Online Self-Supervised Thermal Water Segmentation for Aerial Vehicles

Connor Lee, Jonathan Gustafsson Frennert, Lu Gan, Matthew Anderson, and Soon-Jo Chung

Caltech

Background

Today, aerial robotics equipped with thermal cameras still lack semantic perception capabilities in nighttime littoral environments due to:

- Lack of in-domain (setting and modality) datasets for model training.
- Difficulty of capturing diverse data due to geographic limitations and local flight regulations.

This work: we apply online self-supervised learning to segment aerial near-shore thermal imagery *without seeing thermal images prior to test time*. This enables UAVs to perform operations such as:

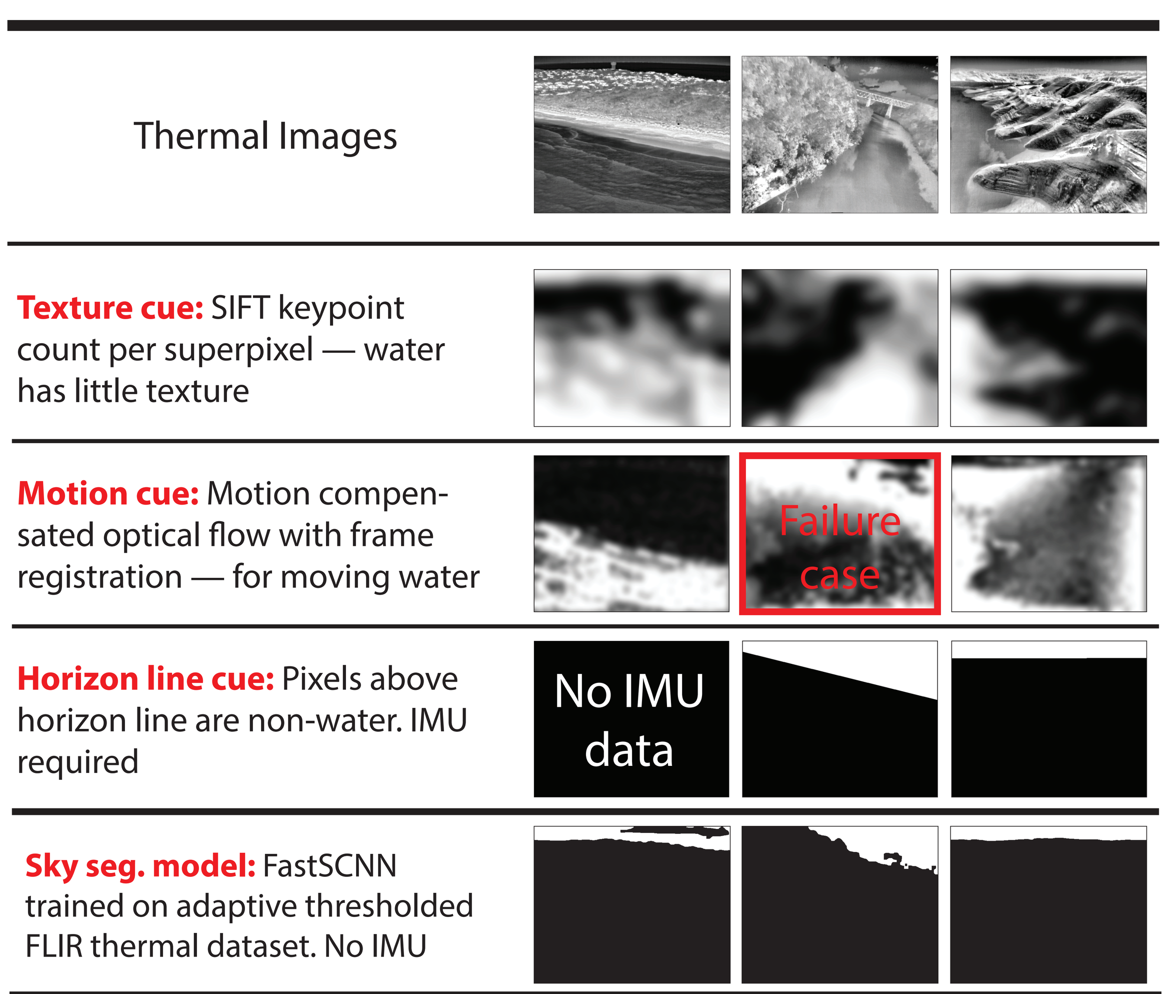
- UAV visual navigation in riverine and coastal settings
- Bathymetry for surface vehicle path planning

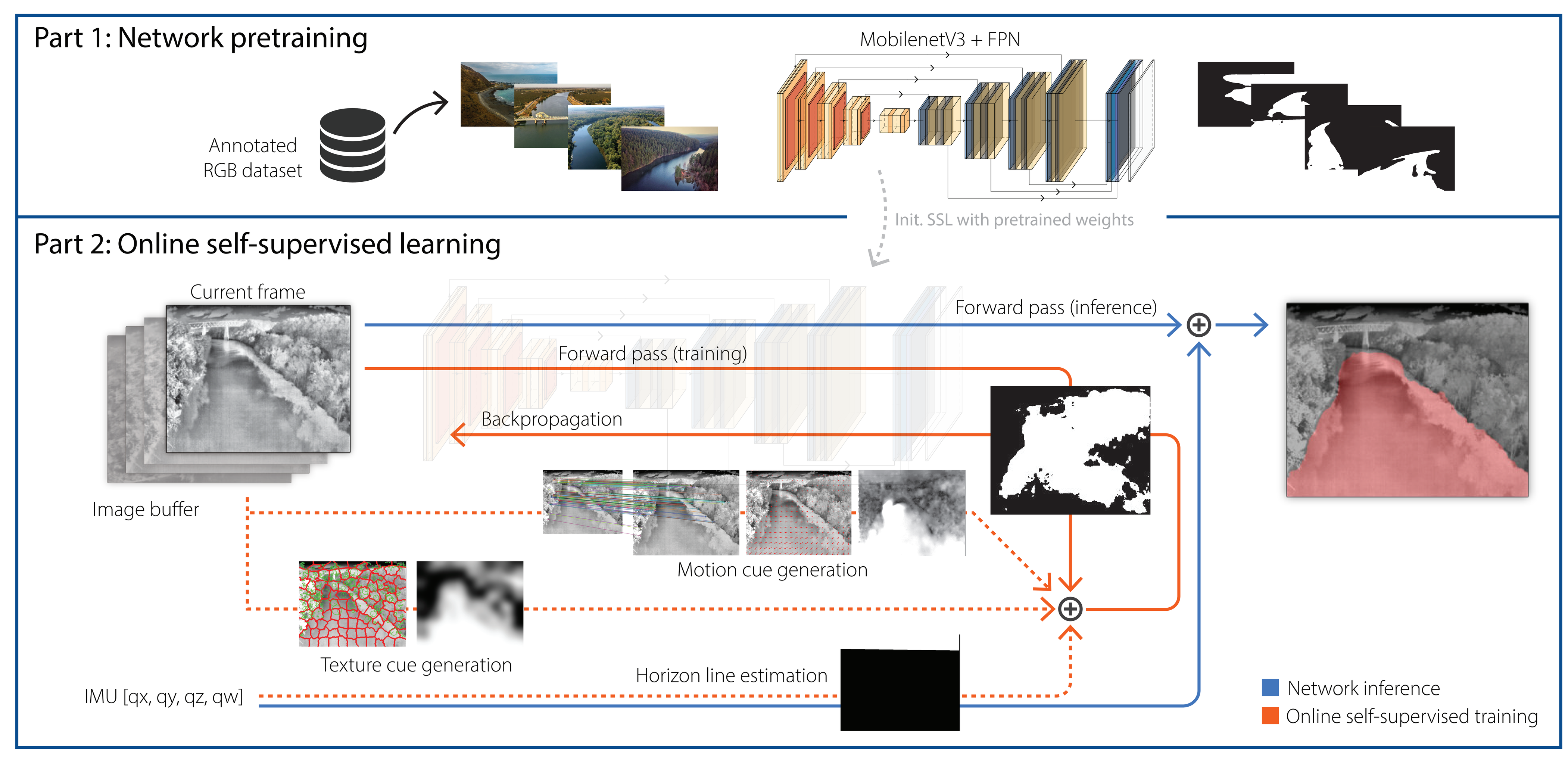
during nighttime and in other degraded visual conditions.

Approach

Our online thermal segmentation algorithm leverages water and non-water cues as supervisory signals to adapt an RGB-pretrained segmentation network during test time.

1. **Network pretraining:** We pretrained a MobileNetV3+FPN on ADE20k, COCO-stuff, and Flickr-scraped water-related images. Annotated RGB imagery were preprocessed to 1-channel via decorrelated random channel averaging.
2. **Online SSL with water cues:** During test time, we learn from pseudolabels created using water cues based on texture and motion characteristics of water and land.





Datasets

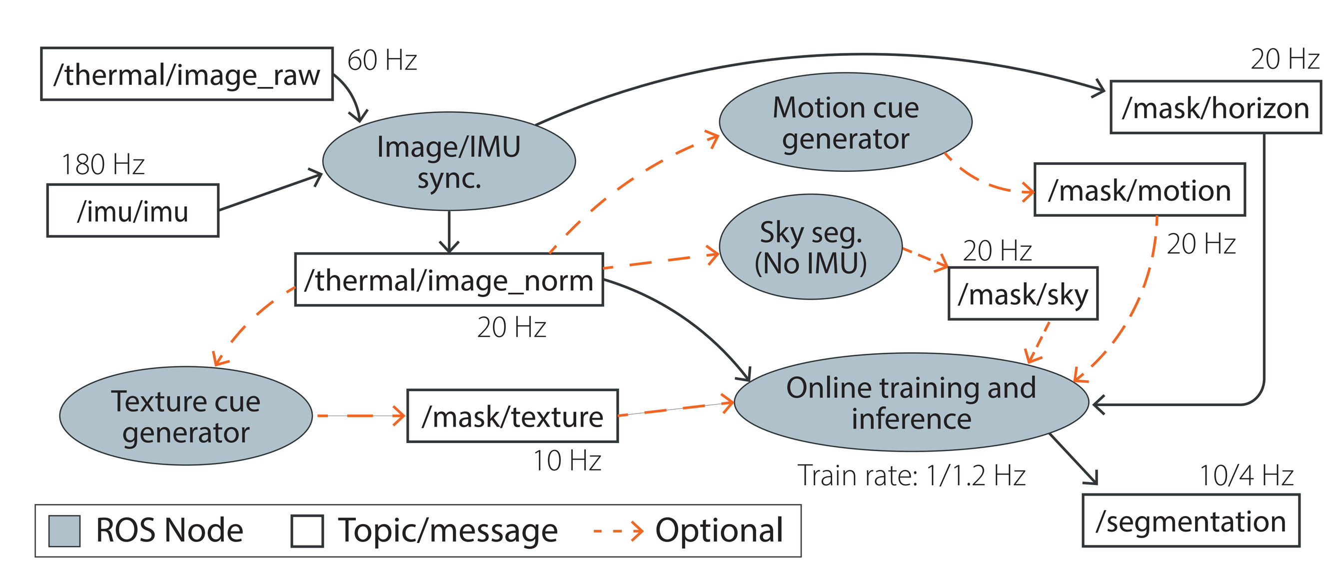
- We collected aerial (40-80 m) and ground thermal sequences covering riverine, coastal, and lake scenery.
- Images sampled for water/non-water annotation at least 2 seconds apart.

Dataset	Near-shore Category	Capture Method	# Images	# Annot.	# Seq.
Kentucky River, KY	River	UAV Flight	7826	94	1
Colorado River, CA	River	UAV Flight	84,993	659	2
Duck, NC†	Coast	UAV Hover	4143	68	7
Castaic Lake, CA	Lake	UAV Flight	101,999	128	2
Big Bear Lake, CA	Lake	Ground	48,676	282	8
Arroyo Seco, CA	Stream	Ground	7	7	–

† Captured and stored in processed 8-bit data.

ROS Implementation on Nvidia Jetson AGX Orin

- Parallel processes enable inference at 10 Hz with periodic online training.
- Maintain separate inference and training copies of the network.



Results

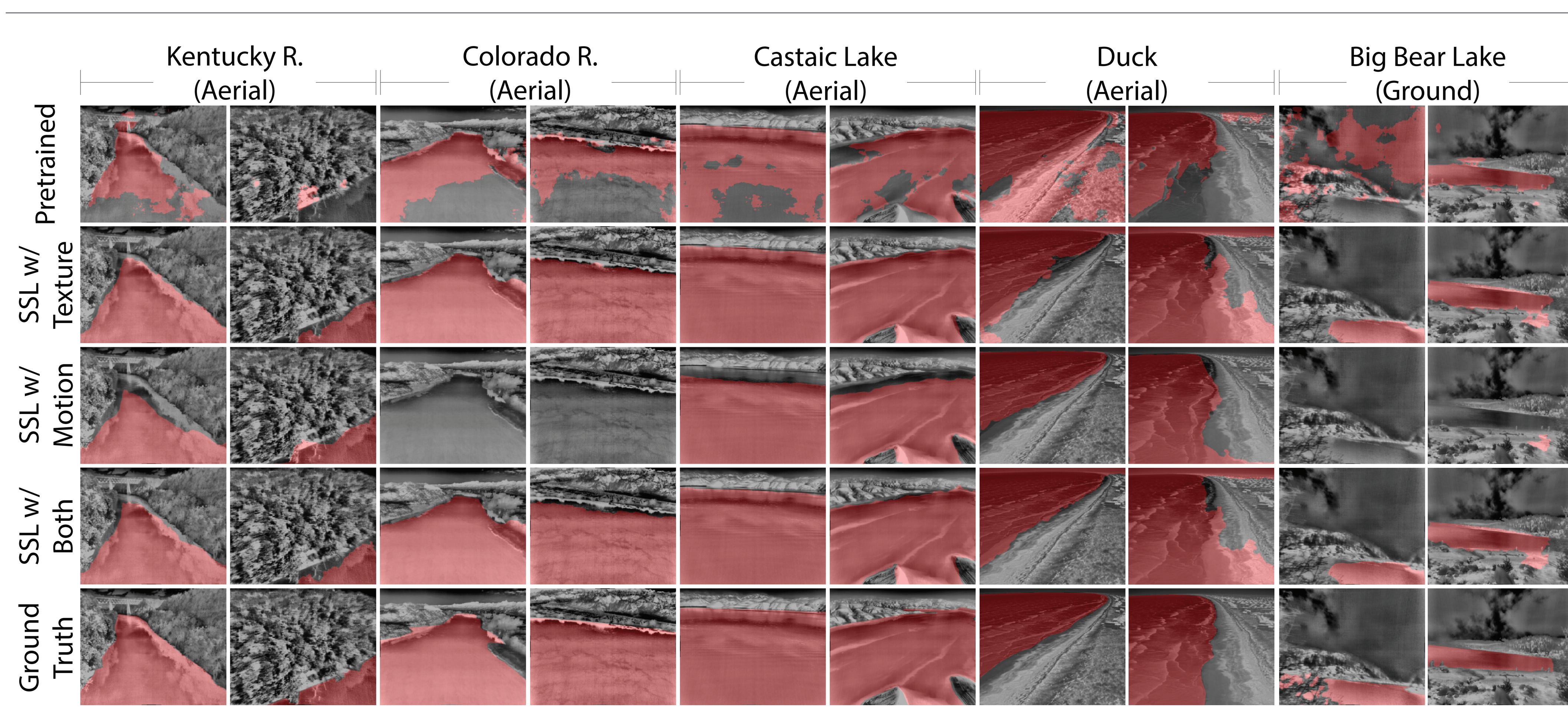


Table 1. Performance in target aerial settings vs. fully-supervised thermal networks.

Method	Training Set	Aerial Test Setting mIoU		
		River	Lake	Coast
MobilenetV3 + FPN	Arroyo Seco	0.62	0.56	0.58
MobilenetV3 + FPN	Big Bear Lake	0.69	0.53	0.64
MobilenetV3 + FPN	Big Bear Lake + Arroyo	0.79	0.63	0.72
MobilenetV3 + FPN	Colorado River	–	0.75	0.44
MobilenetV3 + FPN	MassMIND [1]	0.45	0.31	0.45
Online SSL (PCA) + TC	–	0.90	0.89	0.61
Online SSL (PCA) + MC	–	0.47	0.75	0.81
Online SSL (PCA) + All	–	0.88	0.91	0.65

Table 2. Near-shore water segmentation ablation in different thermal sequences.

Setting	PT	TC Only	MC Only	w/o Sky Seg. nor Horizon Est.			w/ Sky Segmentation			w/ Horizon Est.		
				PT + TC	PT + MC	PT + All	PT + TC	PT + MC	PT + All	PT + TC	PT + MC	PT + All
Aerial River	0.571	0.794	0.474	0.880	0.473	0.865	0.892	0.469	0.877	0.895	0.474	0.878
Aerial Lake	0.438	0.802	0.469	0.889	0.746	0.893	0.847	0.275	0.857	0.889	0.746	0.909
Aerial Coast	0.583	0.574	0.610	0.611	0.805	0.654	0.557	0.649	0.600	–	–	–
Ground Lake	0.600	0.357	0.410	0.388	0.534	0.409	0.660	0.554	0.662	0.745	0.567	0.751

PT – Base Pretrained Network TC – Texture Cue MC – Motion Cue

Conclusion

- We train a water segmentation network via online SSL with basic water cues that demonstrates robustness, outperforming fully-supervised networks.
- Potential applications include nighttime bathymetry and coastline mapping.

References

[1] Shailesh Nirgudkar, Michael DeFilippo, Michael Sacarny, Michael Benjamin, and Paul Robinette. Massmind: Massachusetts maritime infrared dataset. 42(1-2):21–32, 2023.