Njord: An out-in-the-wild real world fish vessel catch analysis dataset

Riegler, Michael A. - Presenting author[1]; Johansen, Dag - Co-author[2]; Juliussen, Bjørn Aslak - Co-author[2]; Schmidt Nordmo, Tor-Arne - Co-author[2]; Ovesen, Aril Bernhard - Co-author[2]; Halvorsen, Pål - Co-author[3]; Johansen, Håvard D. - Co-author[2]; Rui, Jon Petter - Co-author[4]

[1]SimulMet & University of Tromsø¸, [2]UiT University of Tromsø, [3]SimulaMet & OsloMet, [4]University of Bergen

Abstract text

Introduction
Catch surveillance is an integral part of managing the fishery industry. Most surveillance system today are, however, set up to record a wide view of the boat deck. Considering the privacy rights of the fishermen, we argue that the focus of surveillance must shift from recording personnel towards recording the product: the landed fish.
Accurate automated analysis of fish caught on boats is a key factor for an effective and trustful illegal fishing detection system. The big challenge is that existing publicly available dataset of fish images are superficial and not really representing the real-world conditions on fishing boats. Thus, they are not suitable for training high-quality automatic detection systems.
A public high-quality dataset would also help attract more researchers to join the effort, promoting open and reproducible results that can be applied also outside Norway.

Current Datasets
We have performed and extensive search for datasets containing fish images. Though we found several types of fish related datasets, none depicted fish as it is when caught on the boats. The images we found typically focus on single fish, often under specific light and position settings. These are not suitable for training a system that can work on a boat under real circumstances.
Further, ground truth data needed for training, like the weight or size of the fish, is not included.
Finally, the fishes in the datasets we found are often from warmer regions which are different from fish living in the Norwegian sea. Our main finding from the survey on current dataset are that they are not large enough, do not depict real world conditions and do not contain the information needed for our purpose.

Data
From the insights of the previous presented survey we concluded that there is a need for such a dataset and that this dataset needs to be collected under specific conditions. Thus, we suggest the following procedure for the collections of the data.
First, we need to make sure that the data collected is acquired under and open license that allows sharing (Creative Commons). We also propose making the dataset available on the national open research data repository DataverseNO, hosted by UIT.
The data of the caught fish should be videos or images. Important is that the videos or images are taken on the boat. There should be only fish in the data (no humans). Moreover, all additional meta data available should be collected.