

# NjordVid: A Fishing Trawler Video Analytics Task

Tor-Arne S. Nordmo<sup>1,\*</sup>, Aril B. Ovesen<sup>1</sup>, Håvard D. Johansen<sup>1</sup>, Dag Johansen<sup>1</sup> and Michael A. Riegler<sup>1,2</sup>

<sup>1</sup>UiT: The Arctic University of Norway, Norway

<sup>2</sup>SimulaMet, Oslo

## Abstract

Fishing is one of the most important food sources globally. Commercial fishing can potentially be more efficient, precise, and accountable, and if artificial intelligence should be one of the remedies for improvement, one needs a better understanding of inner details and what processes are happening on a fishing trawler. The NjordVid task aims to encourage researchers to tackle this challenge in addition to preserving the privacy of the people working on these boats. The participants are asked to detect events in videos taken on the fishing trawler and to enhance privacy for the people visible in the videos.

## 1. Introduction

Surveillance on board fishing vessels has been argued to be a necessity for sustainable fishing practices and for our ability to fight fraud in the fishery industry [1, 2]. Fishing vessels are secluded environments where a small group of people work and live together in a constrained space, often for several weeks at a time. Introducing video surveillance in such environments, in particular combined with machine learning, has raised new privacy and data protection aspects that need to be addressed. This task provides a unique opportunity to gain insight into the inner workings of a commercial fishing vessel while at sea, its part in the food production pipeline, and the living and working conditions of the crew onboard. Understanding these elements are essential for the development and usage of practical automated surveillance systems.

With this competition we hope to achieve a better understanding of the processes that happen on a fishing trawler and in addition we want to encourage the community to work on this important topic.

## 2. Dataset

The Njord dataset [3] contains surveillance videos from the Hermes fishing trawler that were live-streamed online in 2019 as slow-TV entertainment. The dataset consists of 71 videos that have been annotated so far and 127 videos that are not annotated. The videos have a resolution of  $1,280 \times 720$  and run at 25 frames-per-second. The videos have varying lighting conditions with complex, moving backgrounds due to the trawler being at sea. The videos consist of eight different fixed-camera scenes plus a view with a manually-operated camera for showing particularly interesting events, such as whale observations and other boats. The cameras are changed between on a fixed schedule but can also be manually changed by the captain. This sometimes results in scenes having varying durations. There are overlays that sometimes appear on-screen. These show general information about what is being caught, information about the

---


*MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online*

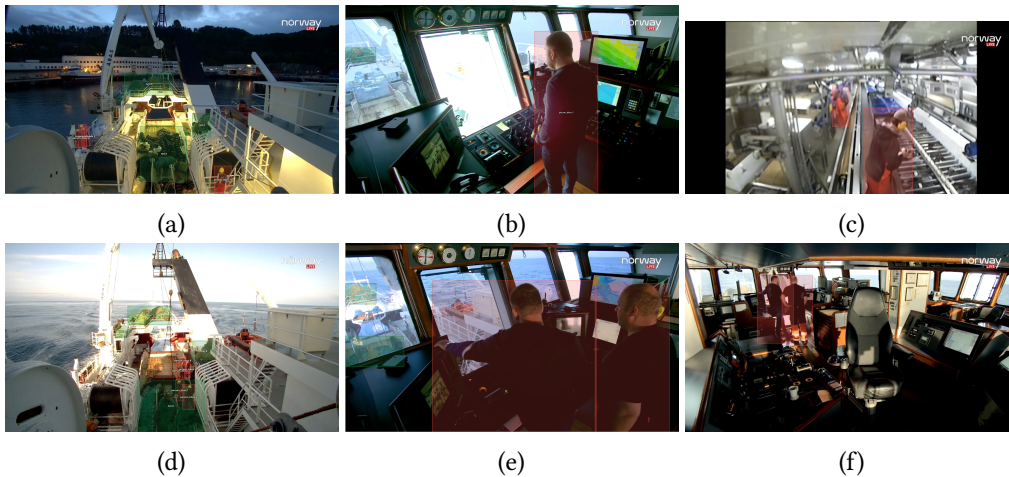
\*Corresponding author.

✉ [tor-arne.s.nordmo@uit.no](mailto:tor-arne.s.nordmo@uit.no) (T. S. Nordmo)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1:** Sample frames from the dataset with overlaying bounding box annotations.

vessel in general, and statistics related to the catch. They also sometimes show a map overlay with the current location of the trawler along with its speed and orientation.

For each video, we have labeled bounding boxes around people, other boats, nets, and fish. The temporal annotations consist of when scene changes occur, when overlays are turned on and off, when Events of Interest (EoI) occur, and when the intro plays. We also have labels that denote whether it is daytime or nighttime, and, due to the videos being from a live-stream, labels for parts of the videos that are before the introduction and after the end of the relevant live-stream. The bounding boxes for fish label groups of fish due to the scenes on deck showing fish being far away from the camera. The bounding boxes for the nets both label nets in use and those lying in heaps on deck.

The dataset is organized as follows. The videos directory contains a subdirectory for each annotated video that contains the video in .mp4 format and two annotation files, one file for the bounding box annotations and one file for the timeline annotations. The two annotation files are structured as csv-formatted files using a semi-colon as the delimiter. The bounding box contains one line per bounding box annotation with the following seven values; class, frame number, center x-position, center y-position, the bounding box's width, and the bounding box's height. The width and height have been normalized by dividing each by the video's width and height, respectively. The timeline annotation file contains one line per annotated class and includes the following two values; the class of the frame and the frame number of the corresponding video. The videos directory also contains an unannotated subdirectory containing all videos that have not been annotated yet.

The dataset Njord is publicly available under the CC BY-NC 4.0 International license.

### 3. Tasks

The NjordVid task consists of two different subtasks, which can be tackled independently depending on your research area of interest. The dataset consists of 198 surveillance videos from a fishing trawler, of which 71 are annotated with bounding boxes and temporal annotations. The goal of the task is to both gain insight into what is happening on fishing vessels and also investigate methods for preserving the privacy of the fishing crew.

**Table 1**

Baseline experiments for the detection of events and people in the development dataset.

Model	Precision	Recall	mAP_0.5	mAP_0.5:0.95
YOLOv5n	0.698	0.502	0.527	0.265
YOLOv5s	0.732	0.545	0.543	0.271
YOLOv5m	0.697	0.552	0.569	0.277
YOLOv5x	0.621	0.570	0.550	0.264

### 3.1. Subtask 1

Detection of events on the boat: The participants are asked to detect events on the boat like people moving, fish caught, etc. In addition to simple detection of the events we also ask the participants to provide an interestingness score which relates to how uncommon the event is. The score should be between 0 to 1 where 0 determines a very common event and 1 a very uncommon event.

### 3.2. Subtask 2

Privacy of onboard personnel: For this task the participants are asked to develop methods to preserve the privacy of the people working on the boat, which includes anything that can identify the person (face, nametags, etc). At the same time the privacy preserving measurements should have as little impact on the analysis as possible.

## 4. Evaluation

For the evaluation of subtask 1 we will use the standard metrics Precision, Recall, F1 score and Matthew correlation coefficient. The interestingness score provided by the participants will be used to weight the resulting scores.

For subtask 2 we will have a group of manual evaluators checking the privacy aspects on the test dataset (basically is the person still identifiable by a human observer or not). In addition we will calculate some metrics before and after the method was applied. Specifically, we will apply an object detection model and evaluate with classic regression metrics before and after the privacy-preserving method is applied.

## 5. Baseline Results

In this section we present some baseline results obtained by training a simple object-tracking model using YOLOv5 on the development dataset. Table 1 shows the performance metrics based on the ground truth given in the development dataset and Figure 1 provides some example images with resulting bounding boxes.

## 6. Discussion and Outlook

The task focuses on the exploration of a completely unknown area where automatic multimedia analysis can have an important impact. We hope that the task will lead to new insights and research questions in addition to inspiring researchers to work on this important topic. For the

future we envision a more complex and multimodal dataset that also contains sensor readings and other additional information.

We particularly thank Hermes staff and owners for relevant discussions, meetings, and for allowing us to annotate and publish the Njord dataset and use it for MediaEval.

## References

- [1] Ministry of Trade, Industry and Fisheries, Framtidens fiskerikontroll, NOU 19:21 (2019).
- [2] P. Release, Fishing rules: Compulsory CCTV for certain vessels to counter infractions, 2021. <https://www.europarl.europa.eu/news/en/press-room/20210304IPR99227/fishing-rules-compulsory-cctv-for-certain-vessels-to-counter-infractions>.
- [3] T.-A. S. Nordmo, A. B. Ovesen, B. A. Juliussen, S. A. Hicks, V. Thambawita, H. D. Johansen, P. Halvorsen, M. A. Riegler, D. Johansen, Njord: a fishing trawler dataset, in: Proceedings of the 13th ACM Multimedia Systems Conference, 2022, pp. 197–202.