UCF CRCV REU – Week 9

By: Felipe Monsalvo

Detection of Coastal Features from Satellite Imagery using Computer Vision

UCF CRCV REU - Project 29

Lead: Harry Newton, Ph.D., Leidos Data Scientist

REU Student: Felipe Monsalvo



Enabling AI to interpret satellite imagery with near-human accuracy



Supports maritime logistics & national safety systems

Cumulative Completion (Weeks 1–9)

- Designed and implemented a detection + segmentation pipeline (Grounding DINO + SAM 2) running on Docker with GPU acceleration
- Integrated prompt engineering, IoU filtering, and dynamic thresholding to improve detection quality
- Drafted a research poster and research paper (final stages)

Completed (This Week)

- Calculated full-pipeline metrics (precision, recall, F1, mIoU) across all test imagery
- Interpreted confusion matrices and error patterns to refine analysis
- Designed and structured the research poster with clear methodology and results
- Integrated graphs, sample outputs, and key findings into the final poster draft

Automated Detection of Coastal Features in Remote Sensing Imagery

Felipe Monsalvo, Mentor: Dr. Harry Newton
University of Central Florida

Center for Research in Computer Vision

Abstract

This poster presents a two-stage computer vision pipeline for detecting and segmenting submerged and semi-submerged coastal hazards in high-resolution RGB satellite imagery. Grounding DINO performs text-prompted open-vocabulary detection, while SAM 2 refines these outputs into precise, polygon-ready masks. Through tailored prompt engineering, dynamic thresholds, and IoU filtering, Grounding DINO achieved 92% recall and 77% precision, delivering highly aligned bounding boxes for irregular targets such as rocks, shipwrecks, and reefs. Guided by these boxes, SAM 2 reached 90% F1-score and 84% mIoU, accurately tracing object boundaries even in challenging scenes. This modular pipeline demonstrates strong spatial accuracy and adaptability, providing a robust foundation for advancing geospatial computer vision in complex maritime environments.

Introduction

Satellite optical imagery is vital for mapping submerged and semi submerged coastal features such as rocks, wrecks, and shifting shorelines that affect navigation safety and environmental monitoring. These features are difficult to detect because of water distortion, camouflage, and noisy imagery. While ship detection in Synthetic Aperture Radar (SAR) imagery has advanced with deep learning, SAR primarily captures metallic objects and misses many coastal hazards. High resolution RGB imagery offers richer detail, but accurately extracting hazards requires specialized processing.

Our work introduces a two-stage pipeline that pairs text prompted detection with instance segmentation. Grounding DINO generates open-vocabulary detections from engineered prompts, and SAM 2 converts them into pixel level masks. With prompt tuning, dynamic thresholds, and IoU based filtering, the pipeline delivers high quality polygon ready outputs and provides a flexible framework for advancing geospatial computer vision and future multimodal research.

Methodology

Two-Stage Pipeline

We built a modular pipeline combining Grounding DINO for text-prompted detection with SAM 2 for pixel-level segmentation. Input imagery is pre-processed, detected, segmented, and post-processed to produce polygon-ready masks.

Detection (Grounding DINO)

- ❑ Open-vocabulary detection using engineered text prompts.
- ❑ Prompts iteratively refined with synonym stacking and context descriptors (e.g., *"submerged rock formation," "isolated rock in water"*).
- ❑ Dynamic thresholds: box confidence ≈ 0.18, text relevance ≈ 0.75.
- ❑ IoU-based filtering and area constraints remove redundant or oversized boxes.

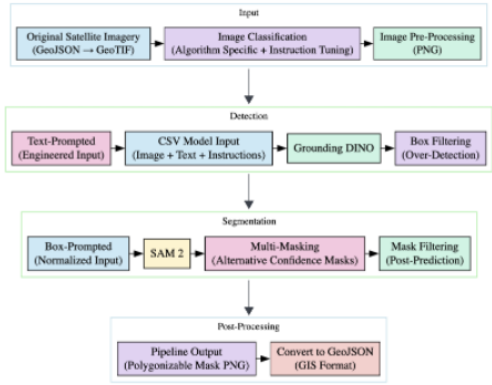
Segmentation (SAM 2)

- ❑ Filtered bounding boxes fed directly as box prompts.
- ❑ Multi-mask output enabled (3 per box), for highest-to-low-confidence mask selection.
- ❑ Masks exported as PNG and optionally converted to GeoTIFF for GIS integration.

Implementation

- ❑ Pipeline run in Docker on WSL 2 (Ubuntu 20.04) with CUDA 11.8 and Python 3.10 for GPU-accelerated processing.

Model



Results & Discussion

Detection (Grounding DINO)

- ❑ High recall indicates most hazards were detected, while moderate precision reflects some over-detection.
- ❑ False positives often came from shadows, algae patterns, or sand features that mimicked hazards.
- ❑ Strong spatial alignment between predictions and ground truth is reflected in the high mIoU.
- ❑ **Figure 1**: Grounding DINO Evaluation Metrics Graph

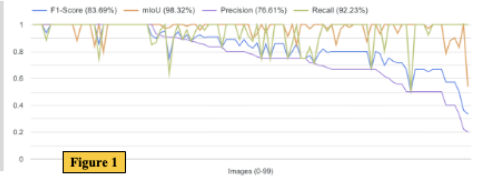


Figure 1

Segmentation (SAM 2)

- ❑ High F1 and mIoU scores show the masks closely matched object boundaries.
- ❑ Accurate bounding boxes led to clean segmentation, while weak or incomplete boxes caused most misses.
- ❑ Performance dipped in low-contrast regions where edges were difficult to distinguish.
- ❑ **Figure 2**: SAM 2 Evaluation Metrics Graph

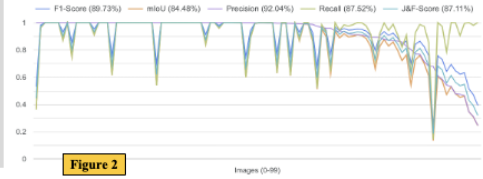
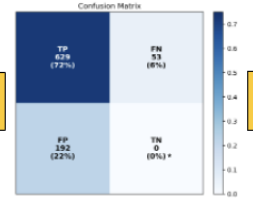
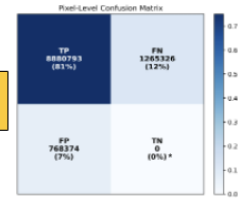


Figure 2

Grounding DINO Box Prediction Confusion Matrix



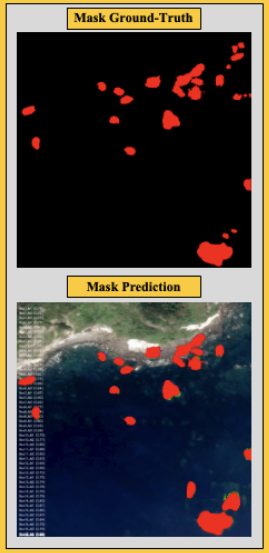
SAM 2 Mask Prediction Confusion Matrix



*TN not applicable: detection and segmentation is evaluated only on annotated objects, not on background pixels or non-target objects of interest.







Ablation Study

Detection Prompt Sensitivity (Grounding DINO)

We evaluated different text prompt types for Grounding DINO. Location/Context aware prompts (e.g. "rock near brownish coastline right side of image") produced the highest **mean Average Precision (mAP%)**. Outperforming both generic (e.g., "rock") and domain-specific (e.g., "isolated submerged rocks") prompts by capturing additional spatial cues in imagery.

Prompt Type	mAP (%)
Generic	74%
Domain-Specific	77%
Location/Context Aware	81%

IoU-Based Filtering and Threshold Tuning

We analyzed detection thresholds and IoU filtering. A box confidence threshold around 0.18 balanced precision and recall, while IoU-based suppression reduced overlapping boxes and false positives. Stricter thresholds increased precision but missed smaller targets.

Box-Threshold	mAP (%)	False Positives
0.13 (Low)	79%	High
0.18 (Optimal)	77%	Moderate
0.23 (Strict)	71%	Low

Chart 3 – Segmentation Prompting and Mask Selection (SAM 2)

We compared mask selection strategies in SAM 2. Selecting the best mask from three candidates (Best-of-N) achieved the highest mIoU, especially in images with overlapping or complex structures, outperforming single-mask or averaged results.

Mask Strategy	mIoU (%)
Top 1	78%
Avg Masks	82%
Best-of-N (N=3)	85%

Conclusion

Contribution

We present a two-stage computer vision pipeline that detects, and segments submerged, and semi submerged maritime hazards in high resolution RGB satellite imagery. Grounding DINO provides text prompted open-vocabulary detection, and SAM 2 refines these detections into instance masks.

Key Findings

The pipeline offers strong control and interpretability for localizing hazards, enabling detailed evaluation of prompt engineering, dynamic threshold tuning, and IoU based filtering. SAM 2's multi mask selection further improves accuracy, producing high quality polygon ready outputs.

Future Work

Multi-Modal Integration

We plan to integrate SAR and hyperspectral data with RGB imagery to improve detection under cloud cover, low light, and challenging conditions. SAR adds all-weather imaging, while hyperspectral data provides richer material and habitat information.

References

[1] S. Liu et al., "Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection," *arXiv preprint*, arXiv:2303.05499, 2023. https://arxiv.org/abs/2303.05499

[2] N. Ravi et al., "SAM 2: Segment anything in images and videos," *arXiv preprint*, arXiv:2408.00714, 2024. https://arxiv.org/abs/2408.00714

Acknowledgements

• This material is based upon work supported by the National Science Foundation Award

In Progress

Current Steps:

- Revise and fix the paper report draft for clarity and completeness
- Refine analysis sections to ensure all metrics and results are accurately represented

Next This Week:

- Prepare and practice the 90-second presentation
- Bridge remaining gaps for deliverables, aligning results, visuals, and written analysis

Schedule

Scheduled Start	Scheduled Finish	Actual Finish	Status	Tasks
7/3	7/10	7/10	Completed	Evaluate the performance of the previous pipeline using semantic segmentation to assess accuracy and consistency across key features.
7/10	7/17	7/17	Completed	Focus efforts on a standardized benchmark dataset featuring accessible data to facilitate comparisons and transparent analysis.
7/17	7/21	7/21	Completed	Begin structuring the experimental results and quantitative analyses, including PNG-based mask outputs used for IoU evaluation.
7/22	7/24	7/24	Completed	Posters due for printing.
7/24	7/31		To Do	Bridge remaining gaps between poster preparation and final report drafting, ensuring alignment of results, visuals, and written analysis.

Thank You