Using machine learning for automatic fish species identification from underwater images

Vaneeda Allken
Institute of Marine Research

INSTITUTE OF MARINE RESEARCH
HAVFORSKNINGSINSTITUTTET

Fish stock assessment

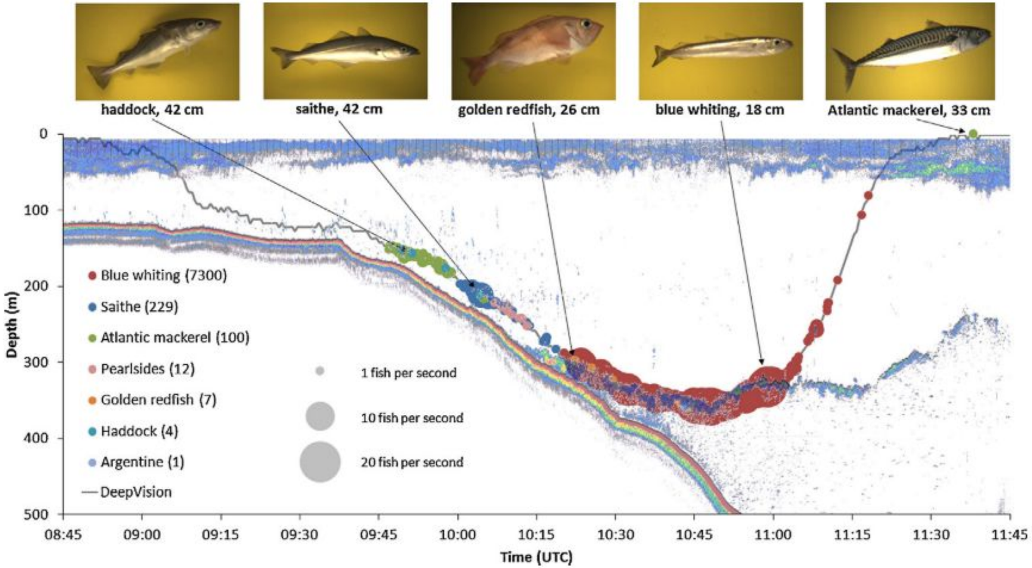
Deep Vision camera

May 2017: Norwegian Spring spawning herring survey

Fish stock assessment

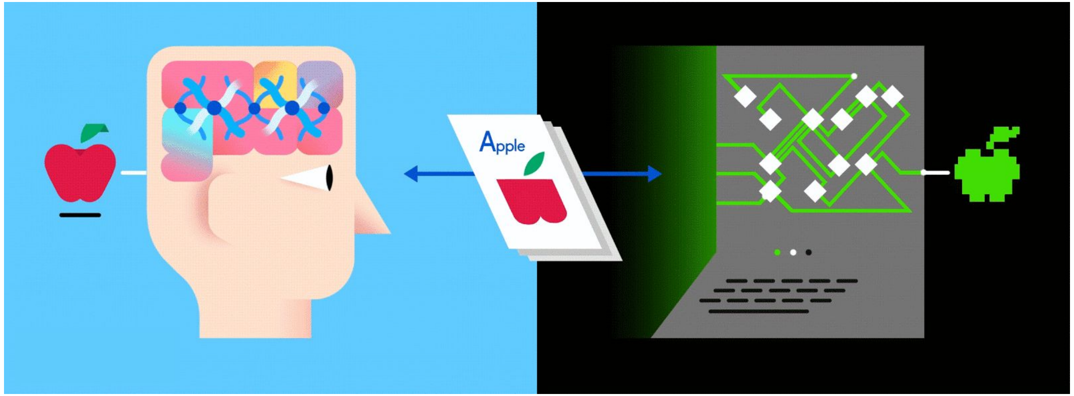
Deep Vision system

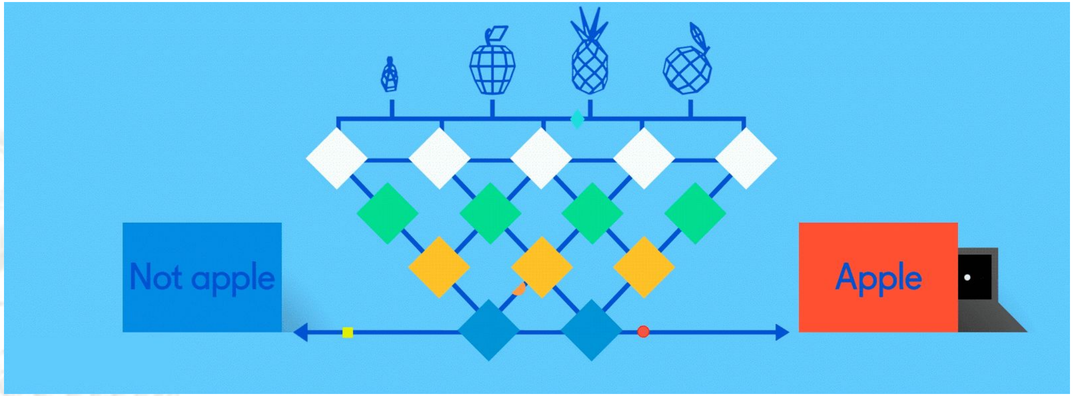
- Underwater trawl camera
- Images at 100 ms interval



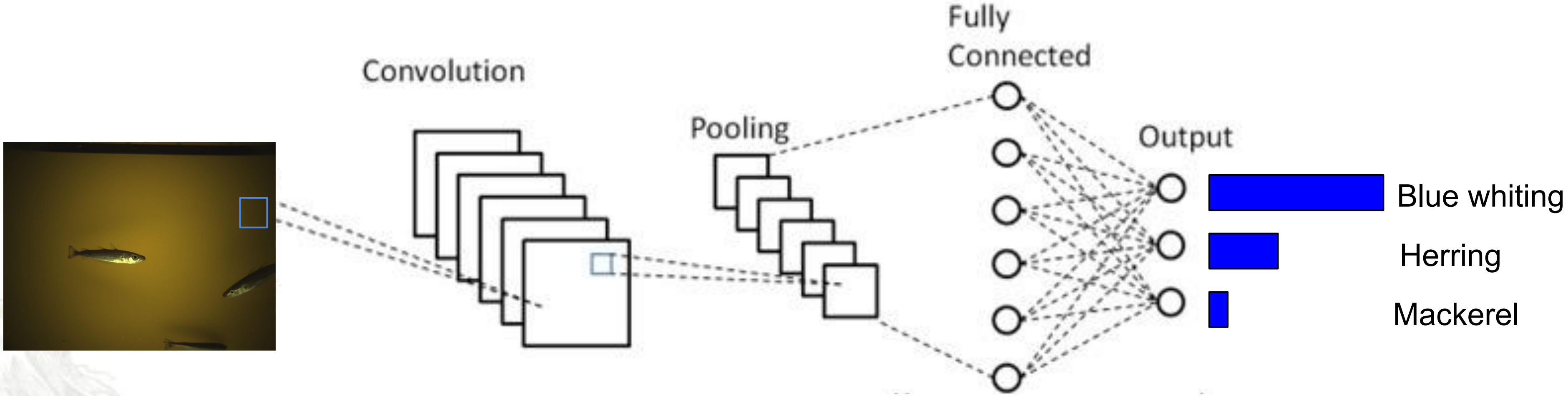
Source: *DeepVision in-trawl imaging: Sampling the water column in four dimensions*. Shale Rosen & Jens Christian Holst. *Fisheries Research, 2013*

Image classification



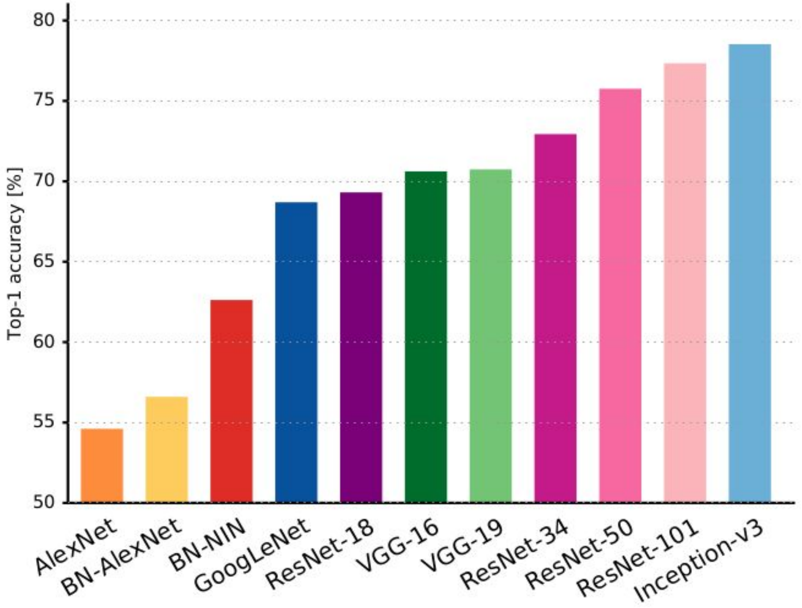


Fish image classification

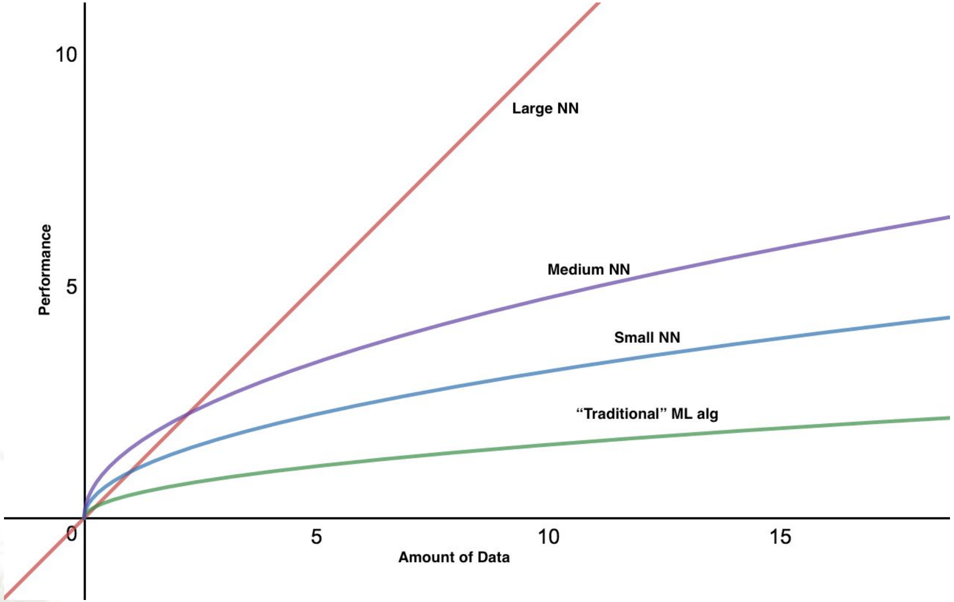


Choice of model architecture

We use TensorFlow with a Keras implementation of Inception-v3



Performance vs amount of data



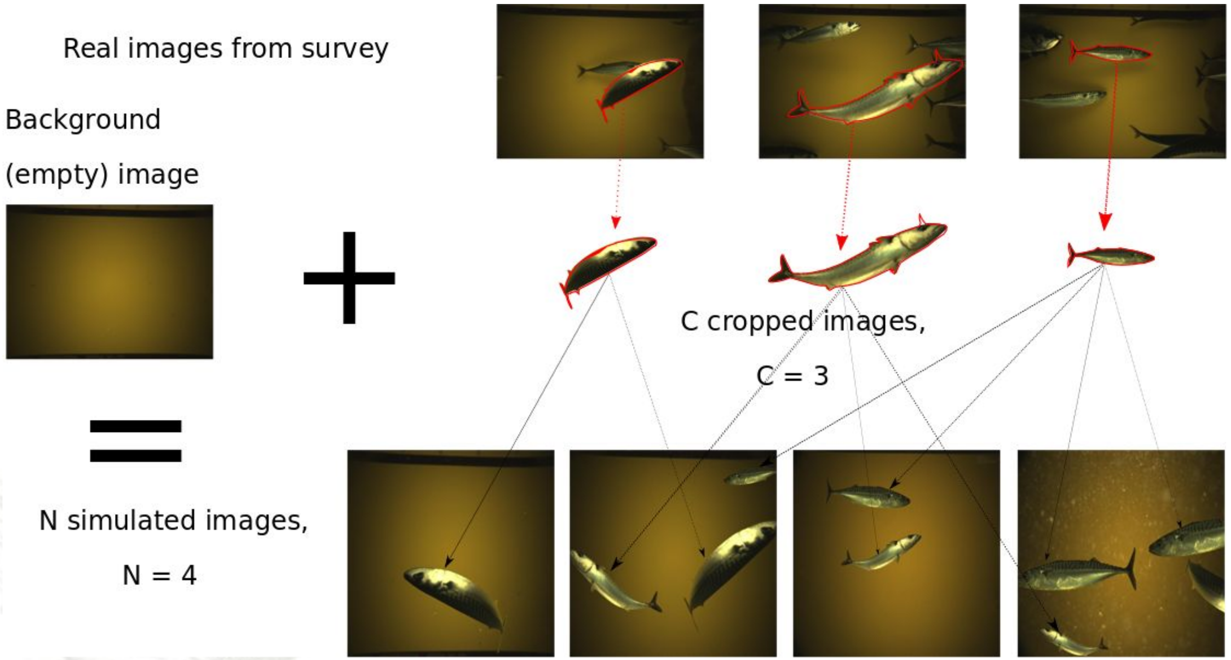
Coping with limited labelled data

Training dataset: 210 images (70 per species)

Validation/test set: 600/2400 images

- Transfer learning
  - Finetuning a model pretrained on ImageNet
- Data augmentation
  - Includes rotation, translation, shearing, flipping and zooming
- Simulating training data

Generating synthetic images



Coping with limited labelled data

Training dataset: 210 images (70 per species)

Validation/test set: 600/2400 images

- Transfer learning
  - Finetuning a model pretrained on ImageNet
  - **Test accuracy (with TL) = 64%**

Coping with limited labelled data

Training dataset: 210 images (70 per species)

Validation/test set: 600/2400 images

- Transfer learning
  - Finetuning a model pretrained on ImageNet
  - **Test accuracy (with TL) = 64%**

- Data augmentation
  - Includes rotation, translation, shearing, flipping and zooming
  - **Test accuracy (with TL + Data Augmentation) = 76.5%**

Coping with limited labelled data

Training dataset: 210 images (70 per species)

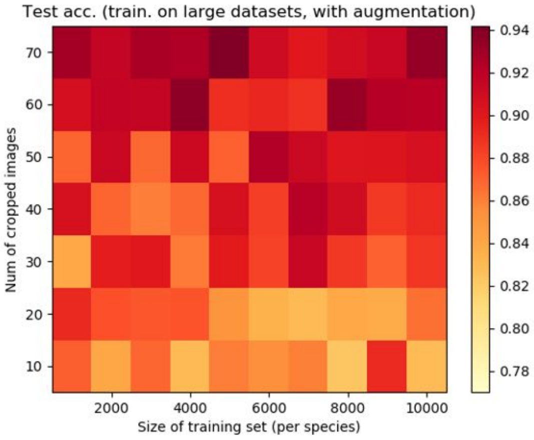
Validation/test set: 600/2400 images

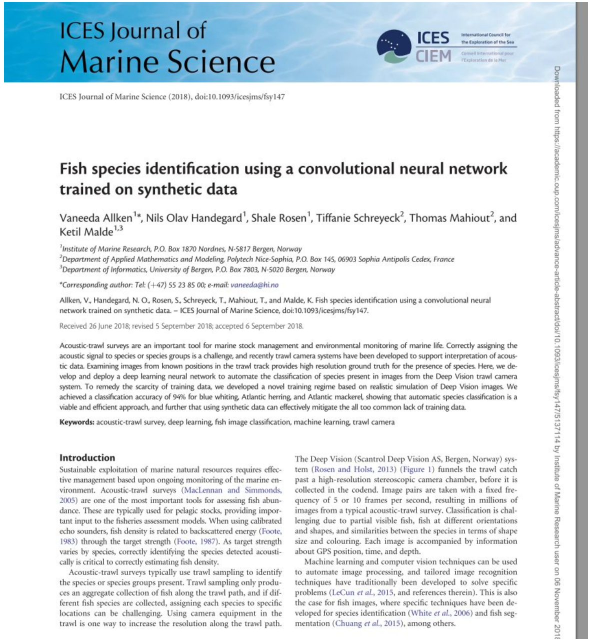
- Transfer learning
  - Finetuning a model pretrained on ImageNet
  - **Test accuracy (with TL) = 64%**

- Data augmentation
  - Includes rotation, translation, shearing, flipping and zooming
  - **Test accuracy (with TL + Data Augmentation) = 76.5%**

- Simulating training data
  - **Test accuracy: Up to 94 %**

Training on synthetic data works!




ICES Journal of Marine Science
ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

ICES Journal of Marine Science (2018), doi:10.1093/icesjms/fsy147

Fish species identification using a convolutional neural network trained on synthetic data

Vaneeda Allken[1]*, Nils Olav Handegard[1], Shale Rosen[1], Tiffanie Schreyeck[2], Thomas Mahiout[2], and Ketil Malde[1,3]

[1]*Institute of Marine Research, P.O. Box 1870 Nordnes, N-5817 Bergen, Norway*
[2]*Department of Applied Mathematics and Modeling, Polytech Nice-Sophia, P.O. Box 145, 06903 Sophia Antipolis Cedex, France*
[3]*Department of Informatics, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway*

**Corresponding author: Tel: (+47) 55 23 85 00; e-mail: vaneeda@hi.no*

Allken, V., Handegard, N. O., Rosen, S., Schreyeck, T., Mahiout, T., and Malde, K. Fish species identification using a convolutional neural network trained on synthetic data. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsy147.

Received 26 June 2018; revised 5 September 2018; accepted 6 September 2018.

Acoustic-trawl surveys are an important tool for marine stock management and environmental monitoring of marine life. Correctly assigning the acoustic signal to species or species groups is a challenge, and recently trawl camera systems have been developed to support interpretation of acoustic data. Examining images from known positions in the trawl track provides high resolution ground truth for the presence of species. Here, we develop and deploy a deep learning neural network to automate the classification of species present in images from the Deep Vision trawl camera system. To remedy the scarcity of training data, we developed a novel training regime based on realistic simulation of Deep Vision images. We achieved a classification accuracy of 94% for blue whiting, Atlantic herring, and Atlantic mackerel, showing that automatic species classification is a viable and efficient approach, and further that using synthetic data can effectively mitigate the all too common lack of training data.

Keywords: acoustic-trawl survey, deep learning, fish image classification, machine learning, trawl camera

Introduction

Sustainable exploitation of marine natural resources requires effective management based upon ongoing monitoring of the marine environment. Acoustic-trawl surveys (MacLennan and Simmonds, 2005) are one of the most important tools for assessing fish abundance. These are typically used for pelagic stocks, providing important input to the fisheries assessment models. When using calibrated echo sounders, fish density is related to backscattered energy (Foote, 1983) through the target strength (Foote, 1987). As target strength varies by species, correctly identifying the species detected acoustically is critical to correctly estimating fish density.

Acoustic-trawl surveys typically use trawl sampling to identify the species or species groups present. Trawl sampling only produces an aggregate collection of fish along the trawl path, and if different fish species are collected, assigning each species to specific locations can be challenging. Using camera equipment in the trawl is one way to increase the resolution along the trawl path.

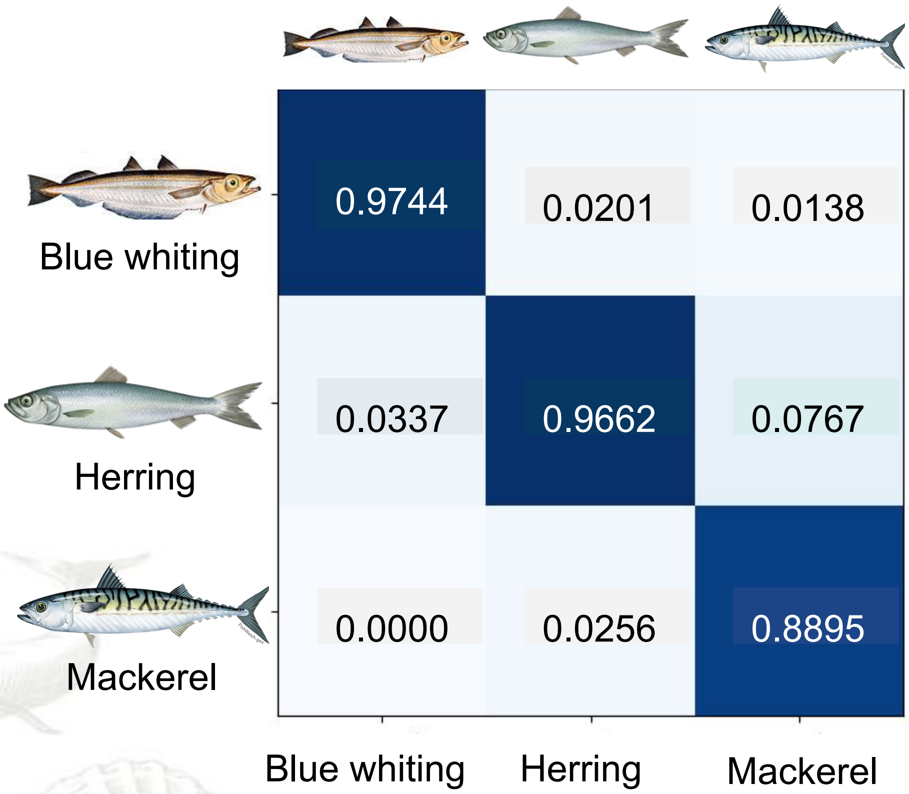
The Deep Vision (Scantrol Deep Vision AS, Bergen, Norway) system (Rosen and Holst, 2013) (Figure 1) funnels the trawl catch past a high-resolution stereoscopic camera chamber, before it is collected in the codend. Image pairs are taken with a fixed frequency of 5 or 10 frames per second, resulting in millions of images from a typical acoustic-trawl survey. Classification is challenging due to partial visible fish, fish at different orientations and shapes, and similarities between the species in terms of shape size and colouring. Each image is accompanied by information about GPS position, time, and depth.

Machine learning and computer vision techniques can be used to automate image processing, and tailored image recognition techniques have traditionally been developed to solve specific problems (LeCun *et al.*, 2015, and references therein). This is also the case for fish images, where specific techniques have been developed for species identification (White *et al.*, 2006) and fish segmentation (Chuang *et al.*, 2015), among others.

Best accuracy: **94.1%**

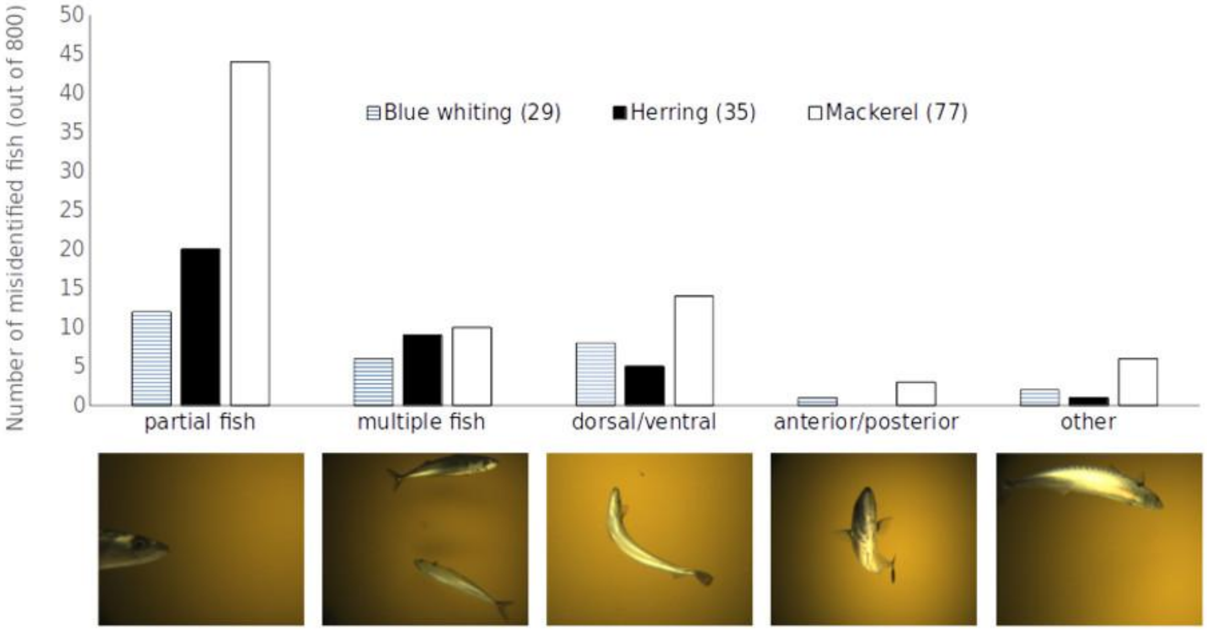
With training set of 5000 synthetic images from 70 fish cutouts per species

Confusion matrix



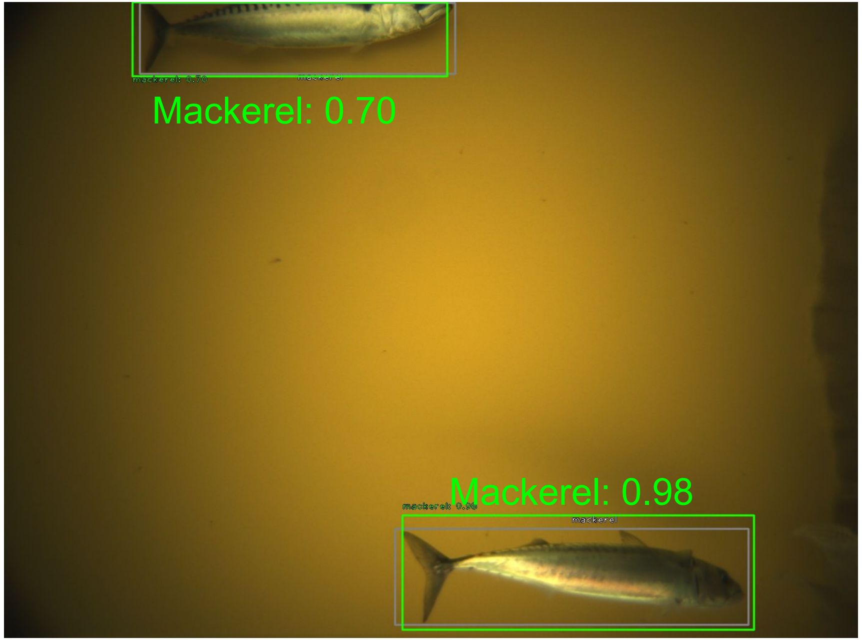
Misclassified images

6% wrong predictions (141 out of 2400 images)



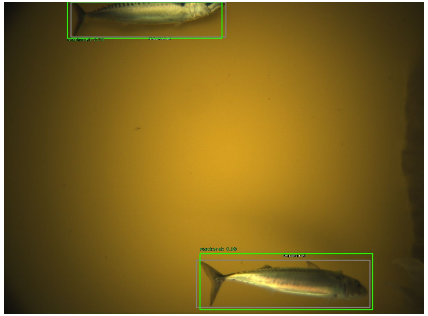
Object (fish) detection

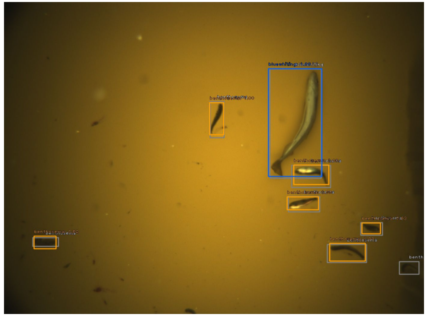
Prediction score



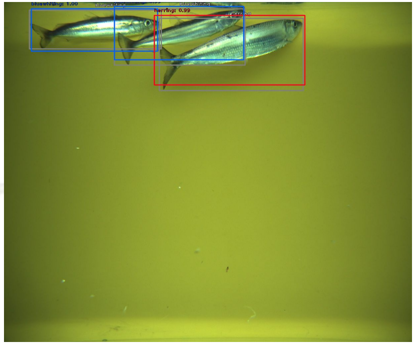
Object (fish) detection

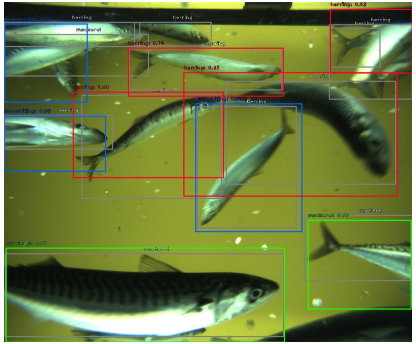
2017



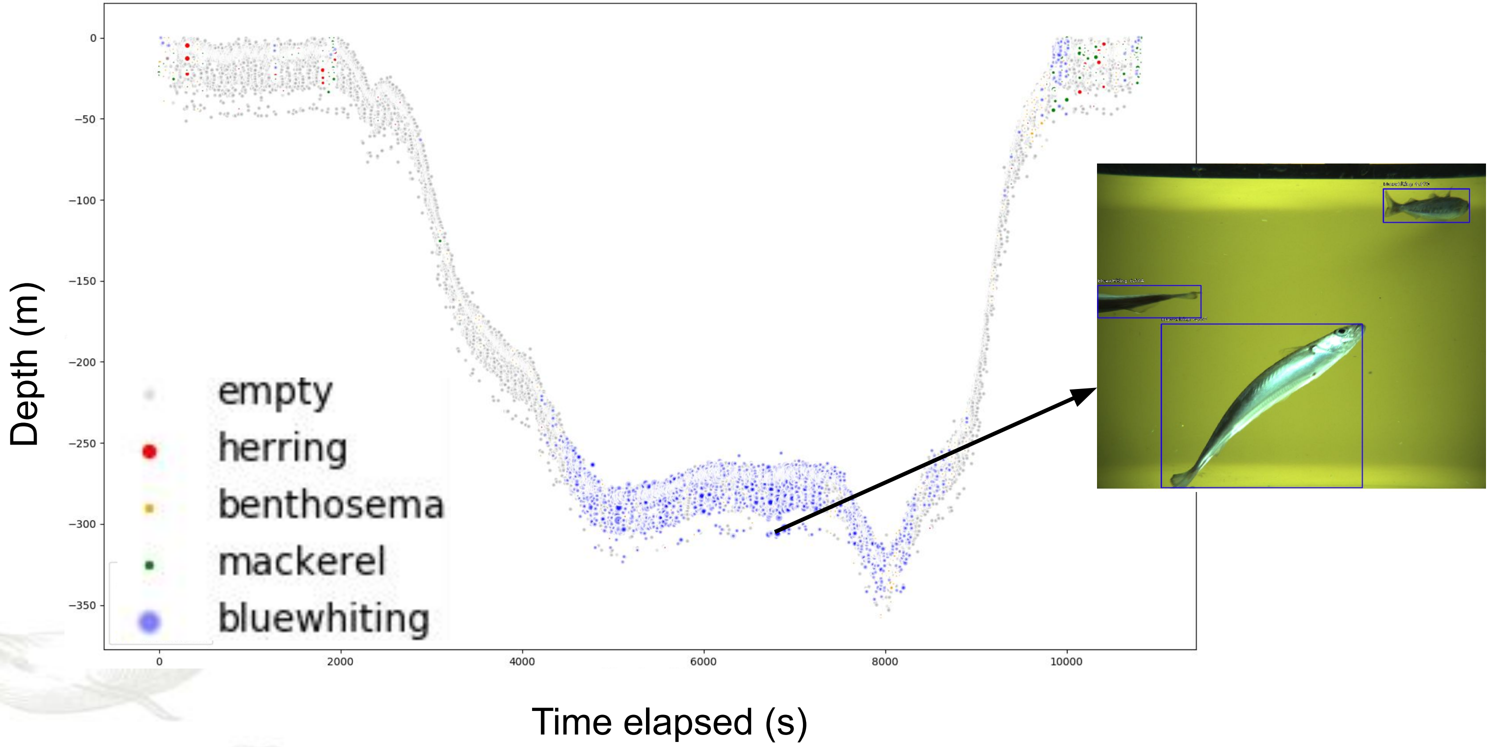


2018

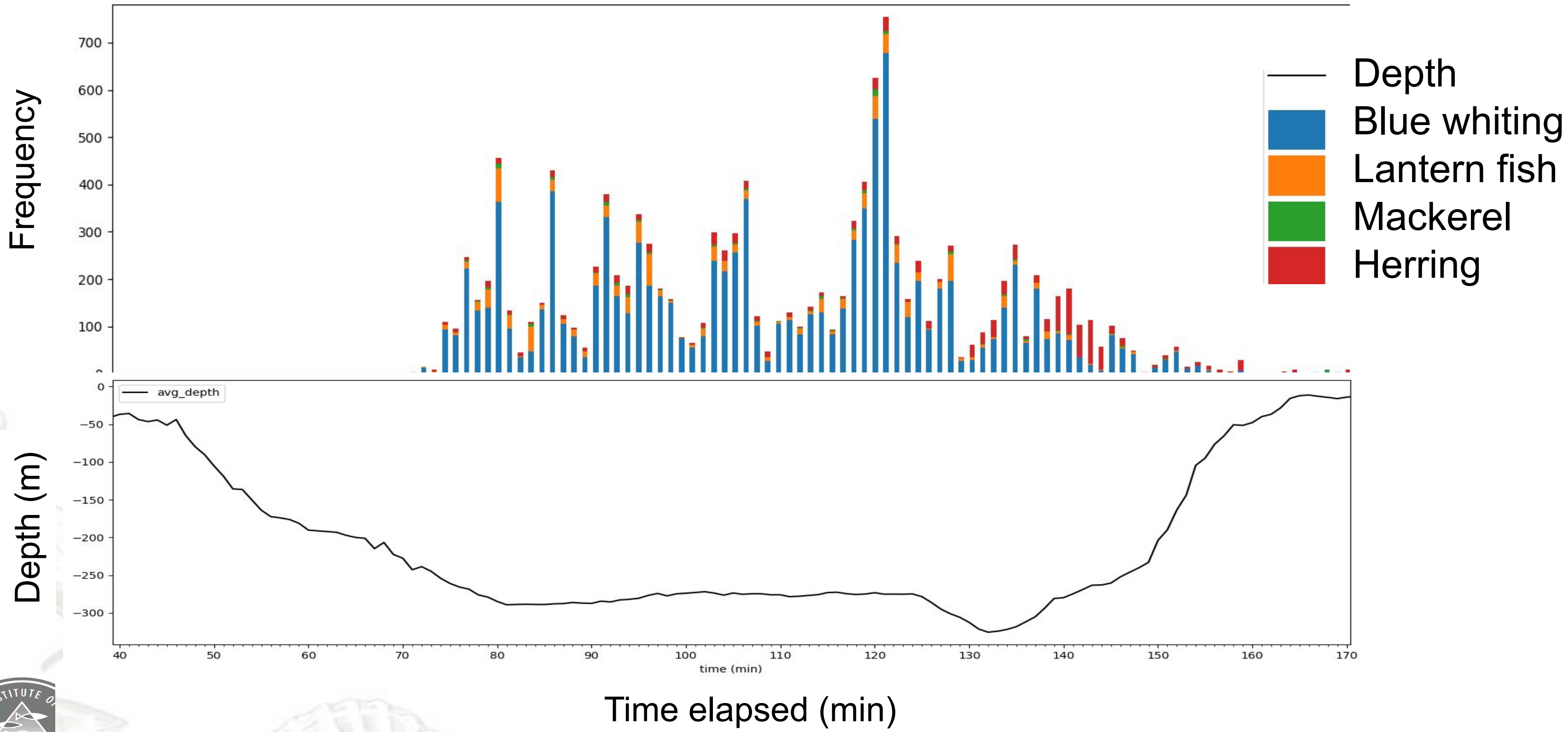




Predicting fish distribution across trawl station



Predicting fish distribution across trawl station

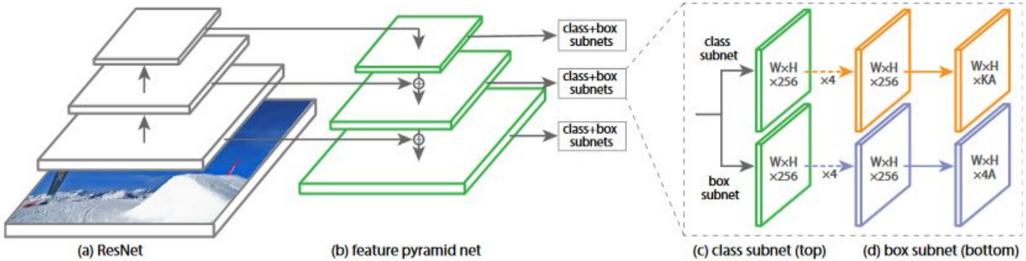


Thank you for your attention!

Object detection : RetinaNet

[Lin et al., Focal Loss for Dense Object Detection, *IEEE 2017*]

- Anchor
- Divide and conquer
  - FPN
- Focal loss
  - pos/neg unbalanced issue

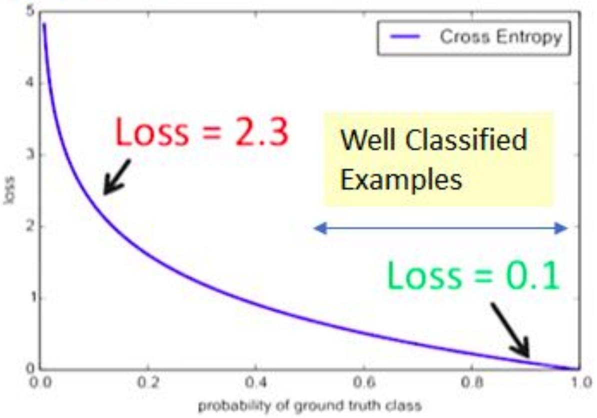


Focal loss

- 100000 easy : 100 hard examples
- **40x bigger loss** from easy examples

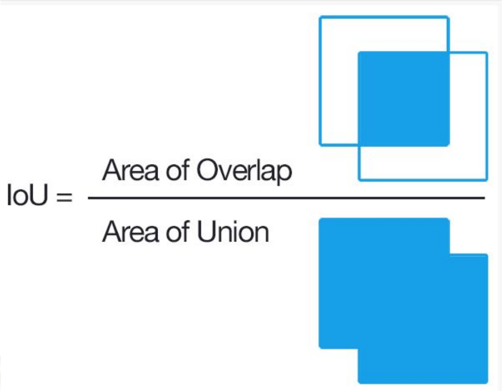
Focal loss

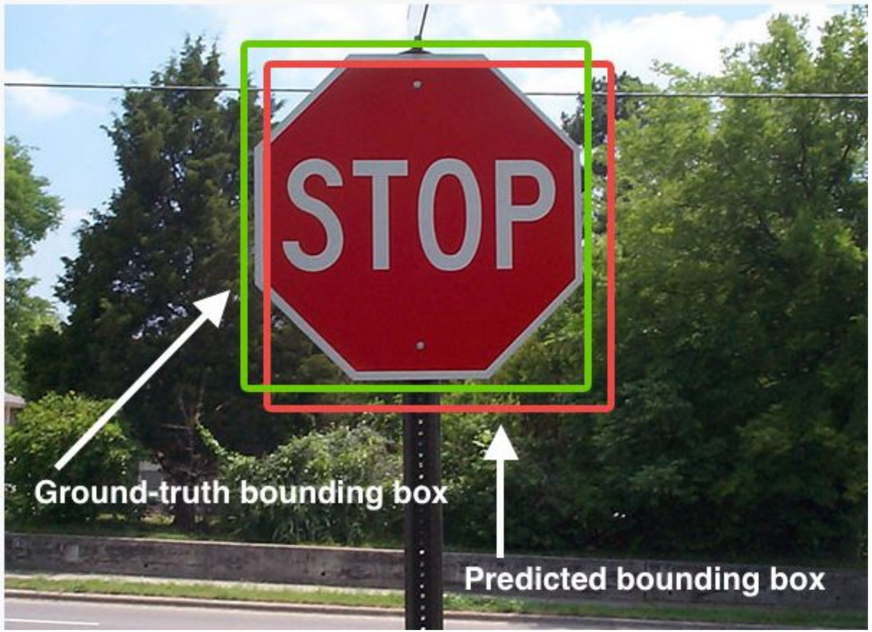
$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t).$$



Example

Intersection over Union (IOU)





Mean Average Precision (mAP)

True positive if IOU $\geq 0.5.$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

TP = True positive

TN = True negative

FP = False positive

FN = False negative

Average Precision = Mean of precision at chosen recall values

mAP = Mean of Average Precision across all classes

Comparison of mAP

Training set	Test mAP (918 images)
Synthetic images (20000)	0.7758
Real images (343)	0.7221
Synthetic + real images	0.8209