Image Labels Are All You Need for Coarse Seagrass Segmentation

Scarlett Raine[1,2], Ross Marchant[3], Brano Kusy[2], Frederic Maire[1], Tobias Fischer[1]

[1]QUT Centre For Robotics [2]CSIRO Data61 [3]Image Analytics

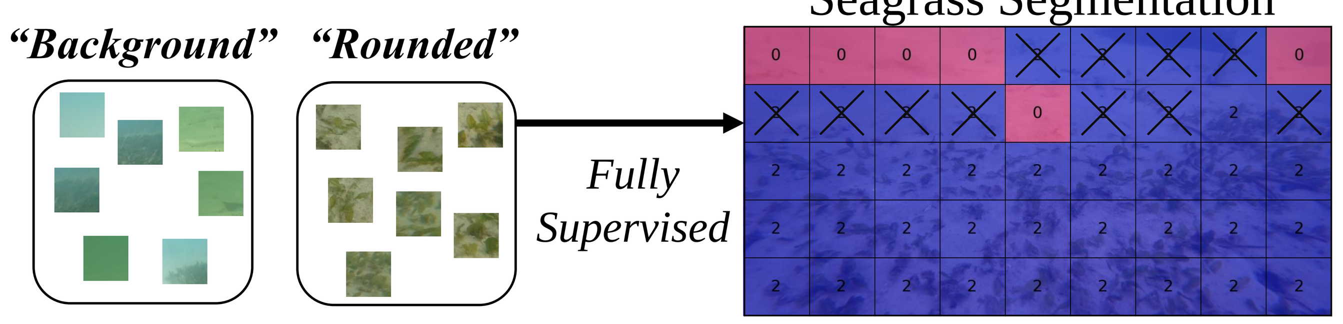
Find our code on GitHub: https://github.com/sgraine/bag-of-seagrass **Email:** scarlett.raine@hdr.qut.edu.au

JAN 4-8 WACV 2024 WAIKOLOA HAWAII

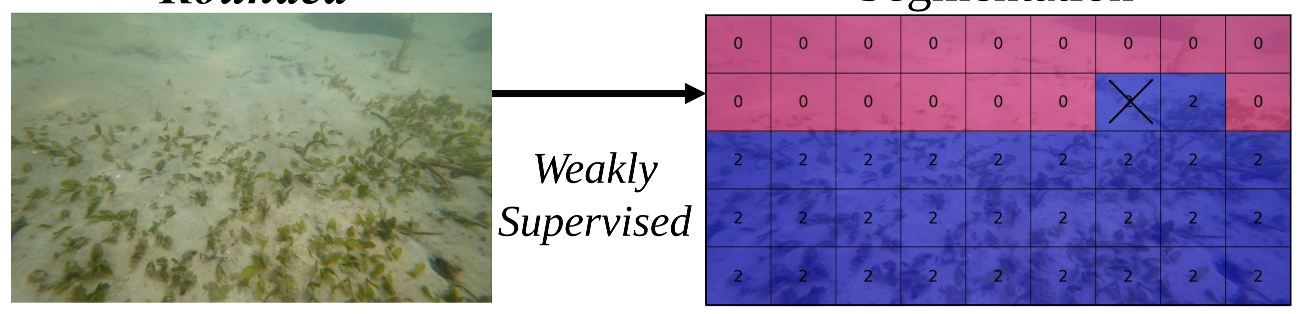
Motivation

- Coastal ecosystems sequester carbon from the atmosphere
- Scientists require data on seagrass meadow extents and species composition to estimate blue carbon sequestration
- Machine learning enables automated processing of images and adaptation of robotic vehicle survey paths in real-time

Prior Approaches



Our Approach: Coarse Seagrass Segmentation



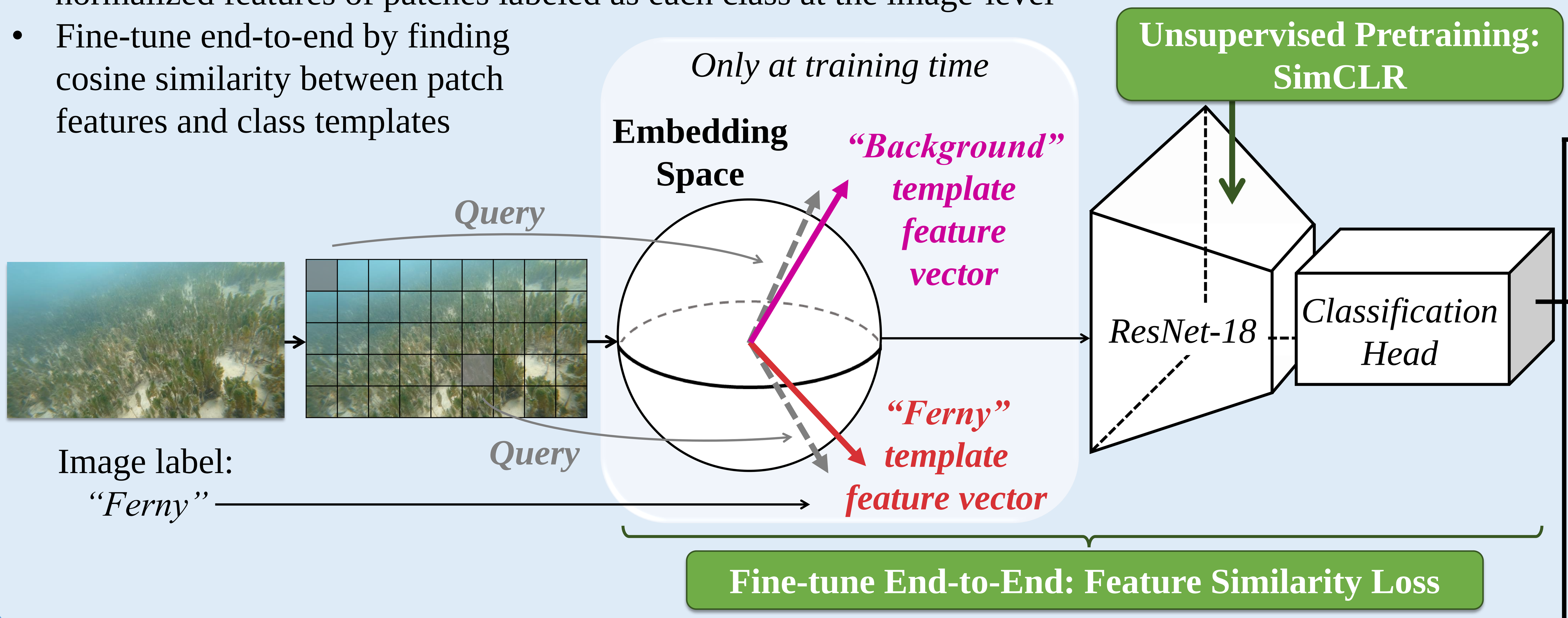
- Seagrass lacks distinct semantic features and has poorly defined boundaries
- Pixel-level labeling is too costly and time-consuming; use image-level labels instead
- Reduces labels required by **96.1%**

Legend: 0 Background 1 Ferny 2 Rounded 3 Strappy ✕ Incorrect

Our Approach

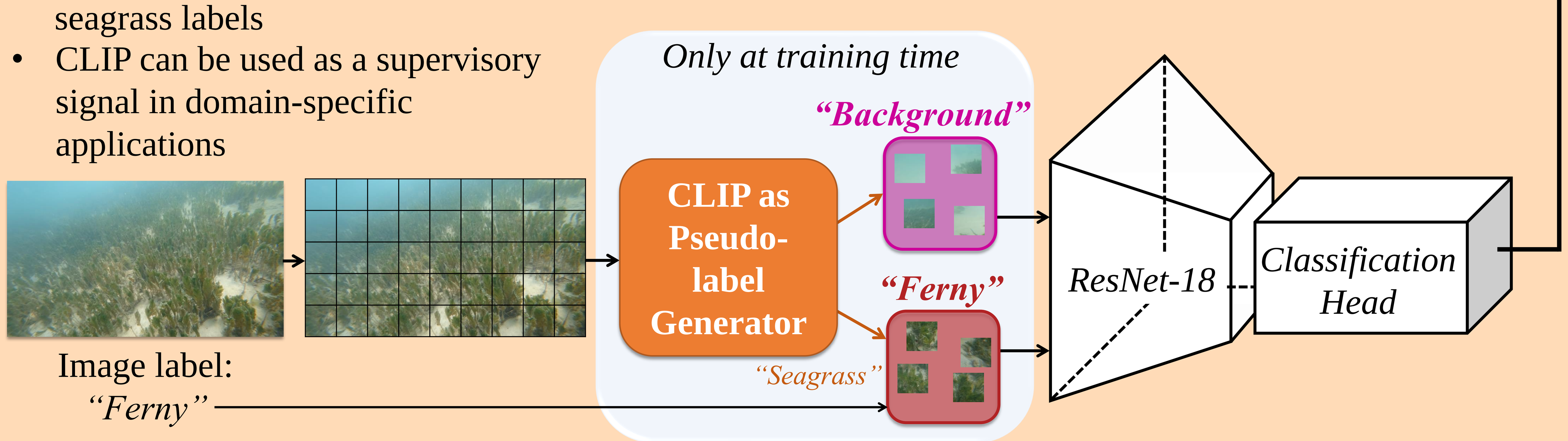
SeaFeats – *weak supervision from feature similarity*

- We obtain per-class template feature vectors (dynamically updated each epoch) by averaging the L2 normalized features of patches labeled as each class at the image-level
- Fine-tune end-to-end by finding cosine similarity between patch features and class templates

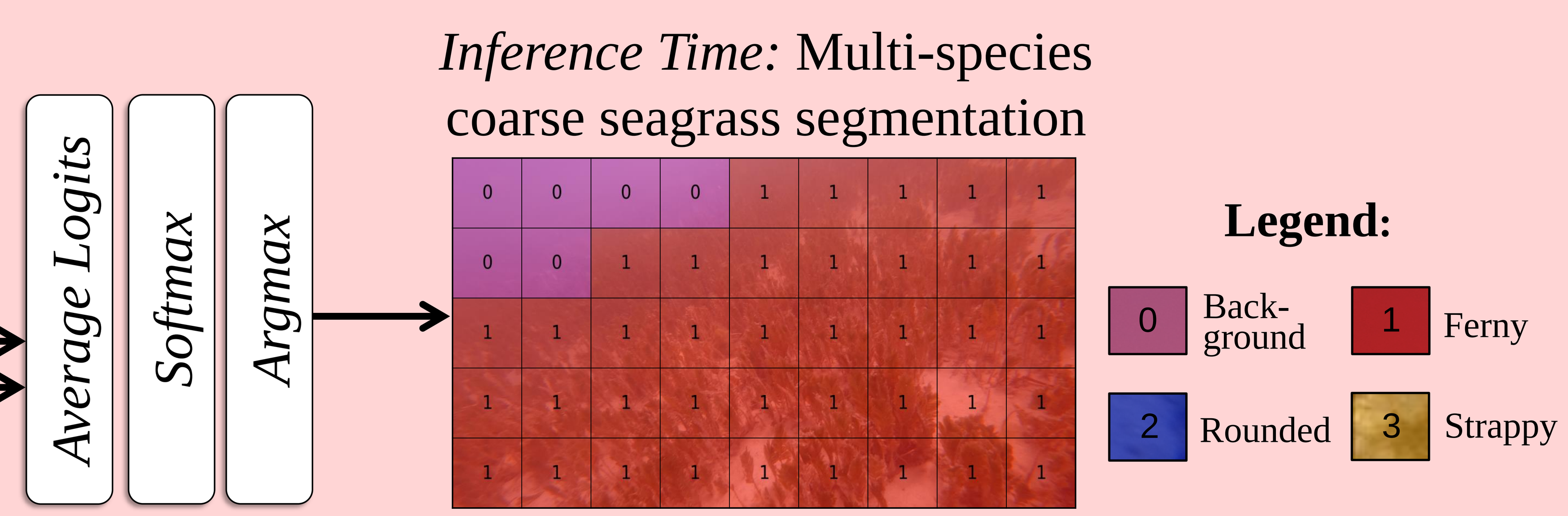


SeaCLIP – *weak supervision from natural language prompts*

- Train ResNet-18 architecture by combining zero-shot pseudo-labels from CLIP with image-level seagrass labels
- CLIP can be used as a supervisory signal in domain-specific applications



Ensemble of Classifiers



- **SeaFeats** is accurate (93.1% F1 score)
- **SeaCLIP** more conservative, less likely to misclassify unclear patches, but lower accuracy overall (87.8% F1 score)
- **Ensemble of SeaFeats and SeaCLIP results in robust performance (95.3% F1 score)**

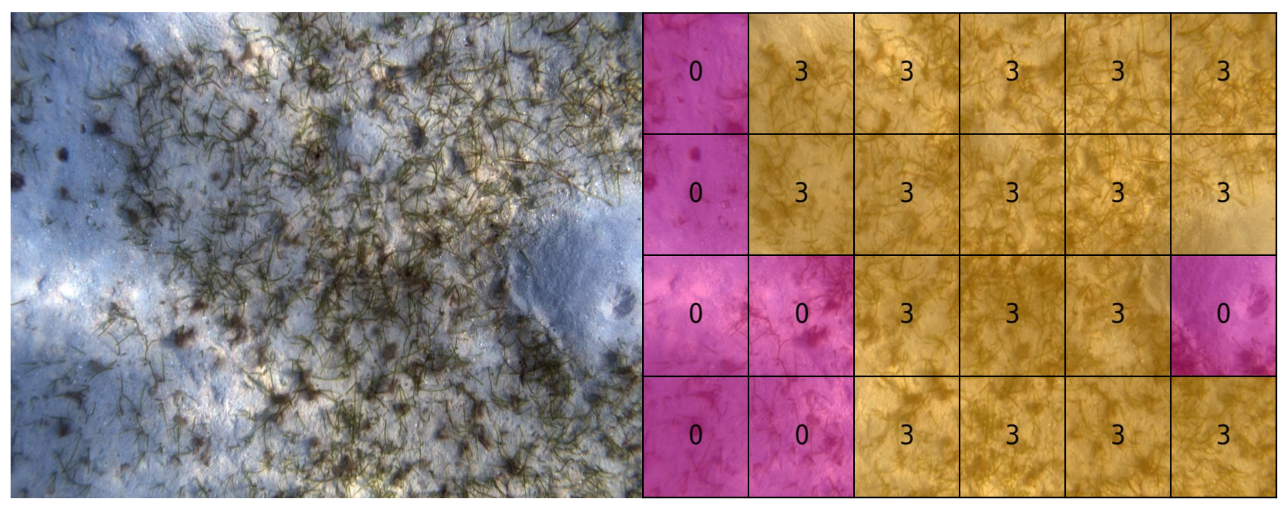
Results

- We outperform the state-of-the-art by **6.8%** (absolute F1 score) on the 'DeepSeagrass' dataset for the multi-species case, and use only **3.9%** of the labels as compared to patch-supervised methods:

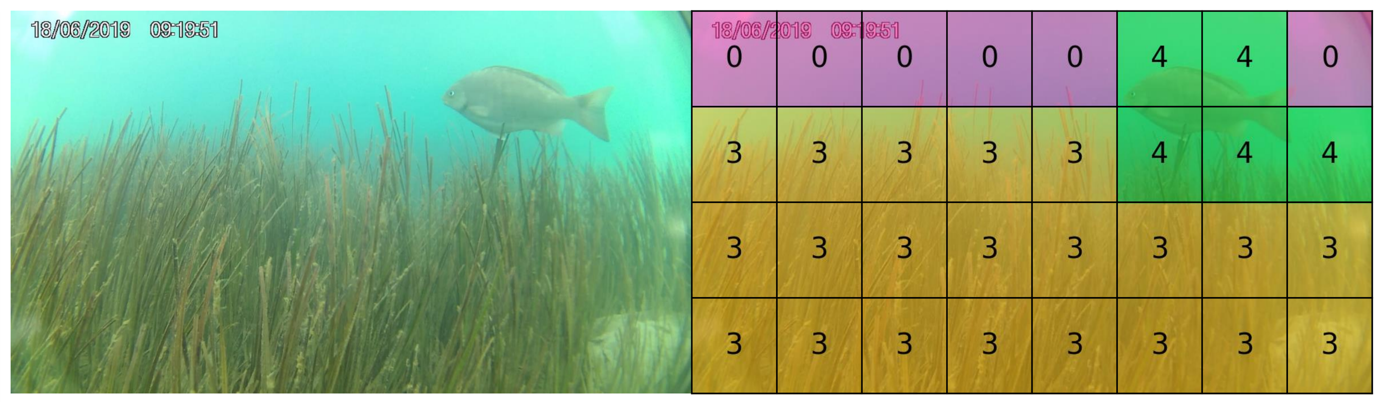
Method	Label Type	Background	Ferny	Rounded	Strappy	Overall
Zero-shot CLIP [33]	Nil	83.41	33.20	42.99	50.76	60.65
SimCLR [2] + Raine et al. [35]	Patch	90.83	65.70	66.96	73.73	77.16
Raine et al. ResNet-50 [35]	Patch	82.94	88.34	91.48	91.56	87.41
Noman et al. EfficientNet-B5 [29]	Patch	84.43	93.18	92.50	86.17	88.52
Ours: SeaFeats	Image	94.93	92.44	89.56	92.12	93.10
Ours: SeaCLIP	Image	86.14	87.82	87.91	91.21	87.84
Ours: Ensemble of SeaFeats + SeaCLIP	Image	**94.97**	**95.42**	**94.69**	**96.29**	**95.33**

Real-world Deployment

- Imagery collected by the 'FloatyBoat' autonomous surface vehicle:
- New data sources containing out of distribution objects:



(a) 'FloatyBoat' Query (b) Inference



(c) 'Global Wetlands' Query (d) Inference

Legend: 0 Back-ground 1 Ferny 2 Rounded 3 Strappy 4 Fish (out of distribution)

Acknowledgements

This work was supported by the Queensland University of Technology (QUT) through the Centre for Robotics, and CSIRO Data61. T.F. acknowledges the ARC Laureate Fellowship FL210100156 and Intel Labs via grant RV3.290.Fischer.