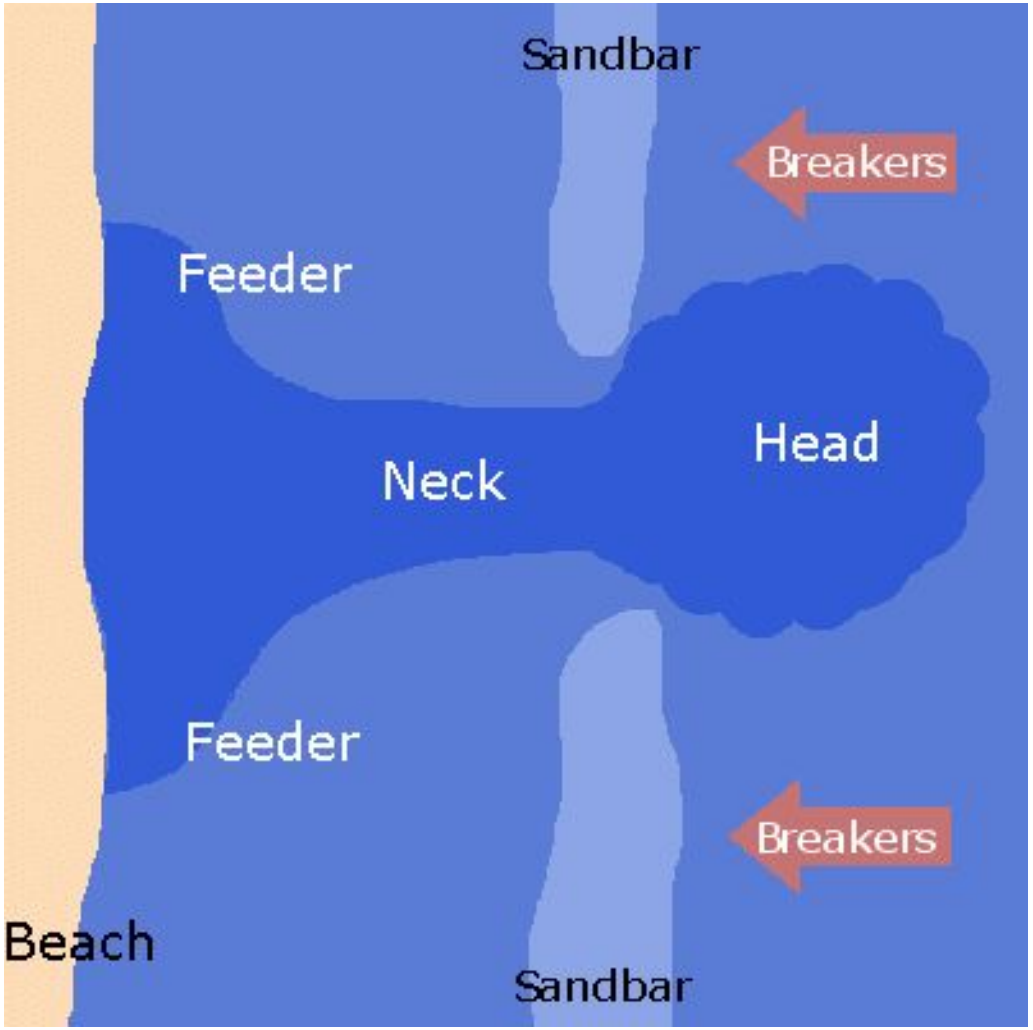
RipVIS: Rip Currents Video Instance Segmentation Benchmark for Beach Monitoring and Safety

Andrei Dumitriu[1, 2], Florin Tatui[2], Florin Miron[2], Aakash Ralhan[1],
Radu Tudor Ionescu[2], Radu Timofte[1]
[1]Computer Vision Lab, CAIDAS & IFI, University of Würzburg, Germany
[2]University of Bucharest, Romania

CVPR 2025
11th-15th of June
Nashville, Tennessee, USA

What is a rip current?

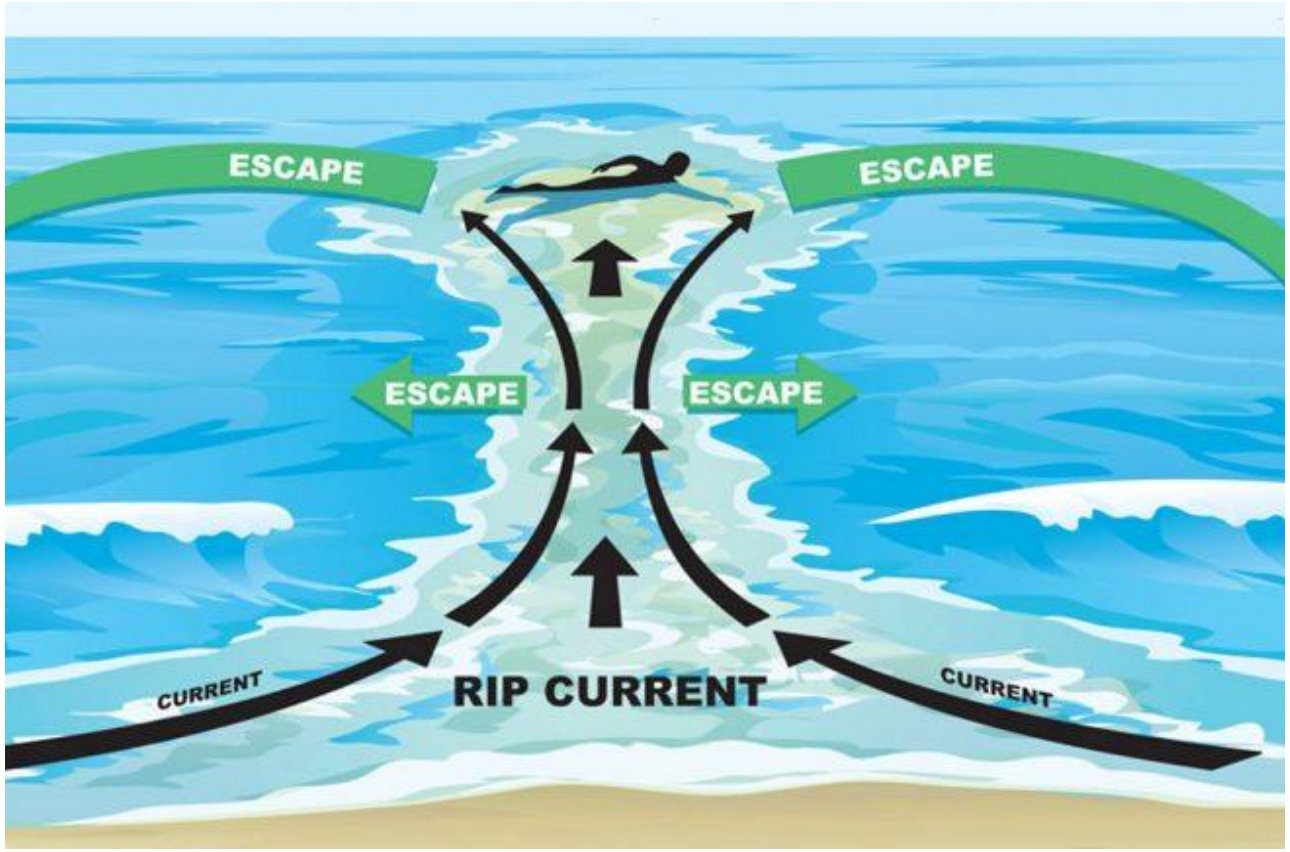
- Strong, narrow flow of water moving away from the shore
- Can flow up to 2.5 m/s (8 ft/s) which is faster than an olympic swimmer



Rip current illustration showing how a rip current works.
Source: https://en.wikipedia.org/wiki/Rip_current

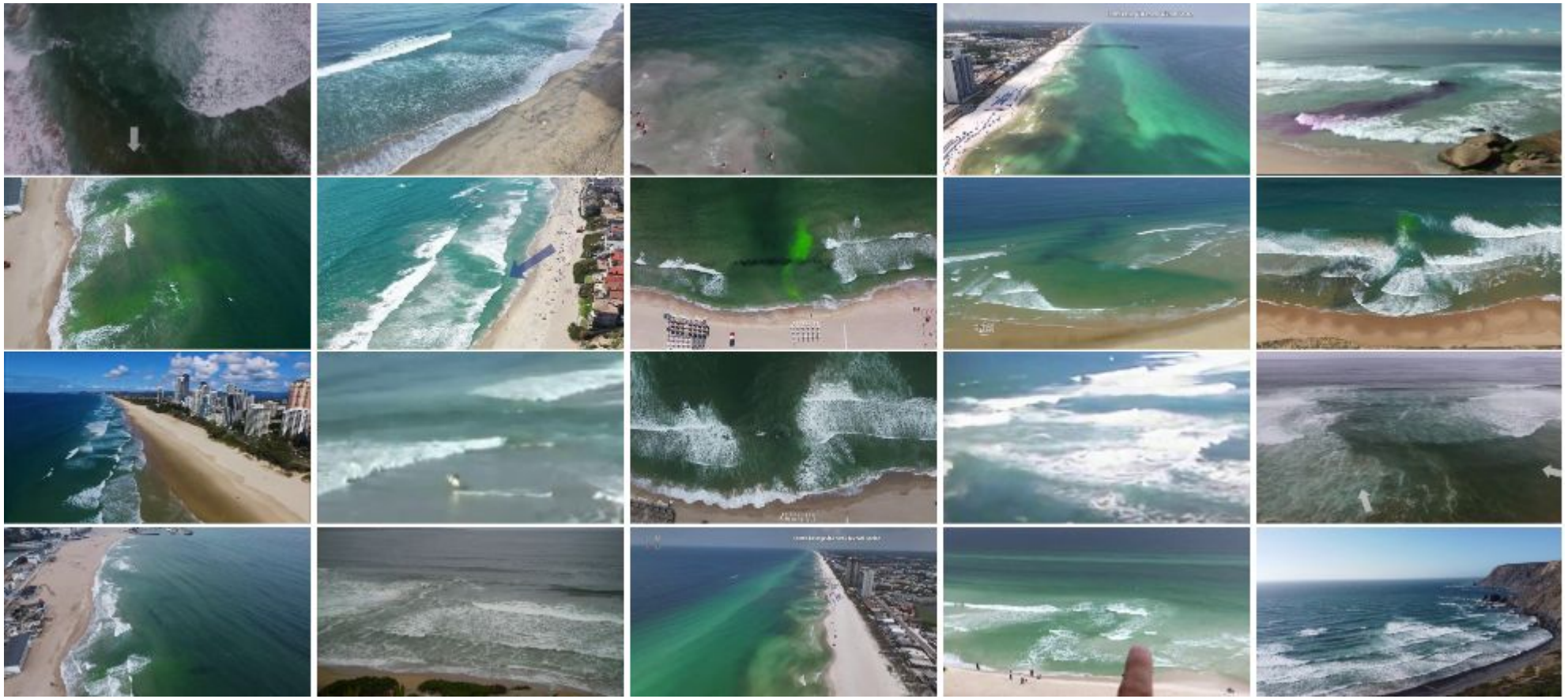
Dangers & Awareness

- Learn to recognize it
- Don’t panic!
- Swim perpendicular to it, not against it!
- Don’t panic!



Rip current illustration showing the direction of the current and how to swim in order to escape it.
Source: https://www.noaa.gov/

Rip Currents in the Wild



Annotated Examples

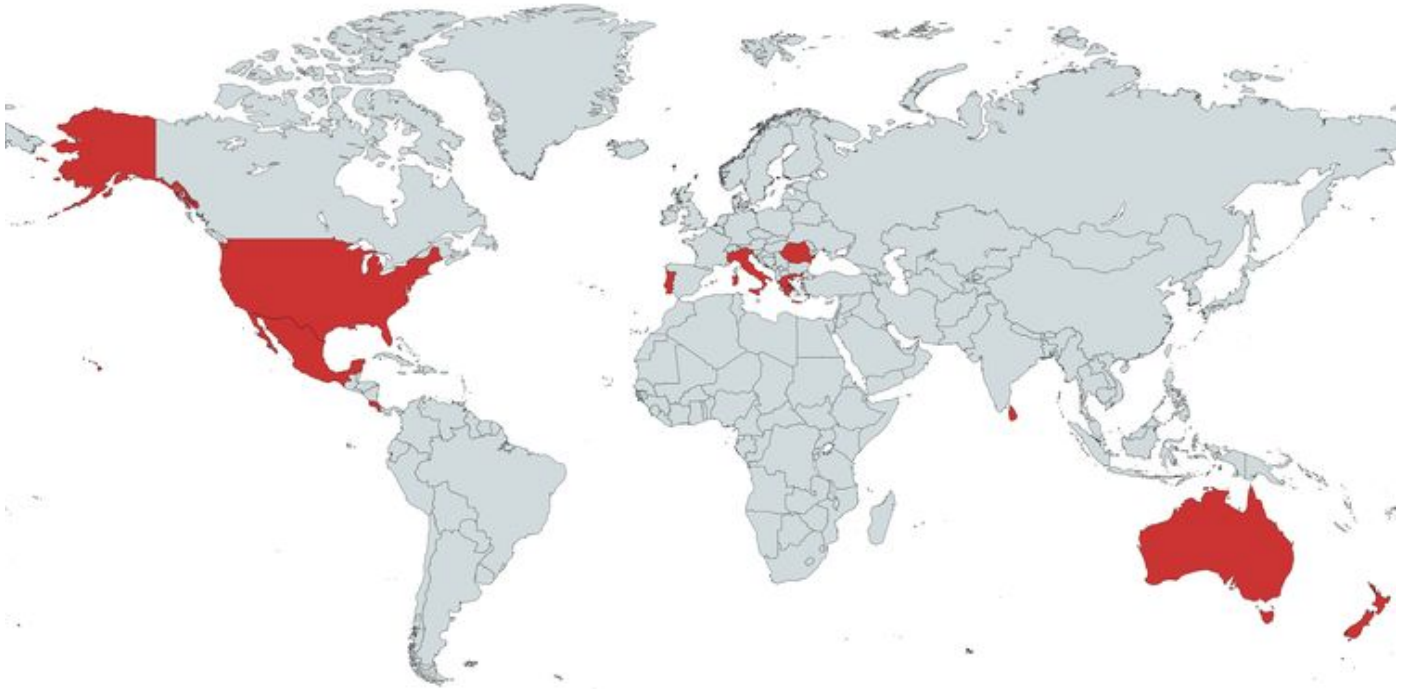
Why Instance Segmentation?

1. The original photo, 2. Existing bounding box annotation, 3. Our ground-truth annotation for instance segmentation, 4. Our prediction.
This example highlights how bounding boxes may exclude relevant parts of the rip currents while also incorporating surrounding noise.

Source: Dumitriu et. al. (2023)

Benchmark Overview

- 184 videos (212,328 frames)
  - 150 videos with rip currents
  - 34 videos without
  - 15,784 annotated frames
- Sourced from at least 11 countries
- 4 viewing angles
- Baseline models and analysis
- Benchmark website and community engagement: https://RipVIS.ai



Known countries for video sources

Various Viewpoints

Existing Datasets vs. RipVIS

Dataset	Total	With Rip Currents	Without Rip Currents	Train	Validation	Test	Segmentation Annotations
Maryan *et al.* (2019) [38]	5,310 images	514 images	4,796 images	4,779 images (10-fold)	-	531 images (10-fold)	✗
de Silva *et al.* (2021) [13]	20,482 images	10,793 images	9,689 images	2,440 images	-	23 videos 18,042 frames	✗
YOLO-Rip (2022) [65]	3,793 images	2,486 images	1,307 images	3,793 images	-	same as de Silva *et al.* [13]	✗
Dumitriu *et al.* (2023) [15]	37,057 frames	26,761 frames	10,296 frames	3,396 images (10-fold)	377 images (10-fold)	25 videos 33,284 frames	✓
RipVIS (ours)	**184 videos 212,328 frames**	**150 videos 163,528 frames**	**34 videos 48,800 frames**	**112 videos 147,802 frames**	**36 videos 32,566 frames**	**36 videos 31,960 frames**	✓

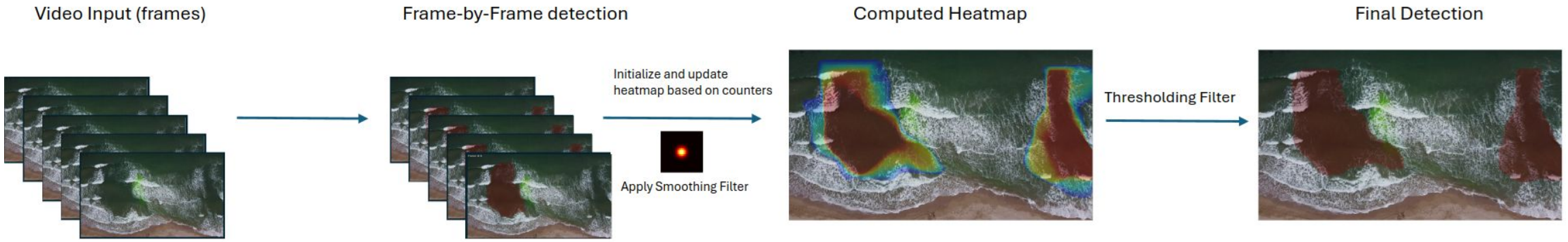
Comparison of public rip currents datasets. As observed, our dataset is an order of magnitude larger than any other publicly available dataset, with increased diversity and a train-validation-test split.

Baseline Models & Metrics

Model	Precision		Recall		AP50		F_1		F_2		FPS	
	Original	+TCA	Original	+TCA	Original	+TCA	Original	+TCA	Original	+TCA	Original	+TCA
Mask-RCNN [23]	0.492	0.538	0.625	0.651	0.530	0.556	0.550	0.589	0.593	0.625	7.84	6.73
Cascade Mask-RCNN [4]	0.606	0.613	0.660	0.686	0.628	0.639	0.632	0.647	0.648	0.670	9.53	7.94
YOLO11n [27]	0.713	0.719	0.558	0.591	0.650	0.648	0.626	0.648	0.583	0.613	128.20	34.48
YOLO11s [27]	0.757	0.752	0.612	0.647	0.705	0.723	0.677	0.696	0.636	0.666	116.27	33.78
YOLO11m [27]	0.739	0.745	0.624	0.648	0.707	0.726	0.677	0.693	0.644	0.665	76.93	29.41
YOLO11l [27]	0.812	0.819	0.588	0.613	0.713	0.729	0.682	0.701	0.622	0.646	57.14	25.98
YOLO11x [27]	0.746	0.742	0.609	0.647	0.682	0.703	0.671	0.691	0.632	0.664	34.01	19.84
SparseInst R-50 [8]	0.520	0.583	0.782	0.807	0.703	0.722	0.644	0.677	0.710	0.749	29.73	18.32
SparseInst PVTv2 [8]	0.683	0.712	0.770	0.798	0.721	0.751	0.724	0.753	0.751	0.780	27.99	17.64

Performance comparison of different models on the test split, with and without TCA. The models are applied on video and the metrics are calculated by evaluating on manually annotated frames. The best result on each metric is highlighted in blue.

Temporal Confidence Aggregation (TCA)

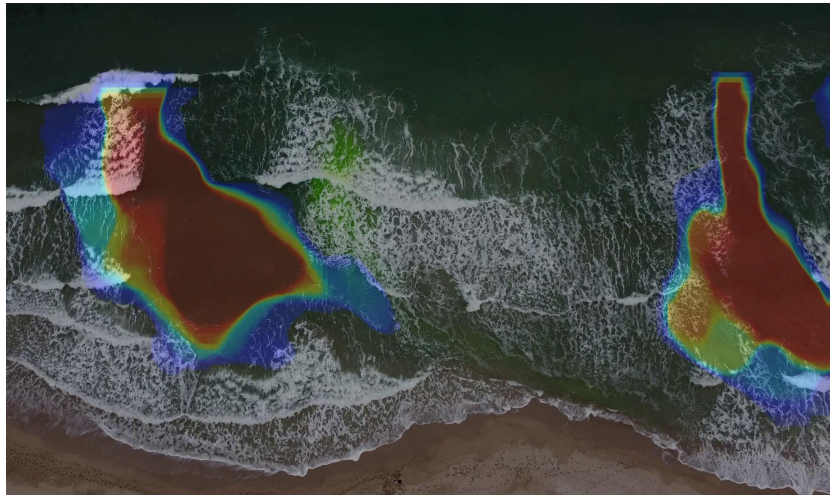


The proposed Temporal Confidence Aggregation (TCA) process, simplified. TCA leverages temporal coherence through down-sampling, instance tracking, temporal smoothing, and hysteresis thresholding to create a stabilized temporal heatmap

TCA: positive impact

Prediction

Pred. + TCA

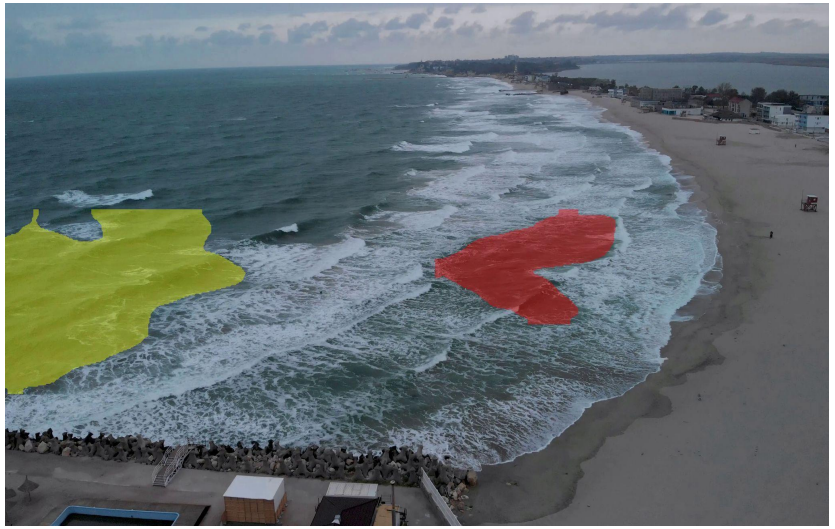


Pred. + Filtered TCA

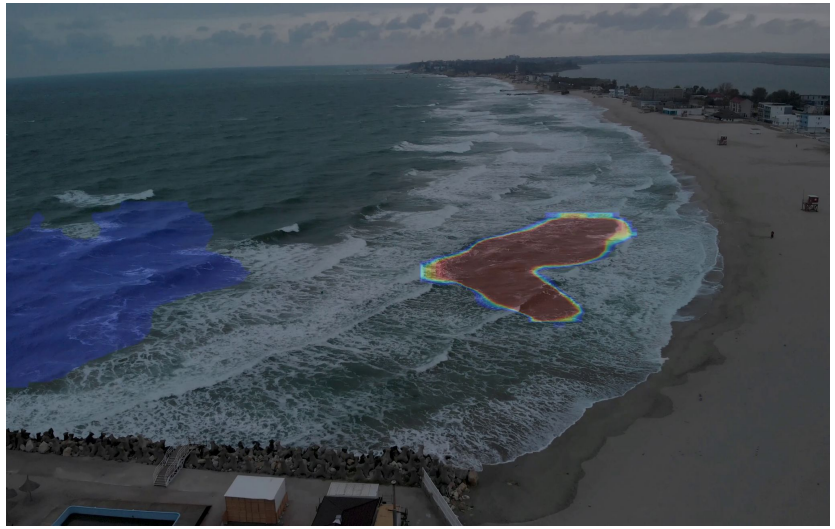
In this situation, TCA manages to filter many false negatives

TCA: positive impact

Prediction



Pred. + TCA

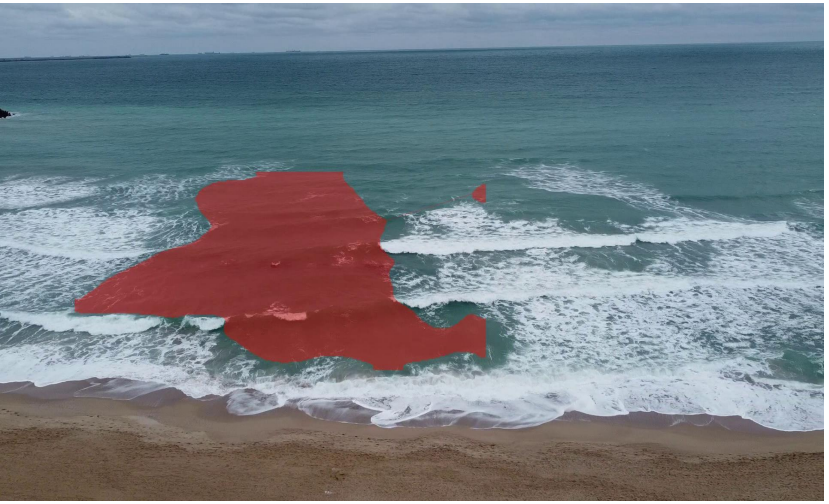


Pred. + Filtered TCA

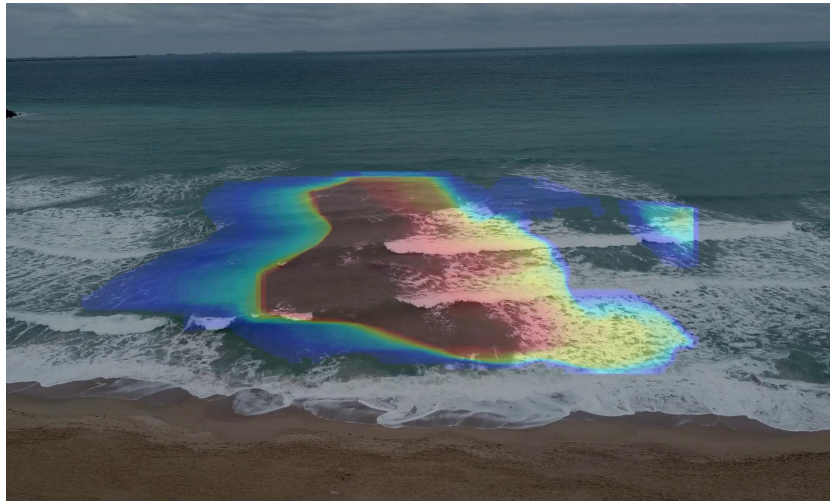
In this situation, TCA manages to filter out many false positives (albeit not all)

TCA: limitations

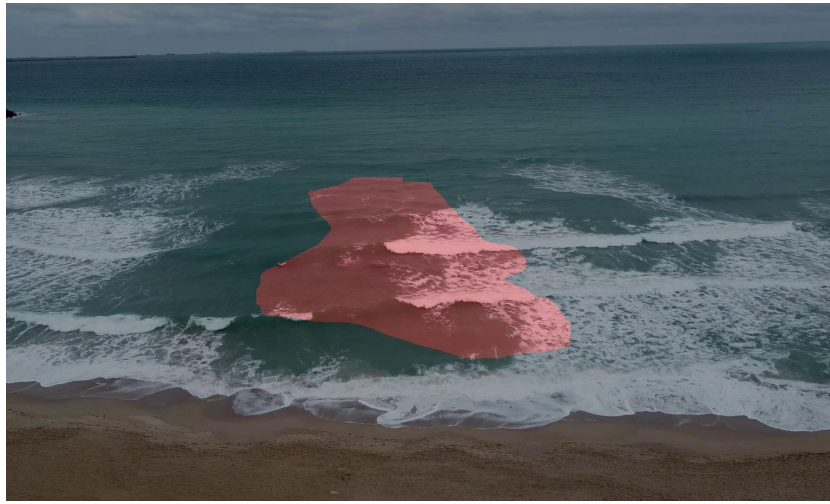
Prediction



Pred. + TCA



Pred. + Filtered TCA



An example were initially TCA helps, but ends up making the results even worse due to sudden movement

Website & Community

- Website: https://ripvis.ai.
- Participate in open challenges.
- Global impact and community.
- Use data under CC BY-NC 4.0 license.
- Submit videos / annotations.
- Community credits.
- Actual impact that can save lives.

Let's work together and make our beaches safer through vision - one frame at a time

Join the challenge • contribute videos • spread awareness