

Marine Imaging Workshop 2026 - Gibraltar

Uncovering Anomalous Events for Marine Environmental Monitoring via Visual Anomaly Detection

**Laura Wehl^{1,4}, Stefan Hein Bengtson^{2,4}, Nejc Novak³,
Malte Petersen^{2,4}**

IT University of Copenhagen¹

Aalborg University²

Anemo Robotics ApS³

Pioneer Centre for AI, Denmark⁴



**AALBORG
UNIVERSITY**



**Anemo
Robotics**



**PIONEER CENTRE FOR
ARTIFICIAL INTELLIGENCE**

Environmental Monitoring

- Static cameras for long-term underwater biodiversity monitoring
- 30 secs recording every 30 mins for months
- Problem: get TB of data, but mainly empty/uninteresting videos



Biodiversity Monitoring

A screenshot of a database application window titled "Database". It features a search bar, a filter dropdown, and an export button. Below these is a table with four columns: "Date", "Video Name", "Duration", and "File Size". The table contains numerous rows of data, representing recordings from January 1st, 2024, to January 11th, 2024. Each row shows a specific recording file name, a duration of 00:30:00, and a file size ranging from 57 MB to 64 MB. The table is scrollable, indicating a large volume of data.

Date	Video Name	Duration	File Size
01/01/2024	Recording_20240101_00-30.mp4	00:30:00	62 MB
01/01/2024	Recording_20240101_30-00.mp4	00:30:00	58 MB
02/01/2024	Recording_20240102_00-30.mp4	00:30:00	64 MB
02/01/2024	Recording_20240102_30-00.mp4	00:30:00	61 MB
03/01/2024	Recording_20240103_00-30.mp4	00:30:00	58 MB
03/01/2024	Recording_20240103_30-00.mp4	00:30:00	63 MB
04/01/2024	Recording_20240104_00-30.mp4	00:30:00	60 MB
04/01/2024	Recording_20240104_30-00.mp4	00:30:00	57 MB
05/01/2024	Recording_20240105_00-30.mp4	00:30:00	65 MB
05/01/2024	Recording_20240105_30-00.mp4	00:30:00	61 MB
06/01/2024	Recording_20240106_00-30.mp4	00:30:00	58 MB
06/01/2024	Recording_20240106_30-00.mp4	00:30:00	59 MB
07/01/2024	Recording_20240107_00-30.mp4	00:30:00	63 MB
07/01/2024	Recording_20240107_30-00.mp4	00:30:00	62 MB
08/01/2024	Recording_20240108_00-30.mp4	00:30:00	60 MB
08/01/2024	Recording_20240108_30-00.mp4	00:30:00	58 MB
09/01/2024	Recording_20240109_00-30.mp4	00:30:00	64 MB
09/01/2024	Recording_20240109_30-00.mp4	00:30:00	61 MB
10/01/2024	Recording_20240110_00-30.mp4	00:30:00	66 MB
10/01/2024	Recording_20240110_30-00.mp4	00:30:00	65 MB
11/01/2024	Recording_20240111_00-30.mp4	00:30:00	57 MB
11/01/2024	Recording_20240111_30-00.mp4	00:30:00	60 MB

Huge Volumes of Data



Empty Videos

Closed-set vs. Open-set



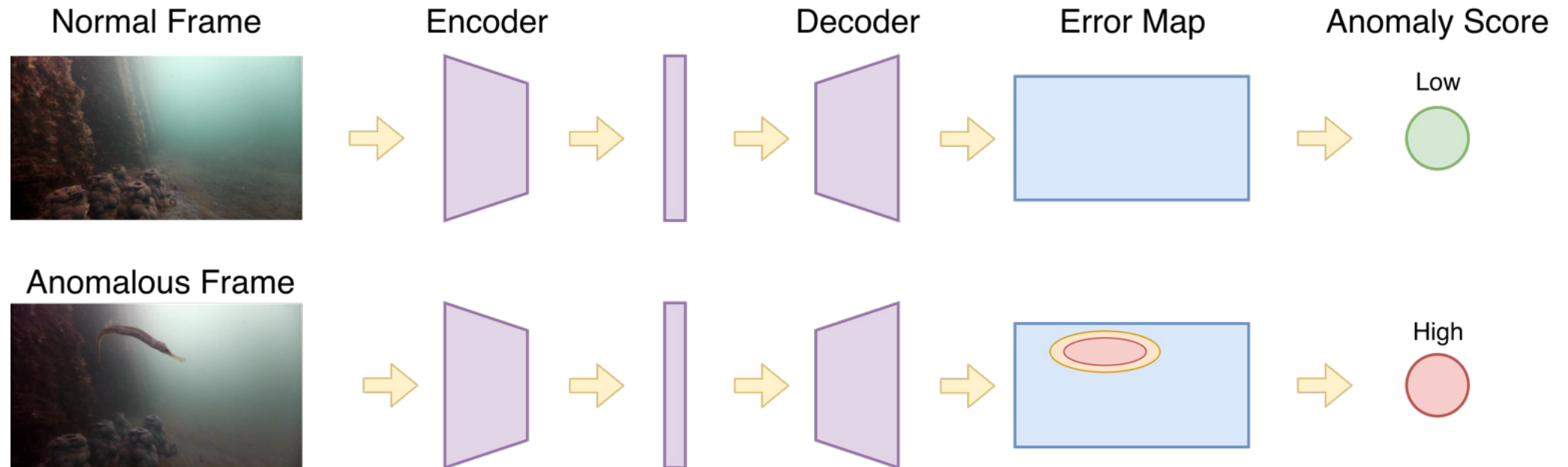
Object Detection



Segmentation

Visual Anomaly Detection (VAD)

- Train on “normal” frames of the scene



Visual Anomaly Detection



Annotation Problem

- **Challenge:** when does an anomalous event start/stop?
- We recruit 16 annotators and use soft labels

Frame 230



Soft label 0.0

Frame 370



Soft label 0.5

Frame 510



Soft label 1.0

The AURA Dataset

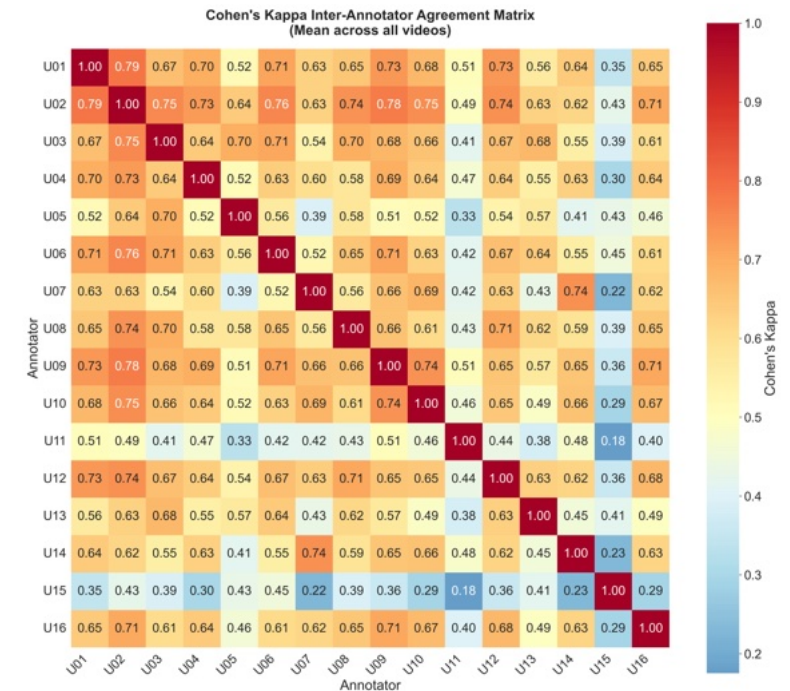
- **New Dataset: Anomalous Underwater Activity**
- 2 scenes, 25 videos, 16 annotators
- Moderate to substantial agreement



Scene A

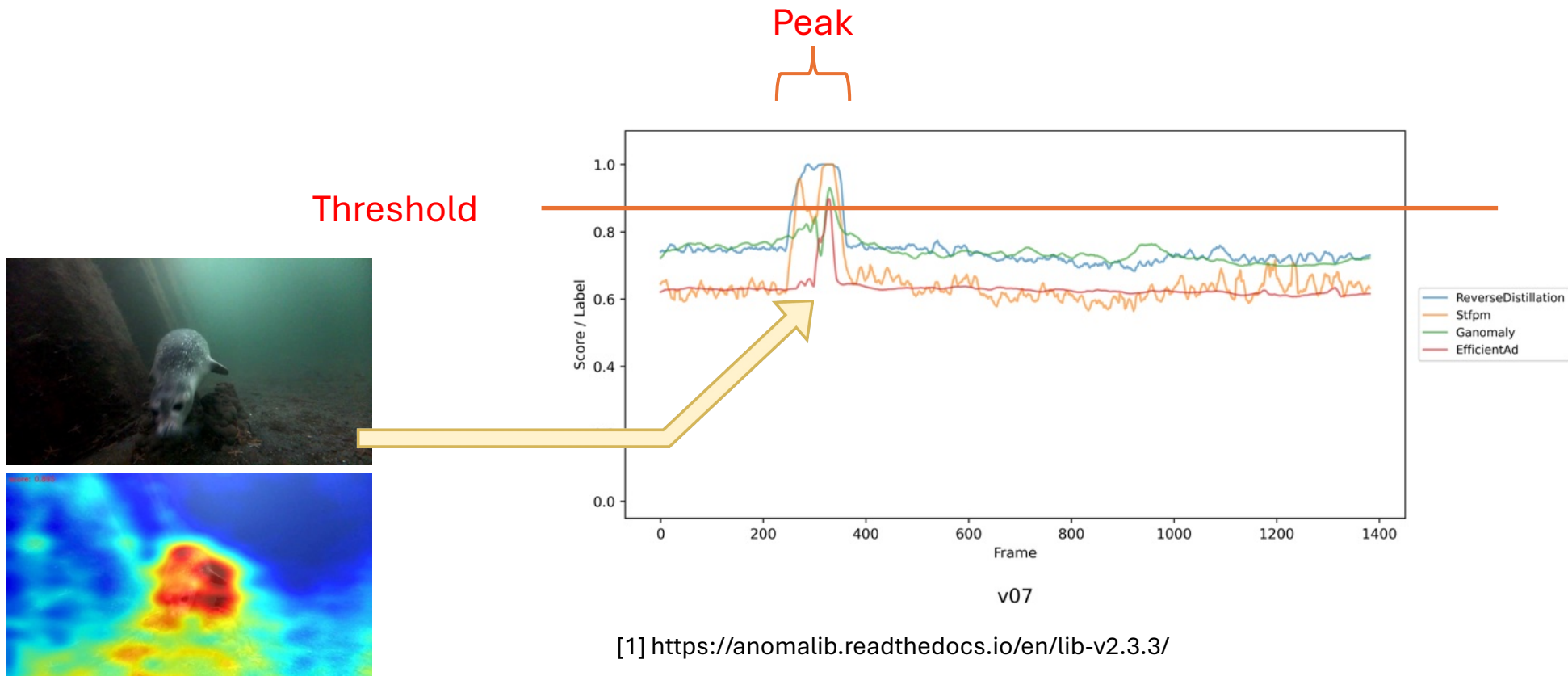


Scene B



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

Conclusions

Underwater VAD

- Consistent/artificial lighting (scene B) makes the scene easier to learn

Event Detection

- Peak-based method performs better than threshold-based extraction

Annotation Guidance

- Performance remains dependent on dataset size, variability, and model architecture



Uncovering Anomalous Events for Marine Environmental Monitoring via Visual Anomaly Detection

Laura Wehl^{1,4}, Stefan Hein Bengtson^{2,4}, Nejc Novak³,
Malte Petersen^{2,4}

IT University of Copenhagen¹

Aalborg University²

Anemo Robotics ApS³

Pioneer Centre for AI, Denmark⁴

AURA Dataset



Full Paper



Published at

ICCV Joint Workshop on Marine Vision, 2025

lw112.github.io

lawe@itu.dk