

WildFin: An in the wild dataset for fish behavioral recognition

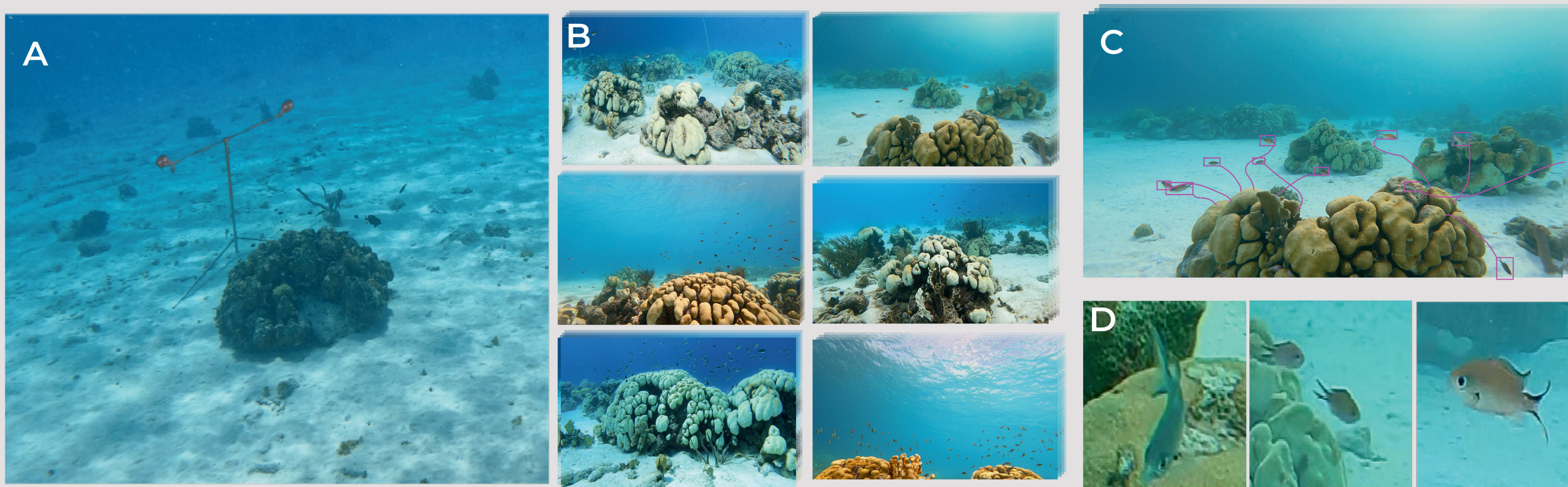
Abigail G. Grassick^{1*}, Jerome Tze-Hou Hsu^{1*}, Ethan Lin^{1*}, Ziang Liu¹, Max Whitton¹, Madelyn Hair², Liam Gutierrez², Haozheng Yu¹, Kristin Branson³, Vivek Jayaraman³, Michael A. Gil², Andrew M. Hein¹, Jennifer J. Sun¹

¹Cornell University, Ithaca NY 14850, ²University of Colorado at Boulder, Boulder CO 80309, ³HHMI Janelia Research Campus, Ashburn VA 20147

CoralCam

1,350

hours of scientist fieldwork

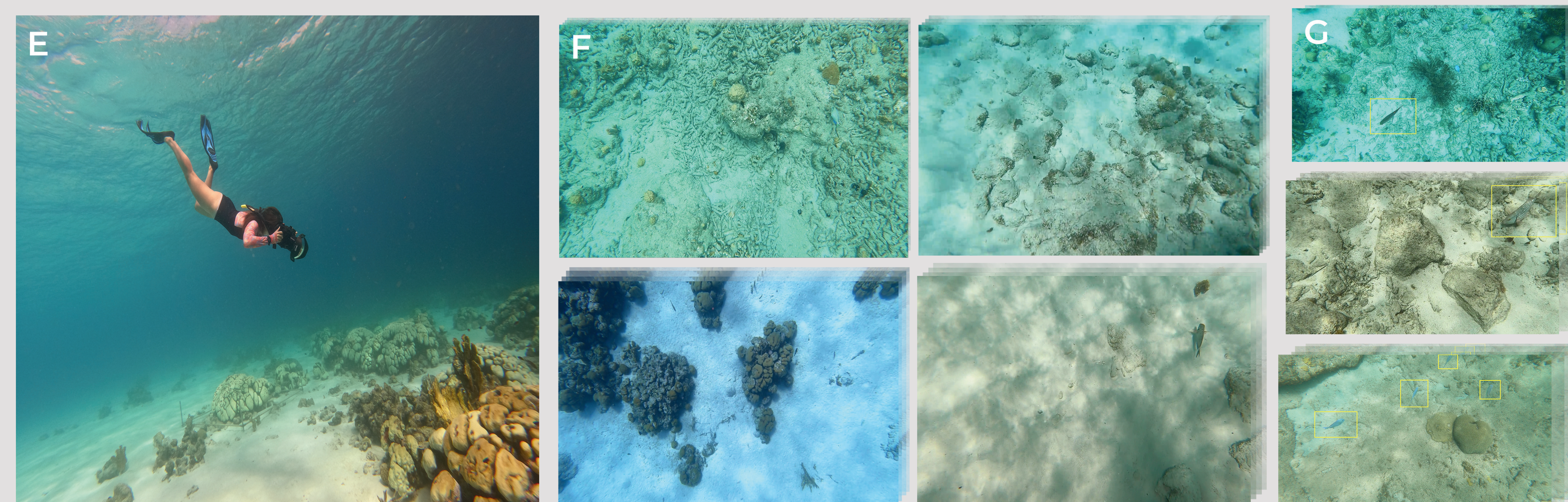


CoralCam data collection and pipeline: (A) stationary cameras which are left in the field to record for a period of 3-5 hours; (B) sampling regimes showing differences in lighting, turbidity, distance to individuals, and background complexity; (C) tracking and bounding box generation; (D) bounding-box crops, from which individual behaviors (biting, charging, fleeing) are determined

600

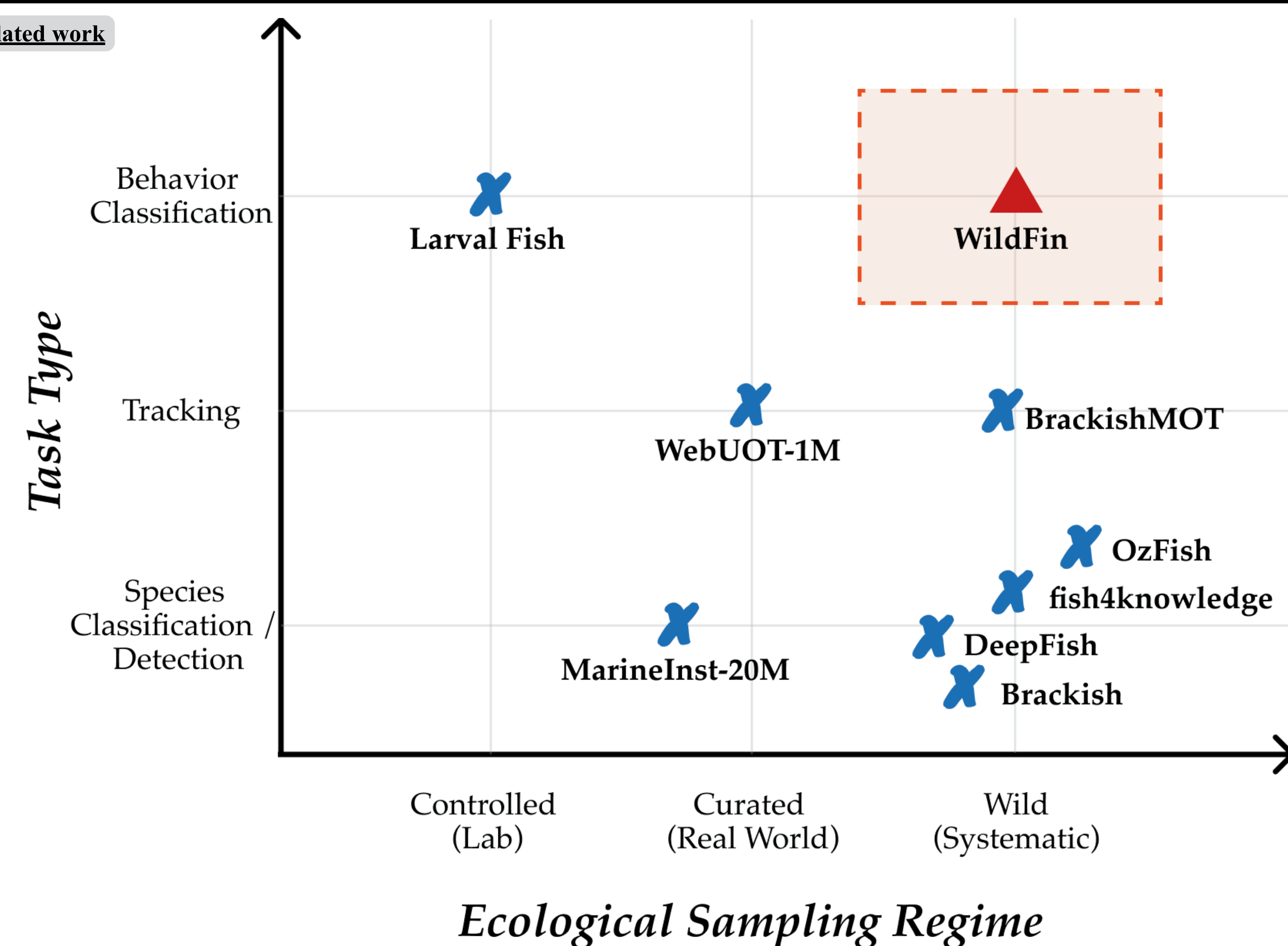
hours of expert annotation

FishFollow



FishFollow data collection and annotation: (E) fish are identified and followed by a diver (F) fish can be tracked through a variety of complex environments, where they are well camouflaged and difficult to localize without the context of their movement; (G) Annotations are performed on the whole frame from an ethogram of 20 behaviors (solo foraging, sand rubbing, and social foraging).

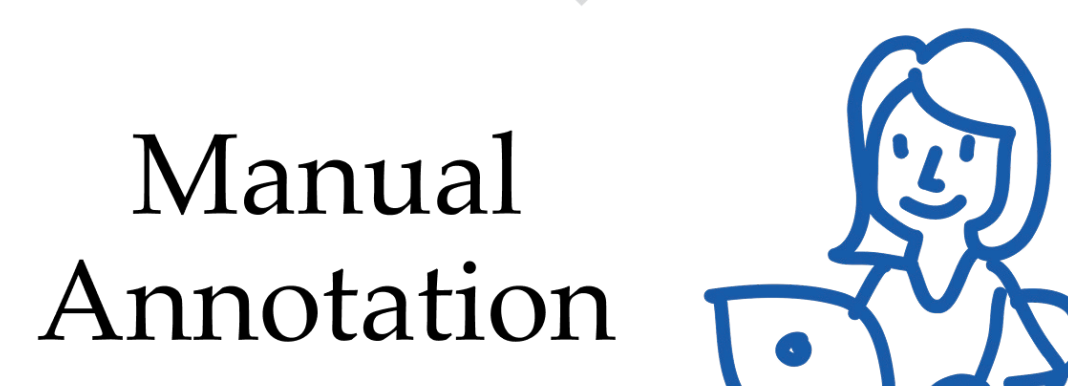
Related work



Key Takeaways

- WildFin fills a gap in underwater/marine behavioral datasets
- Current methods fail to classify both point and state behaviors
- Rare behaviors remain inaccessible to current models

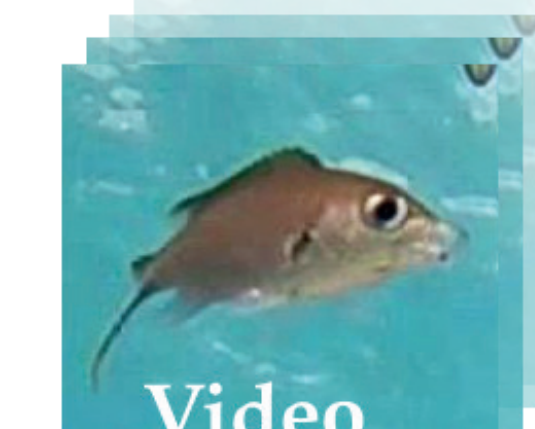
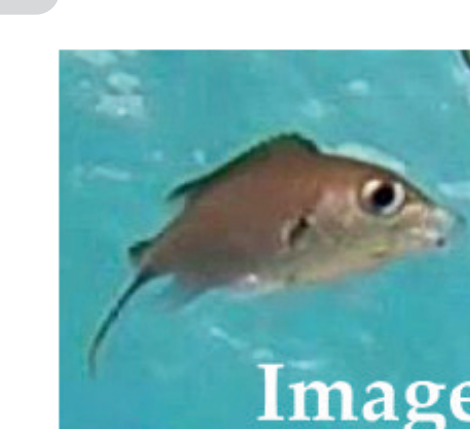
Dataset & code



CV Behavior Classification



Method



Pre-trained Frozen Backbone

Image Foundation Model

Video Foundation Model

Classification Head

Behavior Classification

0.6 Biting
0.25 Feeding
0.05 Charging
0.1 Not Visible
...

0.8 Biting
0.1 Feeding
0.05 Charging
0.05 Not Visible
...

CoralCam

Results

	Overall	Aggression	Biting
DINOv3-B	0.309	0.000	0.464
DINOv3-L	0.329	0.007	0.490
ResNet50	0.286	0.000	0.429
VideoMAE-B	0.362	0.311	0.388
VideoMAE-L	0.450	0.332	0.510
V-JEPA-2-L	0.356	0.294	0.387

Two backbones score 0.000 — they never detect it

FishFollow

	Overall	Habitat	Movement	Bites	Social	Not Vis.
DINOv3-B	0.378	0.582	0.465	0.108	0.055	0.457
DINOv3-L	0.380	0.594	0.458	0.123	0.036	0.458
ResNet50	0.361	0.571	0.460	0.082	0.023	0.405
VideoMAE-B	0.370	0.552	0.454	0.105	0.085	0.486
VideoMAE-L	0.387	0.563	0.472	0.149	0.121	0.411
V-JEPA-2-L	0.375	0.566	0.484	0.119	0.000	0.402

Rare, state-based behaviors: every backbone struggles

Macro-F1 by backbone and behavior category. Color used to indicate model performance by class