Fish Species Detection

Using Convolutional Neural Networks

Group 12: Anna Girerd (leader), Brian Stoss, Natalie Assaad

DS 4002

April 2025

Outline

BACKGROUND

- Our motivation, hypothesis, and data origin

ANALYSIS

- Our model, processes, tricky decisions, and difficulties encountered

RESULTS

- Our findings and conclusions from the data analysis

NEXT STEPS

- Future research and improvements

Motivation & Prediction

Problem

- Overfishing = fishing vessels fish at faster rate than stocks can reproduce

How Data Science can Help

- CNNs can extract features of fish from onboard video footage, helping to monitor and prevent overfishing by fisheries

Hypothesis

Using a **CNN** trained on labeled onboard footage, we aim to **identify fish** species with at least **50% Mean Average Precision** (@0.5), enabling effective fish detection and supporting sustainable fisheries.

Our Data & Model

Fishnet Dataset

- 2,001 JPEGs from longline tuna vessel footage in Pacific
- CSV file (159,119 rows) with bounding box annotations labeling fish as "yellowtail," "rockfish," etc.
- From the Nature Conservancy

YOLOv8

- Object detection models by Ultralytics
- Pretrained on COCO
- Mean Average Precision (mAP) included

Column	Description
img_id	Unique identifier for each image.
label	Label for the object within the bounding box, including the 4 types of tuna as well as 25 additional species names.
center_x	Center point of the X coordinates of the bounding box.
center_y	Center point of the Y coordinates of the bounding box.
width	The width of the bounding box.
height	The height of the bounding box.

Back to the Research Question

Can a YOLOv8-based CNN trained on the Fishnet image dataset achieve 50% mAP@0.5 in detecting and classifying fish species from onboard surveillance footage to support sustainable fishing?

Analysis Plan

Preprocessing

- Convert CSV to YOLO format, normalize pixel values
- EDA

Modeling

- Use YOLOv8 trained on COCO dataset
- Define species within the bounding boxes, use transfer learning to fine-tune model
- Training iterations (batch training, tuning hyperparameters)

Evaluation

- Split 70% train, 15% validation, 15% test
- Evaluate precision and mean average precision
- Analyze false positives and false negatives
- Deploy k-fold validation

Tricky Analysis Decision

- Originally 35,000 jpegs → too large
- 159,119 rows in labeling CSV

Dilemma:

- Do we delete images, labels, or both?

Critical insight:

- YOLOv8 skips missing images in label file automatically

Solution:

- Kept ~2,200 images
- Uploaded reduced dataset to GitHub for group access
- Kept label file as is w/no training crashes

Trade-offs:

- Faster training, easier collaboration but less image diversity

Bias & Uncertainty

Biases:

- Reduced dataset → common species are overrepresented
- Many images are blurry or partially cropped → lowering detection accuracy

Biases Addressed:

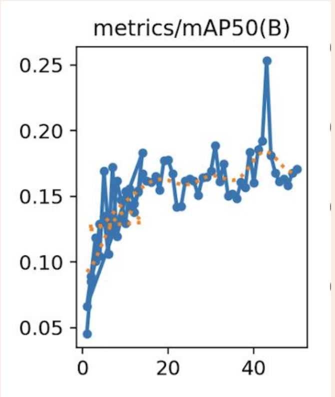
- Deleted images randomly to retain images diversity
- Tracked class distribution in EDA → use class weights

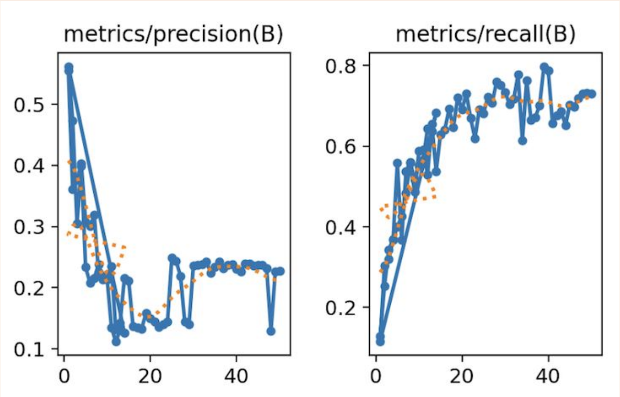
Uncertainties:

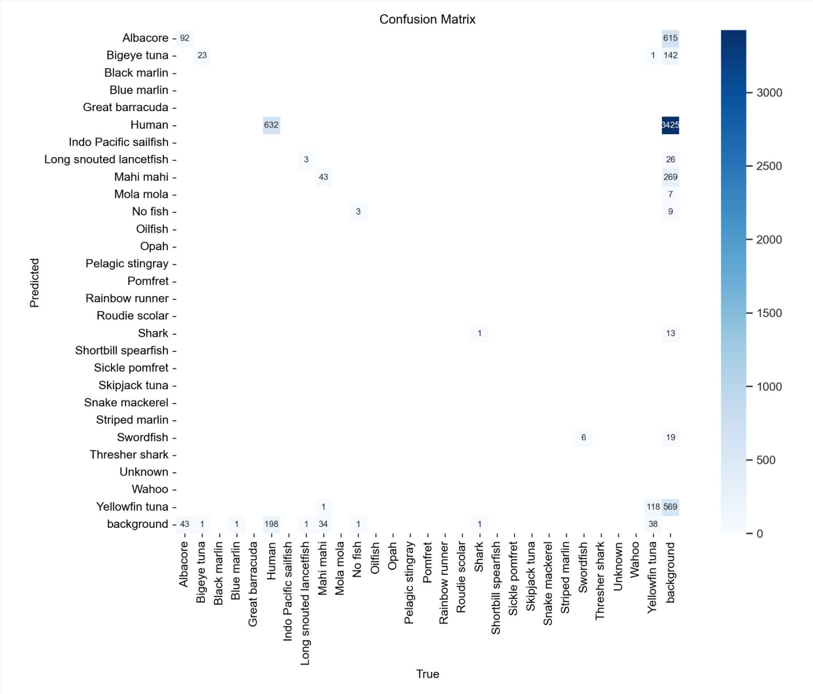
- mAP measures both precision and classification

Results & Conclusions

- Training completed in ~15 min over 50 epochs
- mAP@0.5: 25% → improving, but still below target (50%)
  - Recall is solid; precision improves slightly but remains low
- Confusion matrix: steady performance on common species but misclassifications on rare classes







Next steps

Although our model ran successfully and resulted in mAP@0.5 of 25%, we discovered various avenues for future research and potential improvements in our modeling approach.

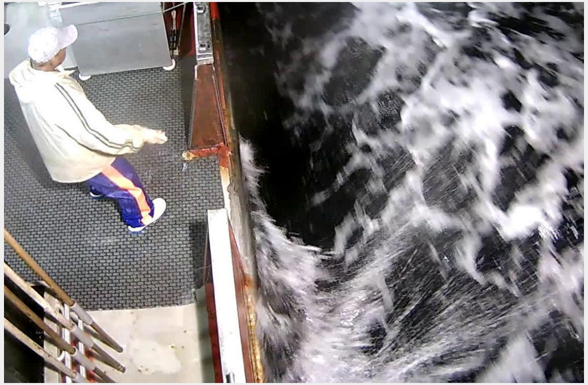
1. Add **more images** for underrepresented species
2. Use **image augmentation** to improve model generalization
3. Experiment with learning rates, batch sizes, and **epochs**

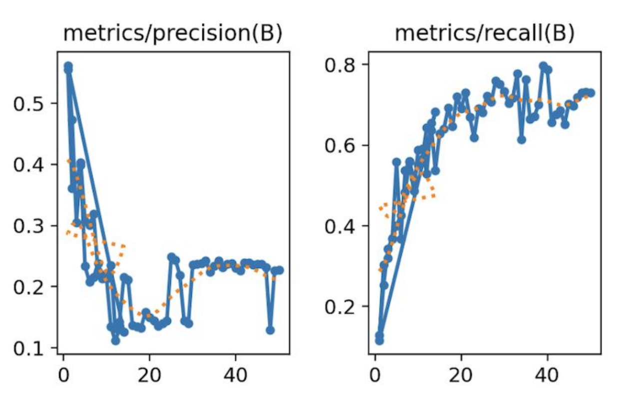
References

- “Description: Overview of the Fishnet Open Images Database,” The Nature Conservancy, https://www.fishnet.ai/description (accessed Apr. 3, 2025).
- J. Kay and M. Merrifield, “The Fishnet Open Images Database: A Dataset for Fish Detection and Fine-Grained Categorization in Fisheries,” Arxiv, https://arxiv.org/pdf/2106.09178 (accessed Apr. 10, 2025).
- “What is overfishing? facts, effects and overfishing solutions,” World Wildlife Fund, https://www.worldwildlife.org/threats/overfishing (accessed Apr. 3, 2025).
- T. Puchner, “Roter Thun, Bluefin Tuna (Thunnus thynnus) in thunfischmast,” Flickr, https://www.flickr.com/photos/tom_puchner/3362791138 (accessed Apr. 22, 2025).

GitHub: https://github.com/bmstoss13/DSProject3

Thank You!




metrics/precision(B)
metrics/recall(B)
0.5
0.4
0.3
0.2
0.1
0.8
0.6
0.4
0.2
0
20
40