E. N. Edinger, Spatial scale and geographic context in benthic habitat mapping: review and future directions. *Mar. Ecol. Prog. Ser.* **535**, 259–284 (2015).

56. M. Yuval, I. Alonso, G. Eyal, D. Tchernov, Y. Loya, A. C. Murillo, T. Treibitz, Repeatable Semantic Reef-Mapping through Photogrammetry and Label-Augmentation. *Remote Sens.* **13**, 659 (2021).

57. Y. Gal, Z. Ghahramani, in *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48* (JMLR.org, New York, NY, USA, 2016), *ICML'16*, pp. 1050–1059.

58. Y. Cao, T. A. Geddes, J. Y. H. Yang, P. Yang, Ensemble deep learning in bioinformatics. *Nat. Mach. Intell.* **2**, 500–508 (2020).

59. M. A. Ganaie, M. Hu, M. Tanveer*, P. N. Suganthan*, Ensemble deep learning: A review. *ArXiv210402395 Cs* (2021) (available at http://arxiv.org/abs/2104.02395).

60. P. Krähenbühl, V. Koltun, in *Proceedings of the 24th International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2011), *NIPS'11*, pp. 109–117.

61. Z. Tian, C. Shen, H. Chen, in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm, Eds. (Springer International Publishing, Cham, 2020), *Lecture Notes in Computer Science*, pp. 282–298.

62. J. Nowosad, T. F. Stepinski, Information theory as a consistent framework for quantification and classification of landscape patterns. *Landsc. Ecol.* **34**, 2091–2101 (2019).

63. K. Riitters, Pattern metrics for a transdisciplinary landscape ecology. *Landsc. Ecol.* **34**, 2057–2063 (2019).

64. J. A. Strong, An error analysis of marine habitat mapping methods and prioritised work packages required to reduce errors and improve consistency. *Estuar. Coast. Shelf Sci.* **240**, 106684 (2020).

65. P. L. Jokiel, K. S. Rodgers, E. K. Brown, J. C. Kenyon, G. Aeby, W. R. Smith, F. Farrell, Comparison of methods used to estimate coral cover in the Hawaiian Islands. *PeerJ.* **3**, e954 (2015).

66. B. Taormina, A. Percheron, M. P. Marzloff, X. Caisey, N. Quillien, M. Lejart, N. Desroy, O. Dugornay, A. Tancray, A. Carlier, Succession in epibenthic communities on artificial reefs associated with marine renewable energy facilities within a tide-swept environment. *ICES J. Mar. Sci.* **77**, 2656–2668 (2020).

67. P. Dumas, A. Bertaud, C. Peignon, M. Leopold, P. Dominique, A "quick and clean"photographic method for the description of coral reef habitats. *J. Exp. Biol. Ecol.* **368**, 161–168 (2009).

68. N. R. Perkins, Image subsampling and point scoring approaches for large-scale marine benthic monitoring programs, 11 (2016).

69. P. Houk, R. Van Woesik, Coral Reef Benthic Video Surveys Facilitate Long-Term Monitoring in the Commonwealth of the Northern Mariana Islands: Toward an Optimal Sampling Strategy. *Pac. Sci.* **60**, 177–189 (2006).

70. B. Taormina, M. P. Marzloff, N. Desroy, X. Caisey, O. Dugornay, E. Metral Thiesse, A. Tancray, A. Carlier, Optimizing image-based protocol to monitor macroepibenthic communities colonizing artificial structures. *ICES J. Mar. Sci.* **77**, 835–845 (2020).

71. J. M. Jurado, L. Pádua, J. Hruška, F. R. Feito, J. J. Sousa, An Efficient Method for Generating UAV-Based Hyperspectral Mosaics Using Push-Broom Sensors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 6515–6531 (2021).

72. M. Moroni, C. Dacquino, A. Cenedese, Mosaicing of Hyperspectral Images: The Application of a Spectrograph Imaging Device. *Sensors*. **12**, 10228–10247 (2012).

73. D. Akkaynak, T. Treibitz, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 1682–1691.

74. Z. Zhong, J. Li, Z. Luo, M. Chapman, Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* **56**, 847–858 (2018).

75. L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, J. Han, On the Variance of the Adaptive Learning Rate and Beyond. *ArXiv190803265 Cs Stat* (2021) (available at http://arxiv.org/abs/1908.03265).

76. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc., 2019; http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf), pp. 8024–8035.

77. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

78. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, the scikit-image contributors, scikit-image: image processing in Python. *PeerJ.* **2**, e453 (2014).

79. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods*. **17**, 261–272 (2020).

Acknowledgments

We would like to thank Carsten John and Oliver Artmann at the Max Planck Institute for Marine Microbiology for their IT support. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 813360 "4D-REEF".

Data availability statement

All data needed to evaluate the conclusions in the paper are present in the paper, the Supplementary Materials and in the data repository <url-pangaea>.

Author contributions

Daniel Schürholz – Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project administration.

Arjun Chennu – Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Tables

Method	Labelspace	Signal type	Overall accuracy	Precision	Recall	Fbeta	Cohen's Kappa
patched	*detailed*	*radiance*	0.767	0.779	0.769	0.769	0.760
		reflectance	0.715	0.723	0.720	0.720	0.708
	reefgroups	*radiance*	0.869	0.871	0.869	0.869	0.829
		reflectance	0.841	0.841	0.841	0.841	0.793
segmented	*detailed*	*radiance*	0.772	0.779	0.771	0.771	0.755
		reflectance	**0.846**	**0.848**	**0.845**	**0.845**	**0.835**
	reefgroups	*radiance*	0.792	0.796	0.794	0.794	0.737
		reflectance	**0.875**	**0.876**	**0.875**	**0.875**	**0.841**

Table 1. Classifiers performance. Comparison of the performance of each ML method in combination with each labelspace and each signal type. The classifiers were tested on disjoint datasets of 50000 patches for the patched method and 15946 segments for the segmented method. The best performing classifiers are highlighted in boldface for the detailed and reefgroups labelspace. In both cases the segmented method with reflectance data provided the best performance. The patched method performed better with radiance data in both labelspaces.

Figures

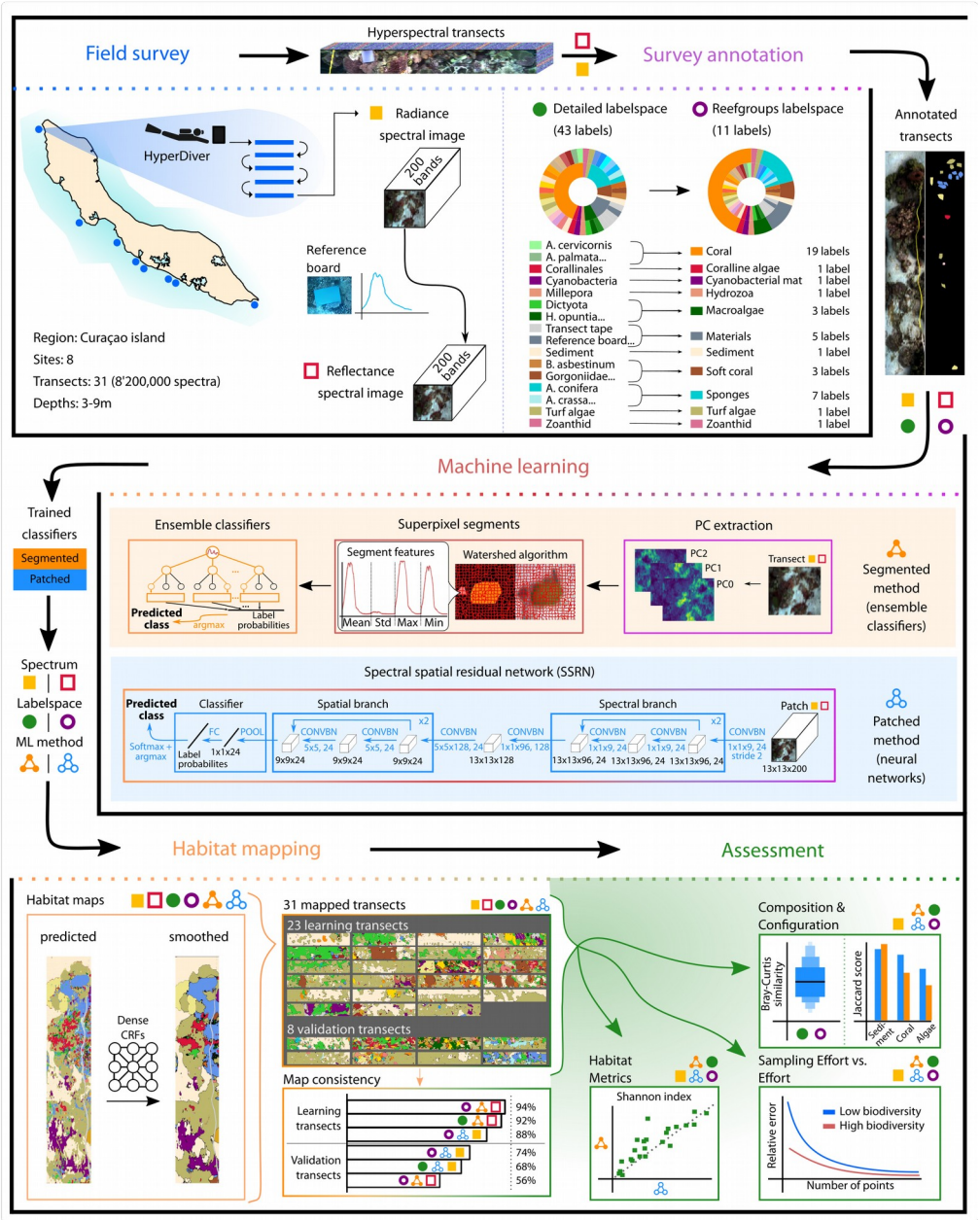

Fig. 1. Schematic of scalable survey-mapping-assessment workflow for digitizing reef habitats. The underwater survey with the HyperDiver at 8 study sites over 147 transects (50 m each) in Curacao produced radiance and reflectance spectral images. A subset of 31 transect images were annotated into the detailed (43 classes) labelspace and aggregated into the reefgroups (11 classes) labelspace. Annotated regions from 23 transects were used in two separate machine learning methods to classify each region of the spectral images into each labelspace. The segmented method used ensemble learning on image superpixel features, while the patched method used spatial-spectral neural network learning of image patches. The classifiers were used to predict the label probabilities at each of 500+ million pixels in all 31 transects. The classifier-predicted probability maps were smoothed with conditional random fields, and converted into densely-labelled categorical maps which capture the habitat structure in the transect. Consistency of the habitat maps (see percentages) were assessed in each expert-annotated region in the the learning transects and in each colony-or-organism annotation in the validation transects, which were not used for machine learning. The maps produced by both ML methods were assessed by comparing habitat metrics such as Shannon index and areal coverage of main functional groups. The ability to recover the compositional and configurational structure of the transects by each ML method was assessed/compared. The effort-vs-error relationship for point-count sampling of the reef habitats was assessed using virtual sampling of the dense habitat maps.

Fig. 2. Detailed and reefgroups labelspaces. (A) Reefgroups labelspace consisting of 11 labels for reef functional groups and abiotic elements. Some classes are divided into two or more classes in the detailed labelspace. Shapes within the color boxes are provided for easier identification between labelspaces. **(B)** Detailed labelspace consisting of 43 classes, representing the deepest possible taxonomic definition of targets in visual annotation of underwater transects. The labelspace consists of 19 genera and species of corals, 1 family of hydrozoa, 3 genera and species of soft coral, 1 family of zoanthid, 9 species of sponges, 1 order of coralline algae, 3 genera and species of macroalgae, 3 substrate classes of turf algae, cyanobacterial mat and sediment, and 5 classes of survey material.

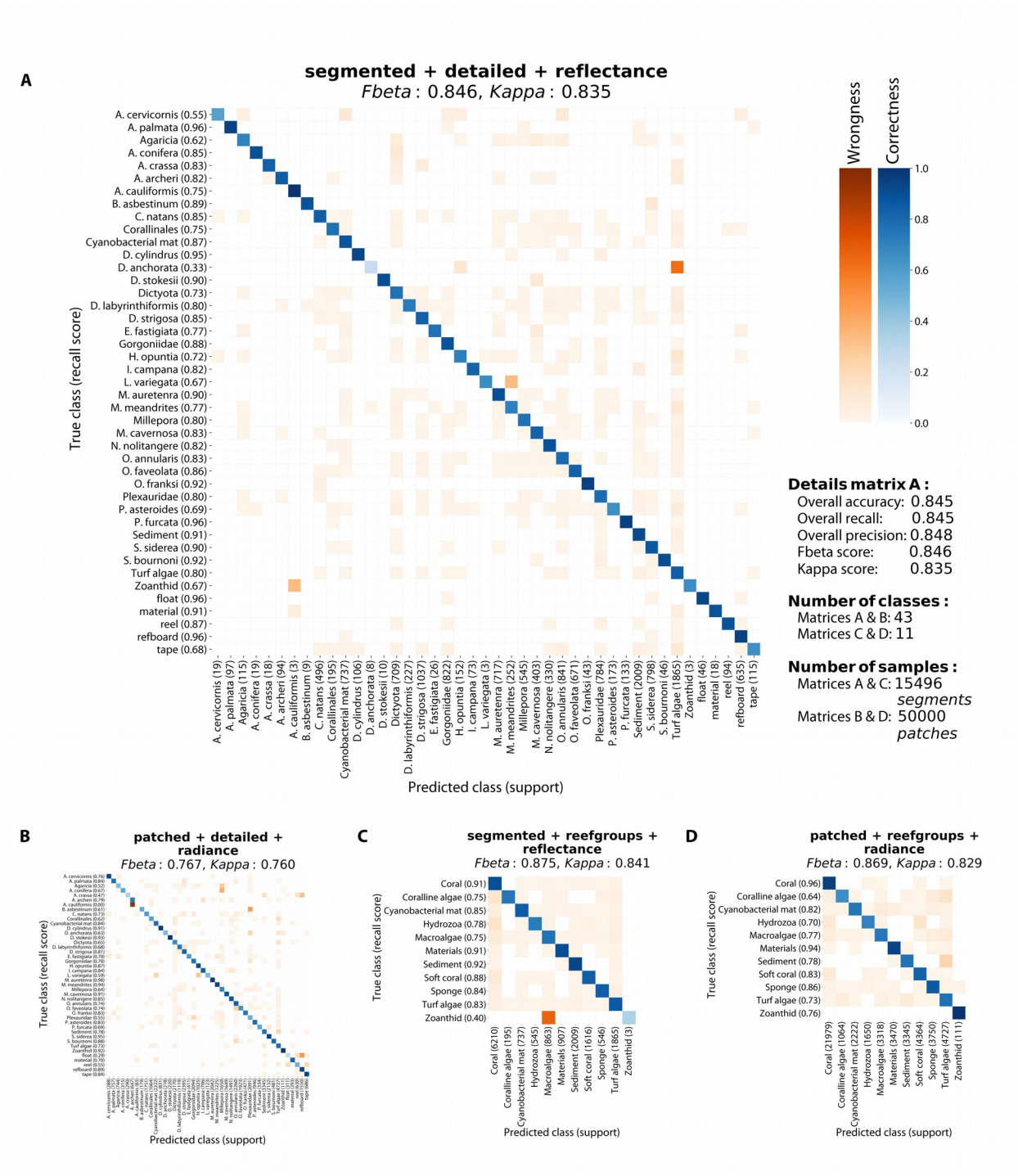

Fig. 3. Performance evaluation of classifiers in detailed and reefgroup labelspaces. Confusion matrices of classifiers from a subset of ML method+labelspace+signal type combinations were used to asses performance on a held-out testing set. The Fbeta score – the harmonic mean of recall and precision – is a better metric for multi-class unbalanced classification than accuracy. Cohen's Kappa score measures classification improvement beyond the baseline of unbalanced data. (**A**) Recall confusion matrix for the segmented classifier predicting into the detailed labelspace from reflectance images. The Fbeta score and other metrics (see side notes) showed excellent performance at 84.6%. There was generally little (<3%) to minor (<20%) confusions for the rare classes such as *Zoanthid* and *D. anchorata*. (**B**) Recall confusion matrix for the patched classifier predicting into the detailed labelspace from radiance images. With an Fbeta score was 76.7%, there was little (<3%) to minor confusions (<20%) with rare classes such as *A. cauliformis* and *refboard*. (**C**) Recall confusion matrix for the segmented classifier predicting into the reefgroups labelspace from reflectance images. The Fbeta score was 87.5% with some relevant confusions shown between the *Zoanthid* and *Macroalgae* classes. (**D**) Recall confusion matrix for a patched classifier predicting into the reefgroups labelspace from radiance images. Corals were classified with a recall of 96%.

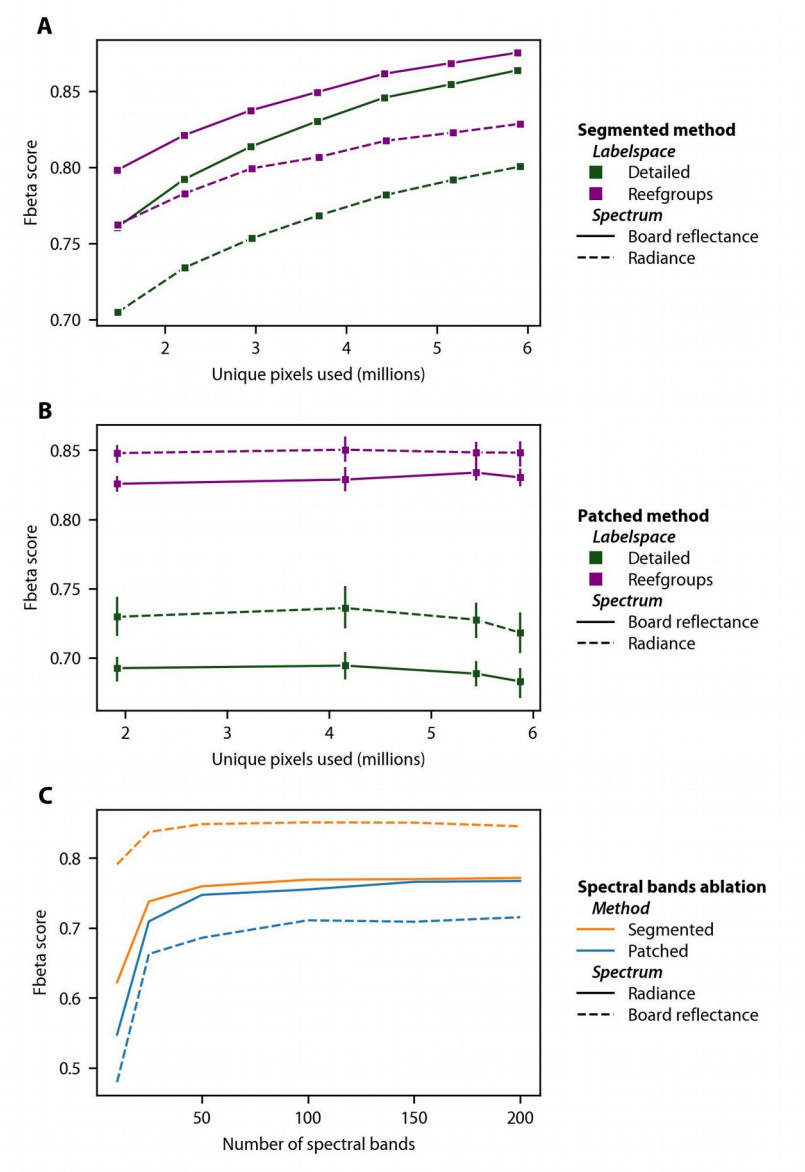

Fig. 4. Effect of quantity and quality of spectral data on classifier performance. The effects of increasing the unique number of pixels in the data, but training with same computational effort, was measured (Fbeta score) for both methods and labelspaces. The segmented method (**A**) showed improved performance whereas the patched method (**B**) showed little change in performance. Both methods performed better at predicting into the reefgroups labelspace (with 11 classes) than the detailed labelspace (with 43 classes) irrespective of the spectral signal type. (**C**) Both methods performed better when using increasing number of (uniformly sampled) spectral bands for training, with limited improvement beyond 50 spectral bands.

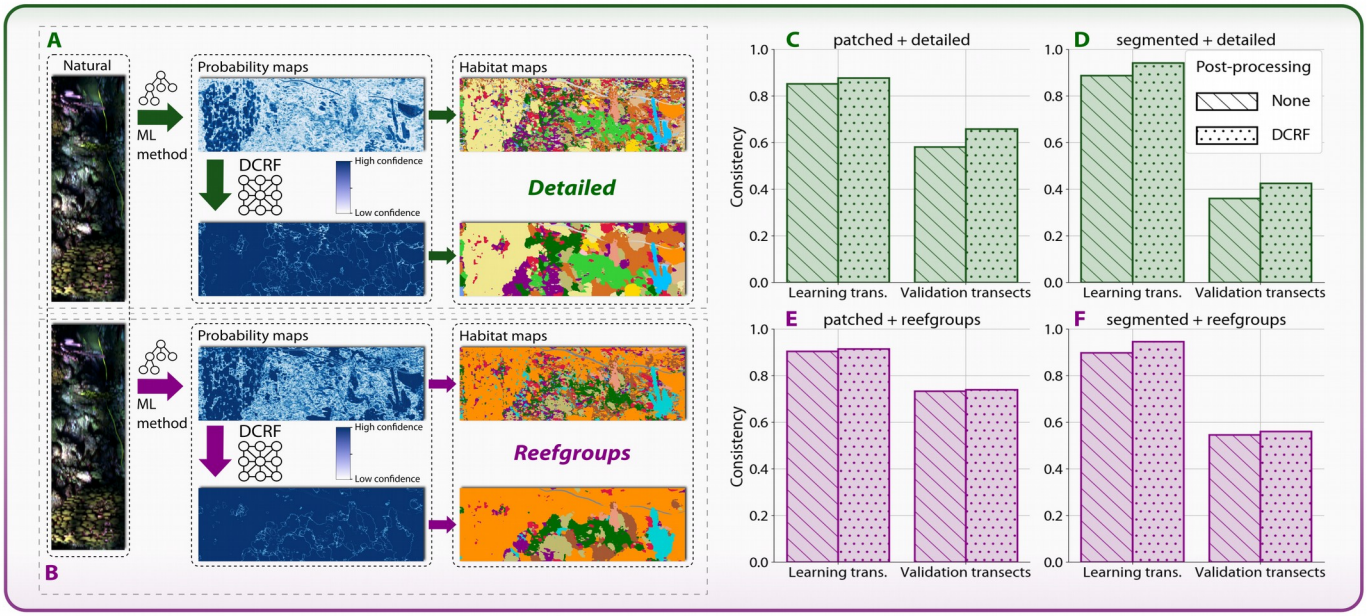

Fig. 5. Smooth and consistent habitat mapping. The process of producing smoothed habitat maps in the detailed (**A**) and reefgroups (**B**) labelspaces. Classifiers from both ML methods produced label probabilities for each pixel. Habitat maps produced from the classifiers' probability maps exhibit small-scale spatial noise and regions of low confidence. Smoothing the label probabilities with DCRFs renders spatially cohesive maps. Note the generally higher confidence in the smoothed probability images, which retain low confidence at the perimeters of adjacent targets. (**C-F**) For each ML method and labelspace, the consistency of the raw and smoothed habitat maps was measured on the learning and validation transects separately. The use of DCRFs slightly improved the map consistency in all cases. Both ML methods presented a drop in consistency when mapping the validation transects in comparison to the learning transects. Nonetheless, the patched method showed a better generalization ability and was more accurate at mapping the validation transects, which were completely disjoint from the learning transects used for training the ML classifiers.

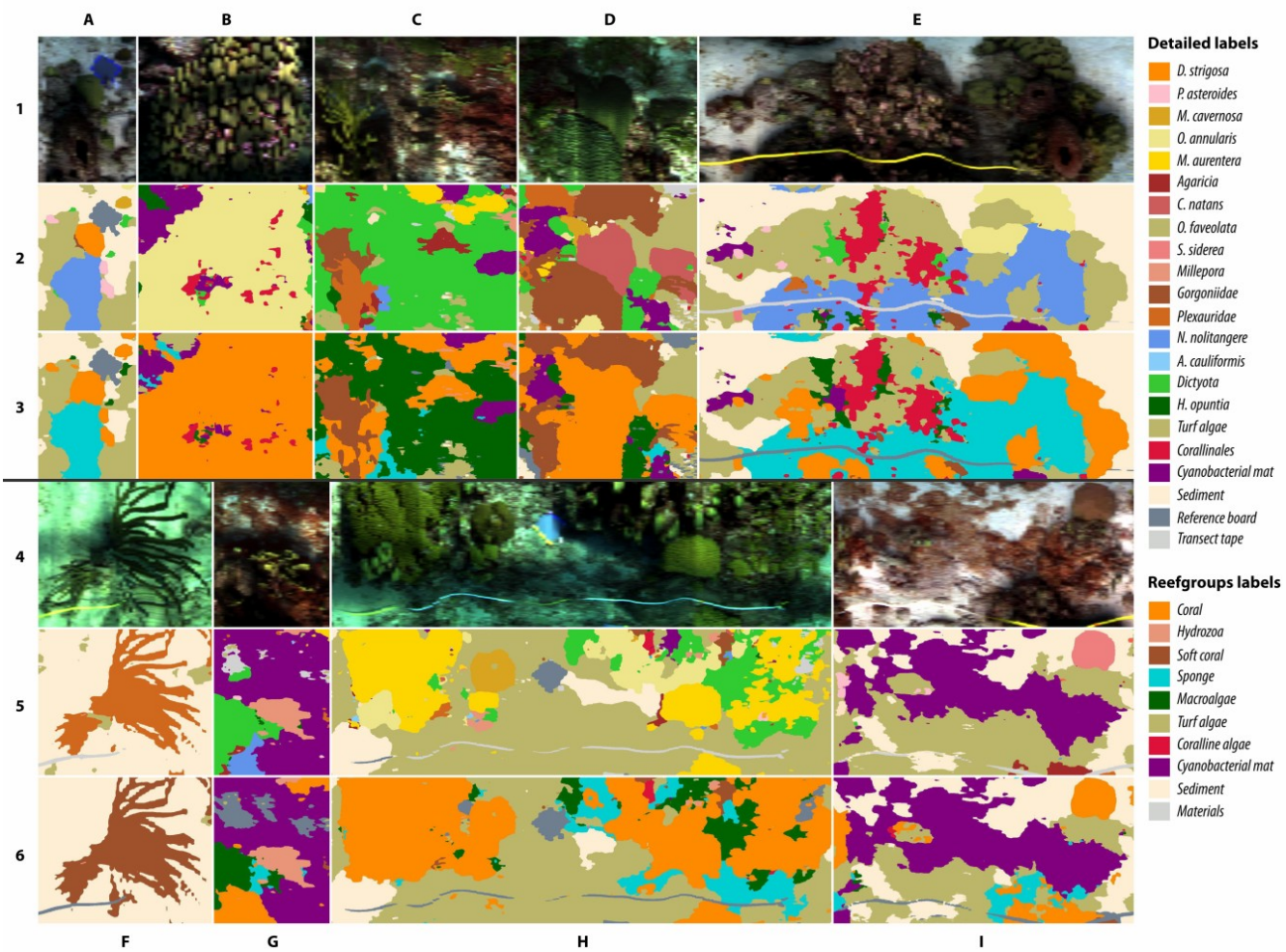

Fig. 6. The rich structure of a digitized reefscape. Sections of habitat maps produced with the patched ML method, with colours corresponding to classes from each labelspace (see legend). Rows 1 and 4 are the natural view, as would be seen by a human observer. Rows 2 and 5 show the sections in the detailed labelspace and rows 3 and 6 in the reefgroups labelspace. Our proposed workflow accurately discerns amongst a large labelspace and delineates complicated shapes of reef biota. Detailed descriptions are in the *Results*.

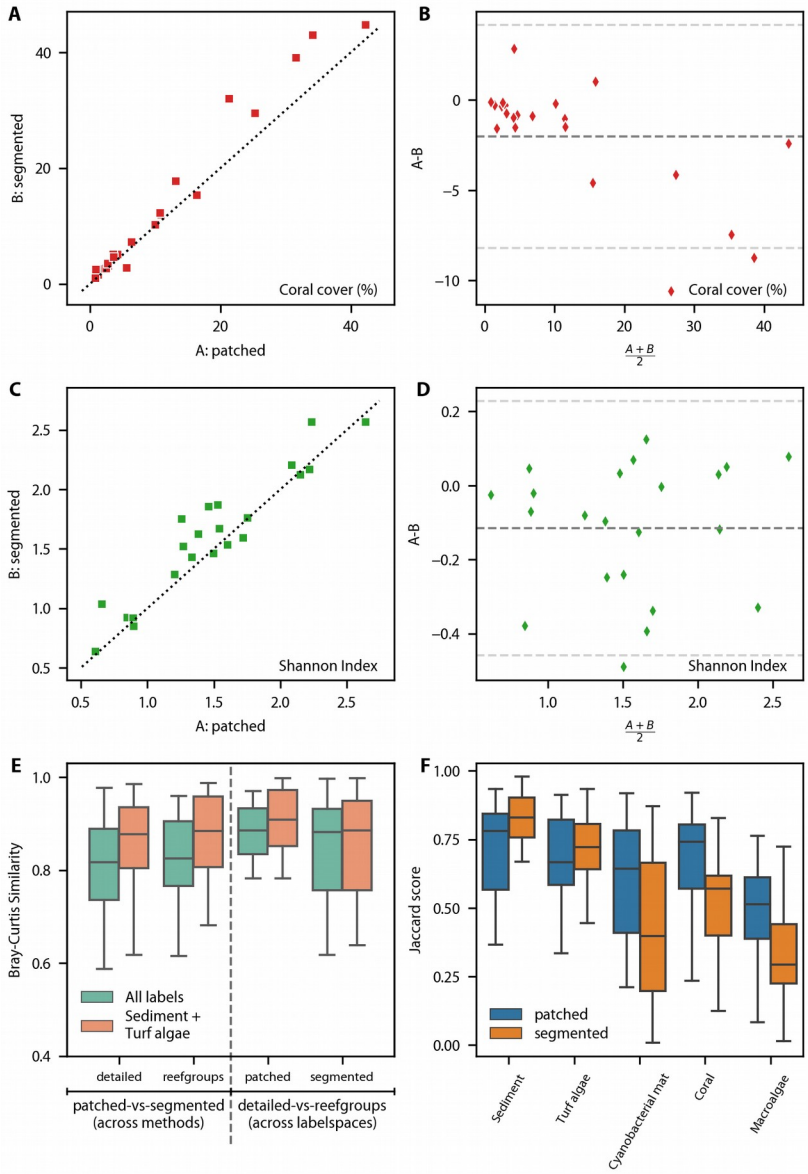

Fig. 7. Assessing the benthic habitat structure from dense maps. (**A**) Coral cover of 22 learning transects was compared from the habitat maps for each ML method. The segmented method predicted slightly more coral cover than the patched method in the transects. (**B**) No clear bias was noted for either method as shown in the Bland-Altman plot. (**C**) The Shannon index values for 22 compared transects were highly correlated between the maps from the segmented and patched methods. The segmented method produced habitat maps with slightly higher Shannon diversity. (**D**) Bland-Altman plot showed no significant bias in either ML method for the Shannon index comparison. (**E**) The median Bray-Curtis similarity between the mapped communities was ~80% across ML methods was and ~88% across labelspaces. Note that the orange bars refer to a labelspace where *Sediment* and *Turf algae* were combined into a single class, resulting in a higher compositional similarity. (**F**) Configurational similarity was assessed using the Jaccard index between the reefgroups and detailed-to-reefgroups habitat maps, for both ML methods. The top-five dominant labels are shown. Given the hundreds of millions of pixels in this assessment, the maps showed very good consistency in configuration.

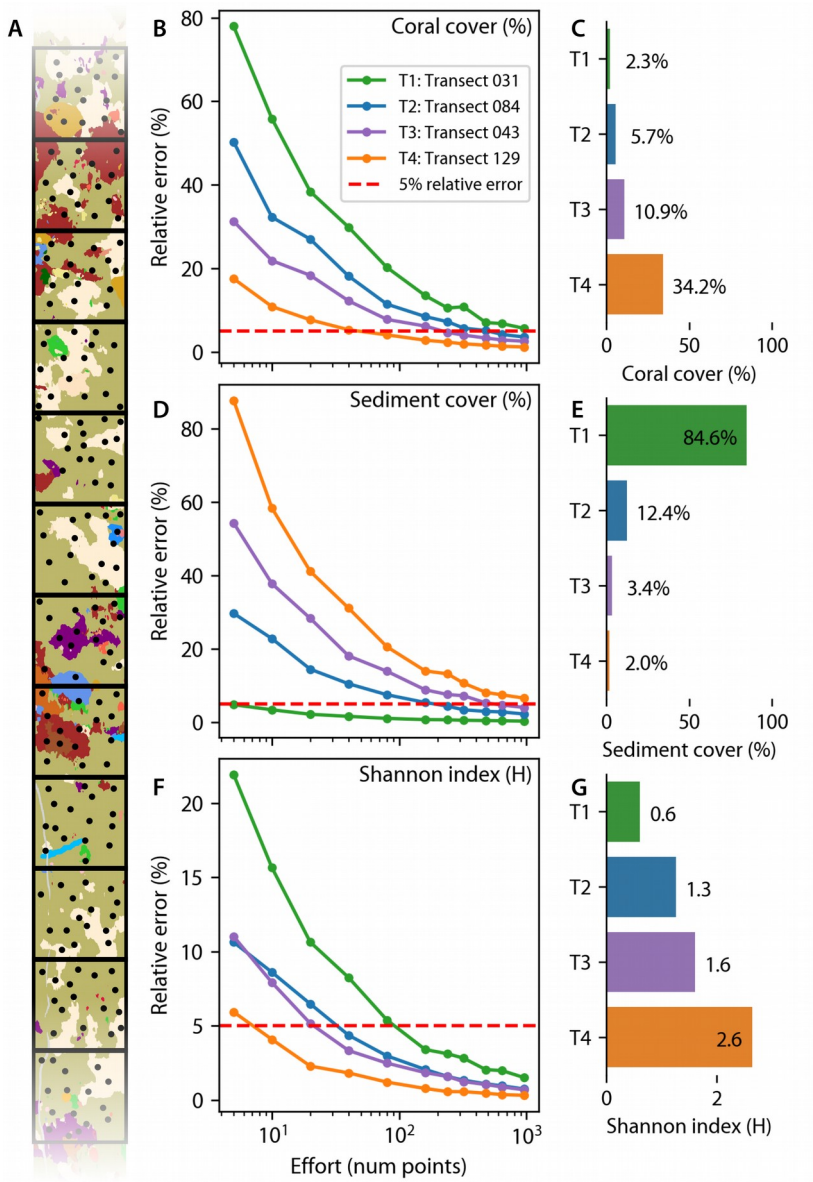

Fig. 8. Effort-vs-error analysis for point-count sampling of reef habitat structure. (**A**) The dense habitat maps were used to conduct simulations of quadrat-wise random sparse sampling to estimate habitat metrics (coverage or Shannon index) through point-count estimates. The sampling effort in taking N = {5, 10, 20, 40, 80, 160, 240, 320, 480, 640, 960} random points per quadrat was compared against the error in the habitat metric relative to the metric of the full map. The number of point samples required to achieve a relative error lesser than 5% (dashed line) was assessed. Four transects were selected with a range of diversity H = {0.61, 1.26, 1.6, 2.64} values based on the full map. (**B**, **C**) The transect with the least coral coverage (T1) required more than 960 points per quadrat to estimate coral coverage within the 5% error limit, whereas the transect with the highest coral coverage (T4) required less than 80 points per quadrat. (**D**, **E**) The transect with the least sediment coverage (T4) required more than 960 points per quadrat to estimate sediment coverage within the 5% error limit, whereas the transect with the highest sediment coverage (T1) required only 5 points per quadrat. (**F**, **G**) The least diverse transect (T1) required 160 points compared to 40 points per quadrat for the more diverse transect (T4) to estimate the Shannon diversity index within the 5% error limit. These results suggest that rarer species require more sampling effort and that at least 160 points per quadrat should be used to estimate habitat metrics in reef transects where the expected diversity or coverage is not previously known.