Computer Vision-Based Detection of Schools of Herring from Acoustic Backscatter Time Series

Alireza Rezvanifar¹, Tunai Porto Marques¹, Melissa Cote¹, Alexandra Branzan Albu¹, Alex Slonimer², Thomas Tolhurst², Kaan Ersahin², Todd Mudge², and Stéphane Gauthier³

¹University of Victoria, Victoria, BC, Canada. E-mail: aalbu@uvic.ca
²ASL Environmental Sciences Inc., Victoria, BC, Canada
³Fisheries and Oceans Canada - Institute of Ocean Sciences, Sidney, BC, Canada

PICES 2019 Annual Meeting

W15: Application of machine learning to ecosystem change issues in the North Pacific
Victoria, BC, Canada
October 17-18, 2019
Context

- Study of acoustic backscatter:
  - Thorough, non-invasive approach
  - Allows to monitor underwater sites for ecosystem changes
- Data:
  - Acquired via multifrequency echosounders (e.g. AZFPs)
  - Visualized as 2D images (echograms)

Sample echogram (67 kHz)  Sample echogram (455 kHz)
Context

● Challenges:
  ○ Echograms typically analyzed via manual or semi-automatic methods:
    ■ Time consuming (tons of data to analyze)
    ■ Prone to errors and inconsistencies
    ■ Expensive third-party software (e.g. EchoView)

● Solution:
  ○ Machine learning can improve data processing and interpretation!
Context

- Collaborative project:

  - (Computer vision research lab)
    - Computer engineers
      ⇒ Develop ML tools

  - (Institute of Ocean Sciences)
    - Biologists
      ⇒ Acquire data

  - University of Victoria

  - Remote sensing specialists and acousticians
    ⇒ Develop echosounders (AZFPs)
Context

● Goal:
  ○ Explore novel ways to detect visual patterns from echosounder data using computer vision and machine learning techniques

● Case study:
  ○ Automatic detection of schools of herring from AZFP measurements
Contributions

1. We propose a dual paradigm approach for fish detection from echograms
   - Classical machine learning paradigm
   - Deep learning paradigm: novel application that goes beyond the few existing works

2. Our framework automates acoustic survey analyses
   - Will reduce processing times, required man-power, and inconsistencies in the results
   - Potential to be scaled to handle additional underwater species (e.g. salmon, zooplankton, etc.)
Proposed Method: Overview

Training Phase

Echogram dataset

ROI extraction

Negative samples background

Annotated Schools

Positive samples schools of herring

Hand-crafted features

Convolutional Neural Networks-based features

SVM classifier

DL-based classifier
Proposed Method: Overview

Inference Phase

Unseen Echogram

For each ROI:
- Hand-crafted features (Orientation, eccentricity, circularity, axis length, mean intensity)
- Convolutional Neural Networks-based features

ROI Extraction

SVM classifier
- Classification result

DL-based classifier
- Classification result
ROI Extraction

Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix
- Filtering by size
- Filtering by orientation

Detected ROIs
ROI Extraction

Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix
- Filtering by size
- Filtering by orientation

Detected ROIs
ROI Extraction

Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix
- Filtering by size
- Filtering by orientation
- Detected ROIs
ROI Extraction

Counts Images (4 frequency channels)

Denoising

Adaptive Thresholding

Opening+Closing

Score Matrix

Filtering by size

Filtering by orientation

Detected ROIs

Original

Opening

Closing
Computer Vision-Based Detection of Schools of Herring

Counts Images (4 frequency channels)

Denoising

Adaptive Thresholding

Opening+Closing

Score Matrix

Filtering by size

Filtering by orientation

Detected ROIs
Computer Vision-Based Detection of Schools of Herring

Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix
- Filtering by size
- Filtering by orientation
- Detected ROIs
Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix

**Green Boxes:**
- Filtering by size
- Filtering by orientation

**Result:**
- Detected ROIs
Computer Vision-Based Detection of Schools of Herring

Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix
- Filtering by size
- Filtering by orientation

Detected ROIs
Counts Images (4 frequency channels)

- Denoising
- Adaptive Thresholding
- Opening+Closing
- Score Matrix
- Filtering by size
- Filtering by orientation
- Detected ROIs
Classical ML: Classification

- **Features:**
  - The objective is to engineer the best set of features based on contextual information of the schools.
  - Features should reflect the appearance and geometry of the schools.
  - Selected features are:
    - Mean intensity of regions
    - Ratio between the minor axis to the major axis of an ellipse that has the same normalized second central moments as the region
    - Eccentricity: how much the center of mass differs from the center of the circumscribed circle
    - Circularity: specifies the roundness of object

- **Classifier:**
  - The still popular Support Vector Machines (SVM) classifier with linear kernel is utilized.
Deep-Learning Classification

The use of deep learning frameworks can automate the classification task by computing discriminant features, regardless of object class.

1. Use a Convolutional Neural Network (CNN)-based architectures for the automatic extraction of features
2. Use the extracted features as inputs of a fully connected network (FCN) that generates predictions
3. Calculate the loss based on the ground truth data
4. Use backpropagation to update network parameters, yielding better predictions
1. Use a Convolutional Neural Network (CNN)-based architectures for the automatic extraction of features.
2. Use the extracted features as inputs of a fully connected network (FCN) that generates predictions.
Deep-Learning Classification

3. Calculate the loss based on ground truth data.
4. Use backpropagation to update network parameters, yielding better predictions.
Experimental Results - Dataset

**Ground truth dataset**

- **100** echograms
- **145** samples of schools of herrings

Samples are used for the extraction of hand-crafted features (SVM) and the training of the deep learning-based classifier.

Echograms with annotated samples (yellow bounding boxes)
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

1. Regions of Interest (ROI extractor output):
   Black bounding boxes
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

2. Use SVM (handcrafted-based features) or deep learning-based approach to classify each ROI: white bounding boxes represent prediction of *schools*
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

3. Compare detection with the ground truth: yellow bounding boxes
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

\[
\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}
\]

If a detection has an IoU > threshold: true positive.

Image retrieved from [https://tinyurl.com/y3gn7mtl](https://tinyurl.com/y3gn7mtl)
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

3. Compare detection with the ground truth: yellow bounding boxes
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

3. Compare detection with the ground truth: yellow bounding boxes

2 conditions:
- Classified as “herring”
- IoU value > IoU threshold
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

3. Compare detection with the ground truth: yellow bounding boxes
How to determine if an ROI is a true positive?

3. Compare detection with the ground truth: yellow bounding boxes
Experimental Results - Evaluation

How to determine if an ROI is a true positive?

3. Compare detection with the ground truth: yellow bounding boxes
How to determine if an ROI is a true positive?

4. Calculate that for all samples in the dataset:
   - 100 samples
   - 145 instances of schools of herring
   TP, FP, TN, FN
Experimental Results: Quantitative

- ROI Extractor Evaluation

<table>
<thead>
<tr>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.931</td>
<td>0.292</td>
</tr>
<tr>
<td>0.2</td>
<td>0.917</td>
<td>0.288</td>
</tr>
<tr>
<td>0.4</td>
<td>0.834</td>
<td>0.262</td>
</tr>
</tbody>
</table>

- Entire Framework Evaluation (IoU = 0.4)

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet50</td>
<td>0.77</td>
<td>0.85</td>
<td>0.81</td>
</tr>
<tr>
<td>DenseNet201</td>
<td>0.78</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>InceptionNet</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Baseline (SVM)</td>
<td>0.51</td>
<td>0.78</td>
<td>0.62</td>
</tr>
</tbody>
</table>
Experimental Results: Qualitative

SVM (IoU threshold 0.4)

Correct detections
Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

Correct detections
Experimental Results: Qualitative

SVM (IoU threshold 0.4)

False detections (FP)
Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

False detection are now TN
A new FP
Experimental Results: Qualitative

SVM (IoU threshold 0.4)

False detections (FP)
Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

False detection are now TN
Experimental Results: Qualitative

SVM (IoU threshold 0.4)

False detections (FP)
Experimental Results: Qualitative

Deep Learning: ResNet-50 (IoU threshold 0.4)

False detection are now TN
Conclusion

- We explored machine learning approaches for the automatic detection of schools of herring from echograms created from AZFP data.
- We proposed and compared two different methods to classify regions of interests:
  - hand-crafted features + support vector machine
  - features automatically extracted and classified by CNNs
- Both methods yielded good results, but CNNs performed best (F1-score: 0.82), even though the dataset was small.
- Limitation: performance of ROI extraction, as classifiers can only classify extracted ROIs.
Current/Future Work

● Our collaborative project continues!
  ○ Single deep learning detection pipeline:
    ■ Single network to perform *localization* and *classification*
    ■ More scalable approach
  ○ Extension to other species, structures, and phenomena that can be monitored with echosounders:
    ■ Current: salmon, zooplankton
    ■ Future: suspended sediments, ocean turbulence, etc.
References and Acknowledgment


