University of Southampton

Quantity Over Quality? Investigating The Effects of Volume and Strength of Training Data in Marine Bioacoustics

ANDY NAPOLI, PAUL WHITE, THOMAS BLUMENSATH

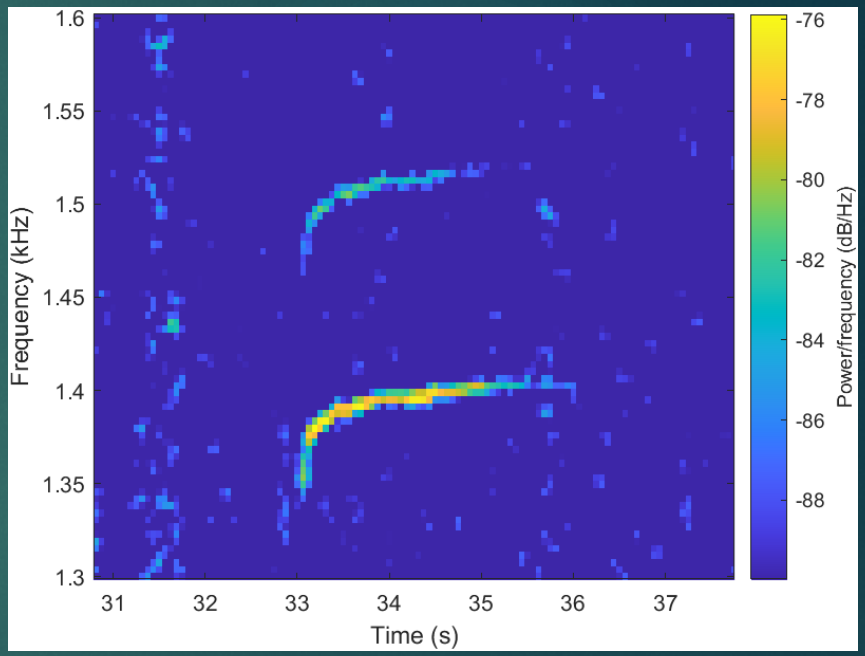
Introduction

- Developing machine learning tools typically requires lots of manually labelled, high-quality training data
- When labelling resources are limited, weak labels are often used instead, resulting in a trade-off between the quality and quantity of the training data created
- Choosing the best labelling strategy requires understanding exactly how performance is affected by these two opposing variables
- We propose a method to jointly investigate the effects of the strength and quantity of labels, and apply it in the context of detecting minke whale vocalisations
- Label strength: The precision to which calls are localised within a longer acoustic recording

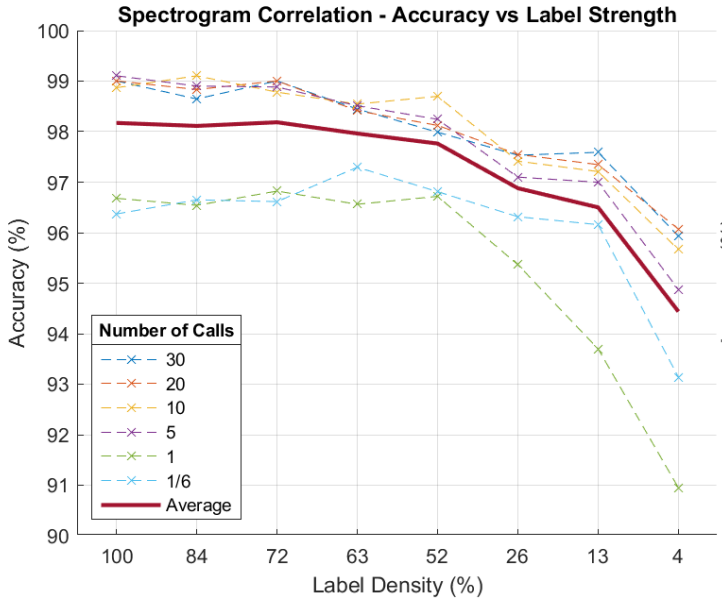
Image courtesy: *The Australian*.

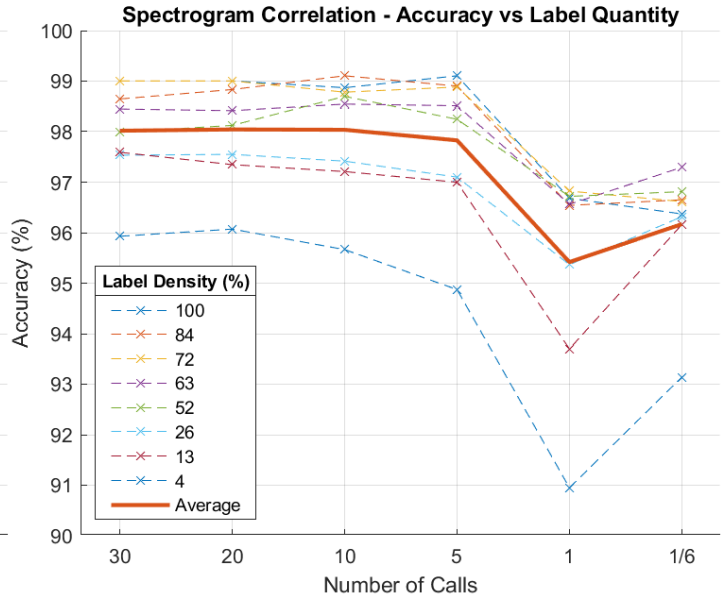
Method

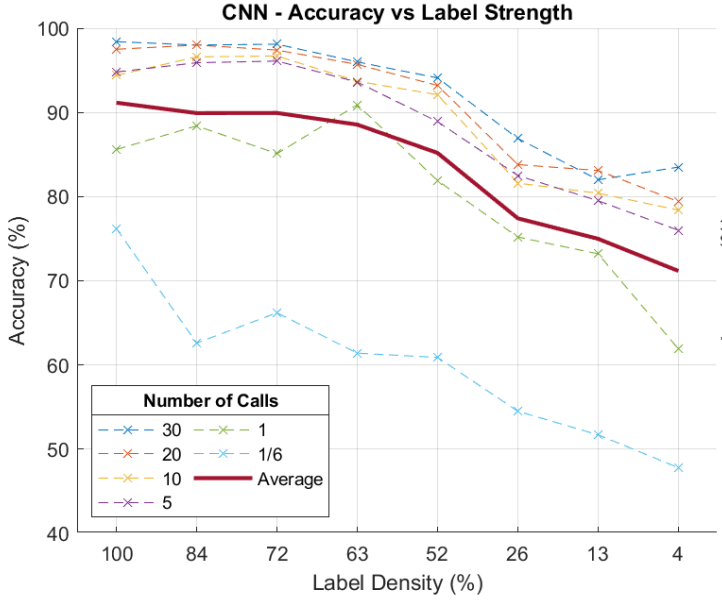
- We label calls to varying levels of precision, from 0.1 to 60 s, by increasing the length of the labels
- More weakly labelled "calls" will contain increasing amounts of non-target signal, in addition to the calls of interest
- The non-target class contains only ambient noise
- $Label\ Density = \frac{Call\ Duration}{Overall\ Label\ \ Length} \times 100\%$
- We then vary the number of calls in each training set, from 30 down to 1. Training on only a single channel of one call is also considered (out of the 6 channels available)
- We compare 2 detection methods: spectrogram correlation and a CNN

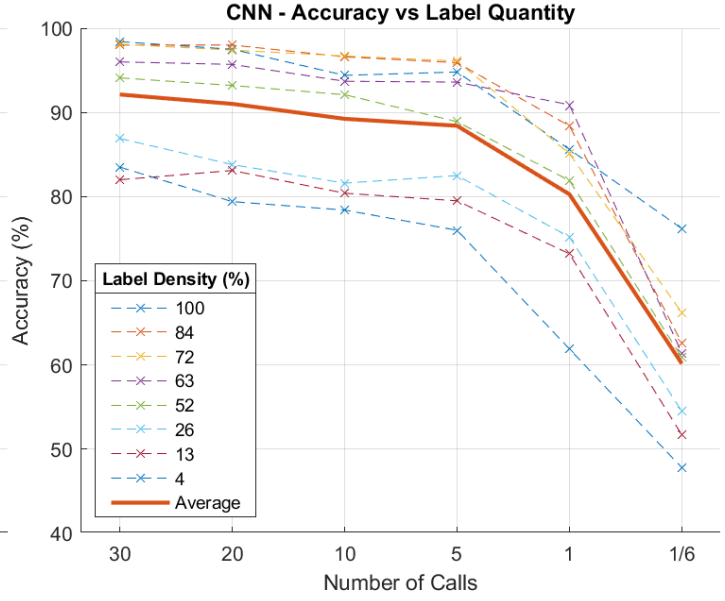


Results


Spectrogram Correlation - Accuracy vs Label Strength
Accuracy (%)
Label Density (%)
100
84
72
63
52
26
13
4
Number of Calls
30
20
10
5
1
1/6
Average


Spectrogram Correlation - Accuracy vs Label Quantity
Accuracy (%)
Number of Calls
30
20
10
5
1
1/6
Label Density (%)
100
84
72
63
52
26
13
4
Average


CNN - Accuracy vs Label Strength
Accuracy (%)
Label Density (%)
100
84
72
63
52
26
13
4
Number of Calls
30
20
10
5
1
1/6
Average


CNN - Accuracy vs Label Quantity
Accuracy (%)
Number of Calls
30
20
10
5
1
1/6
Label Density (%)
100
84
72
63
52
26
13
4
Average

Key Findings

- Quantity over quality holds for the CNN, but not for the spectrogram correlation
- Increasing label strength does not improve the performance of either detector beyond a certain point (60 to 70% label density)
- Performance of the CNN scales better with the size of the training set
- The spectrogram correlation is unable to exploit additional training data beyond the use of 5 calls
- The spectrogram correlation is more robust to fewer training samples, and weaker labels

Key Findings 2

- Using all available audio channels is beneficial to the CNN, but detrimental to the spectrogram correlation
- Interaction effects are observed between label strength and quantity:
  - Stronger labels are more robust to smaller training sets, and larger training sets are more robust to weaker labels
- A possible interaction exists with the length of the analysis frames, especially for the CNN:
  - Longer analysis frames are more robust to weaker labels, but more sensitive to smaller training sets
- Multi-factor analysis is important!

Fin