

Human-in-the-Loop Segmentation of Multi-species Coral Imagery

Scarlett Raine^{1,2}, Ross Marchant³, Brano Kusy², Frederic Maire¹, Niko Sünderhauf¹, Tobias Fischer¹

¹QUT Centre For Robotics ²CSIRO Data61 ³Image Analytics

Email: sg.raine@qut.edu.au

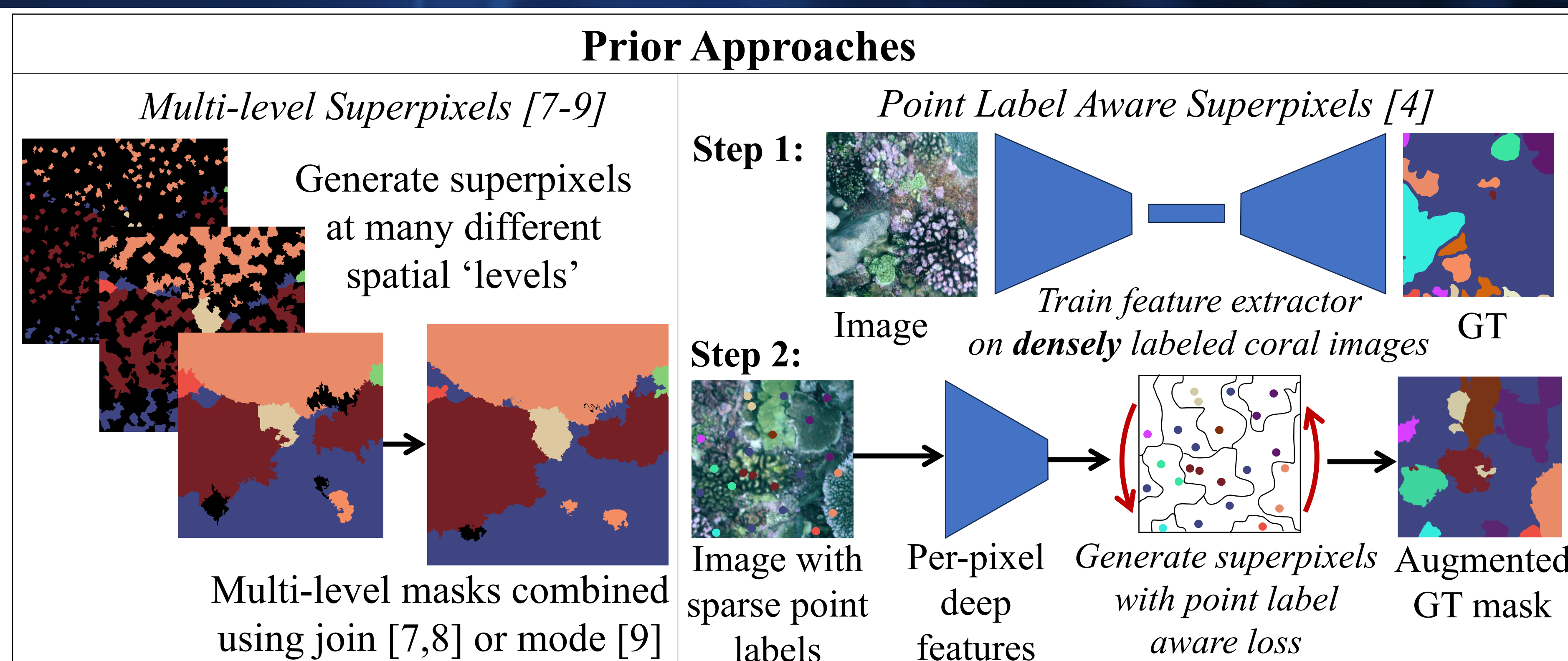
Find our code on GitHub: <https://github.com/sgraine/HIL-coral-segmentation>



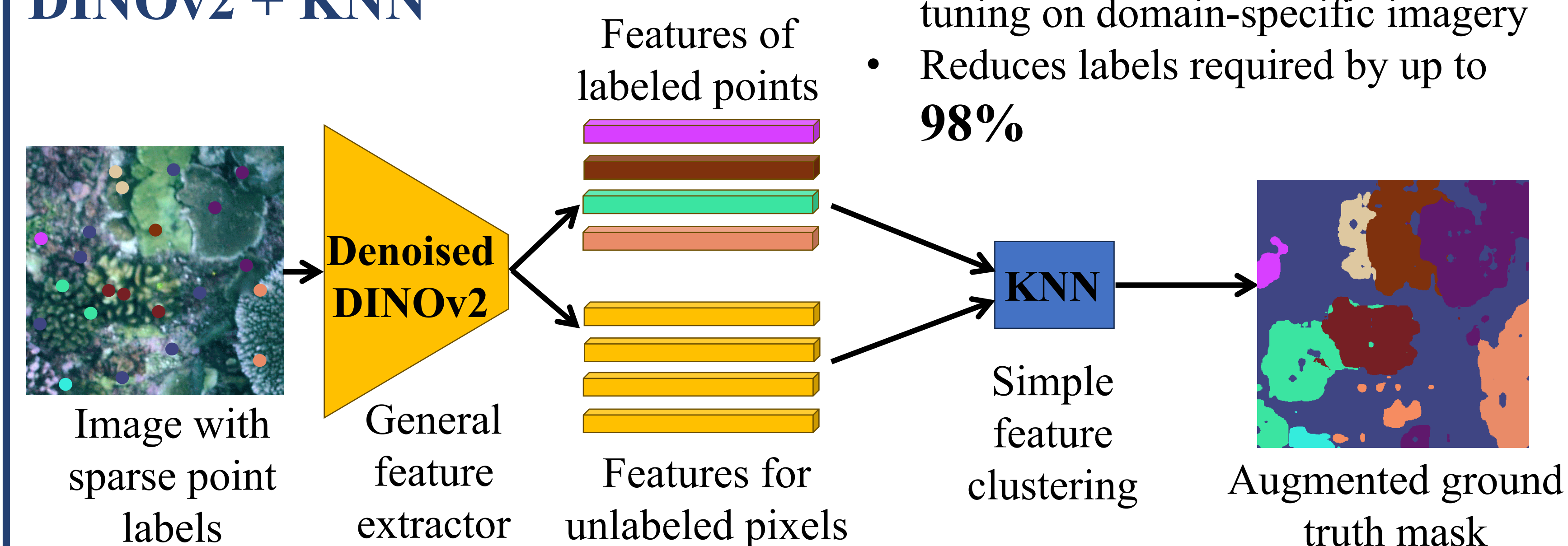
L3D-IVU

Motivation

- Broad-scale marine surveys by underwater vehicles generate large amounts of images which are costly and time-consuming for domain experts to label
- Prior approaches assume 100-300 sparse point labels per image for point label propagation; here we focus on 5-25 points

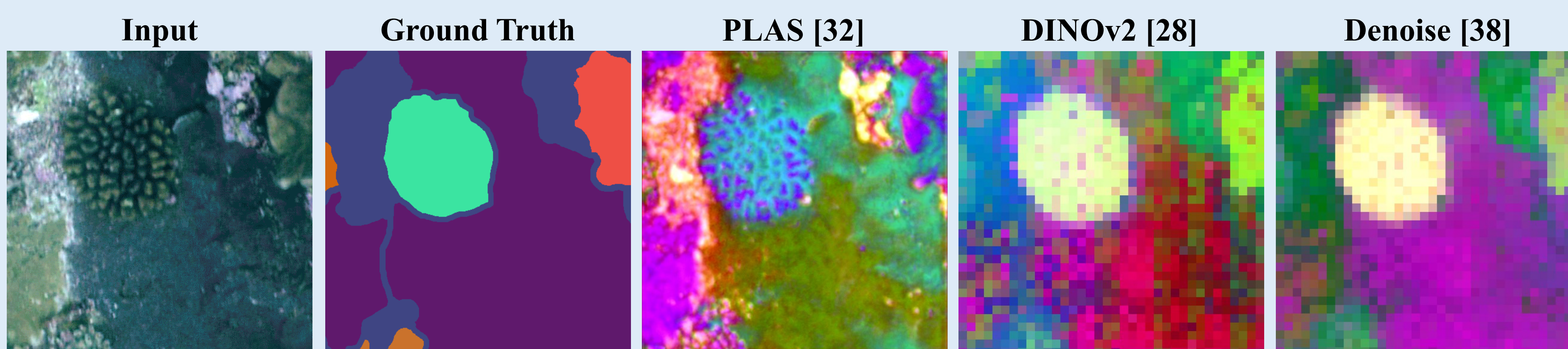


Our Approach: DINOv2 + KNN



Denoised DINOv2 with Clustering in Deep Embedding Space

- Obtain template feature vectors for labeled points from a denoised DINOv2 feature extractor and perform clustering in the deep embedding space to obtain an augmented ground truth mask
- The denoising DINOv2 model groups pixels into meaningful segments and reduces artefacts, resulting in smoother, cleaner features and improved clustering



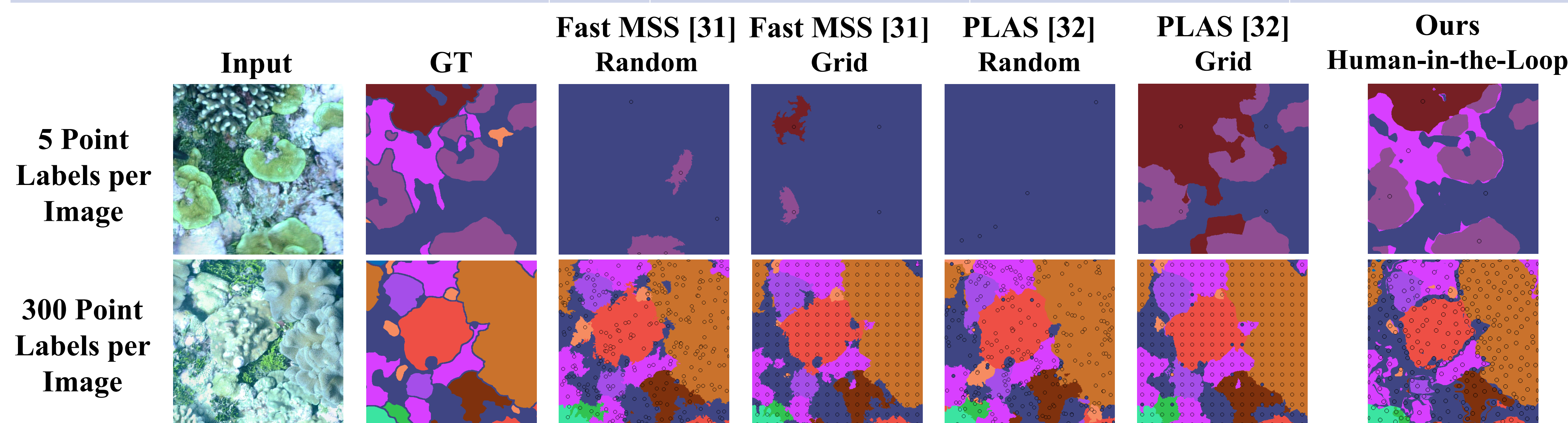
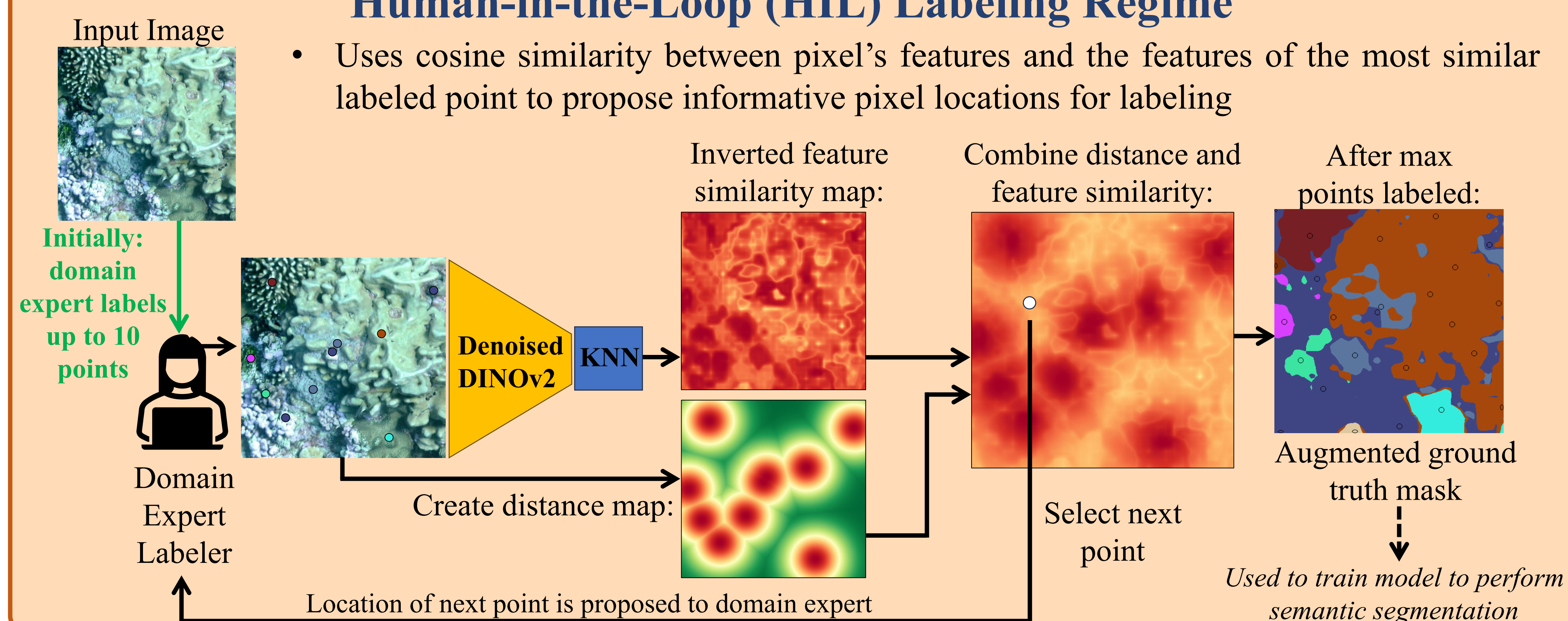
Results

- If only 5 point labels per image are available, we outperform the state-of-the-art by **17.3%** for pixel accuracy and **22.6%** for mIoU on the UCSD Mosaics dataset
- If HIL labeling is not used, we still outperform by **3.5%** for pixel accuracy and **5.7%** for mIoU (5 grid points)
- Grid labels improve point label propagation performance over random labels for all approaches
- For greater than 100 points per image, improvement is marginal with PLAS [32] or our DINOv2+KNN approach
- For very few points per image, it is beneficial to incorporate domain expert knowledge in labeling pixels

Method	Label Style	PA				mIoU				Time per Image (s)			
		5	10	25	300	5	10	25	300	5	10	25	300
Fast MSS [31]	Rand.	7.3	13.5	30.1	86.8	6.6	12.1	28.5	80.1	2.1	2.2	2.2	2.8
Fast MSS [31]	Grid	6.4	13.9	39.2	90.0	5.9	12.8	35.5	86.4	2.4	2.5	2.4	3.0
Point Label Aware Superpixels – Ens. [32]	Rand.	52.7	62.0	71.1	92.5	25.9	35.6	50.5	85.5	4.3	4.6	5.0	5.4
Point Label Aware Superpixels – Ens. [32]	Grid	54.2	65.8	76.3	94.6	30.0	40.3	59.8	89.4	4.1	4.3	5.2	5.3
Ours: Denoised DINOv2 + NN	Rand.	55.7	64.5	75.1	88.8	32.1	42.8	58.0	81.8	4.9	4.6	4.7	4.9
Ours: Denoised DINOv2 + NN	Grid	57.7	69.1	78.7	89.9	35.7	50.6	64.4	85.8	4.8	4.7	4.8	4.7
Ours: Denoised DINOv2 + NN	HIL	71.6	76.4	81.3	89.6	52.6	59.5	68.0	85.0	4.7	5.0	20.0	273.1

Human-in-the-Loop (HIL) Labeling Regime

- Uses cosine similarity between pixel's features and the features of the most similar labeled point to propose informative pixel locations for labeling



Acknowledgements

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