More than a Whistle:

Detecting Delphinids and Anthropogenic Noise with Convolutional Neural Networks

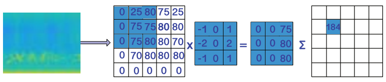
EMSO TIME SERIES CONFERENCE OBSERVING OCEAN SOUND 20-22 OCTOBER 2021 CANARY ISLANDS SPAIN

Ellen White, Msc
Email: elw1g13@soton.ac.uk
Twitter: @Oceanse11en

What and Why

The widespread use of Passive Acoustic Monitoring (PAM) for studying marine mammals has led to an explosion in the scale of acoustic data. The increased volume has created a deficit between data collected and data analysed. Common approaches to tackle this problem are to build detection and classification models which automate the extraction of specific sound sources of interest, making use of signal processing and machine learning algorithms, which isolate the signal of interest. The remaining data is seldom analysed further for species temporal or spatial context, coupled with the lack of understanding of vocal repertoires for many species. Baleen whale vocalisations are well understood, but Odontocetes present a difficult challenge in building detectors due to geographic variation in call types

Convolutional Neural Networks (CNNs) A type of artificial intelligence used in image recognition and processing, designed to process pixel data. The computer 'sees' the input image as a complex composition of edges, lines and corners in order to identify the function that maps the label of the image to the image itself. CNNs are end-to-end, meaning the features required to learn how to map the function are not defined. The flexibility of CNNs means they are ideal for marine acoustic data where soundscapes vary and sound sources are difficult to manually define by features.



Ocean Noise is now listed as an ICUN sustainability goal, and attention to the soundscape as a whole is becoming common practice[1]. Underwater soundscapes can provide indication of habitat quality and evaluate the influence of anthropogenic activities to species of interest. Algorithms which account for more than single-species presence or single call types are critical for monitoring ecosystem health, and extracting ecologically valuable cues from the marine environment, embedded in the soundscape.

CNNs have been used to detect and classify individual species call types[2-6] Limited by the need for large sets of annotated training data to build a CNN model, they have not been exploited to their full potential in the field of marine mammal monitoring. By harnessing the power of existing state of the art CNN model architectures, this study aimed to:

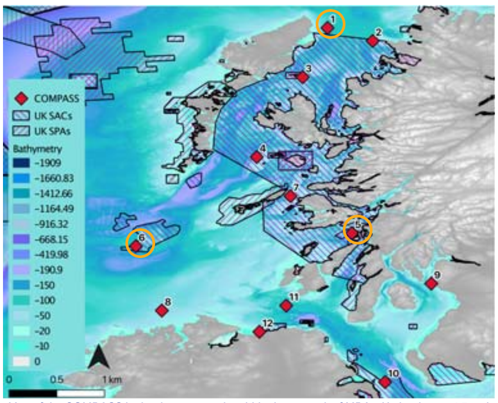
Develop a 'small-scale' CNN for marine sound sources which span the frequency spectra up to 50kHz, with minimal site specific labelled data. We aim to to encourage sustainable ocean monitoring through developing a model capable of deployment on a range of acoustic platforms, to mine existing big data and to be used onboard AUVs

Case Study: The COMPASS project

A network of twelve hydrophones are present across West Scotland as part of the long-term monitoring project COMPASS. Sampling at 96kHz, continuously recording since 2017 has amassed a huge dataset.

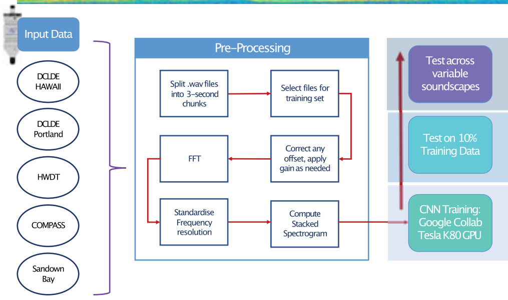
23 species of cetacean are recorded in Western Scotland each year, majority of which belong to the delphinid family and are listed on The Habitats Directive (1992), leading to the designation of several MPAs in the region. The status of species presence within the MPAs remains unknown due to lack of suitable abundance estimates. As a results, it cannot be quantified how successful the MPAs are at protecting and conserving these species. Raw acoustic data from three sites has been made available (1 - Tolsta, 5-Garvellachs, 6-Stanton Banks) for this work. Data is not labelled, our approach aims to develop a CNN model with minimal site specific training data used.

The COMPASS dataset is an ideal opportunity to test if a small-scale CNN is capable of learning to detect a range of sound sources, which are important to marine management and stakeholders. The output will allow for information on regional soundscapes to be delivered on timescales appropriate for conservation and marine management timescales.



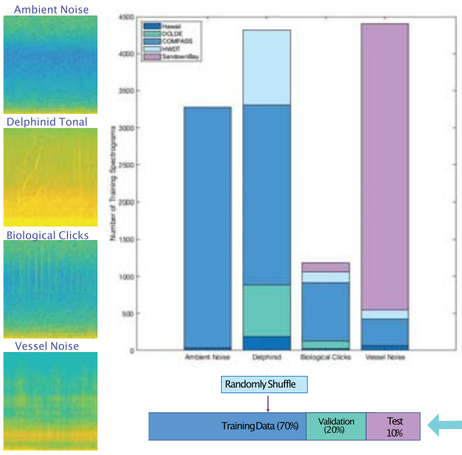
Map of the COMPASS hydrophone network, within the network of MPAs. Hydrophones 1,5 and 6 have been used in this work, with the model built to be used on the remaining array

Methodology

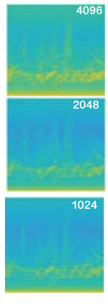


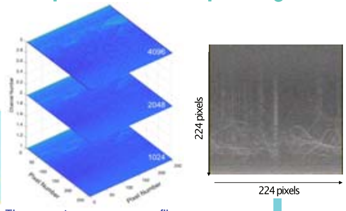
Training Data

Labelled data has been manually annotated from the COMPASS dataset and pooled from existing labelled marine sound databases (DCLDE, MobySound and HWDT charity surveys) to build a training set with diverse temporal and spatial soundscapes, with a range of collection methods, identifying 4 categories:

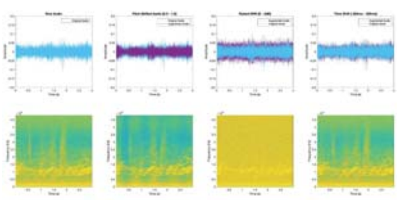


Model Input - Stacked Spectrogram





Three spectrograms per wav file are computed and stacked in an RGB array. R = 1024, G = 2048 and B = 4096. FFT sizes are scaled for the input source sampling rate. All spectrograms are colour scaled -100 0 dB and normalised by pixel

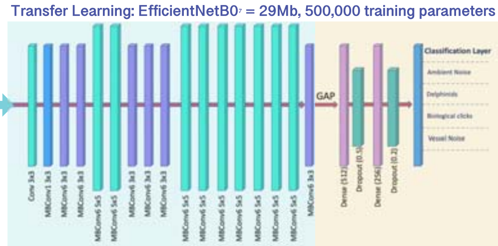


Augmentation is randomly applied to 50% of the training set prior to training to encourage the models ability to generalise to new soundscapes.

During training a horizontal flip of the image is used to augment further.

EfficientNet B0 Adapted Architecture

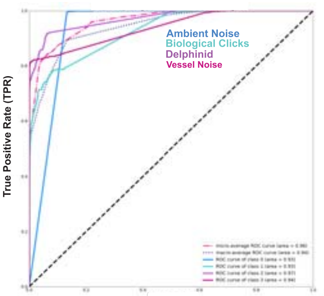
Transfer Learning: EfficientNetB0[7] = 29Mb, 500,000 training parameters



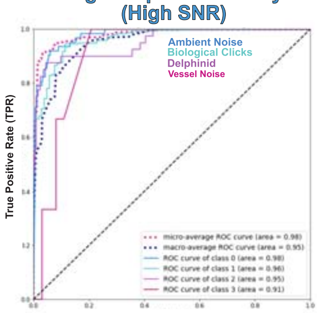
Existing Architecture - Frozen Layers | Custom Layers

Results

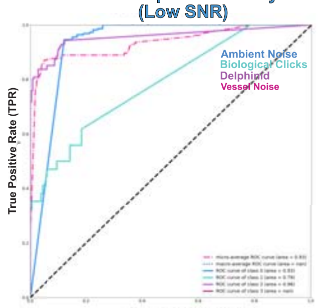
Test Set (10% Training Set)



High Delphinid Activity (High SNR)

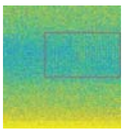


Low Delphinid Activity (Low SNR)



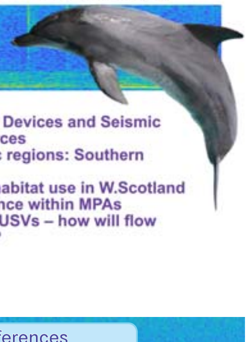
Key Results

- 96% Accurate on Test Set
- 89% Accuracy in low SNR conditions
- 86% + across temporal and spatial soundscape variation



Small imbalanced training set – accuracy remains high
Model = 29MB, suitable for onboard AUVs
FPR common where whistles and clicks present in same frame.
Model able to detect click trains in very low SNR with minimal training examples

What Next?



- Add in Acoustic Deterrent Devices and Seismic Array guns as sound sources
- Test Model in new oceanic regions: Southern Ocean
- Map long-term delphinid habitat use in W.Scotland
- Map anthropogenic presence within MPAs
- Deploy onboard AUV and USVs – how will flow noise effect performance?

Funders and Collaborators

University of Southampton · National Oceanography Centre · COMPASS · Hebridean Whale & Dolphin Trust · INSPIRE · UKRI Natural Environment Research Council

Author Contact Info

Co-Authors:
Prof Jon Bull, bull@noc.soton.ac.uk, University of Southampton
Prof Paul White, P.R.White@soton.ac.uk, University of Southampton
Dr Russell Wynn, russ@wildnewforest, Wild New Forest CIC
Dr Denise Rische, Denise.Rische@sams.ac.uk, Scottish Association of Marine Science
Suzanne Beck, suzanne.beck@afbini.gov.uk, Agri-Food and Biosciences Institute NI
Ewan Edwards, Ewan.Edwards@gov.scot, Marine Scotland

References

1. Ramirez 2017, Science Advances, 3(2), pe1601196.
2. Harvey et al (2018), Google AI Blog
3. Bergler et al (2019), Scientific reports, 9(1),pp.10-17
4. Bermant et al (2019), Scentific Reports, 9(1).pp 1- 10
5. Shiu et al (2020), Scientific Reports, 10 (1), pp.1-12
6. Yang et al (2020), JASA, 147(1), pp.49-55
7. Tan and Lee (2019), Int. Conf. Machine Learning, pp. 6105-6114.