

CoralscapesV2: Panoptic and Fine-Grained Visual Scene Understanding in Coral Reefs

Jonathan Sauder¹, Thomas Ruckli², Gabrielė Strodmskytė²,
Ibrahim Souleiman Abdallah³, Rahma Hassan Abdi³, Djama Goumaneh Awaleh⁴,
Mohamed Houssein Farah⁴, Moustapha Nour⁵, Osama Sharhubil Saad⁶, Mustafa
Mohammed Khalafallah Altaib⁶, Maysoon Kteifan⁷, Farah Alsoqi⁷, Eyad Zgool⁷,
Jafar Al-Omari⁸, Temesgen Gebremeskel Gebreluel⁹, Zekaria Zekeria Abdulkirim⁹,
Meron Ghirmay⁹, Teklehaimanot Beraki¹⁰, Devis Tuia², and Guilhem Banc-Prandi²

¹ MIT

² EPFL

³ University of Djibouti

⁴ Ministry of Environment and Sustainable Development of Djibouti

⁵ Centre d'Études et de Recherche de Djibouti

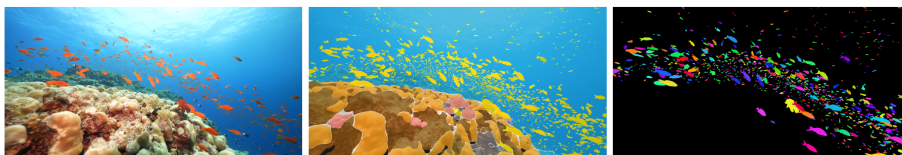
⁶ Red Sea University of Port Sudan

⁷ Aqaba Marine Reserve, Jordan

⁸ United Nations Development Programme Jordan

⁹ Department of Biology, Mai Nefhi College of Science, Eritrea

¹⁰ Ministry of Marine Resources, Massawa, Eritrea



More Scenes

Improved Fine-Grained Segmentation

Exhaustive Fish Instances

Abstract. In order to design conservation and restoration strategies to counter the global decline of coral reefs, ecological monitoring of reefs needs to be scaled up dramatically. Computer vision methods are increasingly used to tackle the vast amount of data: as the paradigm of data collection in reefs shifts from highly standardized and constrained survey images to unconstrained imagery on scalable platforms, it is necessary to design machine learning methods that help to get a fine-grained understanding of reefs from general-purpose reef imagery. This paper provides CoralscapesV2, an extension of the Coralscapes dataset for general-purpose visual scene understanding in reefs. CoralscapesV2 increases the dataset size, scope, label completeness and quality for semantic segmentation, and extends the number of classes from 39 to 95 fine-grained visual categories. Furthermore, CoralscapesV2 provides 65k **exhaustive** fish instance mask annotations, meticulously annotated to completeness by using the video, revealing that annotation of fish based on only static images is insufficient. CoralscapesV2 is the first dataset for panoptic segmentation in coral reefs, capturing a wide range of scenarios in the wild, posing a challenging benchmark for contemporary semantic segmentation and instance segmentation models. CoralscapesV2 is an important step towards general-purpose panoptic segmentation in coral reefs, which has substantial implications for scaling up coral reef monitoring, as it can be employed in a wide range of applications from benthic cover mapping from robot or handheld videos to designing methods for automated quantification and understanding of fish behavior and fish-reef interactions.

1 Introduction

Coral reefs are under unprecedented threat from anthropogenic climate change and local stressors such as overfishing and pollution. Recurrent mass-bleaching events driven by marine heatwaves have caused widespread coral mortality over the past decades [28], restructuring reef assemblages [29] and eroding the capacity of reefs to provide the ecosystem services on which hundreds of millions of people depend [16]. Yet the dynamics of reef decline and recovery remain poorly understood and vary substantially across regions [29], while only a small fraction of the world’s reefs are actively and regularly monitored [63]. Designing effective conservation and restoration strategies therefore requires ecological monitoring that is simultaneously scaled up in spatial and temporal coverage and deepened in the fine-grained ecological detail it captures.

Computer vision is increasingly used in reef monitoring to process the vast amounts of imagery collected. As AI-based vision tools mature, the field is shifting away from highly standardized acquisition protocols (*e.g.* photo quadrats and orthophotos) towards **general-purpose** analysis of reef imagery that can be applied flexibly across survey types and data-collection platforms: diver-held cameras [55,56], ROV/AUV platforms [21,51], and static underwater cameras [36,42].

Most existing datasets target isolated use cases, leaving a real need for high-quality, general-purpose datasets. Coralscapes [57] was the first dataset explicitly designed to benchmark general-purpose segmentation in reefs, providing densely annotated imagery from 35 dive sites across five countries in the Red Sea region. However, Coralscapes emphasizes the benthos, treating fish merely as a single semantic class to be masked out of benthic mapping.

A key component of fine-grained reef understanding, however, is the study of the fish populations reefs host and their entangled dynamics. A variety of datasets and methods for detecting, tracking, and taxonomically identifying fish in images and video have been proposed [8,9,32,34]. However, these are either limited to static images, incompletely annotated, from outside the coral-reef domain, or lacking instance masks, limiting their usefulness for real-world ecological applications such as fish counting or biomass estimation on reefs.

This paper proposes CoralscapesV2, an extension of Coralscapes that bridges this gap and improves the original dataset along every measurable axis:

1. **Increased Size, Completeness, and Quality** We increase the number of annotated images (2075 to 2433), the number of annotated polygons (174k to 270k), and the annotation density (82.9% to 86.8%), while broadening the diversity of scenes across more reef sites in the Red Sea region (35 to 45). Annotation masks of existing images are further refined in terms of density and quality.
2. **Fine-Grained Visual Categories** A refined label set of 95 fine-grained classes captures more taxa, coral morphologies along with their health status, as well as new auxiliary classes.
3. **Exhaustive Fish Instance Annotations** 65k fish are exhaustively annotated on each frame using the video.

Together, these improvements make CoralscapesV2 the first panoptic segmentation dataset for coral reefs. This unlocks applications that were previously out of reach for reef imagery, from joint benthic-cover mapping and fish quantification on unconstrained diver or robot video to the study of fish behavior and fish-reef interactions. The remainder of this paper reviews related datasets (Sec. 2), details CoralscapesV2 and its annotation effort (Sec. 3), and benchmarks contemporary semantic- and instance-segmentation models (Sec. 4).

2 Related Work

2.1 Coral Reef Image Datasets

Annotation of coral reefs at scale is challenging: many corals show high morphological plasticity and taxa can not be identified to species or genus level from imagery. Some genera or species can only be identified with high-resolution close-up images of the corallites, and some simply can not be visually identified. Annotating corals across image quality and water conditions, falling back to the growth form when taxonomy is not clear, and disambiguating other benthic classes (e.g. sponges) requires substantial domain expertise. As such, large-scale annotation of reefs is limitingly expensive, and the landscape of datasets is scattered around conventions, tools, and applications.

Most coral reef scientists and conservation practitioners interact with imagery through platforms such as CoralNet [6], MERMAID [45], or ReefCloud [1], uploading images and annotating them with sparse point labels that are used to train classification models. Consequently, most annotated reef data exists as sparse point labels, as in the Moorea Labeled Corals [6, 7] and Catlin Seaview [22] datasets. A smaller set of datasets provides dense masks rather than sparse points, such as UCSD mosaics [17] or Benthos [66], but these are confined to the orthomosaic domain and limited in scale; CoralSCOP [68] reaches large scale but only distinguishes coral from non-coral, on noisy imagery and labels.

Coralscapes [57] proposed the first general-purpose dense segmentation dataset, with 2075 densely annotated video frames across 39 classes from 35 reef sites in the Red Sea region. It follows the structure of the Cityscapes dataset [12]: each 1024×2048 px image is a frame from a video, with the preceding 20 and following 10 frames available. However, Coralscapes frames were annotated statically, with no temporal context and fish captured only as a single semantic class, instead of as individual instances. CoralscapesV2 directly extends this line of work, refining the label set to 95 classes and, crucially, adding exhaustive fish *instance* masks annotated with video context.

2.2 Fish Datasets

Beyond the benthos, a large and growing body of datasets targets fish. We organize this landscape (Table 1) along the three central axes: whether annotations are per-fish *instance masks*, whether every fish in a frame is *exhaustively* labeled, and whether the data covers *coral-reef* scenes with *video* context.

Table 1: Fish datasets, grouped by annotation type and sorted by size within each group. *Camera*: image-only vs. video (static/moving/mixed). *Exhaustive*: whether all fish in a labeled frame are annotated. † = total over all categories (not fish-only).

Dataset	Ecosystem	Camera	Annotation Type	Images/frames	Annotations	Exhaustive
<i>Bounding boxes</i>						
FishTrack23 [14]	Mixed	Video (mixed)	Box + track	–	~850,000	partial
Orange Chromide [64]	Freshwater	Images	Box	586	10,607	N
Roboflow Fish [62]	Mixed	Images	Box	680	3,142	N
Labeled Fishes [13]	Deep sea	Video (moving)	Box	929	1,005	N
TORSI [59]	Deep sea	Images	Box	1,051	14,414	N
Project Natick [46]	Marine	Video (static)	Box	1,118	998	N
OzFish [44]	Mixed	Video (static)	Box	~1,800	~45,000	Y
Deep Vision [3]	Pelagic	Video (moving)	Box	1,875	4,834	N
AAU Zebrafish [23]	Aquaculture	Video (static)	Box + track	2,224	~6,672	Y
J-EDI JODD [48]	Deep sea	Video (moving)	Box	8,151	15,621†	N
Brackish [50]	Brackish	Video (static)	Box	14,518	12,797	Y
CoReFish [9]	Coral reef	Video (moving)	Box + track	14,580	98,267	partial
FishCLEF-2015 [32]	Coral reef	Video (static)	Box	>20,000	>14,000	N
Marine Detect [49]	Mixed	Images	Box	~21,125	–	N
N-MARINE [5]	Deep sea	Video (static)	Box	23,936	–	Y
OBSEA [19]	Coastal	Images	Box	33,805	69,917	N
Kakadu [31]	Freshwater	Video (static)	Box	44,112	82,904	N
MFT25 [40]	Aquaculture	Video (static)	Box + track	48,066	408,578	Y
River Herring [47]	Freshwater	Video (static)	Box	~60,000	91,482	Y
Puget Sound [18]	Estuary	Video (static)	Box	77,739	67,990†	Y
FathomNet [33]	Deep sea	Images	Box	84,454	175,873†	N
FishNet [35]	Mixed	Images	Box	94,532	114,375	N
Salmon CV [4]	Freshwater	Video (static)	Box + track	532,000	~503,000	Y
<i>Semantic masks (no instance)</i>						
DeepFish [54]	Mixed	Video (static)	Semantic	620	–	N
SUIM [30]	Mixed	Images	Semantic	1,635	–	N
Coralscapes [57]	Coral reef	Video (moving)	Semantic	2,075	~22,000 fish	N
<i>Instance masks</i>						
Fish4Knowledge [34]	Coral reef	Video (static)	Instance + point track	17 videos	–	N
PomerFish [60]	Freshwater	Video (mixed)	Instance	1,115	1,038	N
DeepFish-Inst [20]	Aquaculture	Images	Instance	1,291	7,339	N
WaterMask/UIIS [41]	Mixed	Images	Instance	4,628	–	N
TrashCan [25]	Deep sea	Video (moving)	Instance	7,212	–	N
UIIS10K [39]	Mixed	Images	Instance	10,048	41,862†	N
Fish-Occlusion [67]	Aquaculture	Images	Instance	14,376	144,894	N
USIS16K [26]	Mixed	Images	Instance	16,151	–	N
CoralscapesV2	Coral reef	Video (moving)	Instance	2433	65,738 fish	Y

Annotating fish exhaustively is challenging at scale. Labeling only the obvious, salient fish is comparatively easy, but many important applications – such as fish tracking for abundance estimation or behavior analysis – require that *every* visible fish in an image be recognized. Exhaustive annotation is costly because it must be performed with video rather than single frames, which is slow because many fish appear blurred, occluded, or as tiny objects, and can be hard to distinguish from other suspended matter such as marine snow or under poor visibility. As a result, most fish datasets provide only image-level labels or bounding boxes [35, 69]; several of the largest are exhaustively annotated but remain box-only and outside reefs. Per-instance masks are far rarer and, where they exist, come from freshwater, aquaculture, or domain-generic underwater settings and are non-exhaustive or restricted to salient objects [20, 26, 39, 41, 60, 67]. Within coral reefs, existing fish datasets stop short of exhaustive per-fish instance masks: FishCLEF-2015 [32] and Fish4Knowledge [8, 34] provide only bounding boxes or instance masks from static-camera video, and the recent CoReFish [9] – closest to ours in capturing diver ego-motion reef video – provides bounding boxes and tracks rather than instance masks. As the table shows, no existing dataset combines coral-reef scenes, exhaustive per-fish instance masks, and video context; CoralscapesV2 is, to our knowledge, the first to provide all three.

3 Dataset

CoralscapesV2 improves upon the original Coralscapes along every measurable axis. In principle, the main improvements are **Increased Size, Completeness, and Quality** of the dataset, **Fine-Grained Visual Categories** and **Exhaustive Fish Instance Annotations**. The remainder of this Section details these improvements and the involved annotation effort.

3.1 Increased Size, Completeness, and Quality

CoralscapesV2 increases the size of the Coralscapes dataset along every dimension. The total number of annotated video frames is increased from 2075 to 2433 images (Fig. 1). All existing and new imagery was taken during scientific expeditions and monitoring workshops in reefs of the Red Sea region (Jordan, Djibouti, Eritrea, Sudan, and Israel) between August 2021 and October 2025. Of the 358 new frames, 246 are from reef sites that were present in Coralscapes V1, and the remaining frames are from 10 new reef sites, which are split into train/validation/test splits of 1790/170/473 frames along geographically disjoint

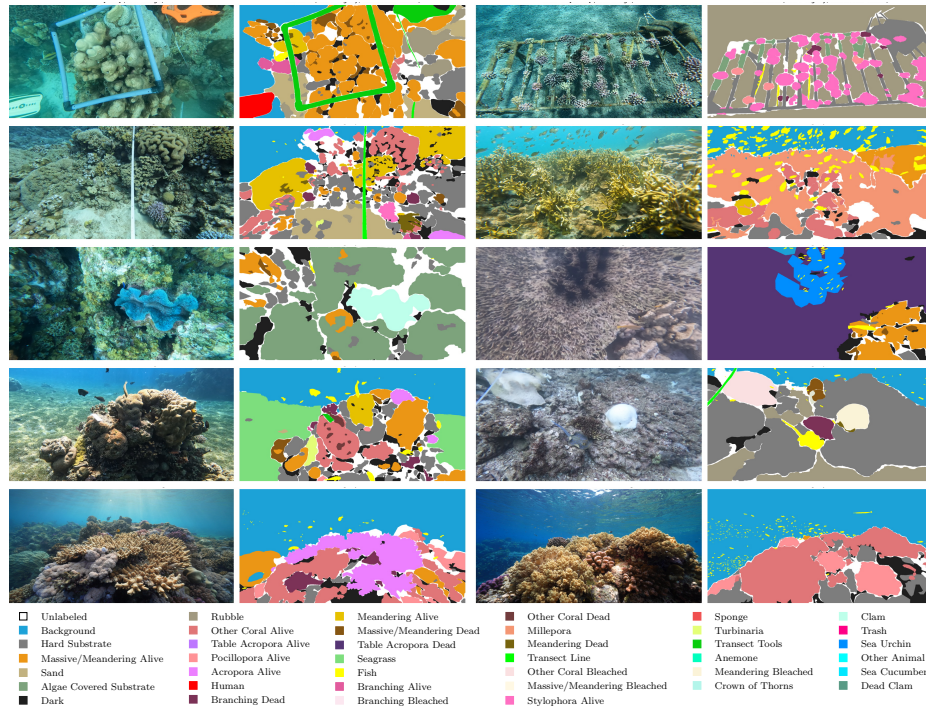


Fig. 1: CoralscapesV2 adds 358 new annotated video frames, with an emphasis on increasing Coralscapes’ diversity of scenes, for example by including photo quadrats and coral restoration infrastructure (top row), or images from a high-quality camera (bottom row).

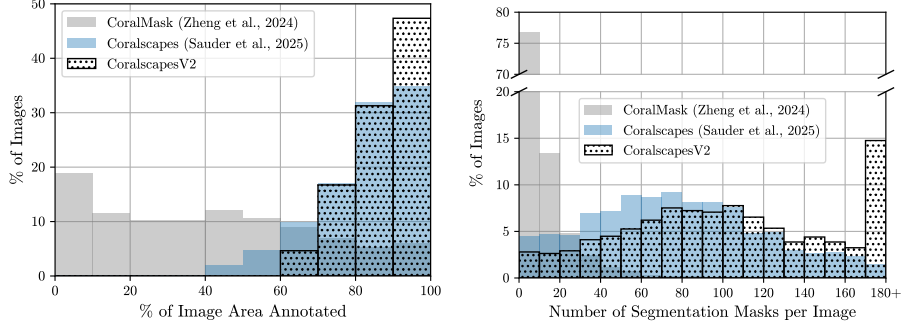


Fig. 2: CoralscapesV2 increases the annotation density and complexity.

reef sites¹. The splits of the reef sites match those of the original Coralscapes dataset (i.e. any image that existed in V1 will be in the same split in V2). Of the new annotated frames, 303 are taken with GoPro Hero 10 cameras (as all of Coralscapes V1), and the remainder with a Sony 7r4 camera (12-24mm f4 G lens), which produces higher-quality footage.

Furthermore, CoralscapesV2 increases the annotation **density**, as shown by Figure 2, across both existing and newly added images. The average area of images that is annotated is increased from 82.9% to 86.8%, and the minimum annotated area for each image from 40% to 60%. In total, CoralscapesV2 increases the total number of user-annotated segmentation masks from 174k to 270k. Besides increasing the number of annotated pixels, every frame’s annotation mask was thoroughly reviewed and potentially improved in terms of **quality**, as showcased in Figure 3. In particular, the annotation of the original Coralscapes dataset was performed with substantial use of Segment Anything Model (SAM) [37], which led to border artifacts, and encouraged over-segmentation: often, the masks for coral colonies included dead coral patches or dark regions, which should be appropriately annotated into their respective classes. Furthermore, the auxiliary ‘background’ and ‘dark’ classes, which are innately subjective but crucial for forcing a non-prediction of benthic classes, are applied more consistently in CoralscapesV2. Beyond that, isolated wrong annotations were corrected and all visible instances of liagora algae (often appearing distinctly white in Coralscapes) were moved from the wrong ‘other animal’ class to the ‘algae covered substrate’ class.

3.2 Fine-Grained Visual Categories

The original Coralscapes dataset summarized all annotated classes into 39 label classes that were selected such that classes are well-represented in the train, validation and test splits. CoralscapesV2 relaxes this requirement, and provides 95

¹ As in the original Coralscapes dataset, the geolocations of individual sites are anonymized in accordance with the sovereignty of local monitoring efforts and to protect reefs from targeted tourism or overfishing pressure.

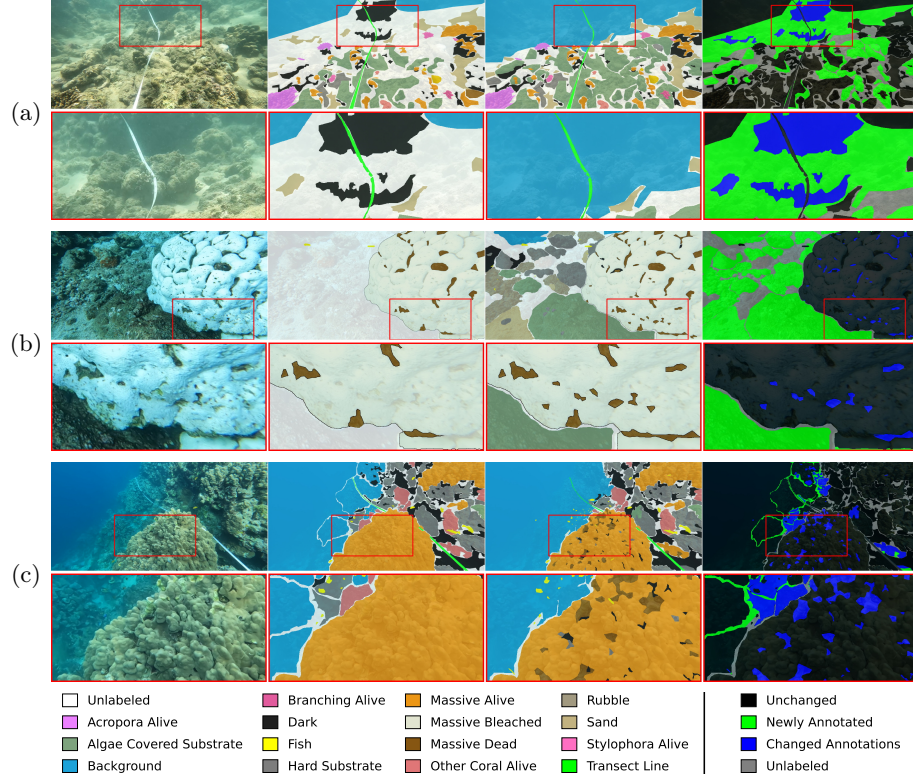


Fig. 3: CoralscapesV2 improves the segmentation targets of existing frames: common improvements are annotating substantially more of the frame area (a, b), applying auxiliary classes like ‘background’ and ‘dark’ more consistently (a), alleviating image-border artifacts from legacy annotations with SAM (b), and filling in small polygons on over-annotated images, including dead and dark patches (b, c). For better visualization, the massive bleached coral mask is outlined in (b).

fine-grained visual categories that are consistently annotated. The remainder of this Section describes the newly added classes.

More Taxa & Morphotypes

CoralscapesV2 separates clearly identifiable coral taxa that were previously summarized. In particular, CoralscapesV2 has classes for visually clearly identifiable taxa, spanning both hard corals (‘lobophyllidae’, ‘porites’, ‘fungiidae’, ‘pavona’, ‘galaxea’, ‘goniopora’, ‘seriatopora’), and soft coral taxa (‘alcyoniidae’, ‘nephtheidae’, ‘xeniidae’). For classes with a clearly distinct morphology but where the taxon can not be identified clearly, new morphologic classes for ‘brain coral’ and ‘thin plate/encrusting coral’ are introduced. Two catch-all classes, ‘hard coral alive’ and ‘soft coral alive’, absorb any coral not assigned a finer label. This guarantees full coverage while preserving the ecologically im-

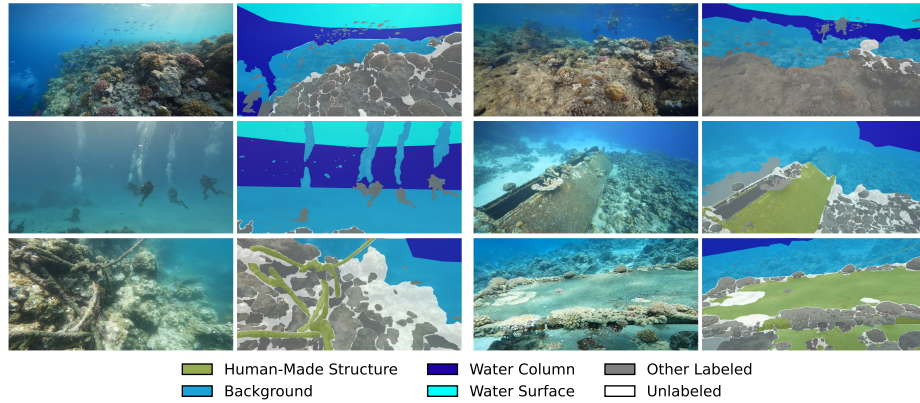


Fig. 4: Samples emphasizing the disambiguated semantics of the new auxiliary classes: ‘water column’ designates image areas with an unobstructed view of the water (no substrate, fish, divers, or marine snow visible), ‘water surface’ denotes shimmering regions from surface waves, and ‘background’ catches all other image areas where blur/visibility prevent a speculation-free assessment of coral taxon or health status.

portant hard–soft coral distinction. Furthermore, classes for the visually distinct black coral ‘cirrhipathes’ and sabellid ‘feather worm’ are included.

The annotation strategy prioritizes correctness over taxonomic depth: labels are given in a non-speculative manner. For example, species in the ‘pavona’ genus can take many growth forms, some of which can be hard to identify as pavona without a high-resolution view of the corallites. In these ambiguous cases, the morphologic label is applied. However, other species (e.g. ‘pavona cactus’ or ‘pavona decussata’) can be confidently identified within the genus pavona, but not distinguished from each other on CoralscapesV2 imagery.

Fine-Grained Coral Composite Health Classes

On top of providing additional taxa, CoralscapesV2 also provides composite health classes (bleached, dead, algae covered) for taxonomic classes.

Auxiliary Classes

CoralscapesV2 provides three new auxiliary classes. All permanent underwater infrastructure including coral restoration nursery tables, concrete blocks, mooring ropes is captured in ‘human-made structure’. Furthermore, the class ‘background’ is disambiguated into distinct semantics, where ‘water surface’ captures light ripples on surface waves, and ‘water column’ explicitly designates areas of the image with a fully unobstructed view into the water. Automatic segmentation of this class is useful for many computer vision applications that estimate optical properties of the water, such as estimating the veiling light term of the backscatter [2, 58], or for automated estimation of the maximum visibility (in meters) for metric-calibrated stereo camera rigs [52]. A disambiguation of these auxiliary classes is exemplified in Figure 4.



Fig. 5: Coralscapes V2 extends the class set from 39 to 95 fine-grained classes, which are shown here along with their polygon and pixel counts in each dataset split.

3.3 Exhaustive Fish Instance Annotations

CoralscapesV2 provides instance-level annotations for every fish in the video frame. The annotators were tasked with using a 4-second video centered around the frame to annotate every single fish instance, even if it is barely visible due to occlusion, small size, poor lighting, camouflage, or far distance from the camera. In particular, even fish that are simply not distinguishable from the static frame of interest, and require the motion from the video, were exhaustively annotated. In total, CoralscapesV2 provides 65k individual fish instances in scenes that are diverse with regards to camera angle, reef, water conditions, number of fish, fish species, and camera distance to the fish, as exemplified in Figure 6. In contrast, in the original Coralscapes, where annotators were also tasked with annotating all fish but only had access to the static frame, 22,233 fish polygons were annotated: comparing the number of fish instances annotated on frames that already existed in V1 reveals that 56.1% of fish instances were missed when annotating with only static frames. For 51% of the matching V1 frames where no fish were annotated in the original Coralscapes, at least one fish is annotated in CoralscapesV2.

The full distribution of fish per image and pixel sizes of annotated polygons is shown in Figure 7, highlighting that CoralscapesV2 reduces the number of frames with 0 or less than 10 fish and increases it across the remainder of the distribution, and that the newly annotated fish concentrate on smaller polygons than those annotated with static frames in V1. CoralscapesV2 also removes false positive labels where marine snow or dark patches in the substrate were annotated as fish, which only becomes apparent when video motion is available.



Fig. 6: Fish instance masks in CoralscapesV2. Each fish is displayed in a random color.

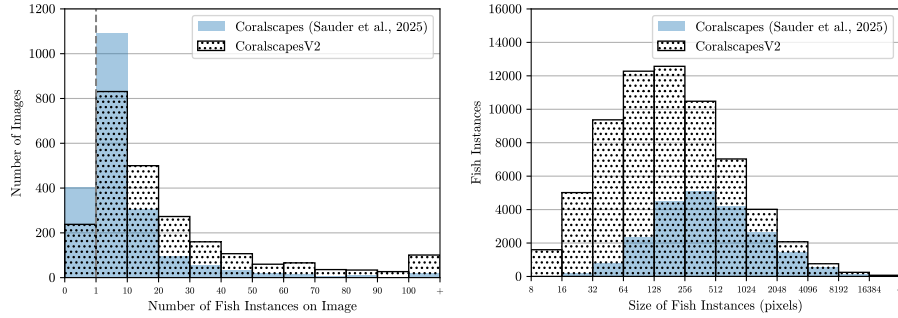


Fig. 7: Histograms showing the number of fish instances on an image (left) and the size distribution of the fish instance polygons in pixels (right).

3.4 Limitations

For some video frames, even with access to the video during annotation, exhaustive precise fish instance annotation is simply not possible. Specifically, these are frames containing large and dense schools of fish that are either small or at the limit of the visibility of the water column. In these cases, annotation was performed to the best ability.

Another limitation is that with the 95 fine-grained classes, it is impossible to find train/validation/test splits where all classes are well-represented directly (Fig. 5). Despite this, we believe this provides an interesting challenge for semantic segmentation with disentangled representations or concept-bottleneck approaches [38]. Furthermore, the fine-grained classes are nonetheless a valuable contribution for real-world coral monitoring applications which could mix CoralscapesV2 with other datasets.



Fig. 8: Examples of images where, even with access to the video, definitive delineation of every single fish instance becomes impossible.

4 Experiments

4.1 Semantic Segmentation

We benchmark contemporary semantic segmentation models on CoralscapesV2. The best-performing model from the benchmarks of Coralscapes V1 (a SegFormer MiT-b5 [65]) serves as a baseline, which is evaluated on CoralscapesV2 without re-training on the 39 mapped classes. The same architecture and training recipe is used to re-train on Coralscapes V2 for both the fine-grained set of 95 classes and the reduced 39 class set. Furthermore, we compare against models consisting of DINOv3-pretrained [61] ViT-B and ViT-L encoders [15] using a DPT head [53] with LoRA [27] applied on the encoder, which we respectively train on V1 (39 classes) and V2 (95 classes). Figure 9 shows that training on CoralscapesV2 improves the accuracy and mIoU for both the test set of V2 and V1, with quantitative results including the 95 class set shown in Table 4. Training details can be found in Appendix B and qualitative samples as well as additional ablations of the prediction resolution and of the effect of either native training on 39 classes or training on 95 classes and re-mapping predictions to 39 classes at evaluation are provided in Appendix C.

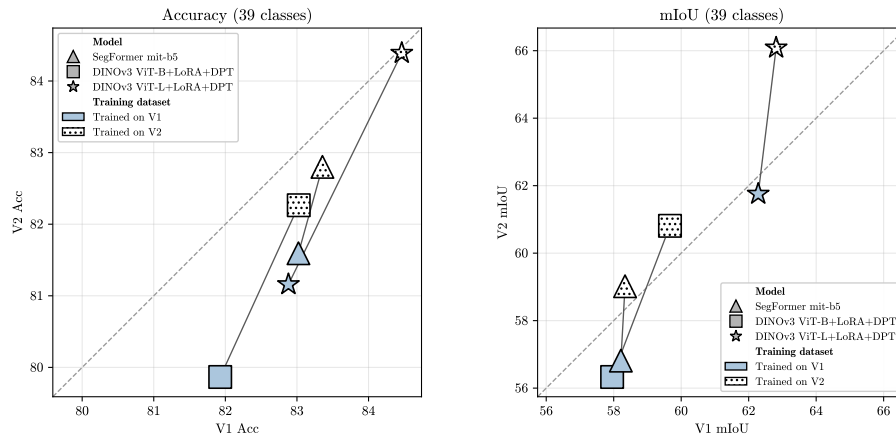


Fig. 9: Visualization of the improvement of training on V2 for the 39 class set.

Table 2: Semantic segmentation results.

Model	Trained on	V2 Acc (95 classes)	V2 mIoU (95 classes)	V2 Acc (39 classes)	V2 mIoU (39 classes)	V1 Acc (39 classes)	V1 mIoU (39 classes)
SegFormer mit-b5	V1	-	-	81.594	56.815	82.761	57.800
DINOv3 ViT-B+LoRA+DPT	V1	-	-	79.866	56.330	81.928	57.953
DINOv3 ViT-L+LoRA+DPT	V1	-	-	81.159	61.754	82.881	62.283
SegFormer mit-b5	V2	80.382	37.177	82.801	59.020	83.356	58.336
DINOv3 ViT-B+LoRA+DPT	V2	80.243	36.918	82.265	60.810	83.025	59.668
DINOv3 ViT-L+LoRA+DPT	V2	81.700	40.221	84.390	66.086	84.464	62.814

4.2 Instance Segmentation

For benchmarking instance segmentation on CoralscapesV2, we train Mask2Former [11] models using ResNet-50 [24] and Swin-L [43] backbones as baselines. In particular, we use the Video-Mask2Former decoder architecture [10] for all models, and evaluate three variants: an image variant (single anchor frame, frame index [0]), a 2-frame video variant (frames [-1, 0]), and a 3-frame video variant (frames [-2, -1, 0]). The results are displayed in Table 3: for the ResNet-50 backbone, video models show a clear trend of improved performance with temporal context. For the Swin-L backbone baseline, the image model outperforms video variants.

Table 3: Instance segmentation and detection performance in image and video settings.

Model	Setting	Segmentation			Detection		
		AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
RN50 + Mask2Former (100q)	Image	12.73	26.8	10.63	12.81	26.46	10.89
RN50 + Mask2Former (100q)	Video (2 frames)	12.77	26.86	10.85	12.86	26.60	10.93
RN50 + Mask2Former (100q)	Video (3 frames)	12.78	26.90	10.66	12.98	26.65	11.08
Swin-L + Mask2Former (200q)	Image	15.41	30.89	13.70	15.23	30.23	13.41
Swin-L + Mask2Former (200q)	Video (2 frames)	14.93	29.93	13.23	14.76	29.24	12.82
Swin-L + Mask2Former (200q)	Video (3 frames)	15.25	30.57	13.48	15.09	30.00	13.26

5 Conclusion & Future Work

This paper presents CoralscapesV2, an extension of Coralscapes that increases the size, annotation density, and quality of the semantic segmentation labels, refines the label set from 39 to 95 fine-grained visual categories capturing more taxa, coral morphologies, health status, and auxiliary classes, and adds 65k exhaustive fish instance masks.

A central takeaway is that fish detection in reefs should be treated as a *video* task rather than an image task. Comparing the fish annotated with video context against the static-frame annotations of the original Coralscapes reveals that 56.1% of fish instances were missed when annotating from static frames alone, and on 51% thought to be devoid of fish, the video context reveals at least one fish. Despite this suggesting that temporal context is essential for exhaustive fish detection in reefs, the corresponding model training gives inconclusive results.

By providing general-purpose panoptic annotations, CoralscapesV2 lays the groundwork for scaling fine-grained reef monitoring beyond constrained survey protocols. Future work includes training improved general-purpose models on this dataset and deploying them on the increasingly diverse platforms used to collect reef imagery, including handheld diver video surveys or robotic AUV/ROV surveys for joint benthic-cover mapping and fish quantification, as well as leveraging the exhaustive fish instance masks for the study of fish behavior and fish-reef interactions.

References

1. AIMS: Reefcloud.ai. <https://reefcloud.ai/> (2023)
2. Akkaynak, D., Treibitz, T.: Sea-thru: A method for removing water from underwater images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1682–1691 (2019)
3. Allken, V., Rosen, S., Handegard, N.O., Malde, K.: A deep learning-based method to identify and count pelagic and mesopelagic fishes from trawl camera images. *ICES Journal of Marine Science* **78**(10), 3780–3792 (2021)
4. Atlas, W.I., Ma, S., Chou, Y.C., Connors, K., Scurfield, D., Nam, B., Ma, X., Cleveland, M., Doire, J., Moore, J.W., Shea, R., Liu, J.: Wild salmon enumeration and monitoring using deep learning empowered detection and tracking. *Frontiers in Marine Science* **10**, 1200408 (2023)
5. Ayyagari, K.D., Drautz, M., Porter, D., Barnes, J., Morris, C., Whidden, C.: Newfoundland marine refuge fish classification dataset (N-MARINE). In: NeurIPS Workshop on Tackling Climate Change with Machine Learning (2025)
6. Beijbom, O., Edmunds, P.J., Kline, D.I., Mitchell, B.G., Kriegman, D.: Automated annotation of coral reef survey images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1170–1177 (2012)
7. Beijbom, O., Edmunds, P.J., Roelfsema, C., Smith, J., Kline, D.I., Neal, B.P., Dunlap, M.J., Moriarty, V., Fan, T.Y., Tan, C.J., et al.: Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. *PloS one* **10**(7), e0130312 (2015)
8. Boom, B.J., He, J., Palazzo, S., Huang, P.X., Beyan, C., Chou, H.M., Lin, F.P., Spampinato, C., Fisher, R.B.: A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. *Ecological Informatics* **23**, 83–97 (2014)
9. Cai, L., Greene, A., Yang, D., Aoki, N., Jarriel, S.D., Ha, J., Formel, N., Mooney, T.A., Girdhar, Y.: Species-level detection and tracking of caribbean coral reef fish. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops (CV4E) (2025)
10. Cheng, B., Choudhuri, A., Misra, I., Kirillov, A., Girdhar, R., Schwing, A.G.: Mask2former for video instance segmentation. *arXiv preprint arXiv:2112.10764* (2021)
11. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1290–1299 (2022)
12. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3213–3223 (2016)
13. Cutter, G., Stierhoff, K., Zeng, J.: Automated detection of rockfish in unconstrained underwater videos using haar cascades and a new image dataset: Labeled fishes in the wild. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW). pp. 57–62 (2015)
14. Dawkins, M., et al.: FishTrack23: An ensemble underwater dataset for multi-object tracking. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (2024)

15. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
16. Eddy, T.D., Lam, V.W., Reygondeau, G., Cisneros-Montemayor, A.M., Greer, K., Palomares, M.L.D., Bruno, J.F., Ota, Y., Cheung, W.W.: Global decline in capacity of coral reefs to provide ecosystem services. *One Earth* **4**(9), 1278–1285 (2021)
17. Edwards, C.B., Eynaud, Y., Williams, G.J., Pedersen, N.E., Zgliczynski, B.J., Gleason, A.C., Smith, J.E., Sandin, S.A.: Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. *Coral Reefs* **36**(4), 1291–1305 (2017)
18. Farrell, D.M., Ferriss, B., Sanderson, B., Veggerby, K., Robinson, L., Trivedi, A., Pathak, S., Muppalla, S., Wang, J., Morris, D., Dodhia, R.: A labeled data set of underwater images of fish and crab species from five mesohabitats in puget sound wa usa. *Scientific Data* **10**, 799 (2023)
19. Francescangeli, M., Marini, S., Martínez, E., del Río Fernández, J., Aguzzi, J., et al.: Image dataset for benchmarking automated fish detection and classification algorithms. *Scientific Data* **10**, 5 (2023)
20. Garcia-d’Urso, N., Galan-Cuenca, A., Climent-Pérez, P., Saval-Calvo, M., Azorin-Lopez, J., Fuster-Guillo, A.: The DeepFish computer vision dataset for fish instance segmentation, classification, and size estimation. *Scientific Data* **9**(1), 287 (2022)
21. Girdhar, Y., McGuire, N., Cai, L., Jamieson, S., McCammon, S., Claus, B., San Soucie, J.E., Todd, J.E., Mooney, T.A.: Curee: A curious underwater robot for ecosystem exploration. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 11411–11417. IEEE (2023)
22. González-Rivero, M., Rodríguez-Ramírez, A., Beijbom, O., Dalton, P., Kennedy, E.V., Neal, B.P., Vercelloni, J., Bongaerts, P., Ganase, A., Bryant, D.E., et al.: Seaview survey photo-quadrat and image classification dataset (2019)
23. Haurum, J.B., Karpova, A., Pedersen, M., Bengtson, S.H., Moeslund, T.B.: Re-identification of zebrafish using metric learning. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). pp. 1–11 (2020)
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
25. Hong, J., Fulton, M., Sattar, J.: TrashCan: A semantically-segmented dataset towards visual detection of marine debris. *arXiv preprint arXiv:2007.08097* (2020)
26. Hong, L., Wang, X., Li, Y., Wang, X.: USIS16K: High-quality dataset for underwater salient instance segmentation. *arXiv preprint arXiv:2506.19472* (2025)
27. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (ICLR) (2022)
28. Hughes, T.P., Kerry, J.T., Álvarez-Noriega, M., Álvarez-Romero, J.G., Anderson, K.D., Baird, A.H., Babcock, R.C., Beger, M., Bellwood, D.R., Berkemans, R., et al.: Global warming and recurrent mass bleaching of corals. *Nature* **543**(7645), 373–377 (2017)
29. Hughes, T.P., Kerry, J.T., Baird, A.H., Connolly, S.R., Dietzel, A., Eakin, C.M., Heron, S.F., Hoey, A.S., Hoogenboom, M.O., Liu, G., et al.: Global warming transforms coral reef assemblages. *Nature* **556**(7702), 492–496 (2018)

30. Islam, M.J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S.S., Sattar, J.: Semantic segmentation of underwater imagery: Dataset and benchmark. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2020)
31. Jansen, A., Walden, D., Walker, S., Buccella, C., et al.: Monitoring tropical freshwater fish with underwater videography and deep learning. *Marine and Freshwater Research* **75**(8), MF23166 (2024)
32. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planqué, R., Rauber, A., Palazzo, S., Fisher, B., Müller, H.: LifeCLEF 2015: Multimedia life species identification challenges. In: International Conference of the Cross-Language Evaluation Forum for European Languages (CLEF). pp. 462–483. Springer (2015)
33. Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., Boulais, O., Cromwell, M., Butler, E., Woodward, B., Bell, K.L.C.: FathomNet: A global image database for enabling artificial intelligence in the ocean. *Scientific Reports* **12**, 15914 (2022)
34. Kavasidis, I., Palazzo, S., Di Salvo, R., Giordano, D., Spampinato, C.: An innovative web-based collaborative platform for video annotation. *Multimedia Tools and Applications* **70**(1), 413–432 (2013)
35. Khan, F.F., Li, X., Temple, A.J., Elhoseiny, M.: FishNet: A large-scale dataset and benchmark for fish recognition, detection, and functional trait prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 20496–20506 (2023)
36. Khiem, N.M., Van Thanh, T., Dung, N.H., Takahashi, Y.: A novel approach combining yolo and deepsort for detecting and counting live fish in natural environments through video. *PloS One* **20**(6), e0323547 (2025)
37. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4015–4026 (2023)
38. Koh, P.W., Nguyen, T., Tang, Y.S., Mussmann, S., Pierson, E., Kim, B., Liang, P.: Concept bottleneck models. In: International conference on machine learning. pp. 5338–5348. PMLR (2020)
39. Li, H., Lian, S., Li, Z., Cong, R., Li, C., Yang, L.T., Zhang, W., Kwong, S.: Advancing marine research: UWSAM framework and UIIS10K dataset for precise underwater instance segmentation. *arXiv preprint arXiv:2505.15581* (2025)
40. Li, W., Liu, Y., Guo, Q., Wei, Y., Leo, H.L., Li, Z.: When trackers date fish: A benchmark and framework for underwater multiple fish tracking. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 6459–6467 (2026)
41. Lian, S., Li, H., Cong, R., Li, S., Zhang, W., Kwong, S.: WaterMask: Instance segmentation for underwater imagery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1305–1315 (2023)
42. Lilkendey, J., Barrelet, C., Zhang, J., Meares, M., Larbi, H., Subsol, G., Chaumont, M., Sabetian, A.: Herbivorous fish feeding dynamics and energy expenditure on a coral reef: Insights from stereo-video and ai-driven 3d tracking. *Ecology and Evolution* **14**(3), e11070 (2024)
43. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF international conference on computer vision (ICCV). pp. 9992–10002. Ieee (2021)

44. Marrable, D., Barker, K., Tippaya, S., Wyatt, M., Bainbridge, S., Stowar, M., Larke, J.: Accelerating species recognition and labelling of fish from underwater video with machine-assisted deep learning. *Frontiers in Marine Science* **9**, 944582 (2022)
45. MERMAID: Marine ecological research management aid (mermaid). <https://datamermaid.org/> (2023)
46. Microsoft: Project natick analysis: Annotated underwater video dataset. https://github.com/microsoft/Project_Natick_Analysis (2018)
47. MIT Sea Grant: MIT sea grant river herring dataset. <https://lila.science/datasets/mit-sea-grant-river-herring/> (2023)
48. Nishio, T., Kawae, Y.: Deep-sea image dataset for organism detection. *Data in Brief* (2026)
49. Orange Business and Ténaka: Marine detect: Object detection models for identifying species in marine environments. <https://github.com/Orange-OpenSource/marine-detect> (2024)
50. Pedersen, M., Haurum, J.B., Gade, R., Moeslund, T.B., Madsen, N.: Detection of marine animals in a new underwater dataset with varying visibility. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. pp. 18–26 (2019)
51. Raine, S., Olivastri, E., Moshirian, B., Fischer, T.: Ai-driven dispensing of coral reseedling devices for broad-scale restoration of the great barrier reef. *arXiv preprint arXiv:2509.01019* (2025)
52. Randall, Y.: Flsea: Underwater visual-inertial and stereo-vision forward-looking datasets. University of Haifa (Israel) (2023)
53. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 12179–12188 (2021)
54. Saleh, A., Laradji, I.H., Konovalov, D.A., Bradley, M., Vazquez, D., Sheaves, M.: A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports* **10**(1), 14671 (2020)
55. Sauder, J., Banc-Prandi, G., Meibom, A., Tuia, D.: Scalable semantic 3d mapping of coral reefs with deep learning. *Methods in Ecology and Evolution* **15**(5), 916–934 (2024)
56. Sauder, J., Banc-Prandi, G., Perna, G., Abdallah, I.S., Saad, O.S., Mohammed, M.A., Al-Sawalimih, A., Meibom, A., Tuia, D.: Rapid consistent reef surveys with deepreefmap. *Scientific Reports* **15**(1), 39145 (2025)
57. Sauder, J., Domazetoski, V., Banc-Prandi, G., Perna, G., Meibom, A., Tuia, D.: The coralscapes dataset: Semantic scene understanding in coral reefs. In: *Proceedings of the International Conference on Computer Vision (ICCV) Joint Workshop on Marine Vision* (2025)
58. Sauder, J., Tuia, D.: Self-supervised underwater caustics removal and descattering via deep monocular slam. In: *European Conference on Computer Vision*. pp. 214–232. Springer (2024)
59. Scoulding, B., Maguire, K., Orenstein, E., Jackett, C.: Tasmanian orange roughy stereo image machine learning dataset. <https://doi.org/10.25919/a90r-4962> (2025)
60. Shi, X., Czerniawski, R., Tanwari, K., Furdyna, A., Krepski, T.: Pomerania fish: A dataset for fishes across pomerania freshwater waterbodies in-situ environments. *Scientific Data* **13**, 83 (2026)

61. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P.: DINOv3 (2025), <https://arxiv.org/abs/2508.10104>
62. Solawetz, J.: Fish object detection dataset. <https://public.roboflow.com/object-detection/fish> (2020)
63. Souter, D., Planes, S., Wicquart, J., Logan, M., Obura, D., Staub, F.: Status of coral reefs of the world: 2020. Global Coral Reef Monitoring Network (2021)
64. Vijayalakshmi, M., Sasithradevi, A.: A comprehensive annotated image dataset for real-time fish detection in pond settings. *Data in Brief* (2024)
65. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: SegFormer: Simple and efficient design for semantic segmentation with transformers. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 34, pp. 12077–12090 (2021)
66. Yuval, M., Alonso, I., Eyal, G., Tchernov, D., Loya, Y., Murillo, A.C., Treibitz, T.: Repeatable semantic reef-mapping through photogrammetry and label-augmentation. *Remote Sensing* **13**(4), 659 (2021)
67. Zhang, C., others, Chen, Y.: An underwater image dataset for occlusion-aware fish instance segmentation. *Scientific Data* (2026)
68. Zheng, Z., Liang, H., Hua, B.S., Wong, Y.H., Ang Jr, P., Chui, A.P.Y., Yeung, S.K.: CoralSCOP: Segment any COral image on this planet. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 28170–28180 (2024)
69. Zhuang, P., Wang, Y., Qiao, Y.: WildFish: A large benchmark for fish recognition in the wild. In: *Proceedings of the 26th ACM International Conference on Multimedia (MM)*. pp. 1301–1309. ACM (2018)

A Additional Samples

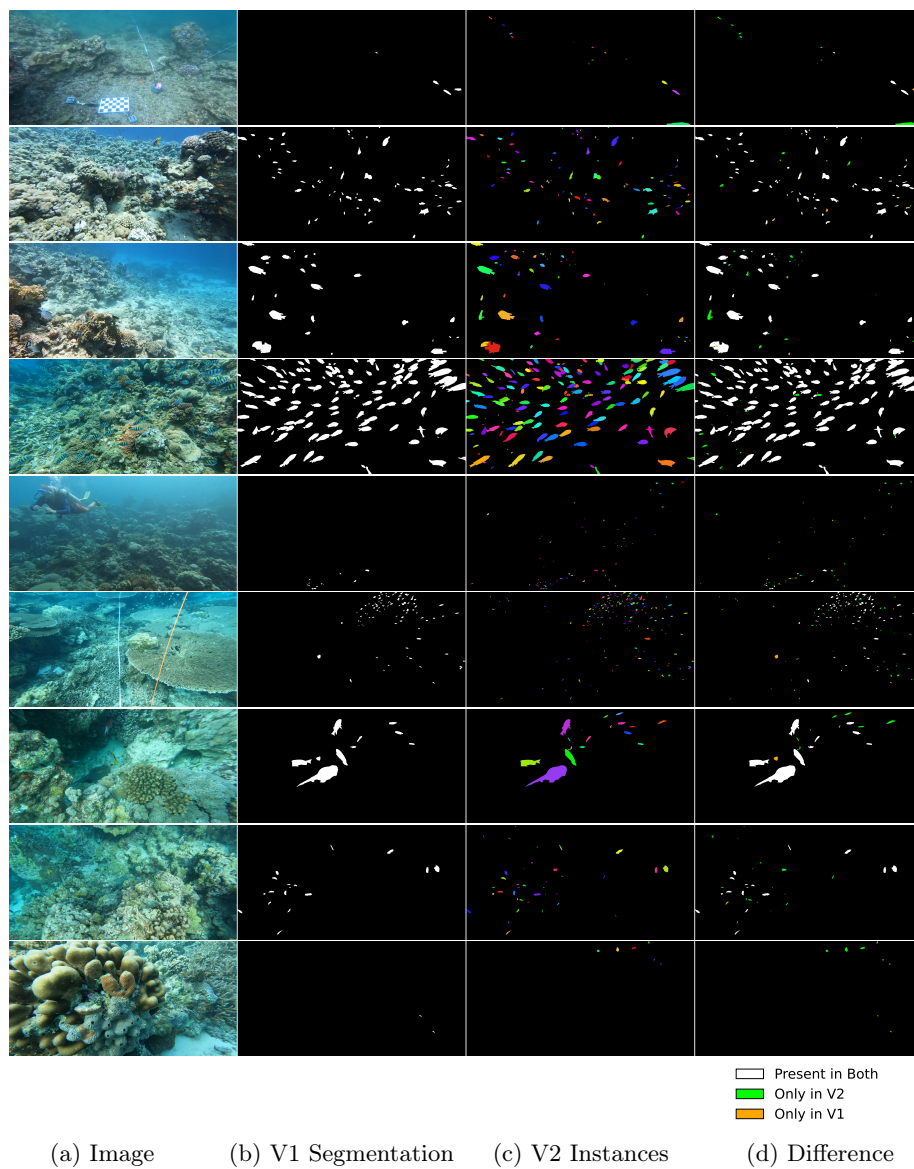


Fig. 10: Images and their fish segmentation masks of V1 and instance masks from Coralscapes V2, along with a visualization of the newly added fish and removed false positives.

B Training Details

SegFormer

We use the best hyperparameter configuration of [57], training the SegFormer MiT-b5 with the AdamW optimizer at a learning rate of 6×10^{-5} and weight decay 0.01, following a polynomial (power 1, i.e. linear) learning-rate decay schedule for a fixed 100 epochs with a batch size of 4 on the combined train+val split, and evaluate on the test split. Training augmentations consist of a random resized crop to 1024×1024 px (scale 0.02–1.0), color jitter (brightness, contrast, and saturation in $[0.8, 1.2]$, hue in $[-0.05, 0.05]$), random rotation of up to $\pm 15^\circ$, and random horizontal flips.

DinoV3+LoRA+DPT

The DINOv3 ViT-B and ViT-L encoders (pretrained on LVD-1689M) are kept frozen and adapted with LoRA ($r = 16$, $\alpha = 32$, dropout 0.05), feeding a DPT head (fusion dimension 256, neck hidden sizes $[96, 192, 384, 768]$, reassemble factors $[4, 2, 1, 0.5]$). We train with AdamW at a learning rate of 5×10^{-5} and weight decay 0.01 with 1000 warmup steps, for a fixed 150 epochs with a batch size of 8 and mixed-precision training. Augmentations use multi-scale training sizes (shortest edge ranging from 736 to 1376px across mixed aspect ratios), a random resized crop (scale 0.2–1.0, ratio 0.5–2.0), random rotation of up to $\pm 20^\circ$, and color jitter (brightness, contrast, and saturation 0.2, hue 0.05). Evaluation is performed at 768×1376 px (the strided 1024×1024 px variant is reported in Appendix C).

Mask2Former

We use a Video Mask2Former [10], built on the universal Mask2Former image architecture [11], with the backbone initialized from COCO instance-segmentation weights, with a single *fish* thing class, 10 transformer decoder layers, and a multi-scale deformable-attention pixel decoder (6 encoder layers, hidden dimension 256). The loss combines classification (weight 2.0), mask (weight 5.0), and dice (weight 5.0) terms with a no-object weight of 0.1, using point sampling with 12544 points (oversample ratio 3.0, importance-sample ratio 0.75). Training follows the configuration of [10], using AdamW at a base learning rate of 10^{-4} , weight decay 0.05, a backbone learning-rate multiplier of 0.1, gradient clipping at 0.01, and mixed precision. We use a warmup polynomial schedule (250 iterations) and an effective batch size of 24 clips (12 clips per step with 2 gradient-accumulation steps). Multi-scale training resizes the shortest edge to 512-1024px and then takes a 512×1024 px crop. We evaluate three variants sharing this recipe: an image variant (single anchor frame, $[0]$), a 2-frame video variant (frames $[-1, 0]$), and a 3-frame video variant (frames $[-2, -1, 0]$); in every variant all clip frames pass through the network but only the anchor frame is supervised and scored. Each model is tuned on CoralscapesV2 train+val for a fixed 15000 iterations.

C Supplemental Results

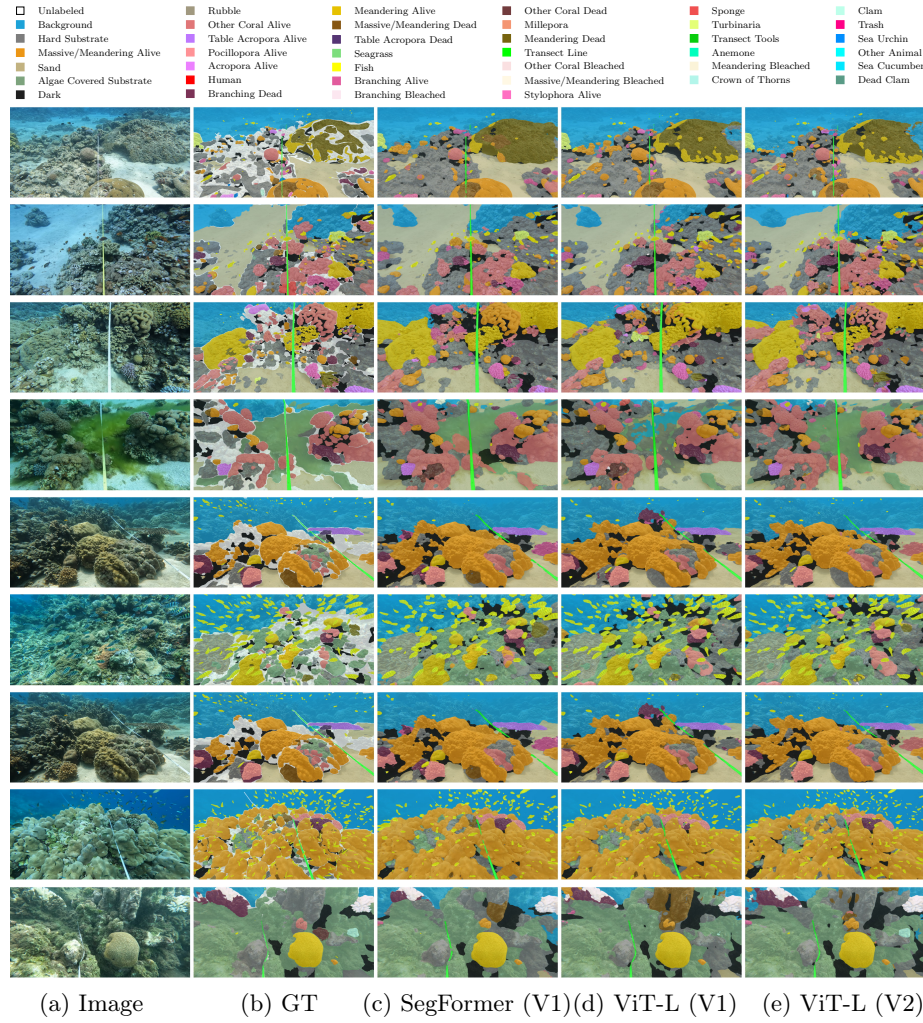


Fig.11: Images from the CoralscapesV2 test set along with their ground-truth segmentation labels and predictions from SegFormer (MiT-b5), and DinoV3 ViT-L+LoRA+DPT trained on V1 and V2.

Table 4: Extended segmentation results, showing the effect of either predicting in a single pass at 768×1376 px resolution, or doing a strided prediction with a 1024×1024 px window at a 1024 px stride as in [57]. Furthermore, the effect of training V2 models on the 39 class set is compared against training on the fine-grained 95-class set, and mapping the predicted results to the 39 legacy classes for evaluation only.

Model	Trained on	Predict at	V2 Acc (95 classes)	V2 mIoU (95 classes)	V2 Acc (39 classes, mapped)	V2 mIoU (39 classes, mapped)	V2 Acc (39 classes, trained)	V2 mIoU (39 classes, trained)	V1 Acc (39 classes, mapped)	V1 mIoU (39 classes, mapped)	V1 Acc (39 classes, trained)	V1 mIoU (39 classes, trained)
SegFormer mit-b5	V1	768×1376	-	-	-	-	80.553	53.516	-	-	81.818	54.755
SegFormer mit-b5	V1	2×1024^2	-	-	-	-	81.294	56.835	-	-	82.761	57.800
DINOv3 ViT-B+LoRA+DPT	V1	768×1376	-	-	-	-	79.269	55.458	-	-	81.296	56.075
DINOv3 ViT-B+LoRA+DPT	V1	2×1024^2	-	-	-	-	79.866	56.330	-	-	81.928	57.953
DINOv3 ViT-L+LoRA+DPT	V1	768×1376	-	-	-	-	80.327	59.189	-	-	82.384	59.060
DINOv3 ViT-L+LoRA+DPT	V1	2×1024^2	-	-	-	-	81.159	61.754	-	-	82.881	62.283
SegFormer mit-b5	V2	768×1376	79.462	36.420	81.937	56.831	82.055	56.033	82.663	56.387	82.371	55.213
SegFormer mit-b5	V2	2×1024^2	80.382	37.177	82.836	59.156	82.801	59.020	83.553	58.615	83.356	58.336
DINOv3 ViT-B+LoRA+DPT	V2	768×1376	79.716	36.317	82.040	58.465	81.822	59.792	82.696	57.675	82.683	58.033
DINOv3 ViT-B+LoRA+DPT	V2	2×1024^2	80.243	36.918	82.584	60.139	82.265	60.810	83.183	59.248	83.025	59.668
DINOv3 ViT-L+LoRA+DPT	V2	768×1376	81.444	40.476	83.950	62.216	84.240	64.472	84.221	60.485	84.361	61.194
DINOv3 ViT-L+LoRA+DPT	V2	2×1024^2	81.700	40.221	84.022	63.617	84.290	66.086	84.287	61.960	84.464	62.814

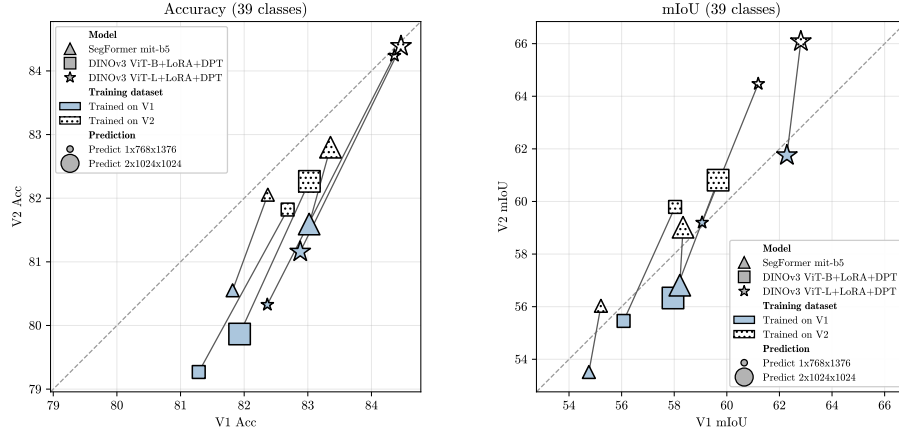


Fig. 12: Visualization of the quantitative metrics of semantic segmentation for models trained on the 39 label class set, showing that training on Coralscapes V2 improves every model's performance on both metrics and datasets.