Hawaiian monk seal vocalization detection and classification using Deep Learning

Kenneth Bremnes*[1], Lucas Hobrak*[1], Rebecca Moen*[1], Colleen Reichmuth[2], Jillian M. Sills[2], Jeppe H. Rasmussen[1,3]

* These authors contributed equally to the work
[1] Center for Artificial Intelligence Research, University of Agder, 4879 Grimstad, Norway
[2] Institute of Marine Sciences, Long Marine Laboratory, University of California Santa Cruz, Santa Cruz, California 95060, USA
[3] Center for Coastal Research, University of Agder, 4604 Kristiansand, Norway

BACKGROUND

The Hawaiian monk seal, *Neomonachus schauinslandi*, is an endangered marine mammal with an urgent conservation priority. Applying passive acoustic methods (PAM) to detect and monitor this species would inform research and management efforts. Underwater vocal behaviour was only recently documented for the species [1].

Manually going through months of recordings is slow, labour intensive and prone to human errors. Deep learning models were developed to automate the process and increase accuracy. This approach is divided into two classification problems with two Convolutional Neural Networks:

1. Detection of monk seal vocalizations
2. Classification of monk seal vocalizations

The underwater recordings used in this project were provided by the Institute of Marine Sciences at UC Santa Cruz; these data were obtained with one adult male monk seal living in human care (NMFS permit 19590; funding from U.S. Navy's Living Marine Resources Program).

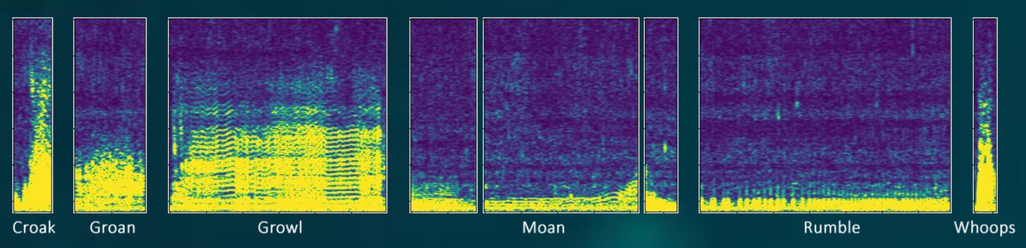

Figure 1: Six previously defined call types [1]

1. DETECTION

- Manual annotations used to create one-minute spectrogram in two categories: Call and no call
- Balanced dataset consisting of 1570 spectrograms
- Data augmentation used to increase ground truth
- Network model based on ResNet50
- 90% used for training 10% for testing

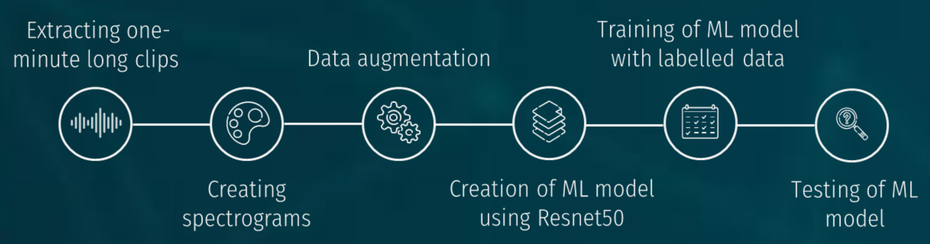

Figure 2: Method for the detection model

RESULTS

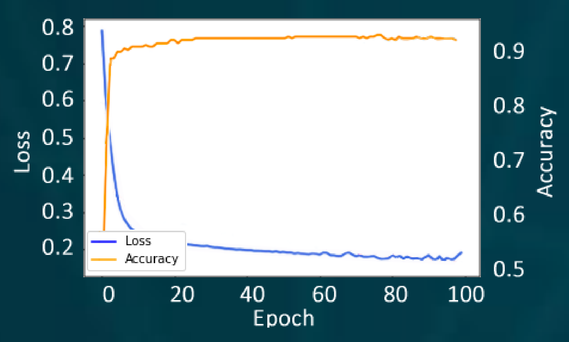

Figure 3: The loss and accuracy for each epoch during training

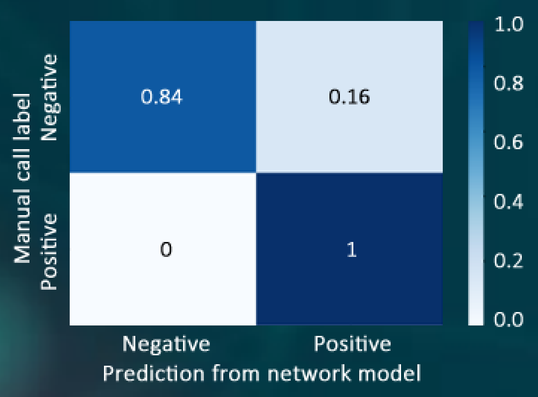

Figure 4: Confusion matrix presenting the binary classification performance

Average detection accuracy of 92%

2. CLASSIFICATION

- 100 hours of PAM, 1200 calls manually annotated into six call types
- Spectrogram created from each isolated call. Zero padding used to ensure uniform image size
- Data augmentation implemented to balance dataset
- Network model based on VGG16
- 80% used for training, 20% for testing

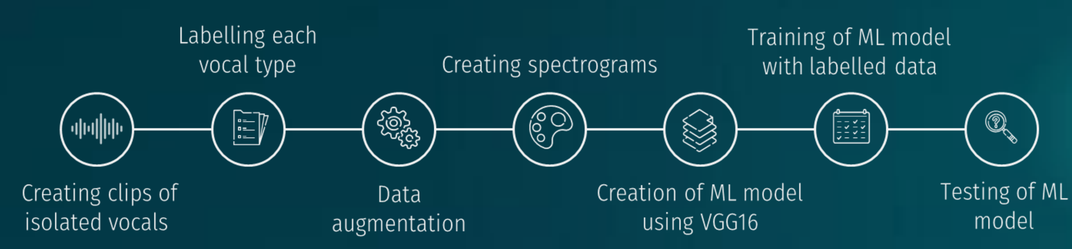

Figure 5: Method for the classification model

RESULTS

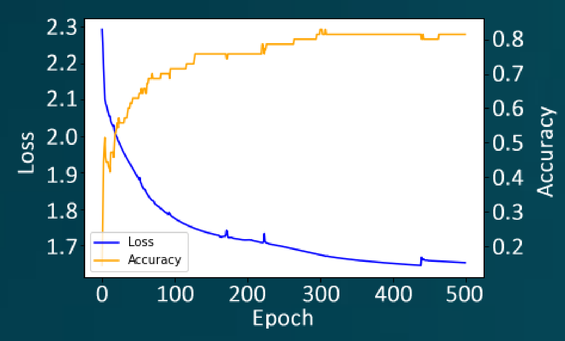

Figure 6: The loss and accuracy for each epoch during training

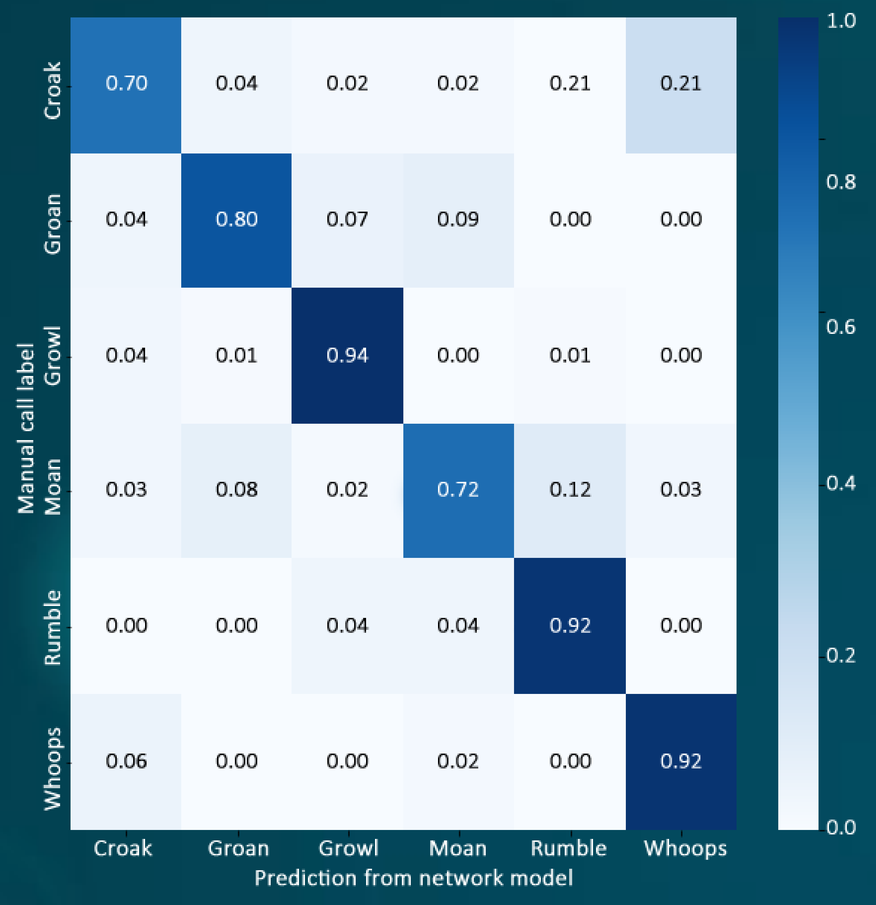

Figure 7: Confusion matrix presenting the correctly classified calls by type

Average classification accuracy of 84%

NEXT STEPS

- Develop rCNN for detection and localization
- Including larger dataset for ground truth
- Including data from more individuals
- Apply approach to field datasets with greater background noise

REFERENCE

[1] J.M. Sills et al. "Underwater hearing and communication in the endangered Hawaiian monk seal Neomonachus schauinslandi." In: *Endangered Species Research* 44 (2021), pp. 61–78. doi: 10.3354/esr01092.