WARD: A Vision-Only Framework for Water-Surface Awareness and Risk Detection

Yiran Zhou, Zheng Wang, Tianai Yan, Sacchin Sundar, Katherine A. Skinner
The Robotics Department, University of Michigan, Ann Arbor, USA

ROBOTICS UNIVERSITY OF MICHIGAN

IEEE Oceanic Engineering Society · OCEANS 2026 Monterey, CA September 21-24

General Poster

Motivation

Low-Cost Autonomous Surface Vehicles (ASV) Constraint

- Small-scale ASVs demand low-cost, vision-only solutions without active radar or LiDAR.

Perception ≠ Safety

- High perception accuracy does not inherently translate into safe navigation.
- Raw detections treat distant buoys and immediate hazards as equally salient.
- The fundamental question: "***Does this obstacle pose an immediate collision risk?***"

Our Work

- WARD fuses **SegFormer segmentation**[1] with **Depth Anything V2 monocular depth**[2].
- An **extended SORT tracker** ensures stable Time-to-Collision predictions.
- A **continuous collision-corridor risk model** reasons jointly about proximity, lateral offset, and semantic class.
- Operating real-time on a consumer-grade GPU, WARD provides discrete safety warnings (**SAFE, CAUTION, IMMEDIATE**) for resource-constrained ASVs.

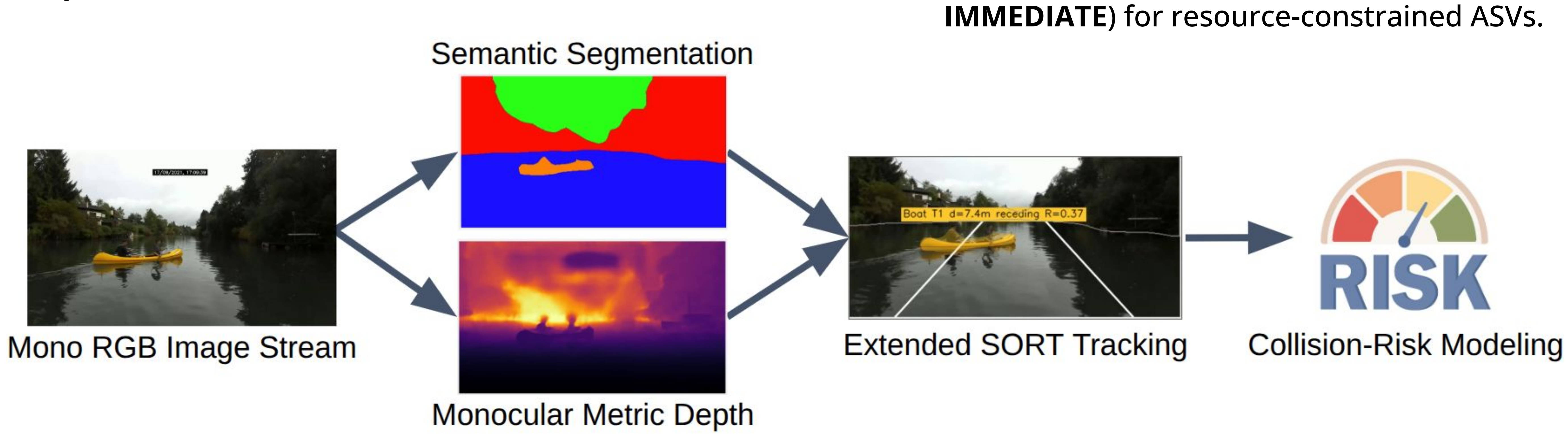

Fig. 1. WARD pipeline from monocular RGB input to actionable collision-risk warnings.

End-to-End Results

Scan for video!

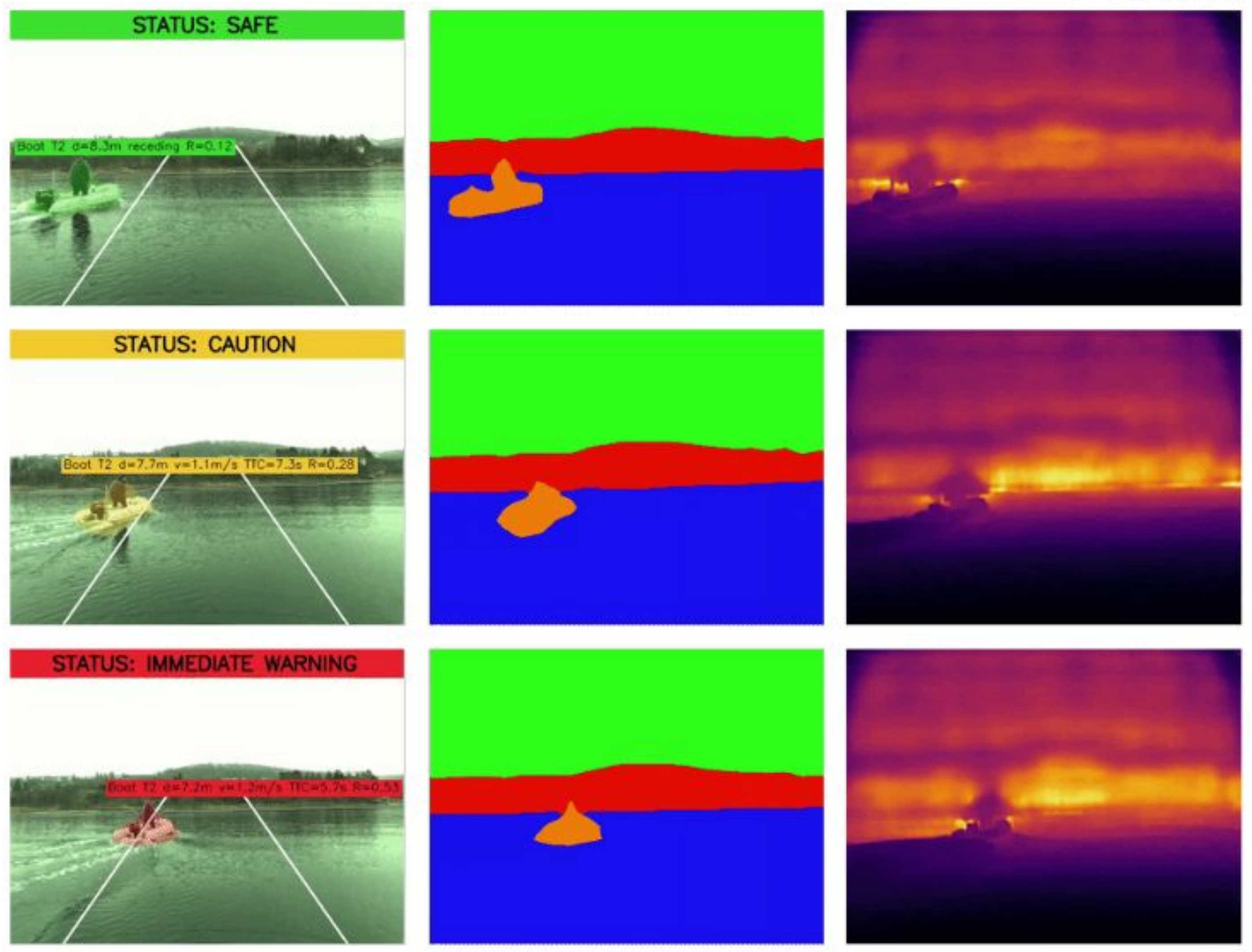

Fig. 2. Qualitative results: WARD escalates warnings as tracked obstacles approach.

On an NVIDIA RTX 4060 GPU:	Average Latency (ms)
Semantic Segmentation (SegFormer)	26.9
Depth Estimation (Depth Anything V2)	38.7
Temporal Fusion & Risk Modeling	7.4
Total	**73.0 (~13.7 FPS)**

Table 3. WARD end-to-end latency on an NVIDIA RTX 4060 GPU.

1. Semantic Segmentation

Method:

- **SegFormer (MiT-B0)**[1] dual-head design on LaRS[3]: a 9-class semantic head and an auxiliary boundary head (τ_b = *0.4*) to split adjacent instances.
- Connected components < 300 pixels are discarded; instance depth is assigned as the **5th percentile of pixel depth** for conservative nearest-point estimation.

Results:

Model	DeepLab[4]	**SegFormer**
Obstacle Recall ↑	0.7983	**0.9651**
Obstacle Prevision ↑	0.9684	**0.9737**
Average Latency (ms) ↓	15.4	**26.9**

Table 1. Segmentation performance of DeepLab and SegFormer on LaRS validation set.

2. Monocular Metric Depth

Method:

- **Depth Anything V2**[2] **(Metric-Outdoor Base variant)** pre-trained foundation model.
- A single offline scaling factor calibrated once against WaterScenes[5] 4D radar ground truth: $d_{cal} = s_{cal}\, d_{raw}$, $s_{cal} = 0.4$ (d_{raw} is raw output from the model, d_{cal} is calibrated metric depth)

Results:

Model	Small	**Base**	Large
MAE (m) ↓	4.434	**4.039**	3.215
RMSE (m) ↓	5.719	**5.363**	4.471
Est/GT (→1.0)	0.787	**0.887**	1.008

Table 2. Depth Anything V2 Small/Base/Large performance within 0–20 m on WaterScenes dataset.

3. Extended SORT Tracking

Method:

- **Extends SORT**[6] to a 9-D Kalman state that tracks bounding box center (c_x, c_y), scale (area) s, aspect ratio r, calibrated metric depth d, and their respective rate of change represented by dotted variables:
- Directly computes **closing velocity** $v_{\text{closing}} = -\dot{d} \cdot f_{\text{fps}}$ and **Time-to-Collision** $TTC = d / v_{\text{closing}}$ without post-hoc exponential moving averages or windowing.

$$\mathbf{x} = [c_x, c_y, s, r, d, \dot{c}_x, \dot{c}_y, \dot{s}, \dot{d}]^\top$$

4. Collision-Risk Modeling

Method:

- Defines a **virtual forward corridor** of width $W_{boat} + 2\,M_{lat}$ (W_{boat} is the ASV beam width and M_{lat} is the lateral safety margin)
- Combines semantic weight w_{class}, forward distance d_{fwd}, forward decay constant D_{safe}, lateral excess e_{lat}, lateral decay constant L_{safe}, amplification factor α_v, closing velocity $v_{closing}$, and reference speed V_{ref} to get continuous risk score:

$$R = w_{class} \exp(-\, d_{fwd}/D_{safe}) \exp(-\, e_{lat}/L_{safe})\,(1 + \alpha_v \min(v_{closing}/V_{ref}, 2))$$

- Maps continuous risk score to discrete, actionable warnings:
  - 🟩 **SAFE** (R < 0.15)
  - 🟨 **CAUTION** (0.15 ≤ R < 0.50)
  - 🟥 **IMMEDIATE WARNING** (R ≥ 0.50)

Conclusion & Future Work

WARD is a lightweight, vision-only framework that bridges raw perception and actionable risk assessment for low-cost ASVs.

Future Work:

- **Field Deployment**: Validate on physical ASVs against vessel heave and pitch.
- **Multi-Camera Setup**: Extend to 360° coverage with cross-camera tracking.
- **Closed-Loop Control**: Feed continuous risk scores directly into autonomous path planning and avoidance maneuvers.

References & Acknowledgement

[1] E. Xie et al., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," NeurIPS, 2021.
[2] L. Yang et al., "Depth Anything V2," NeurIPS, 2024.
[3] L. Žust et al., "LaRS: A Diverse Panoptic Maritime Obstacle Detection Dataset and Benchmark," ICCV, 2023.
[4] L.-C. Chen et al., "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., 2018.
[5] S. Yao et al., "WaterScenes: A Multi-Task 4D Radar-Camera Fusion Dataset and Benchmarks for Autonomous Driving on Water Surfaces," IEEE Trans. Intell. Transp. Syst., 2024.
[6] A. Bewley et al., "Simple Online and Realtime Tracking," ICIP, 2016.

This work was completed as part of the final project for ROB 572 Marine Robotics at the University of Michigan, Winter 2026.