Detecting endangered baleen whales within acoustic recordings using R-CNNs

Mark Thomas[1,2,*], Bruce Martin[2], Katie Kowarski[2], Briand Gaudet[2], and Stan Matwin[1,3]

[1]Dalhousie University [2]JASCO Applied Sciences [3]Polish Academy of Sciences

*mark.thomas@dal.ca

DALHOUSIE UNIVERSITY

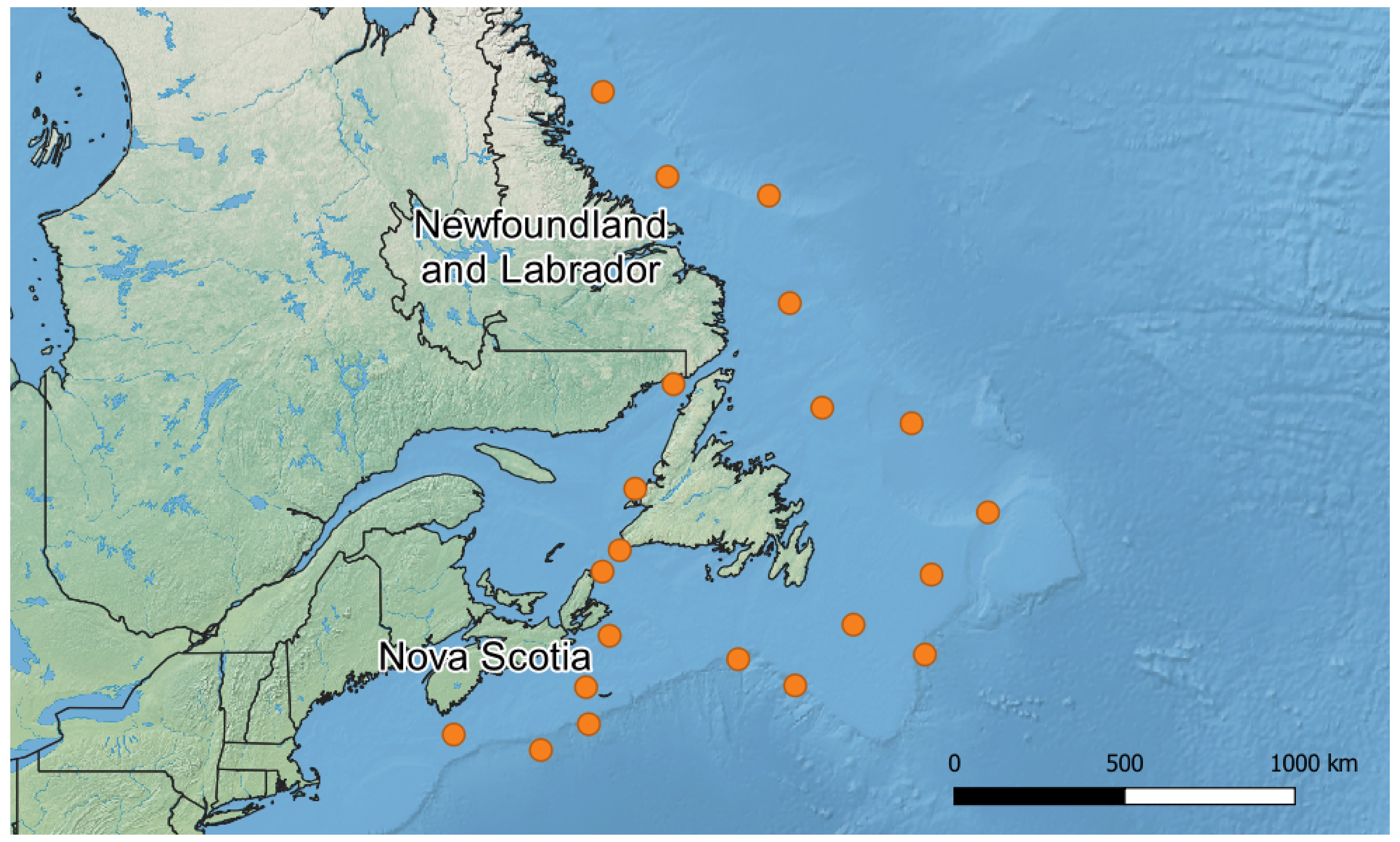
JASCO APPLIED SCIENCES

Introduction

- One of the most common techniques used by marine biologists to determine presence/absence of marine mammals is Passive Acoustic Monitoring (PAM)
- PAM has lead to large quantities of data for which manual analysis is expensive and time consuming
- Most traditional detection algorithms developed for PAM do not generalize well to new sources of noise
- Deep learning provides an opportunity for more generalizable systems [2]

Acoustic Recordings

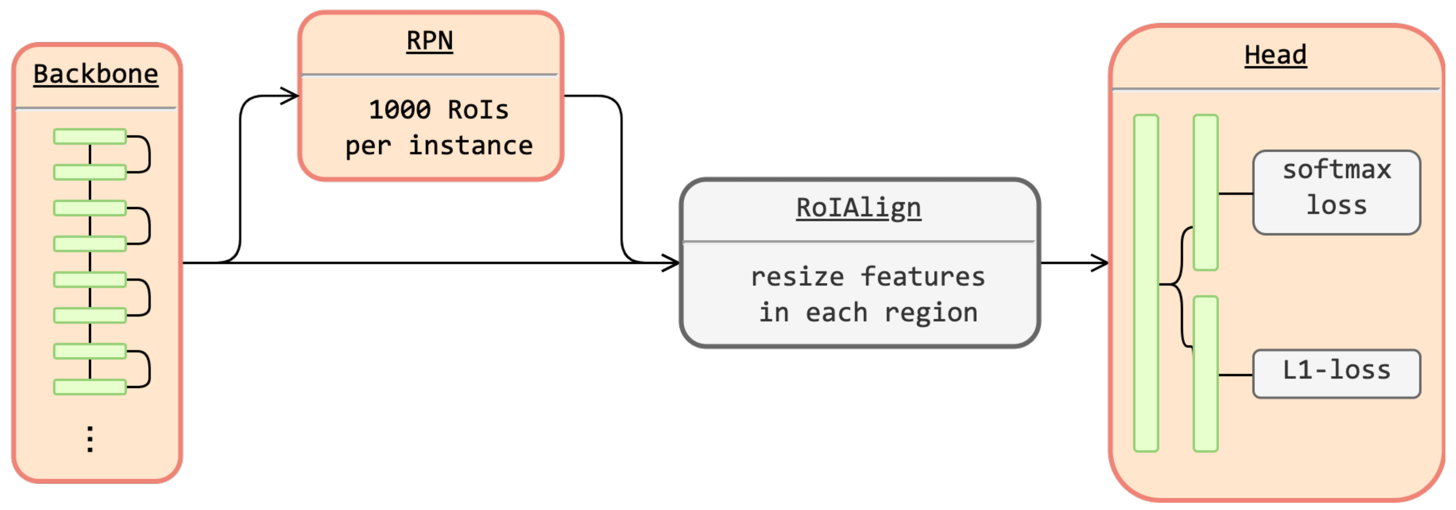
- Acoustic recordings were collected by JASCO Applied Sciences using Autonomous Multichannel Acoustic Recorders (AMARs)
- The devices were deployed off the coast of Atlantic Canada during the fall/summer of 2015 and 2016 surrounding an area of biological interest known as the Scotian Shelf



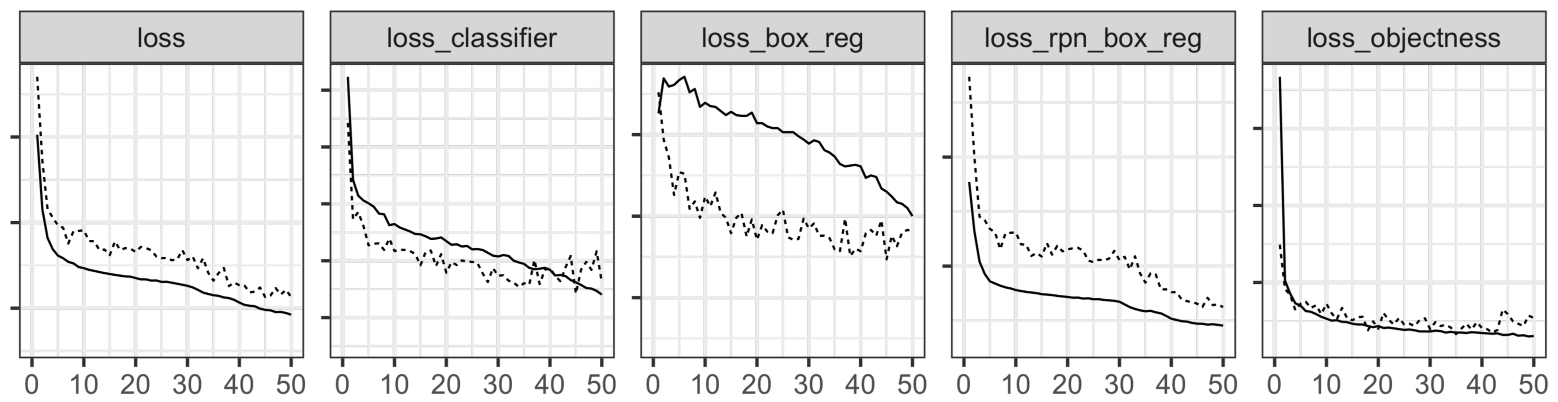
- The recordings were sampled at both 8kHz and 250kHz in order to capture the low frequency vocalizations of baleen whales and high frequency vocalizations of toothed whales
- The acoustic recordings were analyzed by marine biologists to produce annotations in the form of bounding boxes around marine mammal vocalizations
- We focus on identifying the vocalizations of three species of endangered baleen whales: blue, fin, and sei whales

Neural Network Architecture and Training Details

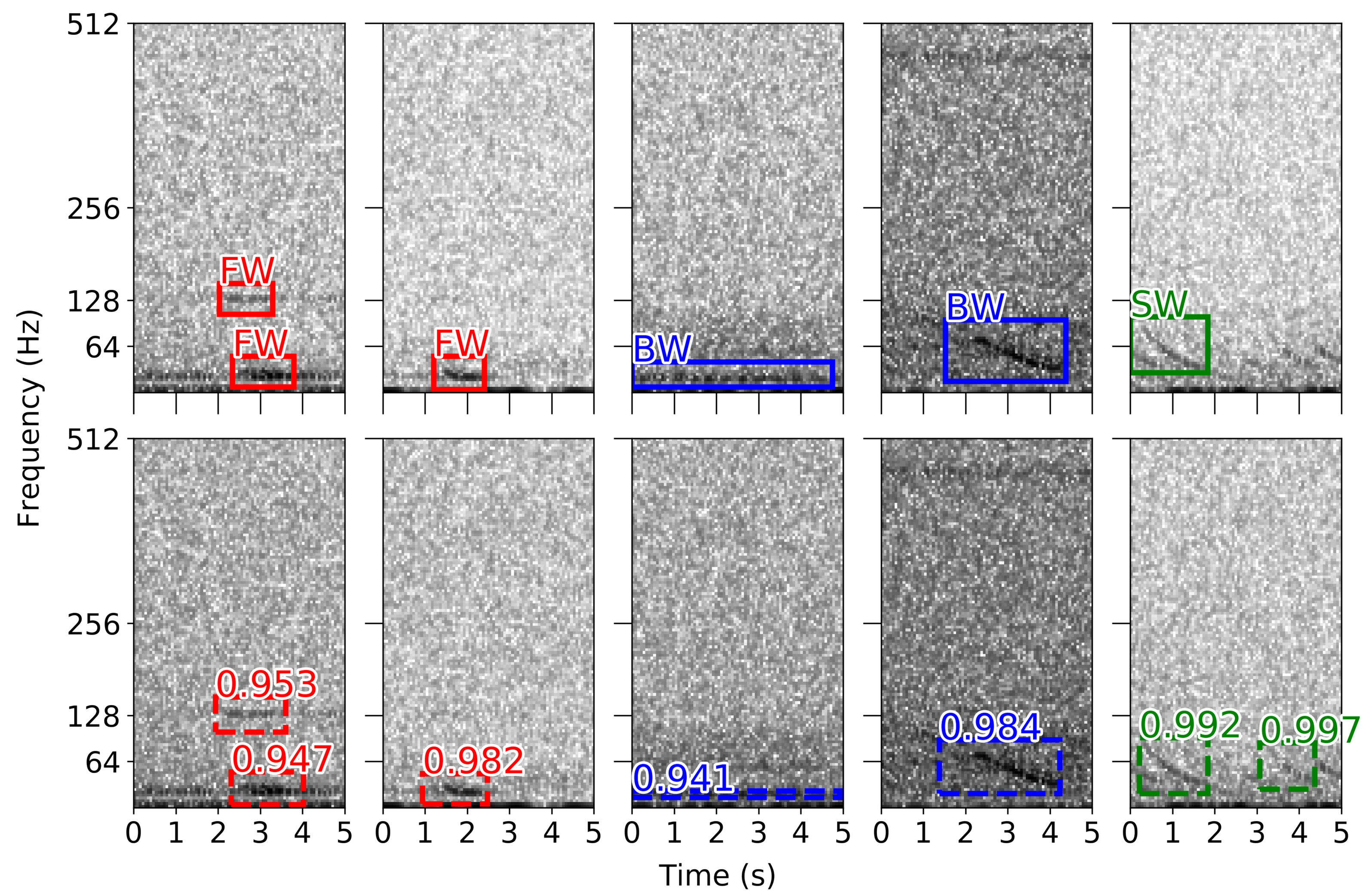
- The underlying neural architecture of the detection system is Mask R-CNN [1]
  - backbone: ResNet-50 + feature pyramid network (FPN)
  - 1000 region proposals per instance
  - RoIAlign for resizing the features in each region of interest
- The network is trained to detect bounding boxes corresponding to marine mammals vocalizations within spectrograms five seconds in length



- Stochastic Gradient Descent (SGD) with momentum=0.9 was used as an optimization routine
- Training was carried out over four NVIDIA P100 GPUs, each with 16GB of memory
- Other training parameters: batch size=4 (one instance per GPU), initial learning rate=0.003 that decayed by a factor of 10 after learning plateaued



Experimental Results



Species	Label	AP@.5	mAP@[.5:.95]	AR@.5	mAR@[.5:.95]
Overall	-	82.1	41.8	91.9	54.8
Blue whale	BW	85.7	52.8	96.2	70.9
Fin whale	FW	75.3	30.8	89.9	40.0
Sei whale	SW	85.4	41.9	89.7	49.4

References

[1] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961--2969, 2017.

[2] Mark Thomas, Bruce Martin, Katie Kowarski, Briand Gaudet, and Stan Matwin. Marine mammal species classification using convolutional neural networks and a novel acoustic representation. *arXiv preprint arXiv:1907.13188*, 2019.

Acknowledgments

The acoustic recordings were collected by JASCO Applied Sciences under a contribution agreement with the Environmental Studies Research Fund.