Njord: A Fishing Trawler Dataset

Tor-Arne Schmidt Nordmo
UiT The Arctic University of Norway

Aril Bernhard Ovesen
UiT The Arctic University of Norway

Bjørn Aslak Juliussen
UiT The Arctic University of Norway

Steven Alexander Hicks
SimulaMet, Norway

Vajira Thambawita
SimulaMet, Norway

Håvard Dagenborg Johansen
UiT The Arctic University of Norway

Pål Halvorsen[*]
SimulaMet, Norway

Michael Alexander Riegler[†]
SimulaMet, Norway

Dag Johansen
UiT The Arctic University of Norway

Abstract

Fish is one of the main sources of food worldwide. The commercial fishing industry has a lot of different aspects to consider, ranging from sustainability to reporting. The complexity of the domain also attracts a lot of research from different fields like marine biology, fishery sciences, cybernetics, and computer science. In computer science, detection of fishing vessels via for example remote sensing and classification of fish from images or videos using machine learning or other analysis methods attracts growing attention. Surprisingly, little work has been done that considers what is happening on board the fishing vessels. On the deck of the boats, a lot of data and important information are generated with potential applications, such as automatic detection of accidents or automatic reporting of fish caught. This paper presents Njord, a fishing trawler dataset consisting of surveillance videos from a modern off-shore fishing trawler at sea. The main goal of this dataset is to show the potential and possibilities that analysis of such data can provide. In addition to the data, we provide a baseline analysis and discuss several possible research questions this dataset could help answer.

CCS Concepts

• **Computing methodologies → Machine learning**; *Cross-validation*; *Supervised learning*; • **Applied computing → Law**.

Keywords

Fishing Trawler, Surveillance, Slow TV, Machine Learning, Artificial Intelligence, Dataset

ACM Reference Format:
Tor-Arne Schmidt Nordmo, Aril Bernhard Ovesen, Bjørn Aslak Juliussen, Steven Alexander Hicks, Vajira Thambawita, Håvard Dagenborg Johansen, Pål Halvorsen, Michael Alexander Riegler, and Dag Johansen. 2022. Njord: A Fishing Trawler Dataset. In *13th ACM Multimedia Systems Conference (MMSys '22), June 14–17, 2022, Athlone, Ireland*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3524273.3532886

[*] Also affiliated with Oslo Metropolitan University, Norway
[†] Also affiliated with UiT The Arctic University of Norway

MMSys '22, June 14–17, 2022, Athlone, Ireland

ACM ISBN 978-1-4503-9283-9/22/06.
https://doi.org/10.1145/3524273.3532886

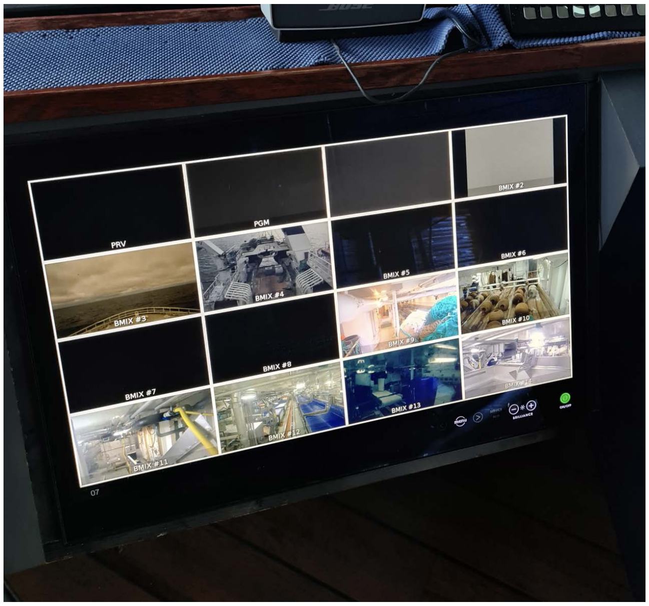


Figure 1: Example of concurrent videos stream observations on the command bridge of the Hermes trawler.

1 Introduction

A modern fishing vessel is infused with high-tech digital technologies. The bridge of a trawler operating in, for instance, the Arctic contains numerous terminals visualizing geographical position and other vessels in the vicinity, weather conditions and predictions, fish finder sonar data and the like. Video streams from the deck and production line under deck are also frequently displayed so that the officer in-charge has real-time information when making operational decisions. Figure 1 is an example of different video streams observed simultaneously on a fishing vessel. The video stream is used, for instance, in a safety context for the crew members alone on deck or working somewhere along the heavy machinery constituting a production line. Accidents in this industry are not an exception and an important problem to consider [6, 12].

The constant collection of voluminous, multimodal data on modern commercial fishing vessels leads to interesting possibilities for the application of advanced analysis methods. For example, using Artificial Intelligence (AI) to analyze this data could lead to new insights supporting more energy-efficient locations of fish to catch, sustainable catching, and a safer working environment for the fishermen. Add to this the potential such technologies can have from a resource control and global management perspective.

AI-relevant technologies are already being applied in this domain. One example is support for sustainable fishing operations [5, 7, 9], another is publishing of fish datasets relevant for developing new models in this specific domain [4, 8, 15]. Fishing vessel and boat detection is yet an example where AI technologies can replace tedious and labor-intensive manual operations [10, 13, 14]. The list is longer, but what is missing is labeled datasets from the internal activities on board a commercial fishing vessel. The work presented in this paper is a first contribution to fill this void.

We present an open and novel dataset called Njord, which was collected from cameras on a high-end commercial fishing vessel operating in the Arctic Ocean and annotated with bounding box and classification annotations. The current dataset contains 71 annotated videos and 127 videos that are not annotated from live-streams that aired in 2019. We envision that the presented dataset can lead to a myriad of new research and a better understanding of a completely unexplored but important area. The dataset is also meant to be updated over time with more videos and additional data sources.

Therefore, the main contributions of this paper are:

(1) We compile and publish a unique, fully open dataset containing surveillance video data based on live-streams from a fishing trawler. A large part of the dataset is thoroughly annotated with both bounding box and classification labels.
(2) We provide domain knowledge about the specific use case including a discussion of legal aspects and current open challenges.
(3) We provide a set of baseline machine learning experiments to benchmark the released dataset and evaluate its technical validity.
(4) We discuss and suggest possible future research directions and application scenarios using the dataset.

The remainder of this paper is organized as follows. In Section 2, we describe how we have structured the dataset. Then, in Section 3, we describe the details of what the dataset contains and how it was collected and annotated. This is followed by Section 4, which gives an overview of legal aspects surrounding the prospect of surveillance in a fishing trawler scenario and datasets in general. After this, we describe potential applications and usage scenarios for the dataset in Section 5. In Section 6, we then describe suggested metrics that are relevant for the dataset and perform some baseline experiments. Finally, we conclude and describe some future work in Section 7.

2 Dataset Structure

The dataset is organized as follows. The root directory contains a *readme.txt* file and a *videos* directory. The *readme.txt* file gives a brief description of the included data and annotations. 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* 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 a *unannotated* subdirectory containing all videos that have not been annotated yet.

3 Dataset Details

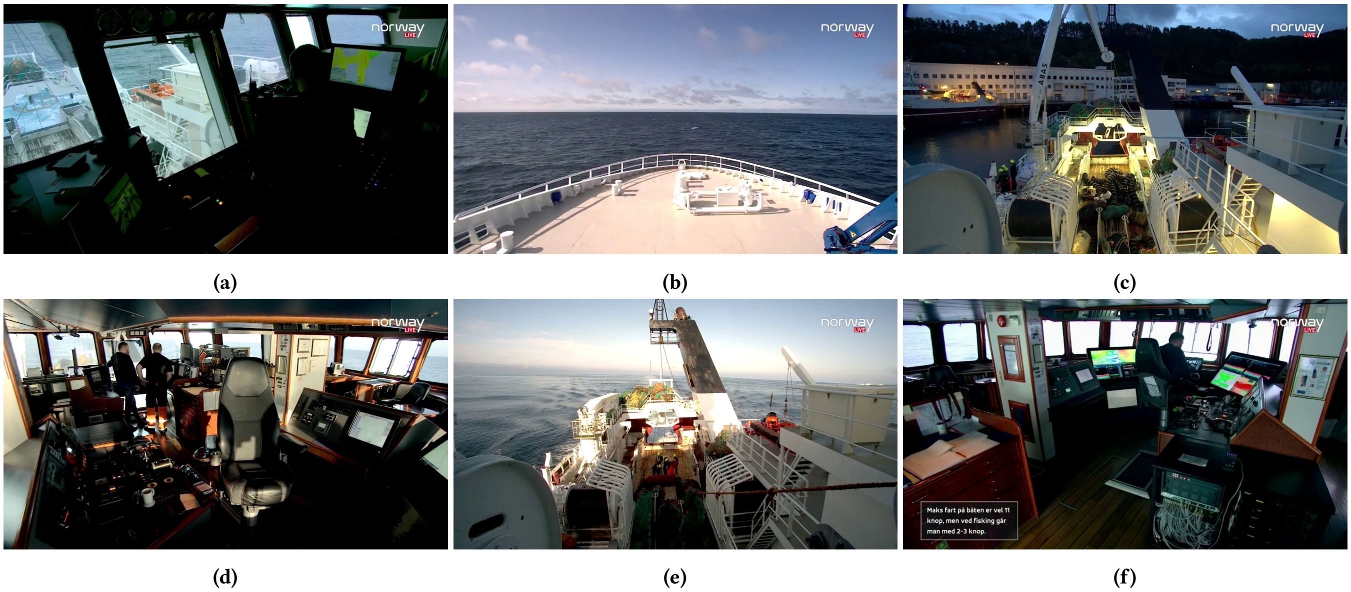
As previously described, the Njord dataset contains surveillance videos from the *Hermes*[1] fishing trawler that were live-streamed online in 2019 as "Slow TV" [1] entertainment. The videos are from a trip from the western shores of Greenland to Norway, documenting their fishing journey. There are a total of 29 live-stream videos that are, on average, 1 hour in duration. These were downloaded from YouTube using the `youtube-dl` CLI tool. They were then split up into 10-minute segments to be easier to deal with both in the labeling and the benchmarking process. At the time of submission, we have annotated a subset of these. This results in a dataset with 71 videos that have been annotated so far and 127 videos that are not annotated. 71 annotated videos, each with a frame rate of 25 fps and a duration of approximately 10-minutes, results in approximately $1,065,000$ frames with annotations. The videos have a resolution of $1,280 \times 720$ and run at 25 fps. 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 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 of it.

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 labels were manually created using Labelbox [3]. Labelbox is a platform for annotating datasets. It has a simple interface that allowed us to label bounding boxes and temporal annotations. For the bounding boxes, it linearly interpolates between keyframes, allowing for a faster annotation operation.

The dataset is anticipated as continuously growing and expanding (in terms of annotations, but also amount of data), and currently,

[1] https://www.hermesas.no


norway LIVE
Maks fart på båten er vel 11 knop, men ved fisking går man med 2-3 knop.

(a) (b) (c)

(d) (e) (f)

Figure 2: Sample frames from different videos of the dataset. (a) - A view from the bridge looking down at the deck, (b) - A view from the front of the vessel, (c) - A view of the deck as the trawler moves from port, (d) - A view of the bridge, (e) - A view of workers on deck, and (f) - A view of the bridge from another angle with an overlay.

it contains 71 fully annotated videos and 127 videos without annotations. The not annotated video can also be useful for unsupervised or self-supervised learning experiments.

Njord is licensed under Creative Commons Attribution Non-Commercial 4.0 International (CC BY-NC 4.0), and is available for download at https://doi.org/10.5281/zenodo.6284673.

4 Legal aspects

A fishing vessel is a secluded environment where people often work and live for several weeks at a time. Introducing video surveillance and video surveillance combined with machine learning in such an environment has privacy and data protection aspects. Besides fundamental privacy rights, the use of surveillance cameras on board vessels needs to comply with European Data Protection Regulations and emerging European AI regulations. Article 4 (1) of the General Data Protection Regulation (GDPR) defines personal data as "any information relating to an identified or identifiable natural person". A picture of a natural person in a surveillance video stream could identify the person and would fall under the definition of personal data in the GDPR. The GDPR requires the processor of personal data to have a valid legal basis (consent, performance of a contract, a vital interest of the data subject, a legal obligation or a public interest) for the processing to be lawful. The lawfulness of the processing of surveillance video data on board a fishing vessel depends on the purpose of the processing. Processing to prevent accidents would likely need to rely on another legal basis than processing to prevent and deter illegal fishing. The legal basis for the processing would also depend on whether, for instance, a fishing company or a public control authority is the processor under Article 4 (8) GDPR.

A proposal for an *Artificial Intelligence Act (AIA)*[2] is currently negotiated in the European Parliament. If surveillance video data is combined with machine learning to detect anomalies in the video stream, the system would be included in the definition of an AI system in Article 3 (1) of the act. The proposed act classifies AI systems after their purpose, where systems that pose a risk of adverse impact on fundamental rights are subject to stricter requirements. An AI system intended to be used by law enforcement authorities for risk assessments or crime analytics is defined as high-risk AI systems under the AIA Article 6 (2). If surveillance video data from fishing vessels is combined with machine learning to report and prevent infringements of fishing regulations, the system might be included in the definition of a high-risk AI system. A high-risk AI system is required to comply with risk assessing procedures throughout its lifetime. Article 10 of the AIA lays down specific quality criteria for datasets applied in high-risk AI systems. According to Article 10 (3) and (4), training, validation and testing datasets shall be "relevant, representative, free of errors and complete." Moreover, datasets applied in high-risk systems shall take into account the characteristics or elements that are particular to the specific geographical, behavioral or functional setting within the high-risk AI system is intended to be used.

Both European data protection regulations and the emerging specific AI regulation in the European Union, in essence, are assessments of the proportionality of the interference in natural persons rights balanced against the purpose of the processing of personal data and the purpose of the AI system. The principle of proportionality requires an interference in a fundamental right, such as the

[2] Regulation of the European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts COM(2021) 206 Final.

right to privacy or data protection, to pursue a legitimate aim, be necessary and suitable to the aim pursued, and the interference and the fundamental right must pass a balance test. For the interference to be proportionate, the reasons to interfere must outweigh the interference in the fundamental right. The lawfulness of a surveillance system on fishing vessels will therefore depend on the aim pursued by the video surveillance, security for the workers, prevention of accidents, automatic documentation of catches or prevention and deterrence of illegal fishing etc., and the extent to which the privacy and data protection rights of the individuals working on board are affected.

Due to these concerns, it can be interesting to look at the proportion of the videos containing people. Based on the videos currently annotated, 44.8% of the frames contain a person. Depending on the use-case, it can potentially be possible to only utilize frames that do not contain people, however, in other use-cases, such as when analyzing fishing procedure, they are crucial. Therefore, looking into anonymization approaches can be useful.

5 Applications and Usage Scenarios

The purpose of publishing this unique dataset is to motivate the machine-learning research community to explore new aspects of the fishing domain. As a starting point, we foresee this dataset to have several applications and usage scenarios. A few examples are presented in the following:

- General object detection with complex backgrounds and lighting conditions;
- automatic documentation of the fishing procedure;
- surveillance of persons and their activities;
- privacy research; and
- detection of Events-of-Interest (EoIs).

Due to the complexity of the backgrounds and the varying lighting conditions, this dataset can be used as a difficult benchmark for object detection. There are multiple scenes where people and other objects overlap. All of this results in a complex scene where an object detection algorithm can be put to the test.

Automatically documenting the progress of the fishing procedure, i.e., from catching the fish to processing and storing of the catch, can be useful for process efficiency. From the different scenes of the videos, one can see all of the different stages of this pipeline. Learning what to focus on in the different scenes can be useful for optimizing the fish processing procedures. With areas of fish being labeled, detecting catch biomass can also be an interesting endeavour. Provided one manages to get a relatively accurate measurement, this can provide an indicator whether the trawler is fishing within their quotas or not.

Surveillance of the fishing vessel and the crew is important, specifically to ensure that action is taken when accidents occur, for example, if a fisherman falls or a net falls on top of them. Surveillance can also document whether a proper procedure regarding the handling of catch and bycatch (i.e., a catch of species that are not allowed to be caught) is being followed. In Norway, all catch and bycatch need to be brought to shore. In addition, the data can also be used to explore privacy aspects of surveillance data and algorithms for privacy-related research by, for example, using it to learn how to obfuscate faces, etc.

Table 1: Evaluation results of the baseline experiments.

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

Considering these videos are from slow-TV live-streams, it might be interesting to detect highlights/events that are the most interesting parts (i.e., EoIs). This could be used to build models that can be used to notify viewers if something interesting happens, etc. Highlight detection has been done previously in sports [11].

Finally, due to the sheer volume of data and labels, applying machine learning pipelines on this dataset is non-trivial. Therefore it can be interesting, from a systems point-of-view, to explore different approaches on how to deal with such voluminous data.

To showcase one possible use case for the data, we perform a set of baseline experiments using the object detection scenario mentioned above in Section 6.

6 Example Use Case Experiments

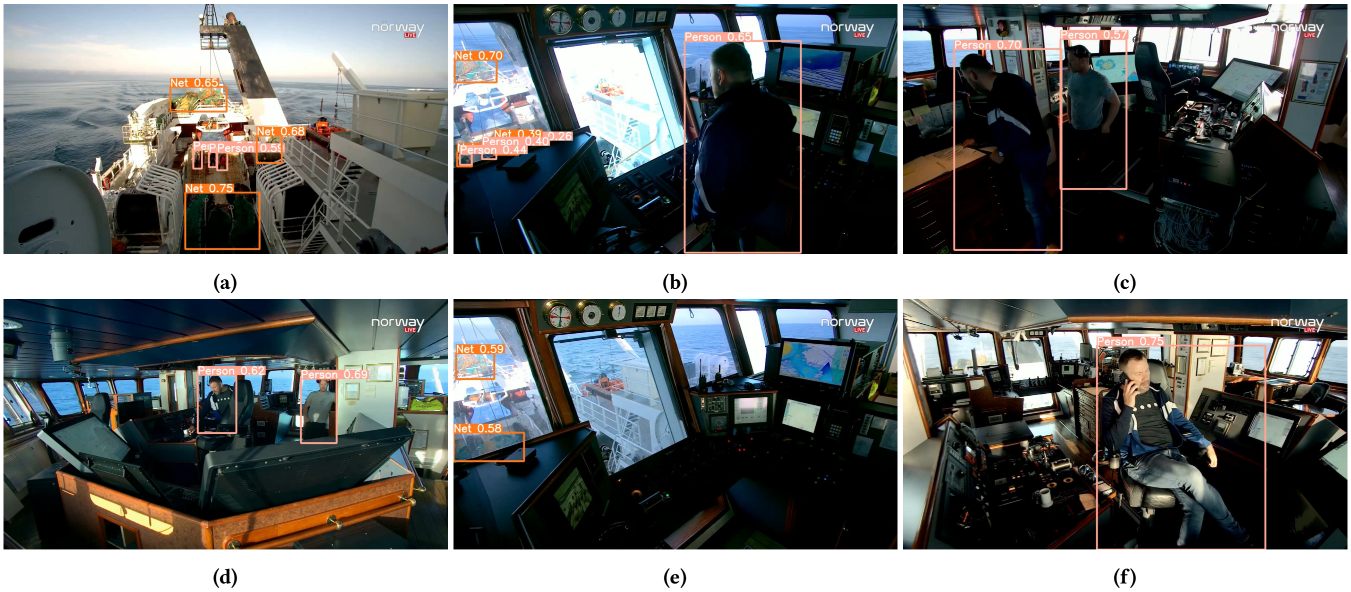
Together with the development and collection of the presented dataset, we performed a series of experiments meant to create a baseline for future researchers to measure against. The experiments use the bounding box annotations included in the dataset to detect specific interest points in the videos. Specifically, we aim to detect people, fishing nets, fish, and passing boats. As the dataset is made up of two different types of annotations, the appropriate metrics used to measure predictive performance vary based on the task. For detection (bounding box prediction), metrics such as precision, recall, and mean average precision (mAP) are most appropriate. For the timeline annotations, classification metrics such as precision, recall, f1-score, and Matthews correlation coefficient should be used.

Not all videos contain bounding box annotations and, for this experiment, we ended up using 58 of the 71 videos. Each of these videos consists of approximately $15,000$ frames. To speed up processing, we decided to only analyze only frame per second of video content, resulting in approximately $6,000$ frames per video.

We use a YOLOv5-based [2] object detection approach, for which the implementation used to perform all experiments is presented in the dataset's official GitHub repository[3] and is based on the official YOLOv5 implementation[4]. We experimented using four different versions of the YOLOv5 architecture (YOLOv5s, YOLOv5n, YOLOv5m, and YOLOv5x), where transfer learning was performed from the official coco weights that are included in the aforementioned YOLOv5 repository. Each model was trained for a maximum of 300 epochs, stopping if the model did not improve $mAP(0.5 : 0.95)$ on the validation dataset for the previous 10 epochs. All experiments were performed using three-fold cross-validation to ensure that each data sample was used in both training and validation. The experiments were run on what can be considered consumer-grade

[3] https://github.com/simula/njord

[4] https://github.com/ultralytics/yolov5



(a) (b) (c)

(d) (e) (f)

Figure 3: Sample predictions made by the YOLOv5m model. (a) - we see several workers on deck where the model is able to detect the workers in addition to the nets, (b) - The captain overlooking the workers on deck for which both the captain and workers are detected, (c) - A view from the trawler bridge with two workers detected by the model, (d) - A view of the bridge with a detected worker and captain, (e) - A view of the deck where the model detects the nets, and (f) - The captain sitting on the bridge which is detected by the model.

hardware consisting of an RTX 3090 Nvidia GPU and an Intel i9 CPU.

According to the results presented in Table 1, the best $mAP(0.5:0.95)$ value of 0.277 and $mAP(0.5)$ of 0.569 are achieved by the YOLOv5m model. However, the YOLOv5s model can obtain a higher precision with a value of 0.732, while the YOLOv5x model shows the best recall with a value of 0.570. Examples of the predictions made by the best model (YOLOv5m) according to the mAP values are depicted in Figure 3. The baseline results are promising and show that the dataset can be used to perform interesting analysis tasks. Nevertheless, the results are far from perfect and future work is needed.

In terms of detection speed, it took on average 192.22 frames per second using the smallest model (YOLOv5n) and on average 86.35 frames per second using the largest one (YOLOv5x). For training, the YOLOv5n model took on average 34 minutes to train per fold, and the YOLOv5x model trained for approximately 2 hours per fold. We see the potential for improvement and several areas for interesting research questions, especially considering the large amount of data being processed.

7 Conclusion and Future Work

In this paper, we describe a novel dataset from the commercial fishing domain, which has not been explored yet. Datasets have previously been published related to sustainable fishing and boat and fish detection, but not specifically about activities and processes happening on board fishing vessels. With this dataset, we contribute to opening up this new and challenging domain of rapidly growing interest. We present a baseline experiment on object detection using the dataset, which shows promising results but holds potential for improvement. In addition, we point at several possible research directions that can take advantage of the Njord dataset.

Specifically for the presented dataset, we will continue to annotate the remaining 127 videos and update the dataset continuously that we have available and also extend the dataset further with new data. There might be other objects or features of these videos that could be interesting to have annotated. Due to the videos coming from a live-stream, the captain sometimes addresses the audience and gives general information about where they are, what the plans are for that day, and other general information. Labeling this could be interesting for general natural language processing tasks such as question answering or translation. For the fish annotations, it would be interesting to have fine-grain annotations such as segmentations per fish that could, for example, be used to train models that automatically approximate the biomass of the entire catch.

Acknowledgements

We particularly thank Hermes staff and owners for relevant discussions, meetings, and for allowing us to annotate and publish this dataset. This work is partially funded by the Research Council of Norway project number 274451 and Lab Nord-Norge ("Samfunnsløftet").

References

[1] Gerard Gilbert. 2014. *Slow Television: The latest Nordic trend.* https://www.independent.co.uk/arts-entertainment/tv/features/slow-television-chess-trains-and-knitting-9122367.html

[2] Glenn Jocher. 2020. *ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements.* https://doi.org/10.5281/zenodo.4154370

[3] Labelbox. 2022. *Labelbox*. https://labelbox.com

[4] Daoliang Li, Qi Wang, Xin Li, Meilin Niu, He Wang, and Chunhong Liu. 2022. Recent advances of machine vision technology in fish classification. *ICES Journal of Marine Science* (2022).

[5] Mi-Ling Li, Yoshitaka Ota, Philip J Underwood, Gabriel Reygondeau, Katherine Seto, Vicky WY Lam, David Kroodsma, and William WL Cheung. 2021. Tracking industrial fishing activities in African waters from space. *Fish and Fisheries* 22, 4 (2021), 851–864.

[6] Alihan Mermer, TÜRK Meral, and Zafer Tosunoğlu. 2022. Occupational health and safety in large-scale fishing vessels registered in Aegean ports. *Ege Journal of Fisheries and Aquatic Sciences* 39, 1 (2022), 18–23.

[7] Jaeyoon Park, Jungsam Lee, Katherine Seto, Timothy Hochberg, Brian A Wong, Nathan A Miller, Kenji Takasaki, Hiroshi Kubota, Yoshioki Oozeki, Sejal Doshi, et al. 2020. Illuminating dark fishing fleets in North Korea. *Science advances* 6, 30 (2020), eabb1197.

[8] Alzayat Saleh, Issam H Laradji, Dmitry A Konovalov, Michael Bradley, David Vazquez, and Marcus Sheaves. 2020. A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports* 10, 1 (2020), 1–10.

[9] Monique Simier, Jean-Marc Ecoutin, and Luis Tito de Morais. 2019. The PPEAO experimental fishing dataset: Fish from West African estuaries, lagoons and reservoirs. *Biodiversity Data Journal* 7 (2019).

[10] Paolo Spagnolo, Francesco Filieri, Cosimo Distante, Pier Luigi Mazzeo, and Paolo D'Ambrosio. 2019. A new annotated dataset for boat detection and re-identification. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–7.

[11] Kaiyu Tang, Yixin Bao, Zhijian Zhao, Liang Zhu, Yining Lin, and Yao Peng. 2018. AutoHighlight : Automatic Highlights Detection and Segmentation in Soccer Matches. In *2018 IEEE International Conference on Big Data (Big Data)*. 4619–4624. https://doi.org/10.1109/BigData.2018.8621906

[12] Jiangping Wang, Anthony Pillay, YS Kwon, AD Wall, and CG Loughran. 2005. An analysis of fishing vessel accidents. *Accident Analysis & Prevention* 37, 6 (2005), 1019–1024.

[13] Tianwen Zhang, Xiaoling Zhang, Xiao Ke, Chang Liu, Xiaowo Xu, Xu Zhan, Chen Wang, Israr Ahmad, Yue Zhou, Dece Pan, et al. 2021. HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), 1–22.

[14] Tianwen Zhang, Xiaoling Zhang, Xiao Ke, Xu Zhan, Jun Shi, Shunjun Wei, Dece Pan, Jianwei Li, Hao Su, Yue Zhou, et al. 2020. LS-SSDD-v1. 0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images. *Remote Sensing* 12, 18 (2020), 2997.

[15] Yue Zhang, Masato Yamamoto, Genki Suzuki, and Hiroyuki Shioya. 2022. Collaborative Forecasting and Analysis of Fish Catch in Hokkaido from Multiple Scales by Using Neural Network and ARIMA Model. *IEEE Access* (2022).