MassMIND: A semantically segmented labeled dataset of long wave infrared images in marine environment

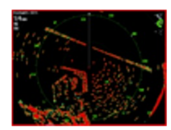
SHAILESH NIRGUDKAR, MICHAEL DEFILLIPO, MICHAEL SACARNY, MICHAEL BENJAMIN, PAUL ROBINETTE

11 AUGUST 2022

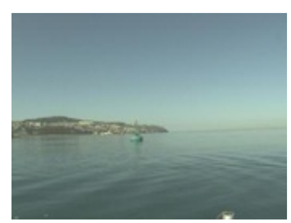
MOOS-DAWG 2022

SENSORS IN MARINE ENVIRONMENT

- Surface vehicles can become autonomous only if they have awareness of surroundings
  - Other boats, bridges, buoys etc.
- Radar as a traditional sensor
  - Slow refresh rate
  - Low fidelity

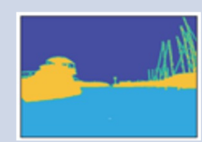


- Optical cameras as a preferred choice
  - Small form factor
  - Excellent results with deep learning
  - Great resolution
  - Limitations
    - Glitter, reflections deteriorate image quality
    - Cannot be used in bright sunlight, low light environment



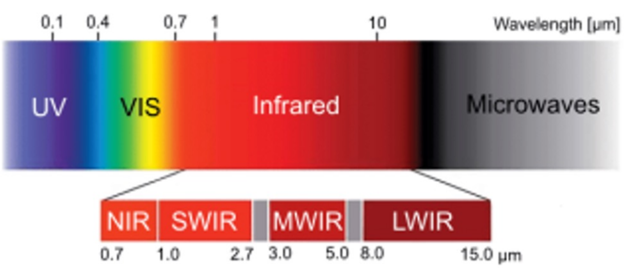
*Credit: Images from MaSTR1325 dataset**

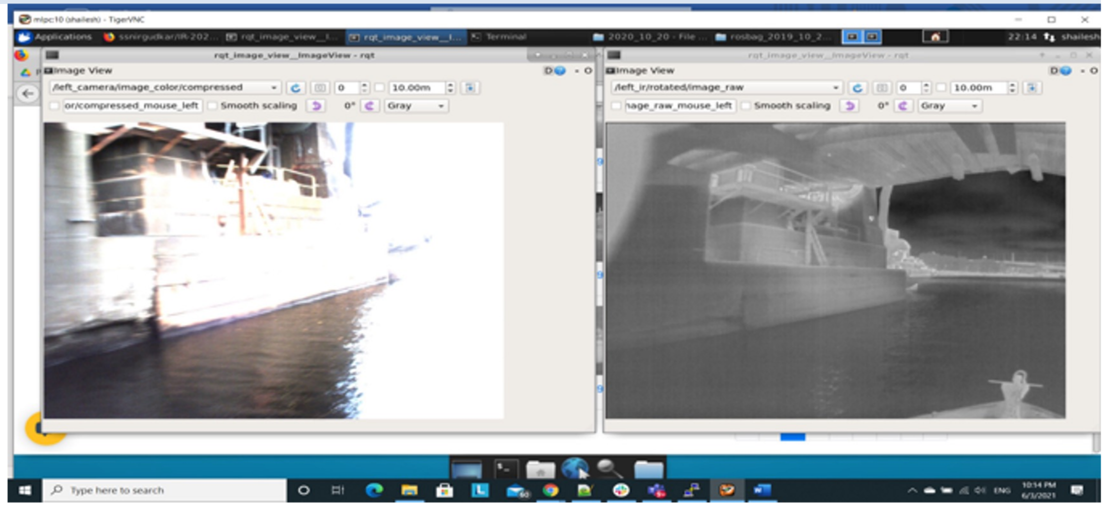
RELATED RESEARCH

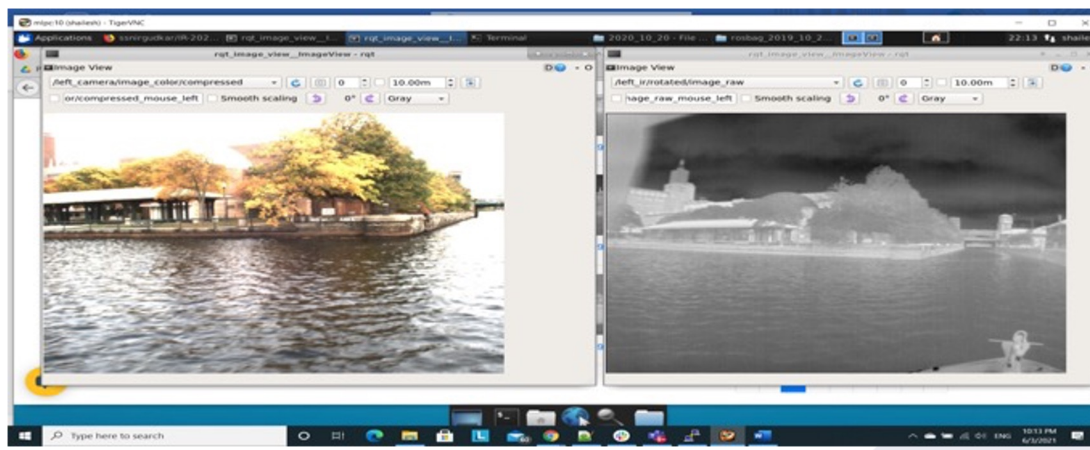
Authors	Title	Brief	
D. D. Bloisi, L. Iocchi, A. Pennisi and L. Tombolini	ARGOS-Venice Boat Classification," *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2015, pp. 1-6	• Boat classification dataset (each image: 800 x 240 pixels) • 14 survey cells covering Grand Canal of Venice • 24 specific categories + water	 
M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf and C. Kanan	VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015	• 1000 paired RGB and infrared images • 6 ship categories • Sensors installed on the pier	 
B. Bovcon, J. Muhovič, J. Perš and M. Kristan	The MaSTr1325 dataset for training deep USV obstacle detection models," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019	• Semantic segmentation of optical images • 3 classes : Sky, water and obstacle • 1325 images, 1278 x 958 pixels	 

COMPARISON OF OPTICAL IMAGES WITH LWIR

- Water dynamics
- Reflection and sun glitter
- Weather conditions







PROBLEM

- The publicly available datasets are mostly created with optical images which has limitations in marine environment.
- The VAIS dataset captured the infrared images from a static viewpoint and has only boat classification labels.

CURRENT RESEARCH

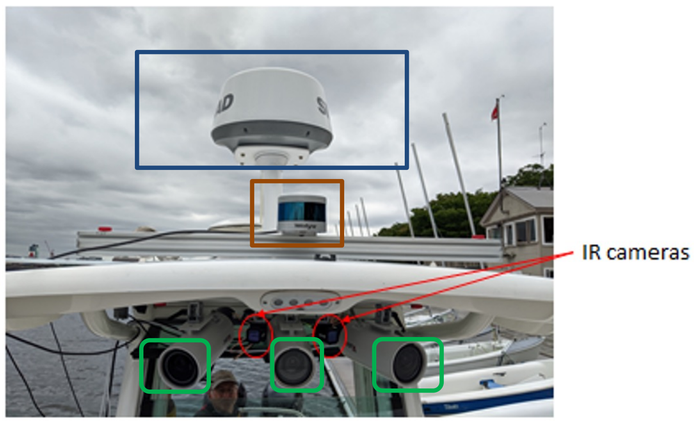
- There is a need to create a dataset which is not impacted by marine environment.

- To be useful for surface vehicle autonomy
  - The dataset should capture real life images from different viewpoints
  - The labeling information should be comprehensive such as semantic segmentation

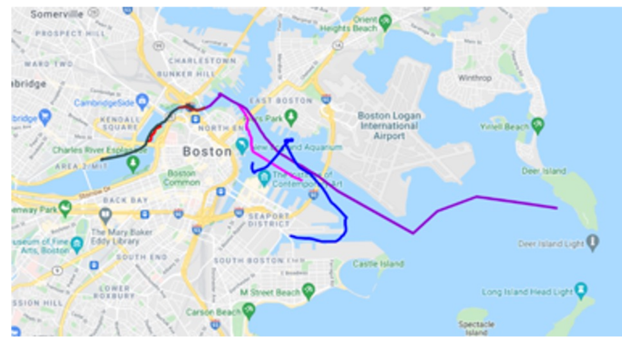
- MassMIND:
  - Around 2900 LWIR images, semantically segmented into 7 classes.
  - Raw images captured over a period of 2 years in and around Boston harbor.

ASV SETUP

R/V Philos



VLP-16 Velodyne 3D Lidar
4G Broadband radar
3 FLIR 1.3 MP cameras
2 FLIR ADK 0.3 MP infrared cameras

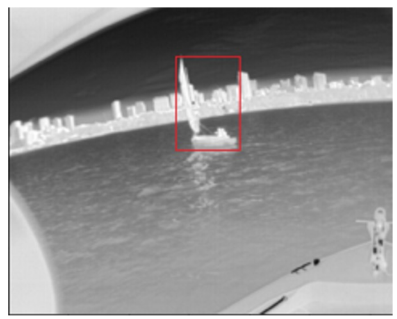


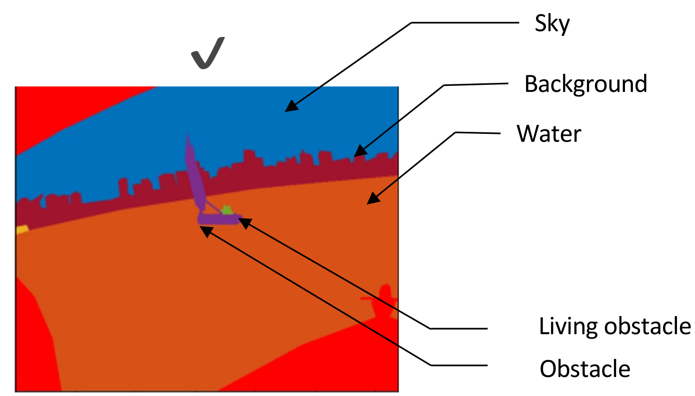
Video recordings were done over a period of 2 years (2019-2021)

Credit: MIT AUV Lab

CHALLENGES FACED IN DATASET CREATION

- What kind of label?

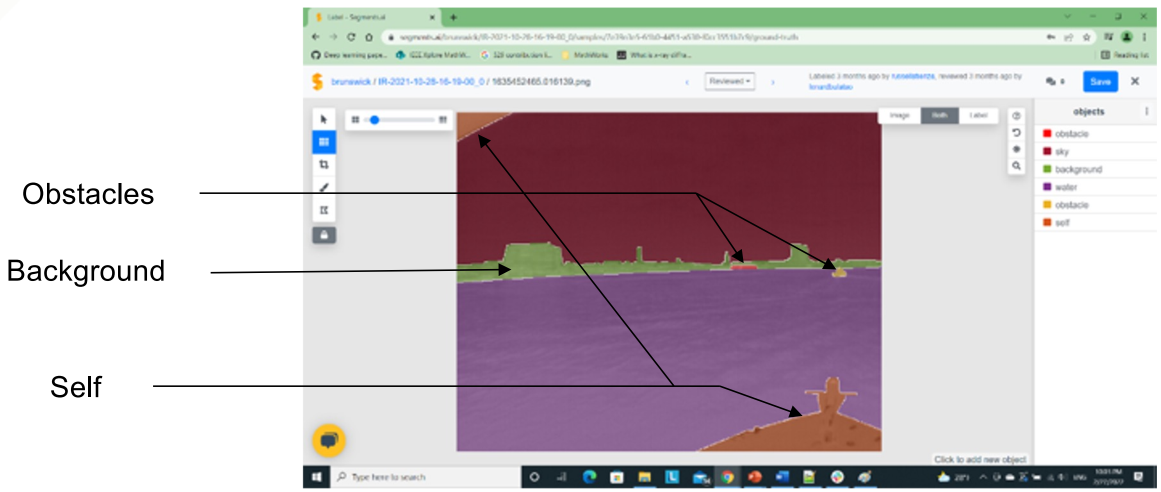




Segmentation mask contains much more information

LABELING PROCESS

- 7 classes – Sky, water, background, bridge, obstacle, living obstacle and self
- Annotation tool



segments.ai

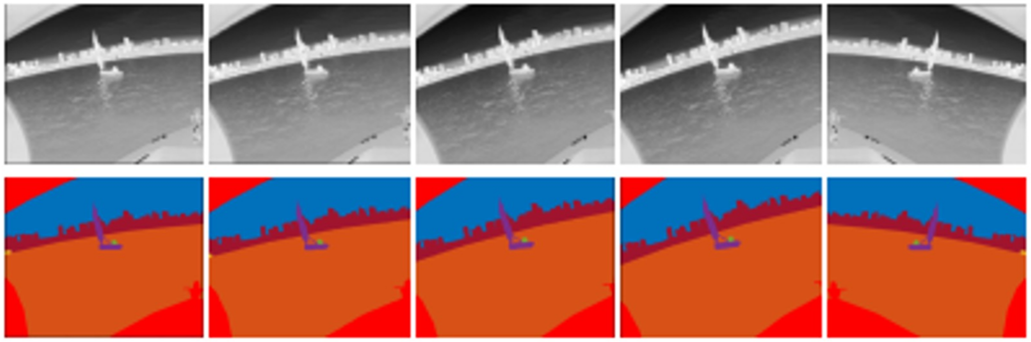
VERIFICATION OF LABELING

- Manual
- Programmatic
  - Check if each pixel is assigned a class id

  - Incorrect label ids (off by 1 difference in API)

DATA AUGMENTATION

- Around 2900 images were labeled
- Dataset is increased by augmentation
  - Rotation and mirror



DATASET METRIC

Spread of classes in MassMIND

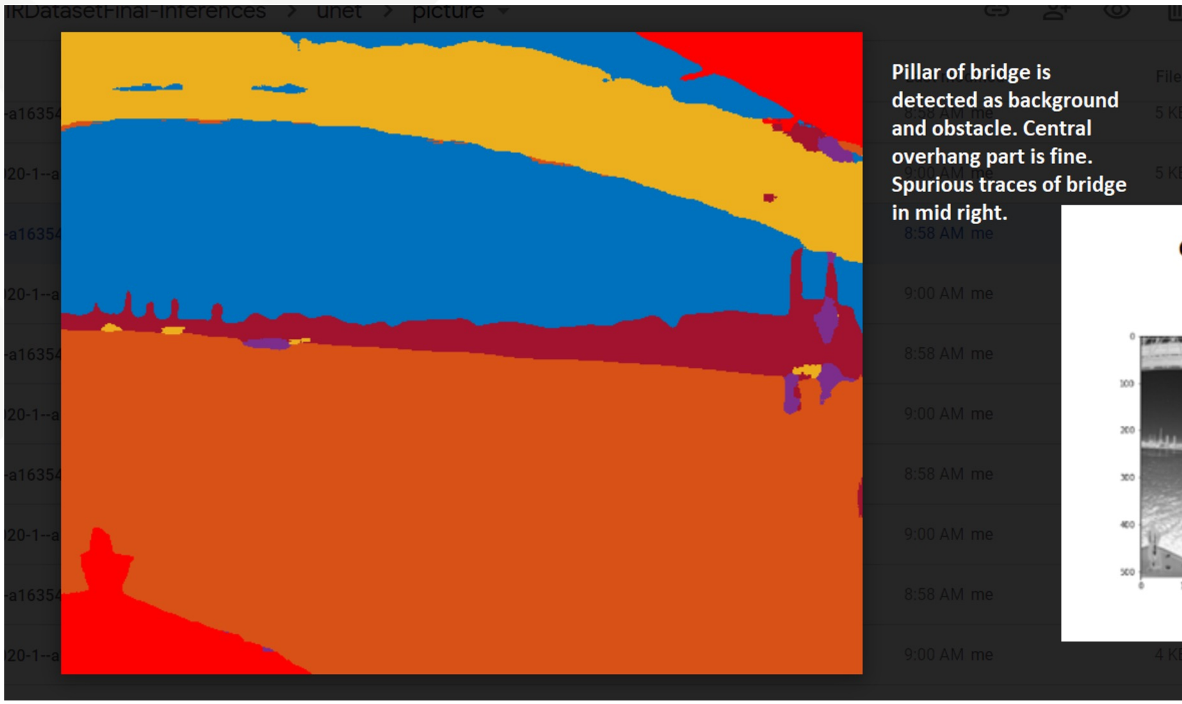
Class	Total number of instances	% distribution of pixel area
Sky	2902	30.58
Water	2916	52.21
Bridge	715	1.67
Obstacles	7120	0.94
Living obstacles	4350	0.05
Background	2860	11.28
Self	1501	3.25

Though obstacles occupied only 1% of area overall, it was adequate to run inference

DEEP LEARNING ARCHITECTURES

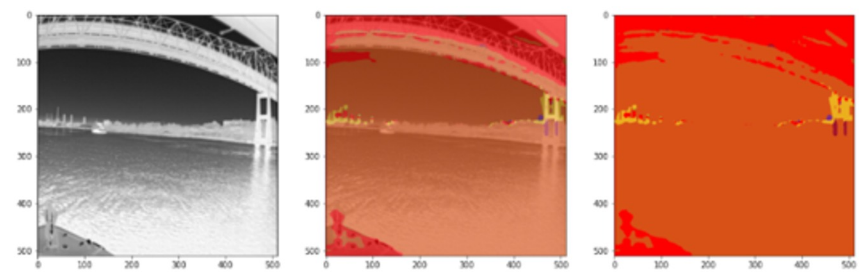
- Unet: One of the first segmentation architectures, simple yet powerful
- PSPNet: Uses global context to predict local regions, better performance
- DeepLabv3: Most advanced, best performance

EARLIER RESULTS: NO SHUFFLING



UNet

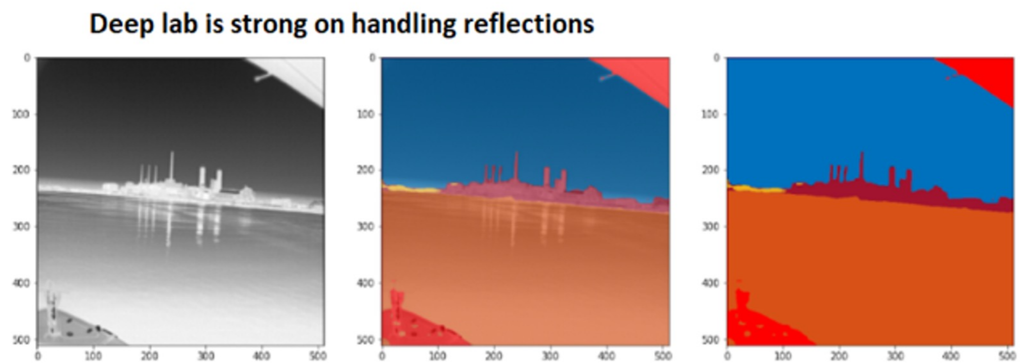
Completely botched up result. Sky, water, self and bridge are mixed



Deeplab

EARLIER RESULTS: NO SHUFFLING

UNet



Deeplab

IMPROVING TRAINING OUTCOME

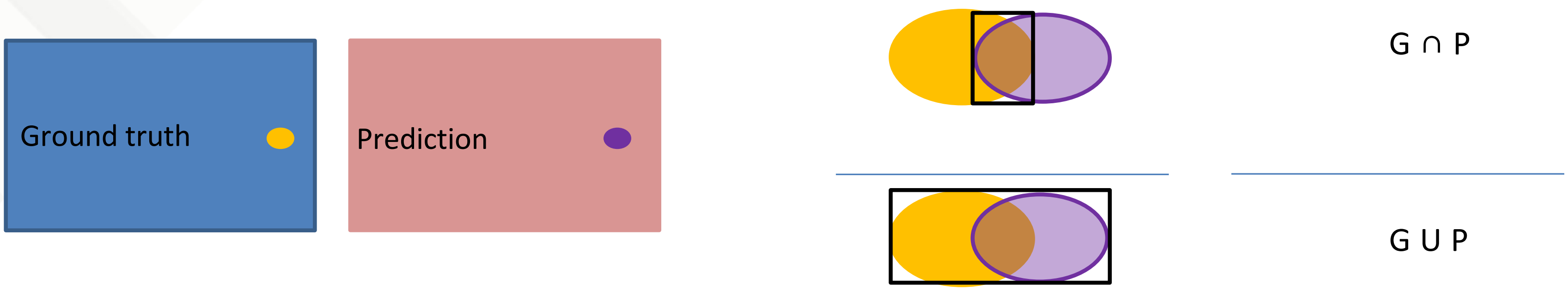
- Images were chronologically fed to the classifiers
  - Shuffled the order

- Rotation of masks had to be done in a certain way to avoid creation of incorrect class labels.

- Smaller size images were used earlier resulting in loss of small obstacles.
  - Original image size (640 x 580) was used

MEASURING THE ACCURACY OF SEGMENTATION

Intersection over Union



For multi-class classification, it is best to consider each class separately.

DATASET EVALUATION

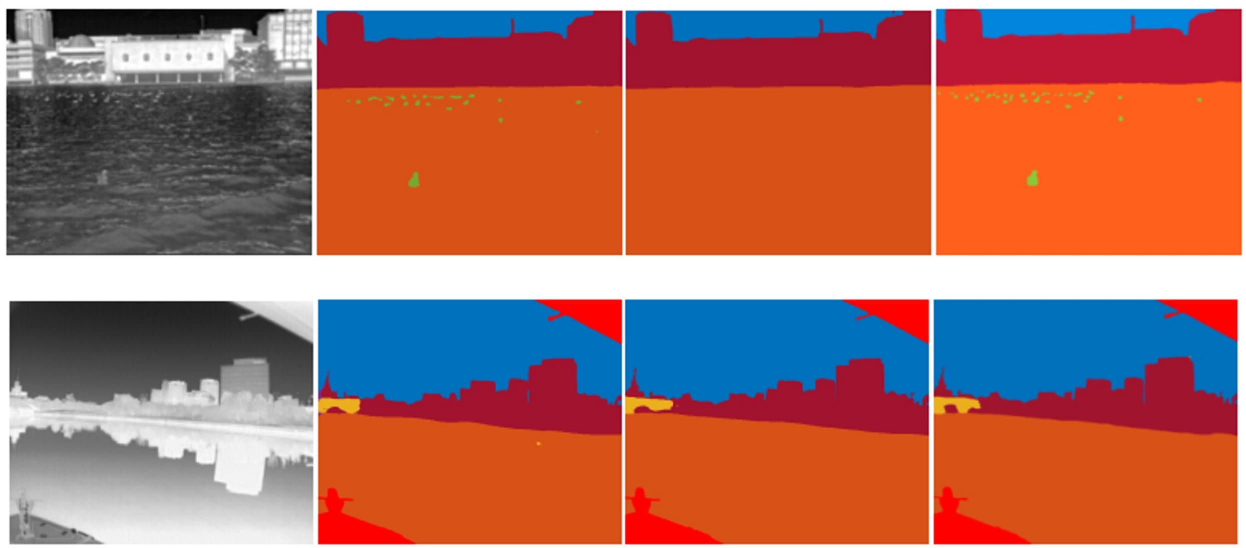
For class 'obstacle'

	Threshold	Precision (%)	Recall (%)	F1
UNet	0.6	55.0	21.9	31.3
	0.3	76.5	58.6	66.4
DeepLabv3	0.6	72.7	40.7	52.2
	0.3	82.7	72.8	77.4
PSPNet	0.6	73.3	41.8	53.3
	0.3	82.9	73.6	78.0

Precision: #true positives/total elements labeled as positive
Recall: #true positives/total elements actually belonging to the positive class

INFERENCE

UNet PSPNet DeepLabv3



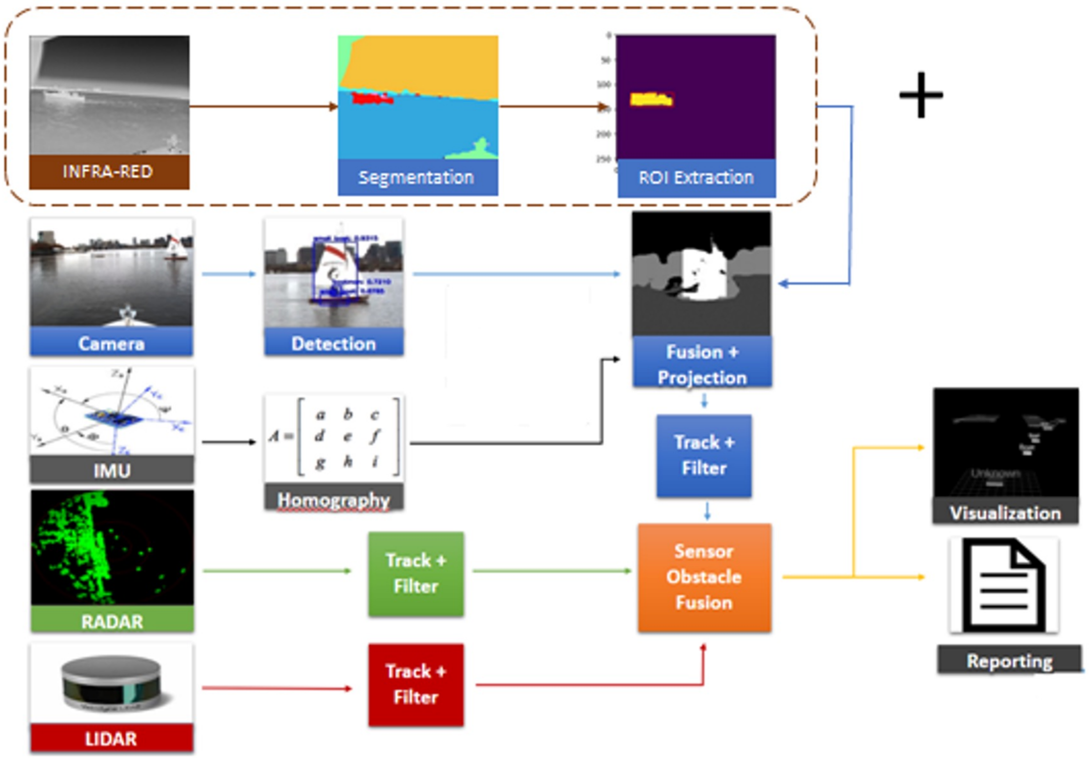
Birds, reflection are handled well

Dataset is published on Github!

- https://github.com/uml-marine-robotics/MassMIND*
- The dataset paper is currently under review

*Reference: https://seagrant.mit.edu/auvlab-datasets-marine-perception-2-3/

PERCEPTION TRACKING*



* ICRA 2022 robotics perception and mapping workshop

PERCEPTION TRACKING: DEMO

Next steps

- Auto calibration between camera sensors and radar
- Integration with MOOS

Acknowledgement

This work is partially sponsored by the Brunswick Corporation