A) Motivation
Our camera, model SONY RX1R II, captures images at a high resolution, but reflections can cause many false detections, quickly overwhelming the downstream processing.

B) Challenges
- Surface reflections
  - Images acquired by the UAV camera capture signals from multiple sources.
  - These reflections are caused by objects other than the water surface and can be classified as specular or diffuse.
- Small objects
  - Water-leaving radiance. Solar elevation of less than 30 degrees results in low water-leaving signals.
  - Inpainting
  - A fast method than SAHI leverages a fine-grained saliency map based on the DINO transformer foundational model and the T2T Vision Transformers.

C) Approach: Detection
- SAA (Surface Albedo Against), which incorporates a method that refines the detection of small objects in underwater images.
- The frequency of observation of different classes varies dramatically, with some common in our data. The long-tail problem in real-world data. A small portion of the classes form the majority of data, while other classes are rare and difficult to detect.

D) Approach: Clustering DINO + HDBSCAN
- Our approach uses a method that combines DINO for object detection and HDBSCAN for clustering. This method has been successfully applied to various datasets, including a real-world dataset of small objects.
- The number of clusters for each class can be automatically determined by the algorithm.

Figure 5. Results SAHI. Figure 6. (left) original image, (right) fine-grained saliency map}

References
- DINO
- HDBSCAN

Accelerating Marine UAV Drone Image Analysis with Sliced Detection and Clustering (MBARI SDCAT)

Monterey Bay Aquarium Research Institute (MBARI), Moss Landing, California USA (duanie@mbari.org)

How to cite

Figure 7. Cluster Algorithm Workflow

Figure 8. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 9. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 10. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 11. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 12. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 13. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 14. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 15. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 16. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 17. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 18. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 19. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 20. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 21. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 22. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 23. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 24. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 25. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 26. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 27. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 28. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 29. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 30. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 31. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 32. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 33. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 34. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 35. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 36. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 37. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 38. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 39. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 40. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 41. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 42. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.

Figure 43. Example saliency maps. The final layer of the DINO H10 IIT (clustering) output for the input image. The final layer is a pixel-level saliency map that is used to guide the object detection process.