Leveraging Unlabelled Data through Semi-supervised Learning to Improve the **Performance of a Marine Mammal Classification System**

Mark Thomas^{1,2,*}, Bruce Martin², and Stan Matwin^{1,3}

*mark.thomas@dal.ca

¹Dalhousie University ²JASCO Applied Sciences ³Polish Academy of 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
- A considerable proportion (\approx 97%) of PAM data sets remain unanalyzed by human experts [2]

Acoustic Recordings

- We make use of two PAM data sets collected by JASCO Applied Sciences
 - Set A "Bay of Fundy": used for model training/validation
 - Set B "Atlantic OCS": used during testing as a proxy for out-of-distribution examples



• We are particularly interested in the "pulse train" vocalization of minke whales



Semi-supervised Learning of PAM Data

- We adapt the semi-supervised learning algorithm **MixMatch** [1] to leverage unlabeled data and improve the performance of a convolutional neural network (CNN) used to classify spectrograms containing minke whale vocalizations
- We replace the traditional image augmentation routines used in MixMatch with **SpecAugment** [3]



• The training data is highly unbalanced in favour of possible false alarms made by other species/sources:

		Se	et A	Set B
Acoustic source	shorthand	training	validation	testing
Minke whale pulse train	MW	556	56	336
Ambient noise	AB	5560	620	_
Fin whale	FW	3383	422	_
Humpback whale	HB	5773	597	_
North Atlantic right whale	RW	462	49	_
Non-biological noise	NN	-	-	266
Sei whale	SW	-	-	62

• The majority of the training data is unlabeled (20:1)



 × Possible false alarm Pulse train 	
	station 1
	station 2
	station 3
15 Nov '15	

Experimental Results

- respectively
- lead to well-performing models
- the spectrograms

Performance evaluated on Set A: Validation Set

Training paradigm Supervised Semi-supervised (λ_U =

Performance evaluated on Set B: O.O.D. Examples

Training paradigm precision recall F-1 score 0.75213 0.75178 0.75195 Supervised Semi-supervised ($\lambda_U = 10, \alpha = 0.5$) **0.89658 0.88799 0.89226**

- more susceptible to training bias

References

- Oliver, and Colin A Raffel.
- [2] Katie A Kowarski and Hilary Moors-Murphy. monitoring. Marine Mammal Science, 2020.
- Ekin D Cubuk, and Quoc V Le. recognition. *arXiv preprint arXiv:1904.08779, 2019.*



• Using a grid search we found well-performing values of the MixMatch hyperparameters λ_U and α to be 10 and 0.5,

• Balanced SpecAugment masking hyperparameters were found to

• For our case we use two time masks and two frequency masks with maximum possible widths of roughly 10% the dimensions of

	precision	recall	F-1 score
	0.79700	0.85807	0.82641
= 10, $\alpha = 0.5$)	0.81645	0.90301	0.85755

• The baseline model trained using only labeled data appears to be

• The performance of the semi-supervised model can generalize to acoustic data collected in distinct locations, at varying times and depths, and is less susceptible to unknown acoustic sources

[1] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital

Mixmatch: A holistic approach to semi-supervised learning.

In Advances in Neural Information Processing Systems, pages 5049--5059, 2019.

A review of big data analysis methods for baleen whale passive acoustic

[3] Daniel S Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph,

Specaugment: A simple data augmentation method for automatic speech