Fish detection and classification using deep learning

1st Kameswari Devi Ayyagari
Faculty of Computer Science
Dalhousie University
Halifax, Canada
devi.ayyagari@dal.ca

2nd Corey Morris
Department of Fisheries and Oceans
St. John's, Canada
corey.morris@dfo-mpo.gc.ca

3rd Joshua Barnes
National Research Council Canada
St. John's, Canada
joshua.barnes@nrc-cnrc.gc.ca

4th Christopher Whidden
Faculty of Computer Science
Dalhousie University
Halifax, Canada
cwhidden@dal.ca

***Abstract*—Deep learning models have shown great results at detecting and classifying objects in other domains like medical imaging and robotics, but deep learning is still underutilized in ocean applications such as analyzing underwater video. Working collaboratively with NRC and DFO, we aim to bridge this gap and develop deep learning methods and best practices designed for studying fish health and biodiversity.**

The goals of this research are twofold a) Detection and classification of different species of fish in the underwater video camera data gathered by DFO b) Quantification of the number of samples of different species required to train an efficient classification and detection model.

This abstract presents preliminary results from applying the widely used YOLOv4 object detection and classification model to two species of fish in the dataset. This work is a first step towards better understanding biodiversity and fish health and the impact of human activity such as seismic testing on fish.

***Index Terms*—Object detection and classification, underwater video camera data, fish detection and classification**

I. Introduction

Analyzing underwater video data is expensive, time-consuming, manually intensive, and requires lot of domain expertise. Underwater video data also poses challenges like low light conditions, occluded fish, and background confusion 1 due to vibrant seabed structures that deep learning models have been proven to be sensitive to.

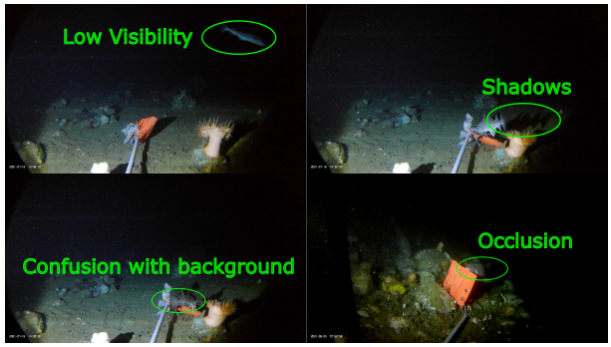


Fig. 1: Challenges in the underwater video camera dataset

This work is funded by NSERC, OFI, NRC and DFO.

The primary objective of the conducted research is to successfully detect and classify different species of fish in underwater video data. A secondary objective of this research is to determine the number of samples of different species of fish required to develop an efficient and accurate deep learning model to inform data collection efforts.

In this preliminary research, we present the results from training a model to successfully classify and detect 2 species of fish: Cod and Striped Wolffish. We argue that 1000 samples of each class are sufficient to train a model that can successfully detect and classify Cod and Striped Wolffish.

II. Related Work

Although deep learning has been widely applied to data in various domains extensively in the last few years, its applicability to underwater video data remains severely underutilised. Species classification has been the major focus in this domain [2]. Although [9] showed that deep learning can be applied to unconstrained underwater video data, their work focused on a fish/no-fish binary detection and classification model. Other works that use deep learning algorithms for fish detection and classification use data acquired from high resolution cameras [7] or data acquired in controlled environments [4] or use different kind of data like acoustic data [3].

[6] describes the camera infrastructure used to acquire this data and explored the feasibility of applying machine learning algorithms to this dataset by classifying and detecting different species of fish in a subset of 200 frames. This research, in its preliminary form extends [6]'s work to the entire dataset, and quantifies the number of each species of fish required to train an efficient classification and detection model.

III. Definitions

A. Intersection over Union (IoU)

IoU is the ratio of area of overlap and area of union between the ground truth annotation and the model's prediction.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{1}$$

B. Mean Average Precision (mAP)

Mean average precision is the average of the average precision(AP) of each class. AP is computed as the area under the precision-recall curve.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{2}$$

where N is the number of classes.

IV. Materials and Methods

The dataset has been acquired by placing cheap inexpensive cameras at an approximate depth of 450-500m along the marine slopes of the Northeast Newfoundland marine refuge. 89 videos, recorded at 30fps, have been annotated by a fish expert at DFO using the VIAME [8] platform. The annotated dataset has 8 species of fish, with the class distribution described in I. The 89 videos are partitioned into training, validation and test splits with a a 75-15-15 ratio, based on the distribution of the frequency of frames with fish in each video.

Species Name	Count
Cod	18052
Roughhead Grenadier	5830
Striped Wolffish	5459
Thorny Skate	2081
Spintail Skate	875
Wolffish	789
Redfish Mentella	172
Turbot	34

TABLE I: Species and their frequencies in the dataset

Yolov4 [1], a standard deep learning classification and detection model, is successfully trained to classify and detect two species of fish: Cod and Striped Wolffish, with pretrained weights from MSCOCO dataset [5] .

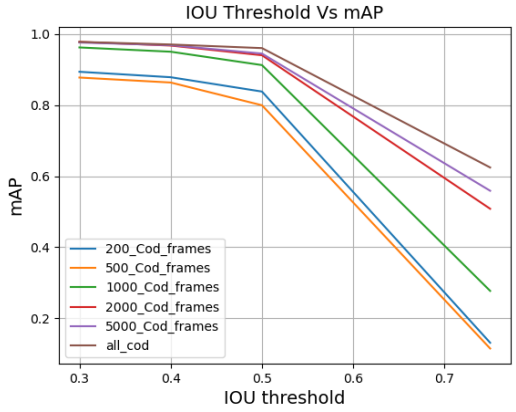
To quantify the number of samples of each species of fish required to train an efficient classification and detection model, subsets of 200, 500, 1000 and 2000 frames of each species of fish are a) randomly and b) sequentially subsampled from the dataset and performance is evaluated on a) validation set with only Cod frames and b) validation with Cod frames and water frames.

V. Results and Discussion

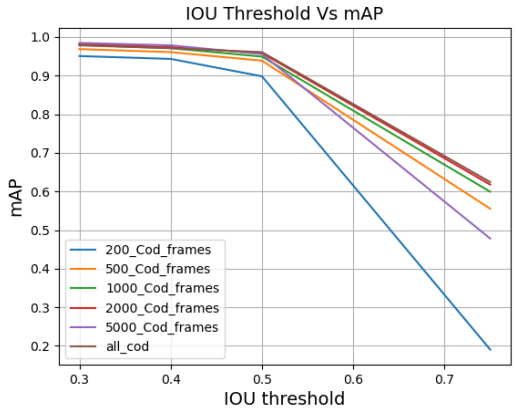
For simplicity, we start by training yolov4 models to classify and detect Cod, and extend the experiments to Striped Wolffish. To compare the performance of the trained models, we observe mean average precision (mAP) across different Intersection over Union (IoU) thresholds.

A. Training on Cod frames; Validation on Cod frames

We expect the performance of supervised learning models to improve with increase in the number of training samples. Contradictorily, in 2a, we observe that the model trained with 500 Cod performs worse than the model trained with 200 Cod. We attribute this degradation in performance to the poor video quality in one of the two videos used to train the 500 Cod model. To mitigate the potential bias that can be caused by a single low quality video and curate a training set that is more representative of the test set, subsets of 200, 500, 1000 and 2000 frames are randomly sampled from the training set. Fig 2b shows the mAP of the randomly subsampled models across different IoU thresholds.



(a)



(b)

Fig. 2: a) Performance evaluation of models trained on sequentially subsampled Cod frames and validated on Cod frames. b) Performance evaluation of models trained on randomly subsampled Cod frames and validated on Cod frames

B. Training on Cod frames; Validation on water frames

The average ratio of frames with water (frames without any annotated objects) to the frames with fish in the dataset is 5.6:1. To test the performance of the trained models on water frames, we curate a validation set with Cod frames and water frames. Fig 3 plots the mAP of models trained on randomly subsampled sets of Cod and validated on Cod and water frames. 10% reduction in mAP is observed.

C. Training with water frames; Validation on water frames

We hypothesise that adding water while training will reduce false positives and improve the performance of the trained models on the validation set with water frames. We train

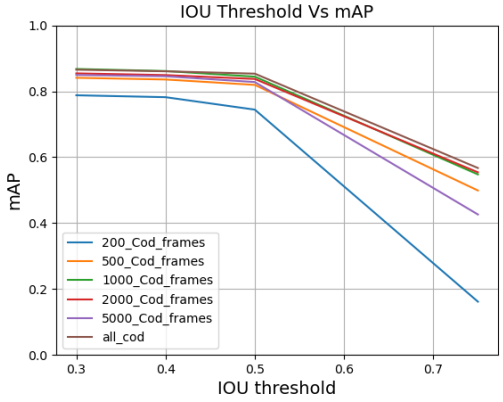


Fig. 3: Performance evaluation of models trained on randomly subsampled Cod frames and validated on Cod and water frames

models by adding 25, 50, 100 and 200 % water frames with Cod frames during training. Although, reduction in mAP is still observed on average, as shown in fig 4a, adding water frames improves the performance of the models trained on smaller subsets. We also note that the models trained with little(25%) water and the models trained with lot(200%) of water do better than other variations.

Similar experiments were performed adding another species, Striped Wolffish to the training set: 4b.

VI. Conclusion

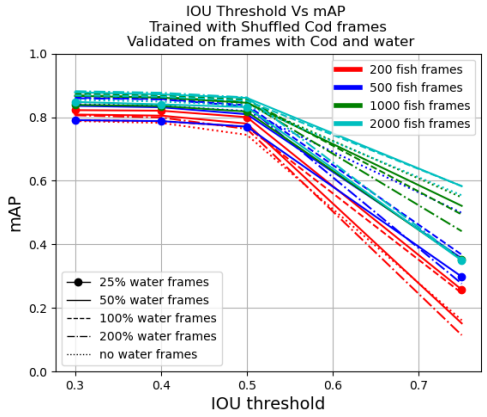
We successfully trained a yolov4 model to classify and detect Cod and Striped Wolffish in deep underwater camera data. We conclude that 1000 samples of each species of fish, and an IoU threshold of 0.4 are ideal parameters to efficiently classify and detect these species of fish in this dataset.

Furthermore, we learn that the models trained with just fish do not generalise well when validated on water frames and adding water frames while training is in itself not sufficient enough to train a model that performs well on the validation set with water frames.

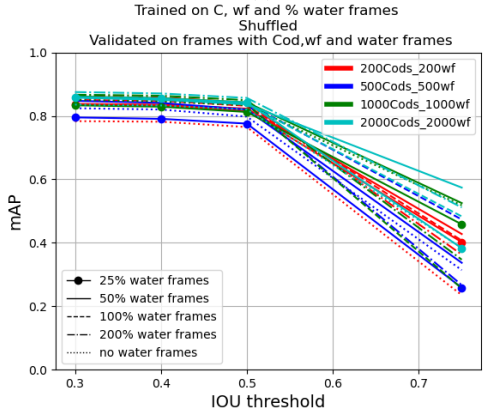
VII. Future work

We will extend our work to detect and classify all 8 species of fish and determine the number of samples of each species of fish required for efficient classification and detection.

Automating detection and classification of fish species will enable monitoring of marine ecosystems at a much larger scale than is manually feasible. This will serve as the first step towards better understanding biodiversity and fish health and the impact of human activity such as seismic testing on fish.



(a)



(b)

Fig. 4: Performance evaluation of models trained on randomly subsampled fish and water frames and validated on fish and water frames. a) Cod and water frames, b) Cod, Striped Wolffish and water frames.

References

[1] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *ArXiv*, abs/2004.10934, 2020.

[2] Muhammad Ather Iqbal Hussain, Zhi-Jie Wang, Zain Ali, and Shazia Riaz. Automatic fish species classification using deep convolutional neural networks. *Wireless Personal Communications*, 116, 01 2021.

[3] Vishnu Kandimalla, Matt Richard, Frank Smith, Jean Quirion, Luis Torgo, and Chris Whidden. Automated detection, classification and counting of fish in fish passages with deep learning. *Frontiers in Marine Science*, 8, 2022.

[4] Jia-Hong Lee, Mei-Yi Wu, and Zhi-Cheng Guo. A tank fish recognition and tracking system using computer vision techniques. In *2010 3rd International Conference on Computer Science and Information Technology*, volume 4, pages 528–532, 2010.

[5] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2014.

[6] Joshua Barnes Dustin Schornagel Christopher Whidden Morris, Corey and Phillippe Lamontagne. Measuring effects of seismic surveying on groundfish resources off the coast of newfoundland, canada. *Journal of Ocean Technology*, 16(3):57–63.

[7] Alzayat Saleh, Issam H. Laradji, Dmitry A. Konovalov, Michael Bradley, David Vazquez, and Marcus Sheaves. A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports*, 10(1):14671, Sep 2020.

[8] VIAME Contributors. VIAME: Video and Image Analytics for Marine Environments, 5 2017.

[9] Dian Zhang, Noel E. O'Conner, Andre J. Simpson, Chunjie Cao, Suzanne Little, and Bing Wu. Coastal fisheries resource monitoring through a deep learning-based underwater video analysis. *Estuarine, Coastal and Shelf Science*, 269:107815, 2022.