

22nd April 2026

Deep Neural Networks for Reliable Underwater Vision

PhD Defense

Laura Weihl





William Thompson 1856 [1]

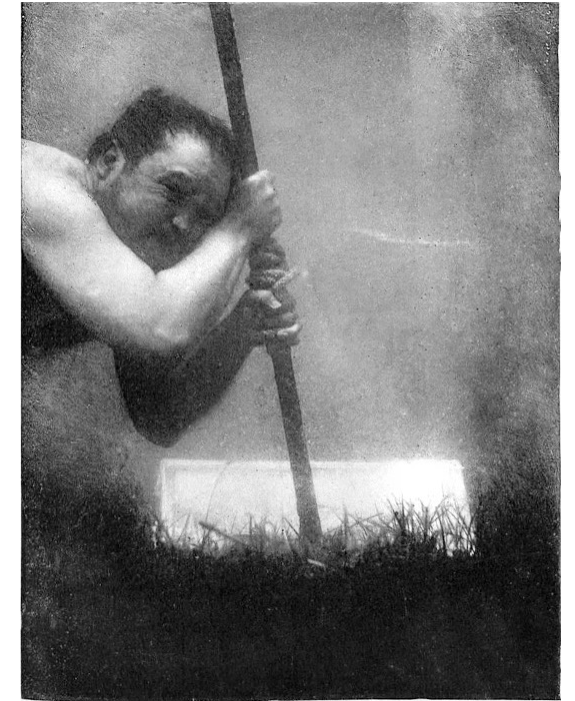
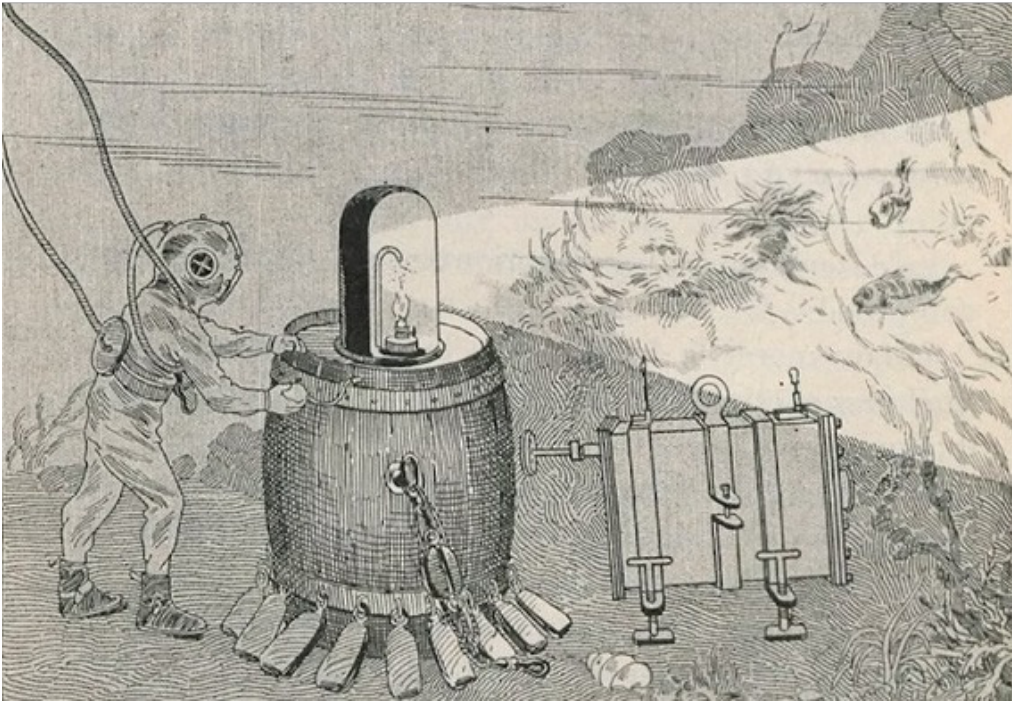
[1] <https://www.artpublikamag.com/post/louis-boutan-pioneered-underwater-photography-gave-the-world-its-first-look-deep-beneath-the-sea>



[2]

[2] <https://camera-wiki.org/wiki/wet-collodion>

Underwater Photography



Louis Boutan in 1899

[1] <https://www.artpublikamag.com/post/louis-boutan-pioneered-underwater-photography-gave-the-world-its-first-look-deep-beneath-the-sea>

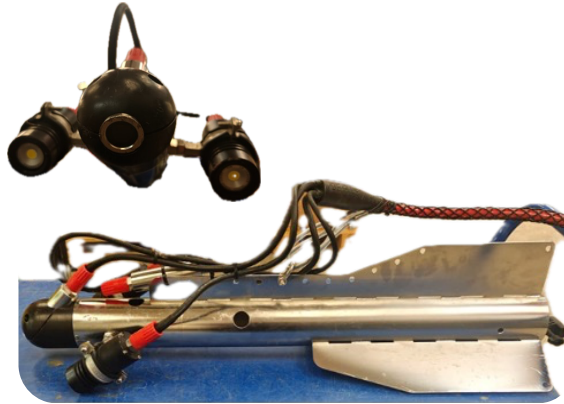
Underwater Tech

Static Cameras



Biodiversity monitoring

Towed Cameras



Environmental monitoring

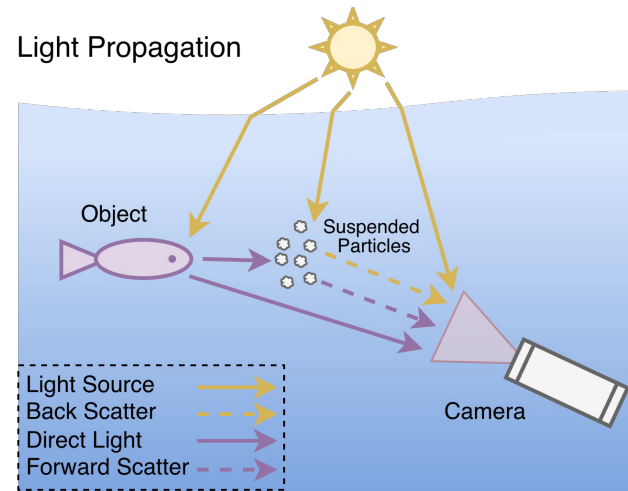
6DoF Cameras



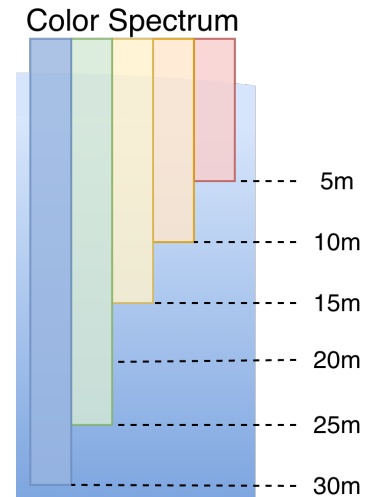
Offshore inspection

Challenges

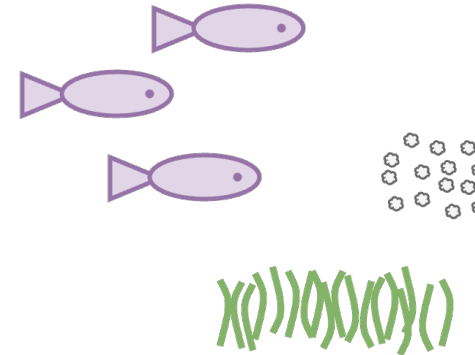
Light Scattering



Color Distortion



Dynamic Elements

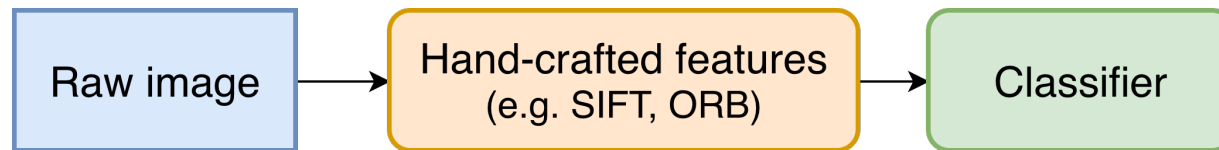


Data Scarcity



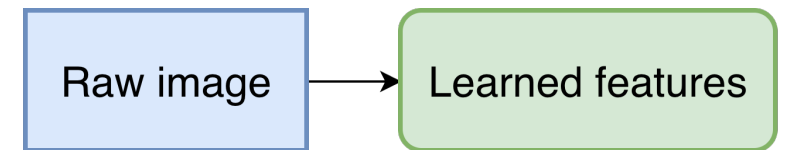
Computer Vision

Traditional Computer Vision



- + small datasets
- + fast inference, no GPU needed
- often domain-specific
- do not generalize well

Deep Learning



- + state-of-the-art accuracy
- + can generalize well
- needs labeled data
- expensive to compute

Can we apply deep learning underwater? Can we get **reliable** results?

Research Questions

RQ1 Image Degradation

How effective are deep neural networks (DNNs) at overcoming underwater image properties?

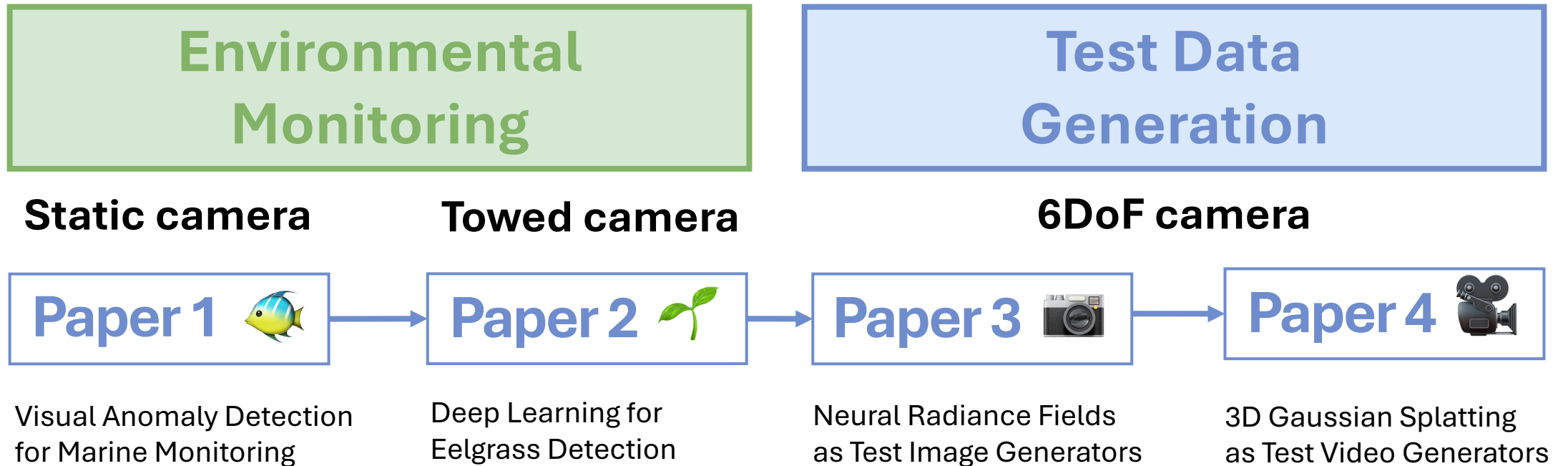
RQ2 Analysis at Scale

How can DNNs improve the capabilities of environmental monitoring?

RQ3 Generating Tests

How effective are 3D scene representations at generating realistic and targeted test data?

Research



... from constrained to unconstrained camera motion

Biodiversity Monitoring

- Static cameras for long-term underwater biodiversity monitoring
- 30 secs recording every 30 mins for months
- Problem: get TB of data



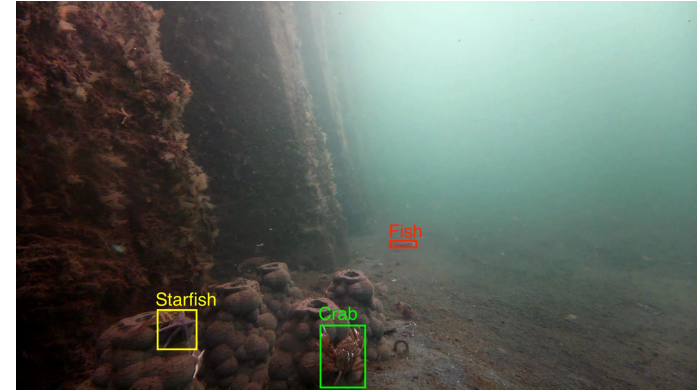
Database				Q Search...	Filter	Export
Date	Video Name	Duration	File Size			
01/07/2024	Recording_20240701_00-00.mp4	00:30:00	62 MB			
	Recording_20240701_00-30.mp4	00:30:00	59 MB			
02/07/2024	Recording_20240702_00-00.mp4	00:30:00	64 MB			
	Recording_20240702_00-30.mp4	00:30:00	61 MB			
03/07/2024	Recording_20240703_00-00.mp4	00:30:00	58 MB			
	Recording_20240703_00-30.mp4	00:30:00	63 MB			
04/07/2024	Recording_20240704_00-00.mp4	00:30:00	60 MB			
	Recording_20240704_00-30.mp4	00:30:00	57 MB			
05/07/2024	Recording_20240704_00-00.mp4	00:30:00	65 MB			
	Recording_20240705_00-30.mp4	00:30:00	61 MB			
06/07/2024	Recording_20240706_00-00.mp4	00:30:00	56 MB			
	Recording_20240706_00-30.mp4	00:30:00	59 MB			
07/07/2024	Recording_20240706_00-00.mp4	00:30:00	63 MB			
	Recording_20240707_00-30.mp4	00:30:00	62 MB			
08/07/2024	Recording_20240702_00-00.mp4	00:30:00	60 MB			
	Recording_20240704_00-30.mp4	00:30:00	58 MB			
09/07/2024	Recording_20240709_00-00.mp4	00:30:00	64 MB			
	Recording_20240704_00-30.mp4	00:30:00	61 MB			
10/07/2024	Recording_20240709_00-00.mp4	00:30:00	66 MB			
	Recording_20240709_00-30.mp4	00:30:00	60 MB			
11/07/2024	Recording_20240701_00-00.mp4	00:30:00	57 MB			
	Recording_20240701_00-30.mp4	00:30:00	62 MB			

Challenge

- Most of the time nothing happens:



Deep Learning Underwater



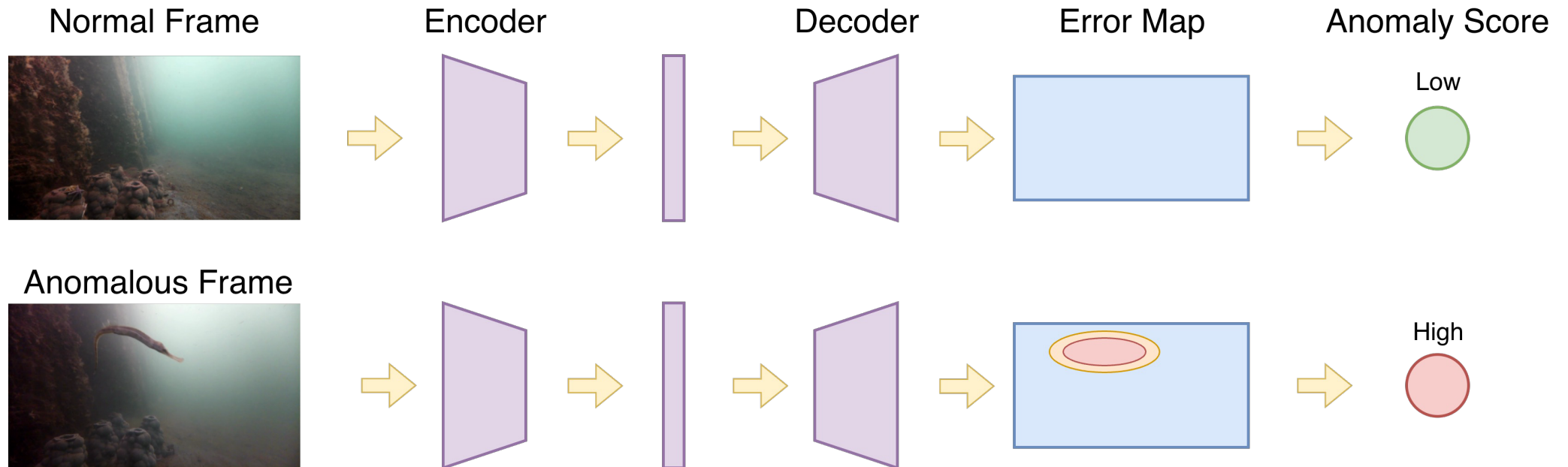
Object Detection



Segmentation

Visual Anomaly Detection

- Train on “normal” frames of the scene



Visual Anomaly Detection

Paper 1 



Annotation Problem

- We recruit 16 annotators and use soft labels



Soft label 0.0



Soft label 0.5



Soft label 1.0

The AURA Dataset

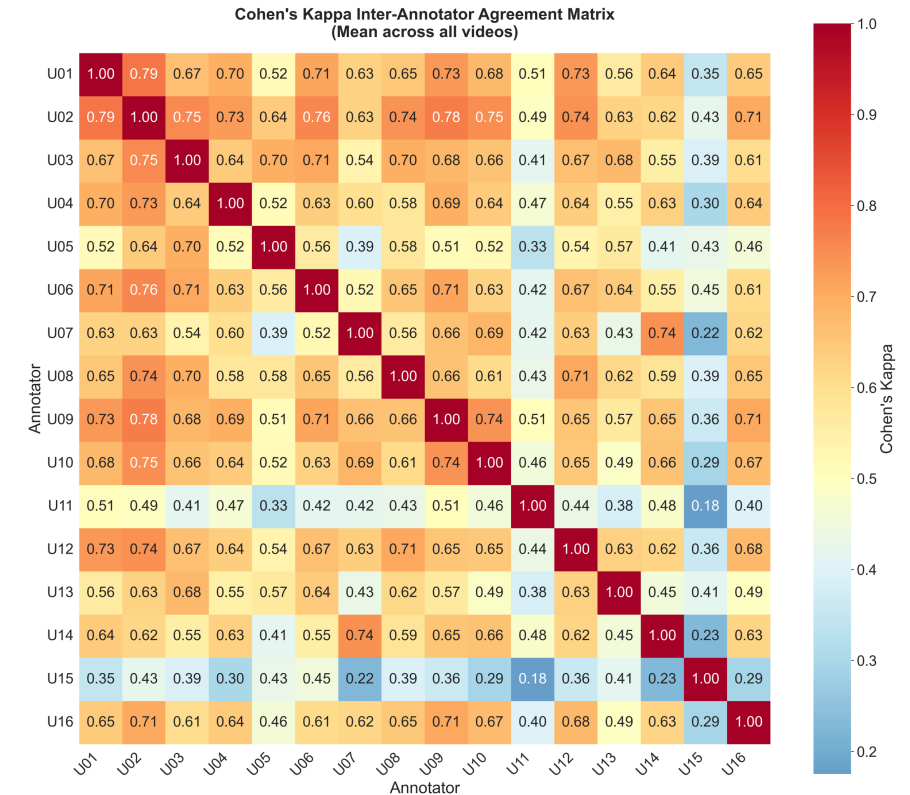
- **Anomalous Underwater R Acticity**
- 2 scenes, 25 videos, 16 annotators
- Moderate to substantial agreement



Scene A

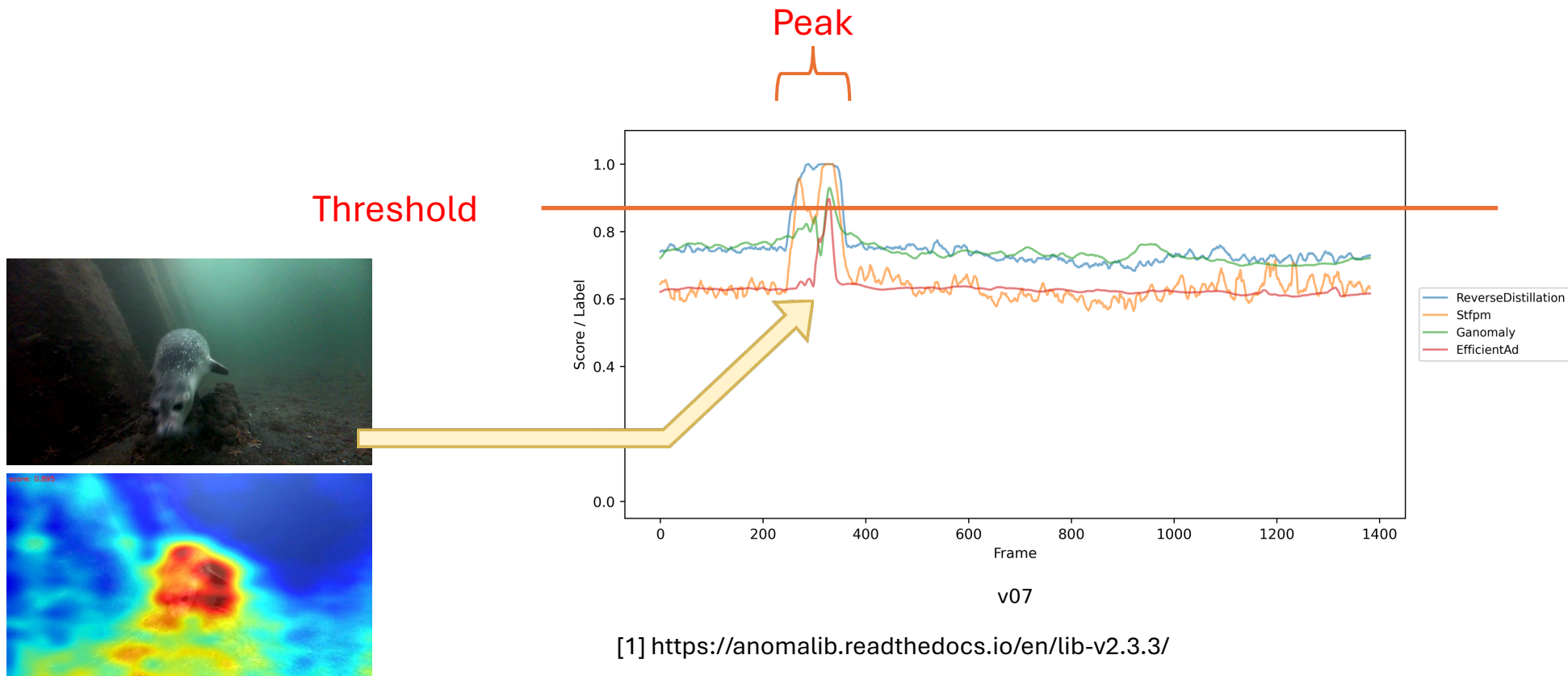


Scene B



Experiments

- Evaluate 4 visual anomaly detection methods in anomalib [1]
- Event extraction: threshold-based vs. peak-based



Anomaly scores vs. soft labels

- Report mean average error ($\downarrow$)

	MAE (Mean $\pm$ Std)	
	Scene A	Scene B
EfficientAD	0.393 $\pm$ 0.101	0.379 $\pm$ 0.048
GANomaly	0.425 $\pm$ 0.127	0.435 $\pm$ 0.099
Stfpm	0.465 $\pm$ 0.107	0.320 $\pm$ 0.058
Reverse Distillation	0.271 $\pm$ 0.151	0.259 $\pm$ 0.041

- Key takeaway:
 - Reverse Distillation aligns best with human annotations in both scenes
 - Scene B is consistently easier across models

Results - tIoU

- t-IoU: temporal intersection over union

		Scene A	Scene B
EfficientAD	Find Peaks	0.347	0.745
GANomaly	Find Peaks	0.242	0.471
Reverse Distillation	Find Peaks	0.673	0.861
Stfpm	Find Peaks	0.301	0.863
EfficientAD	Threshold	0.374	0.709
GANomaly	Threshold	0.083	0.217
Reverse Distillation	Threshold	0.461	0.806
Stfpm	Threshold	0.225	0.748

- Key takeaway:
 - Reverse Distillation performs best at finding anomalous events
 - Peak-based event extraction works better than threshold-based

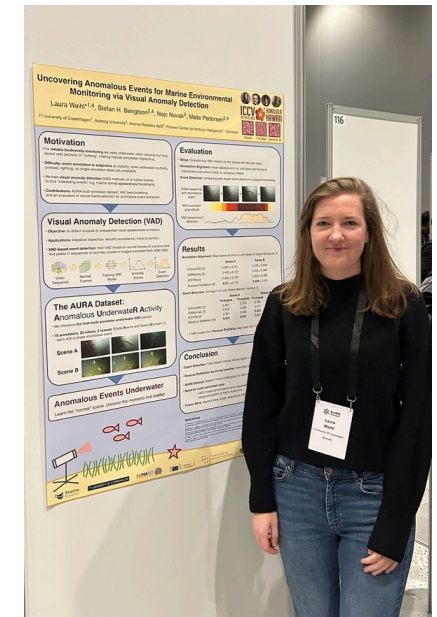
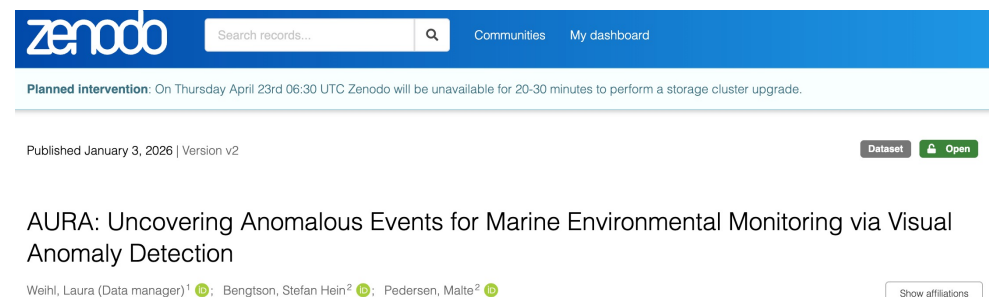
Paper 1



Uncovering Anomalous Events for Marine Environmental Monitoring via Visual Anomaly Detection

Laura Wehl, Stefan Hein Bengtson, Nejc Novak,
Malte Pedersen

Published at
IEEE Computer Society / Computer Vision Foundation
ICCV Joint Workshop on Marine Vision, 2025
Honolulu, Hawaii, USA



Research

Environmental Monitoring

Static camera

Towed camera

Paper 1



Paper 2



Visual Anomaly Detection
for Marine Monitoring

Deep Learning for
Eelgrass Detection

Detecting Eelgrass

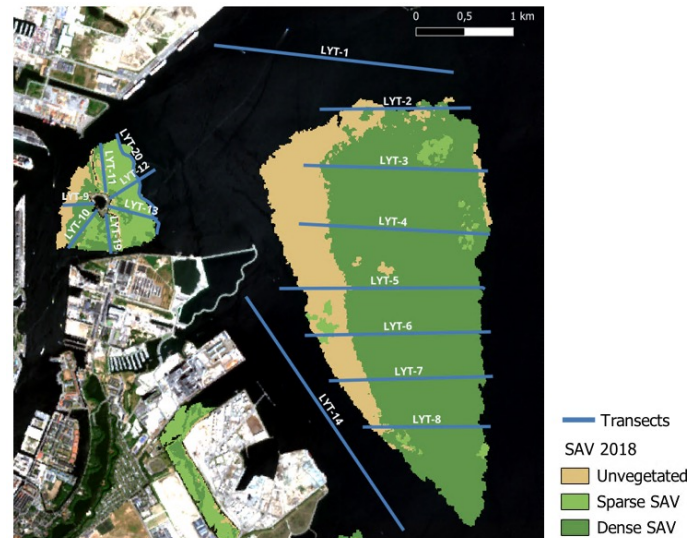


Why Seagrass?

- seagrass health is used as a proxy metric for water quality in EIAs
- transect: a camera is mounted onto a sled a towed behind a vessel
- problem: image data is manually labeled in spreadsheets



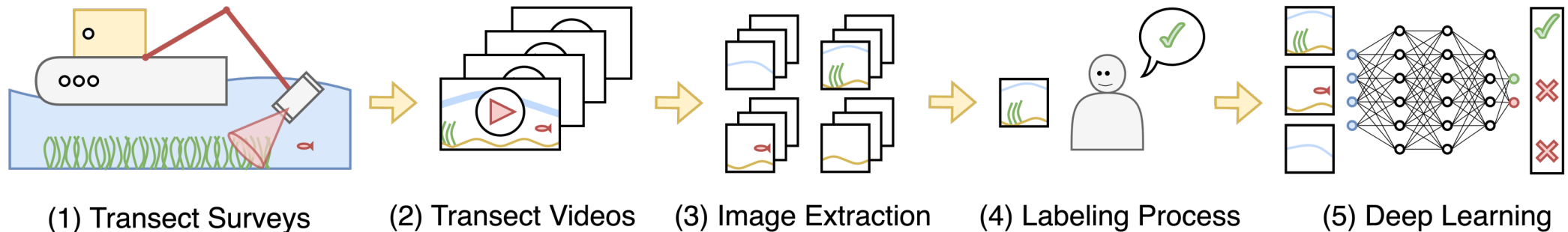
Lynetteholm Copenhagen



SeagrassFinder

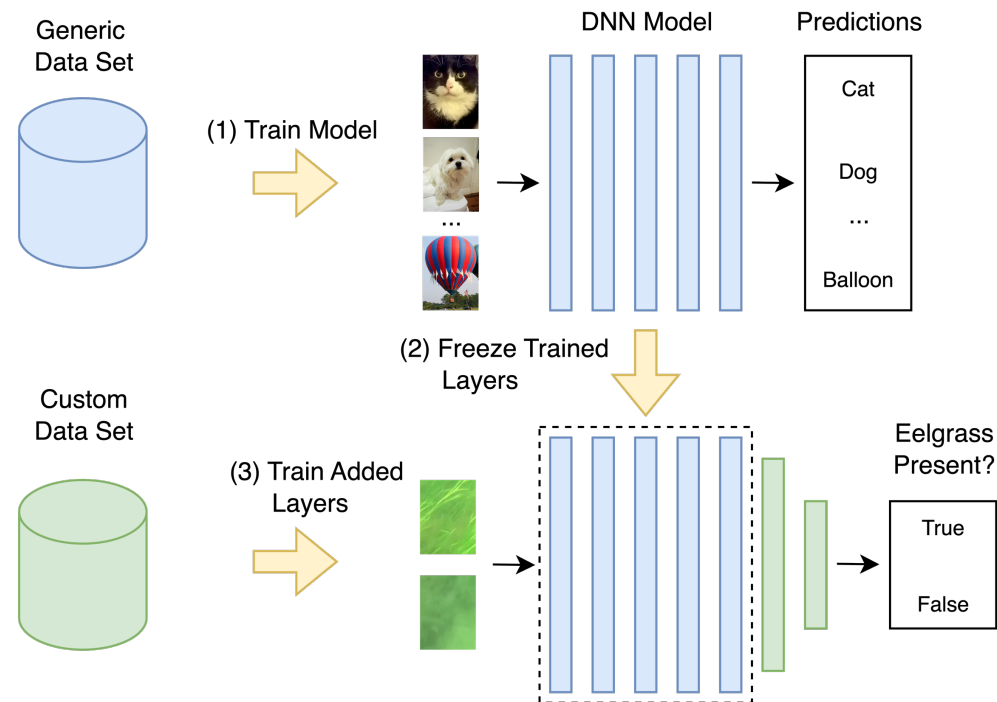
Paper 2 

- we extract images from EIA transect videos
- we deploy an effective annotation platform to get labels
- we apply transfer learning to train four types of DNNs

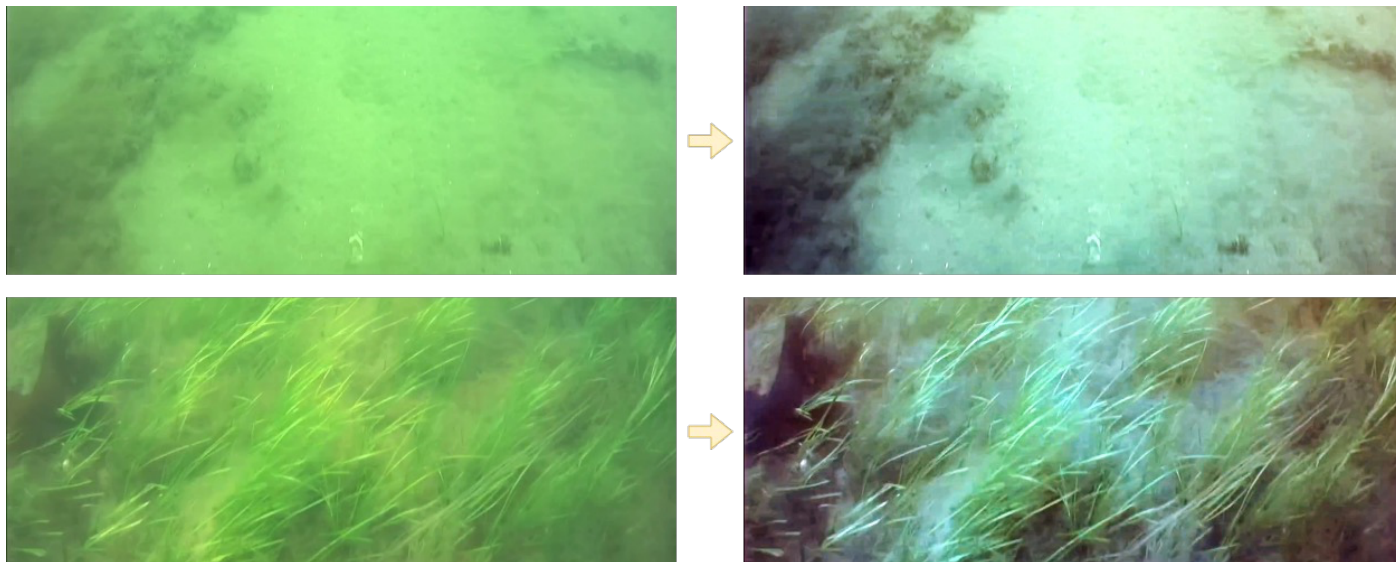


Transfer Learning

- DNNs are pre-trained on **generic** data sets like ImageNet
- then fine-tune on a **custom** eelgrass data set



Underwater Image Enhancement



Sharma, P., Bisht, I. and Sur, A., 2023. Wavelength-based attributed deep neural network for underwater image restoration. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(1), pp.1-23.

Eelgrass Presence Detection

- Binary classification task on four types of DNNs:



Eelgrass present



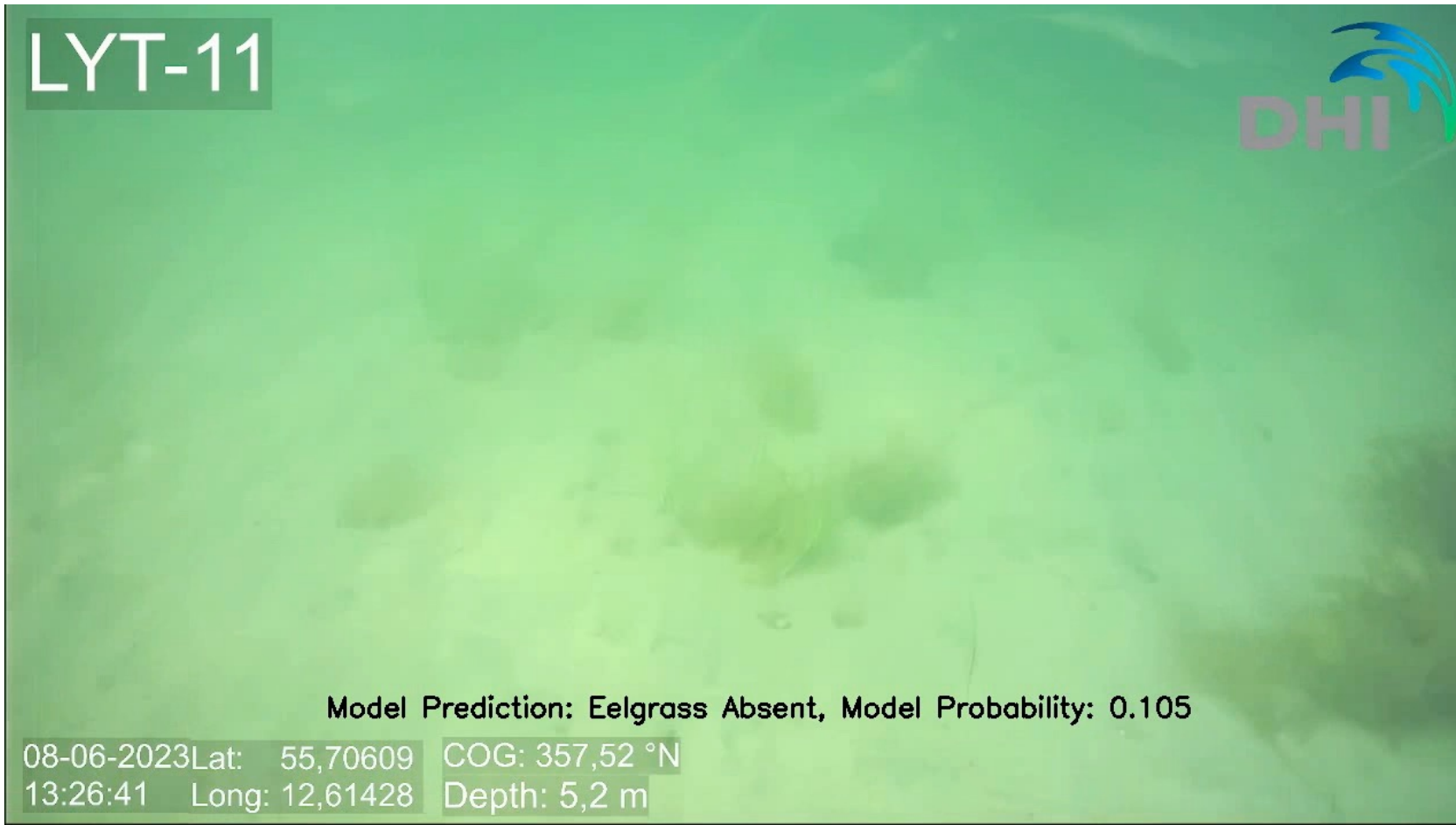
Eelgrass absent

Model	Accuracy (↑)	AUROC (↑)	CE (↓)
ResNet-18	0.870	0.939	0.098
ResNet-18++	0.867	0.936	0.097
InceptionNetV3	0.855	0.936	0.115
InceptionNetV3++	0.868	0.940	0.108
DenseNet-169	0.866	0.942	0.105
DenseNet-169++	0.879	0.943	0.099
DenseNet-201	0.874	0.944	0.105
DenseNet-201++	0.875	0.946	0.095
ViT	0.852	0.938	0.126
ViT++	0.901	0.953	0.082

- Key take-aways:
 - Vision transformer (ViT) shows best performance
 - Image Enhancement (++) improves DNN performance in some cases
 - Seagrassfinder is deployed in multiple projects in Denmark, Germany, Canada

Eelgrass in the Wild

Paper2 



Paper 2



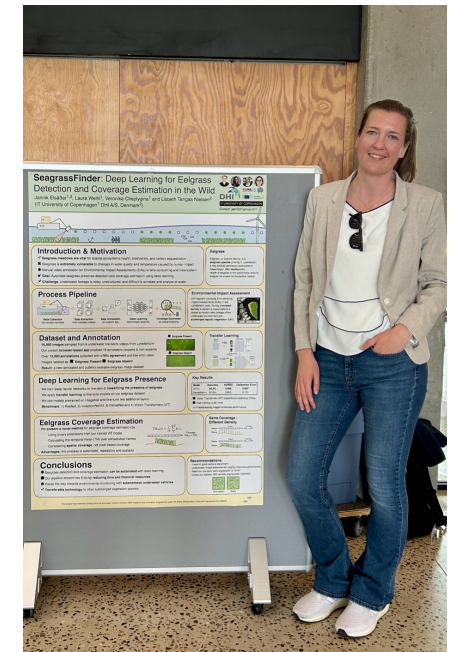
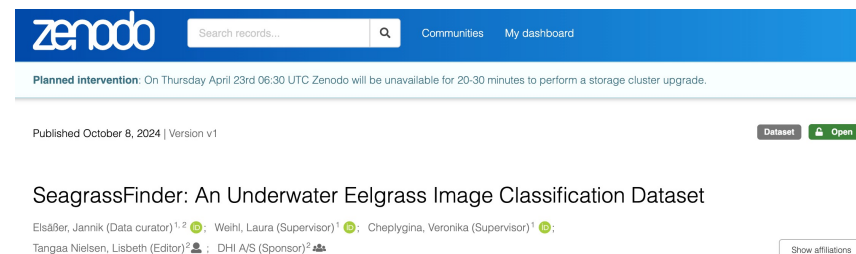
SeagrassFinder: Deep Learning for Eelgrass Detection and Coverage Estimation in the Wild

Jannik Elsäßer^{1,2}, Laura Wehl², Veronika Cheplygina²,
Lisbeth Tangaa Nielsen¹

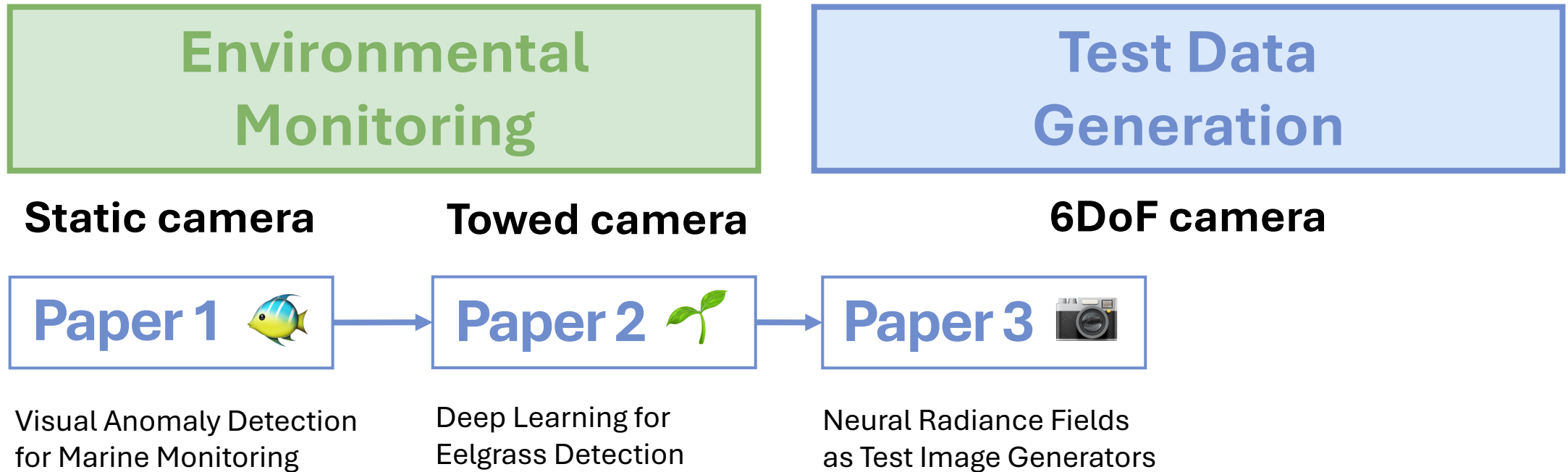
¹DHI, Hørsholm

²ITU, Copenhagen

Published in
Ecological Informatics, 2025
Elsevier



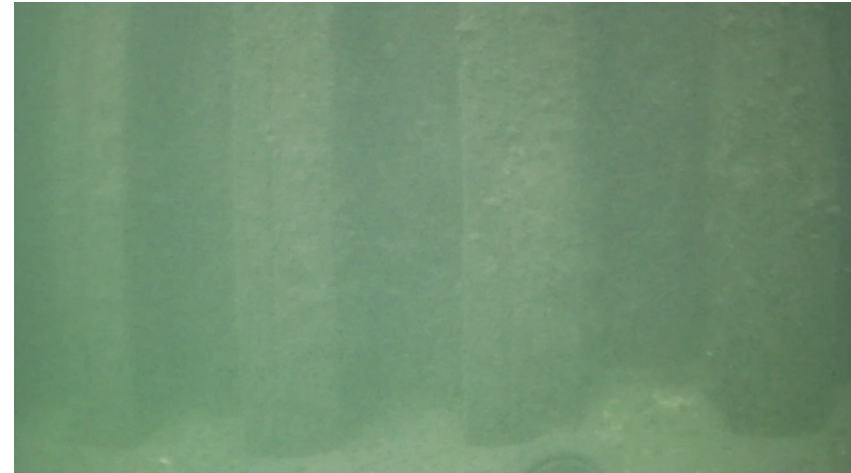
Research



Simulation



Simulation [1]

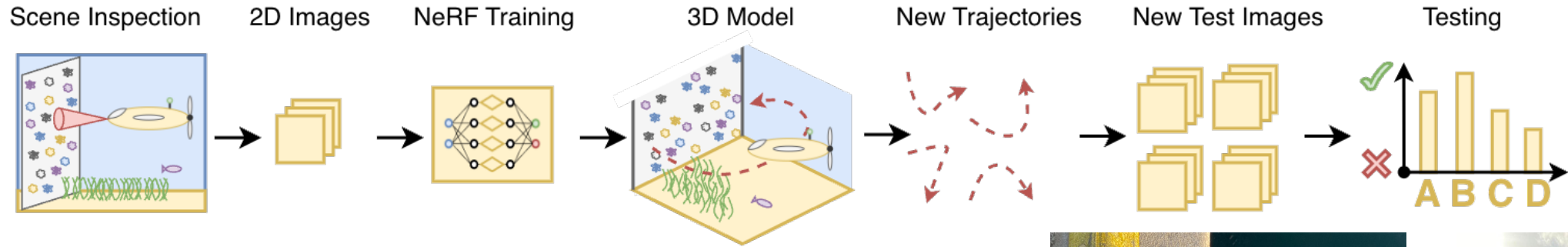


Real World

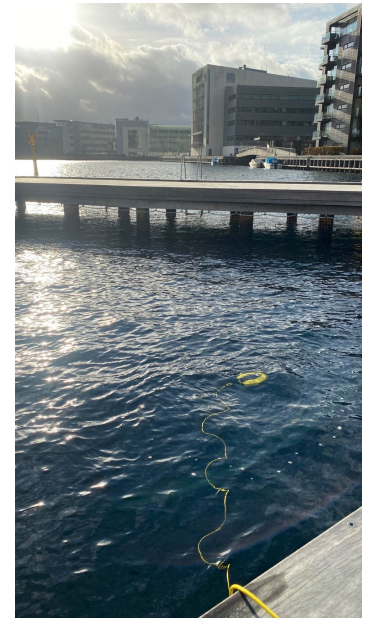
- **simulation-to-reality gap:** good performance in simulation does not guarantee good performance in the real world

NeRF-to-Real Tester

Paper3 



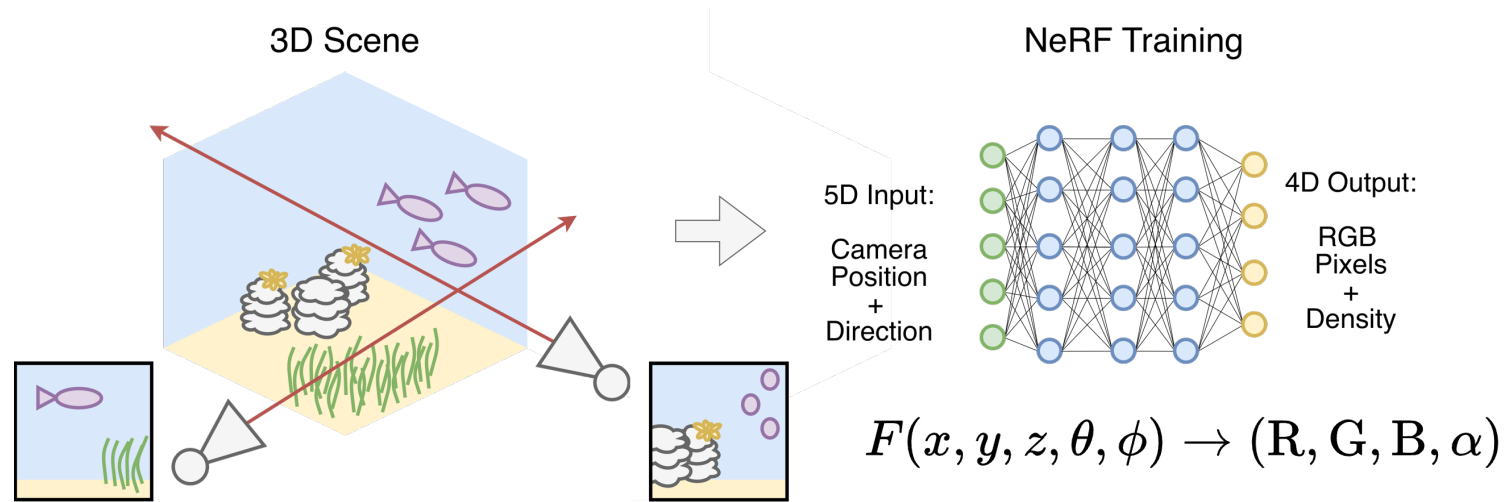
- + light-weight framework
- + based on real images
- + no need for ground truth camera poses



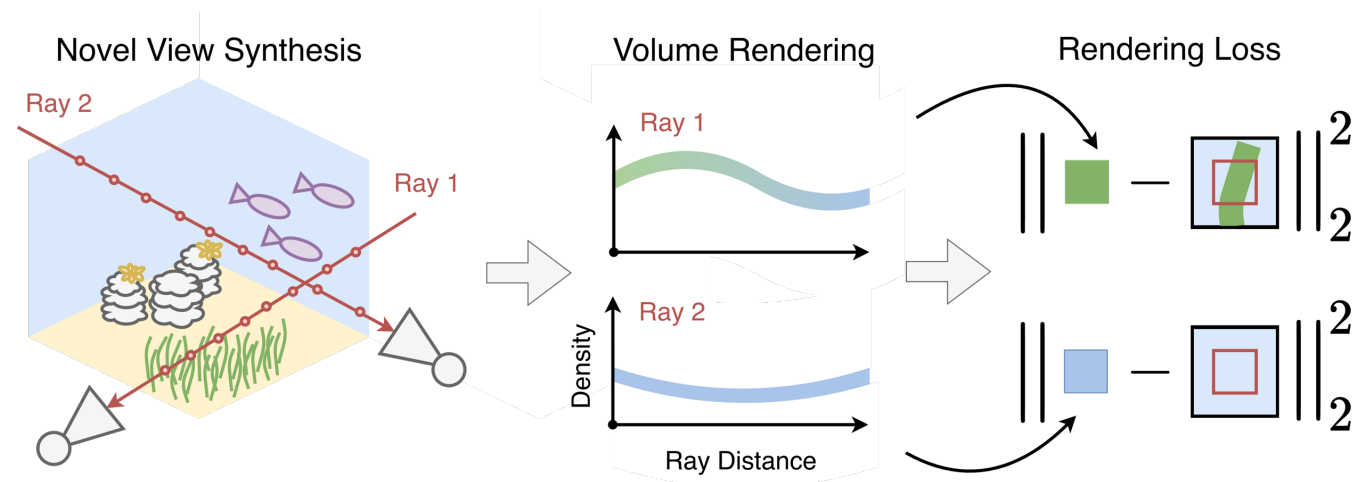
Neural Radiance Fields (NeRFs)

Paper 3 

Training

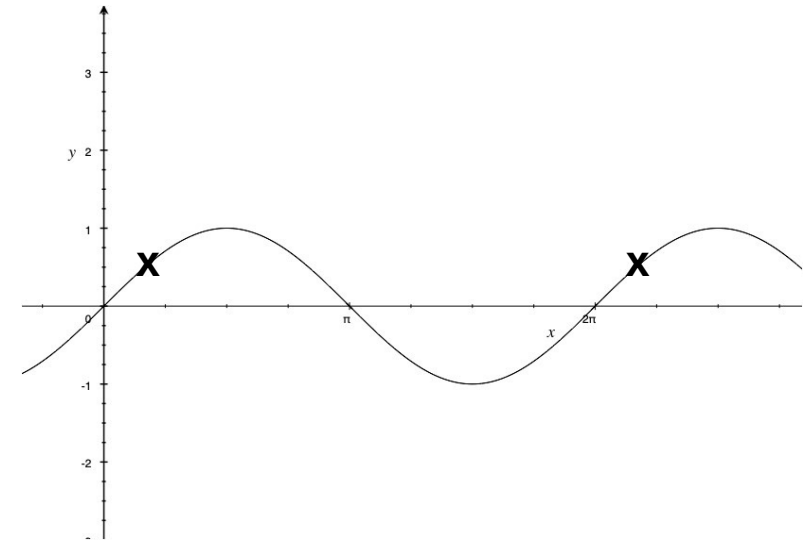
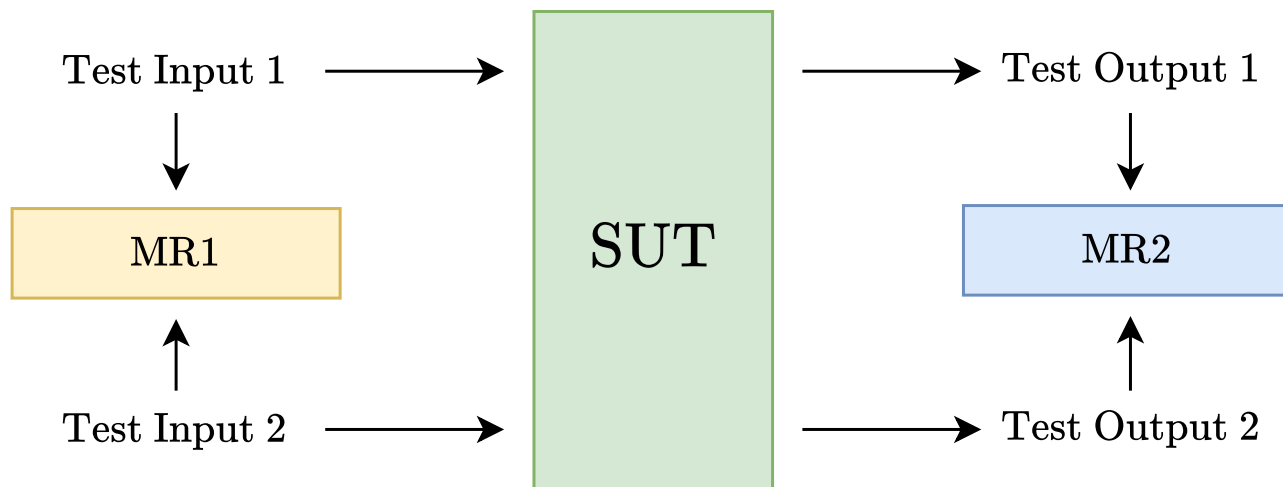


Rendering



Metamorphic Testing

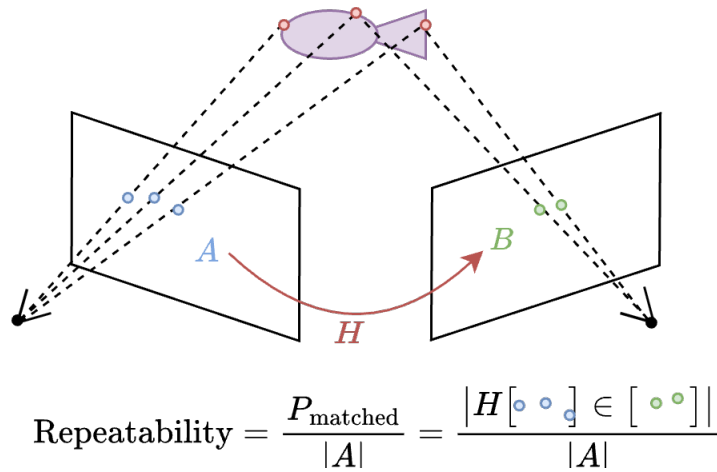
- define **metamorphic relations** between input/output pairs to test for behavioural consistency in a system under test (SUT)



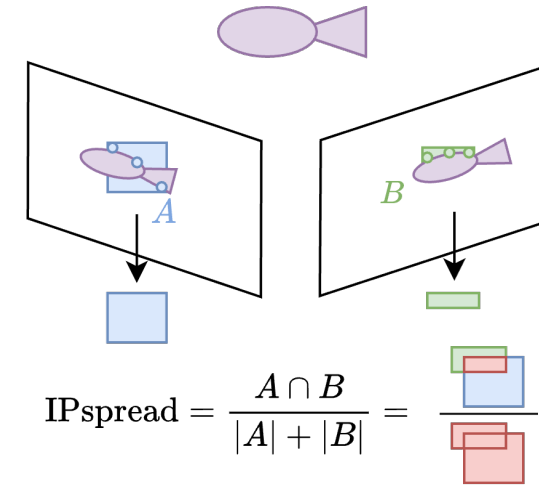
$$\sin(x) = \sin(x + 2\pi)$$

Metrics

Repeatability



Interest Point Spread



Baseline

- DeepXplore [1] mutations imitate camera failures



Dirt on Camera Lense



Loose Cable Connection



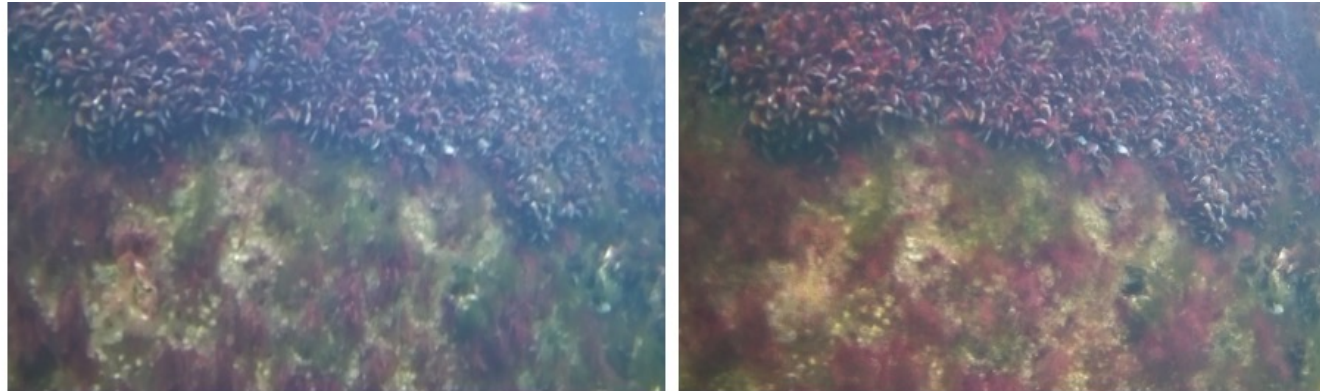
Automated Exposure
Adjustments

[1]. Pei, K., Cao, Y., Yang, J. and Jana, S., 2017, October. Deepxplore: Automated whitebox testing of deep learning systems. In *proceedings of the 26th Symposium on Operating Systems Principles* (pp. 1-18)

NeRF-to-Reality Domain Shift

- NeRF-generated images are visually close to real images

Scene 1



Scene 2



Real

NeRF

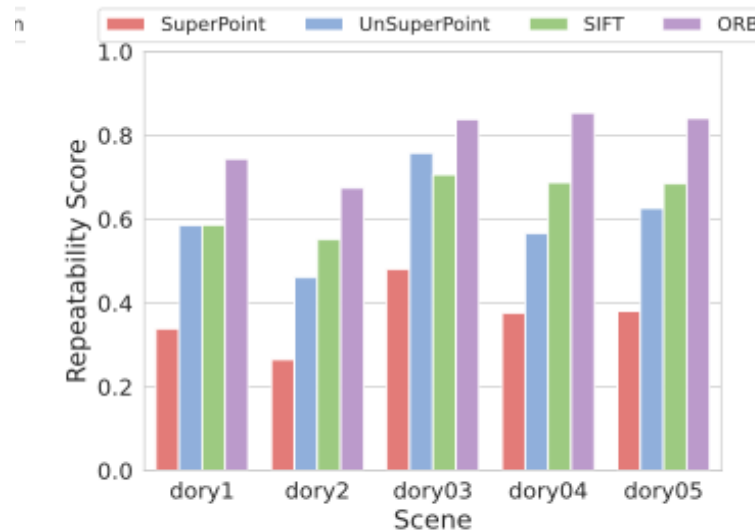
Sensitivity to Domain Shift

Paper3 

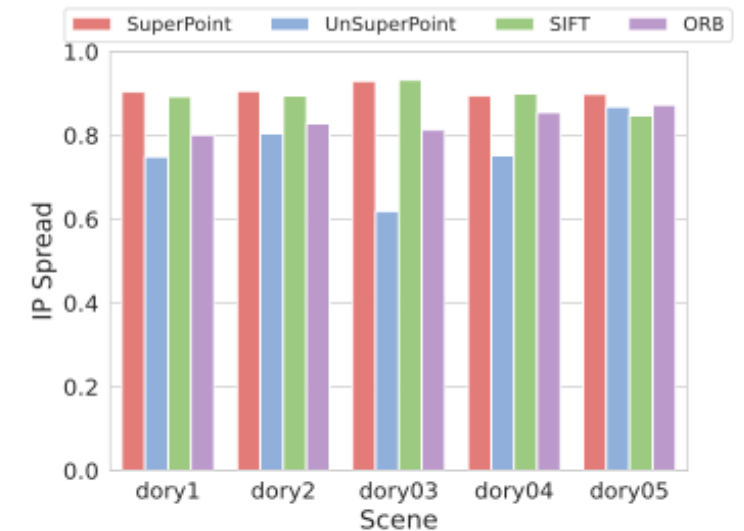
- We test 4 feature extractors on 5 underwater scenes



SuperPoint



Avg. Repeatability

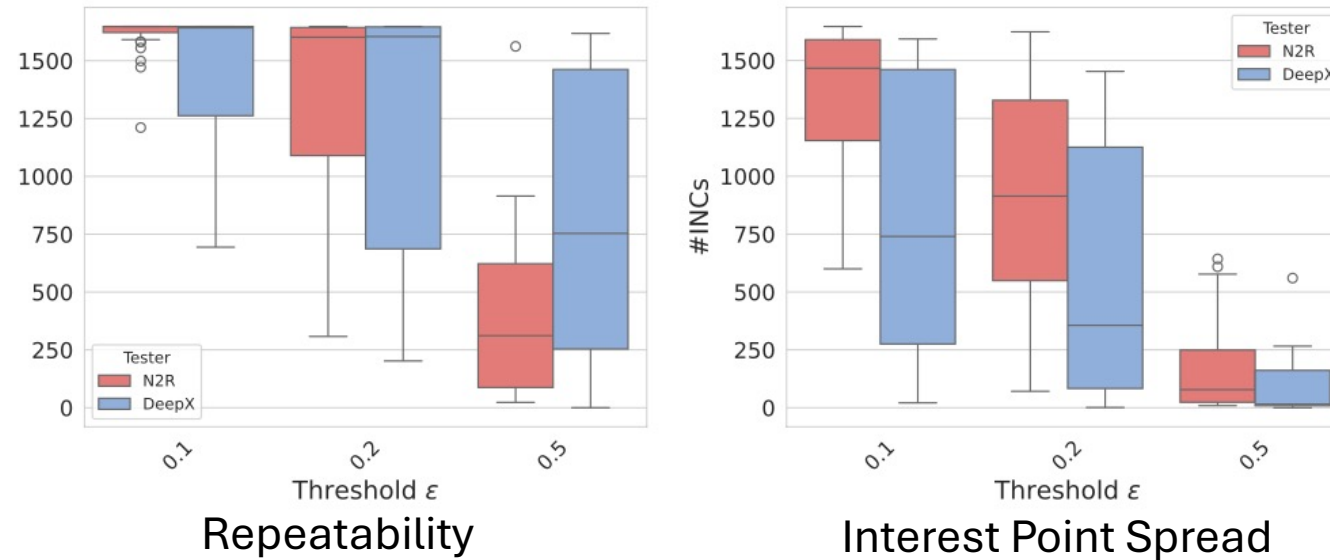


Avg. Interest Point Spread

- Key takeaway:
 - ORB shows the highest repeatability scores
 - different feature extractors have different sensitivities to NeRF images

Inconsistencies

- #Incs compared to DeepXplore mutations:



- Key takeaway:
 - N2R uncovers more inconsistencies than DeepXplore on the strictest test setting

Paper 3



NeRF-To-Real Tester: Neural Radiance Fields as Test Image Generators for Vision of Autonomous Systems

Laura Wehl¹, Bilal Wehbe², Andrzej Wąsowski¹

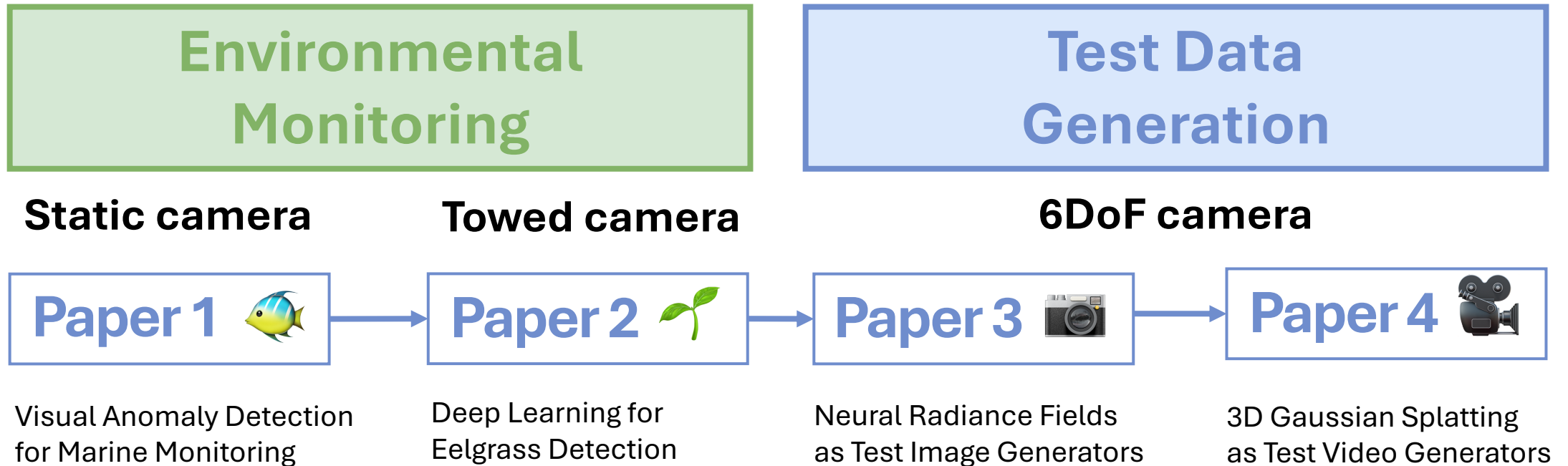
¹ITU, Copenhagen, Denmark

²DFKI, Bremen, Germany

Published at
IEEE Robotics and Automation Society
Engineering Reliable Autonomous Systems Conference, 2025
Worcester, MA, USA

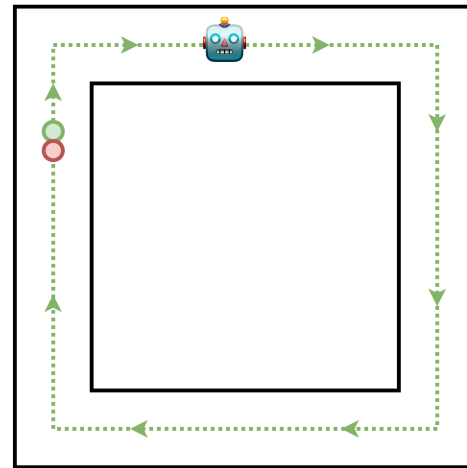


Research

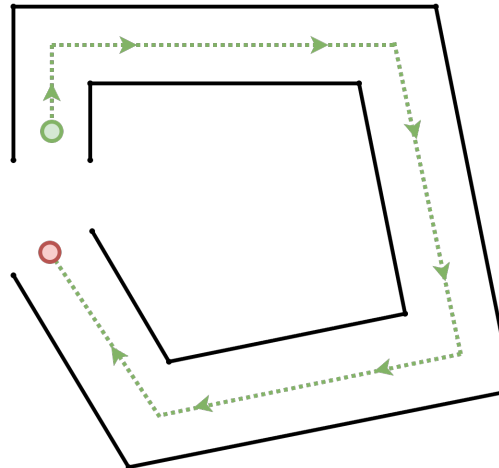


Loop Closure Detection

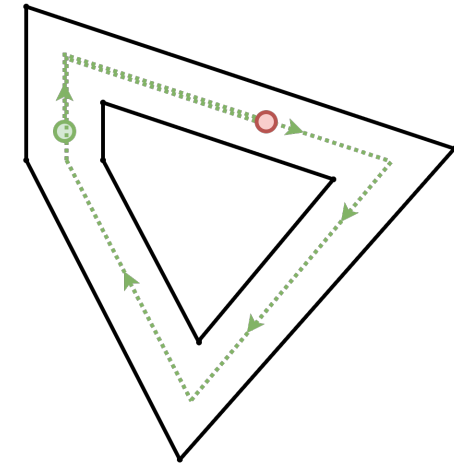
- Start
- Loop Closed
- Path



True Map



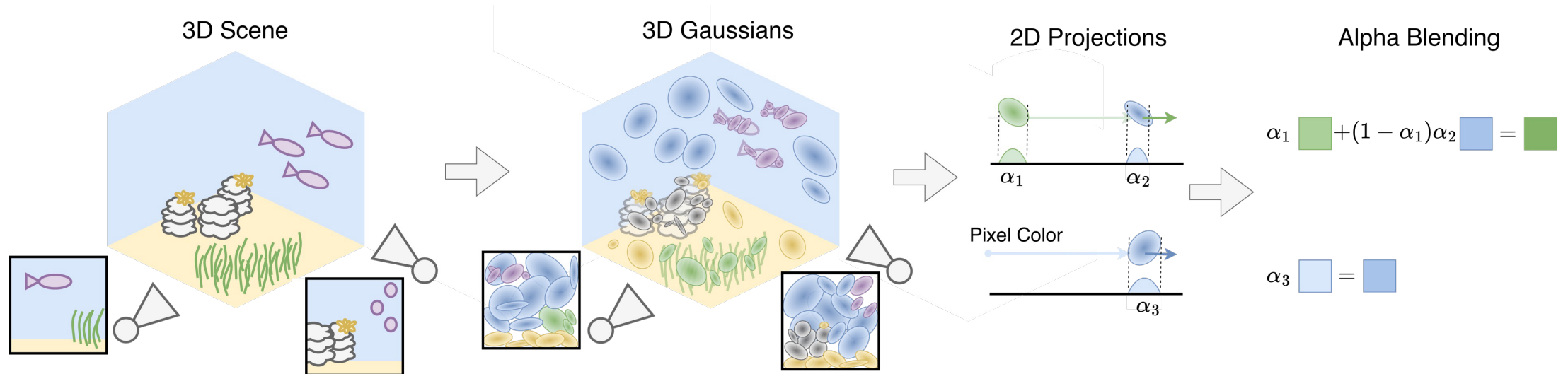
False Negative



False Positive

3D Gaussian Splatting

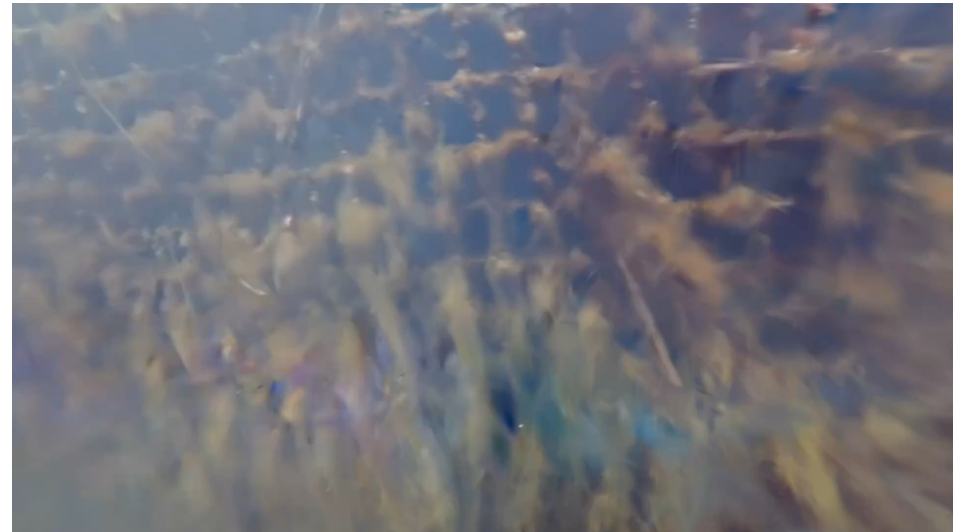
- Learns a set of 3D Gaussians



$$g_i(x) = e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}$$

3D Gaussian Splatting

- we can render videos in a 3DGS model with different loop sizes



Experimental Setup

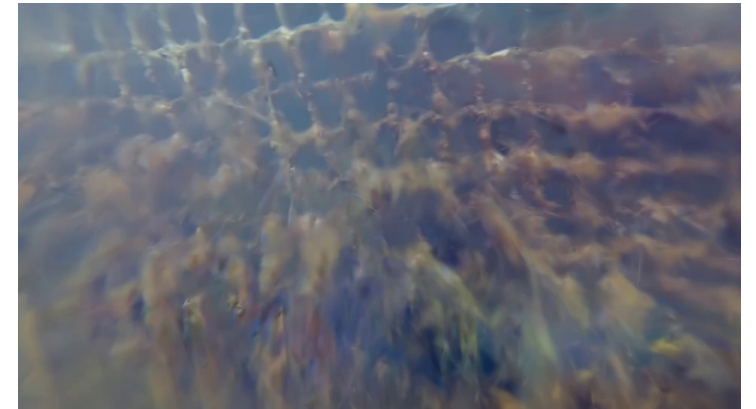
- Generated 30 videos in total in 5 loop sizes
- Generated at different intervals during training
- Evaluate 2 loop closure detection methods: iBoW, DBoW3
- Success: trajectory drift below 0.1m at the loop closure



5k steps



15k steps



30k steps

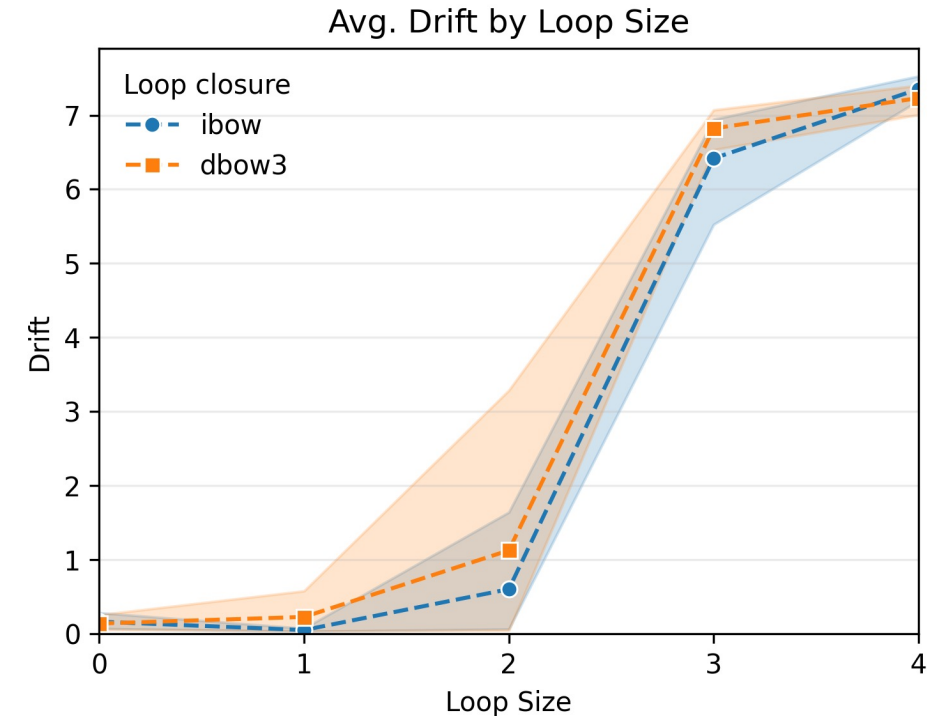
Preliminary Results

- Loop Detection
 - #Proc: number of processed runs
 - #Succ: loop detected

Loop	DBoW3				iBoW			
	#Proc.	#Succ.	Succ.% (cond.)	Succ.% (abs.)	#Proc.	#Succ.	Succ.% (cond.)	Succ.% (abs.)
0	15	10	66.7%	33.3%	18	14	77.8%	46.7%
1	8	6	75.0%	20.0%	13	11	84.6%	36.7%
2	3	2	66.7%	6.7%	7	3	42.9%	10.0%
3	6	0	0.0%	0.0%	8	0	0.0%	0.0%
4	4	0	0.0%	0.0%	10	0	0.0%	0.0%

- Key takeaways:
 - not marginal failures, LCD can break down completely
 - 3DGS can generate targeted, controllable test videos for LCD

- Drift Analysis





Gaussian Splats as Test Video Generators for Visual Loop Closure in Autonomous Underwater Vehicles

Laura Wehl¹ and Bilal Wehbe²

¹ITU, Copenhagen, Denmark

²DFKI, Bremen, Germany

Accepted for publication at
IEEE Oceanic Engineering Society
AUV2026 Conference, 2026
Southampton, UK



Conclusions

RQ1 Image Degradation

Paper 1 

Paper 2 

Paper 3 

Paper 4 

- DNNs can effectively overcome underwater image degradation
- performance remains dependent on dataset size, variability, and model architecture

RQ2 Analysis at Scale

Paper 1 

Paper 2 

- DNNs support automation of environmental monitoring
- reducing reliance on manual annotation while keeping experts in the loop

RQ3 Generating Tests

Paper 3 

Paper 4 

- NeRFs and 3DGS offer promising alternatives to simulators for test data generation
- fuse photorealism of real data with the controllability of simulation

Contributions

New data

- AURA: first multi-annotator underwater VAD benchmark, 70+ downloads
- SeagrassFinder: 8300 annotated images, 100+ downloads

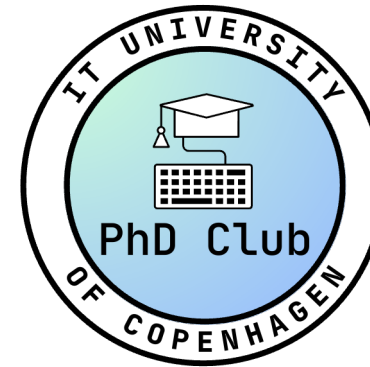
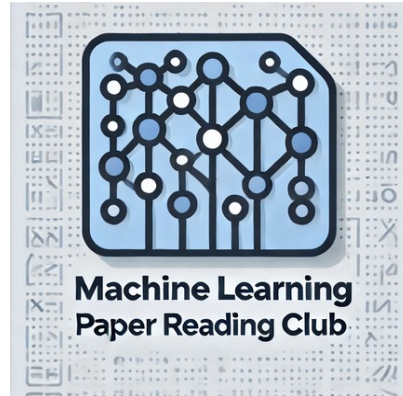
New tools

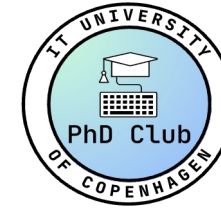
- NeRF-to-Real Tester: first use of NeRF renders as test inputs for vision components
- 3DGS video generator: first use of Gaussian splatting for navigation testing

Milestones

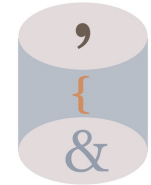
- first VAD application in marine monitoring
- first Vision Transformer for seagrass detection

Other Contributions

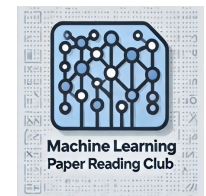




{d,s,&r}



THANK YOU!



Deep Neural Networks for Reliable Underwater Vision

PhD Defense

Laura Weihl

