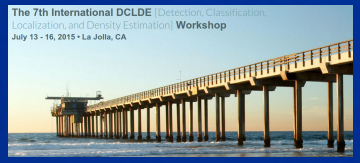
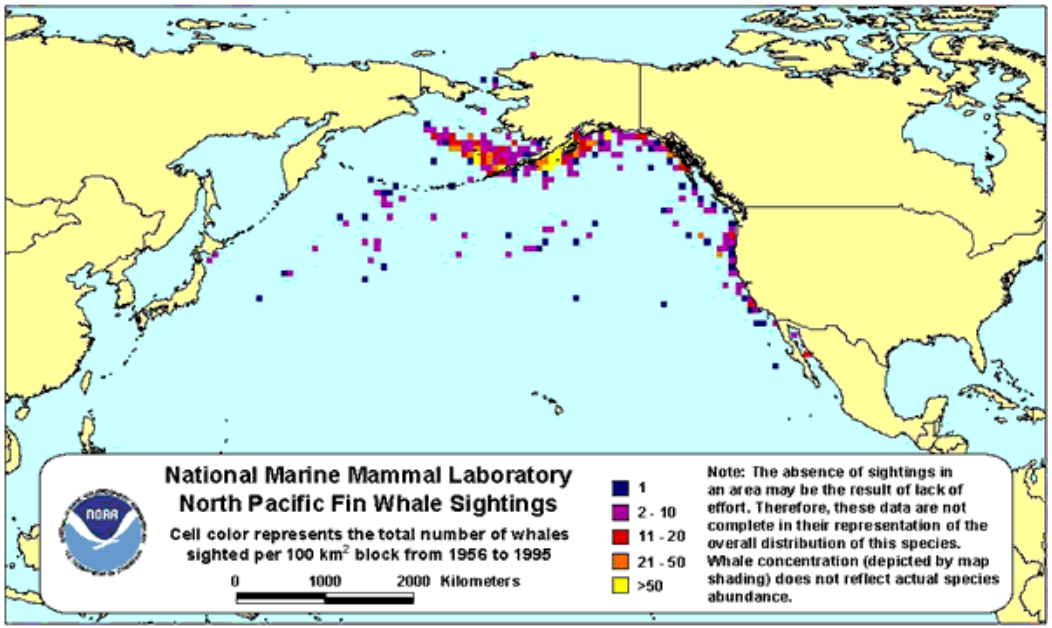
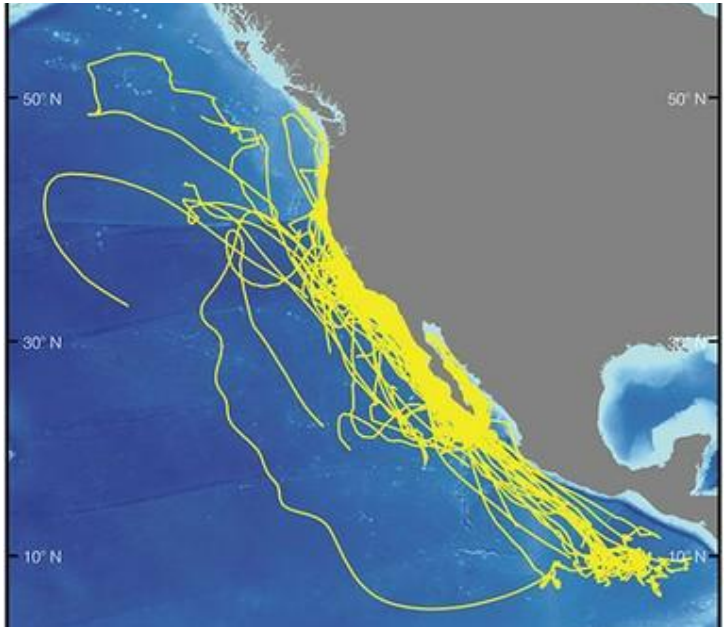
Classification of blue whale D calls and fin whale 40-Hz calls using deep learning

Jeremy Karnowski[1] & Yair Movshovitz-Attias[2]
[1]University of California, San Diego, USA
[2]Carnegie Mellon University, Pittsburgh, USA
jkarnows@cogsci.ucsd.edu @mwakanosya



Passive Acoustic Monitoring





- Blue whale and fin whale population sizes are declining.
- Vocalizations found from passive acoustic monitoring can provide massive amounts of data on population sizes and migratory patterns.

A Problem

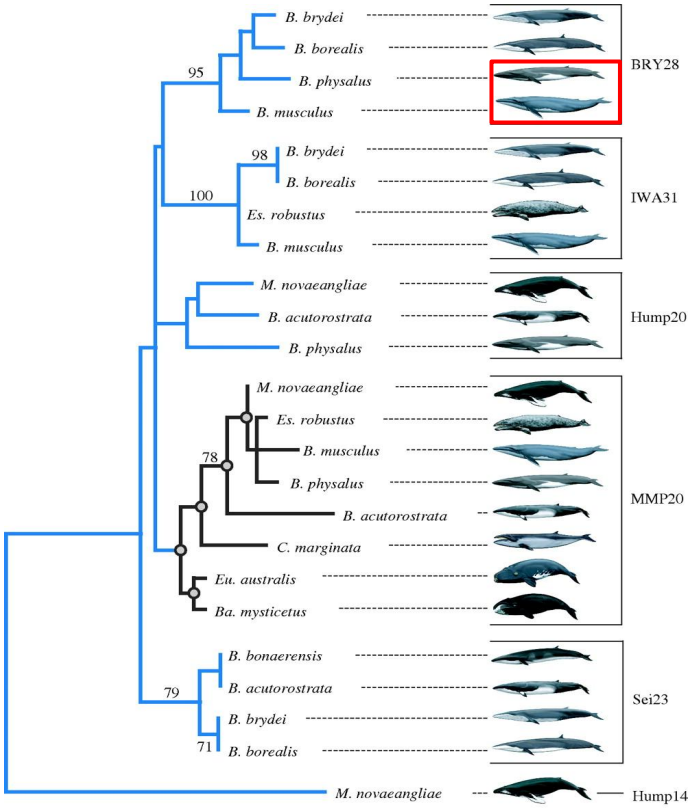
"Since the fin whale detectors can be triggered by blue whale calls, a separate detection algorithm for blue whales is being developed to allow for differentiation between the two."

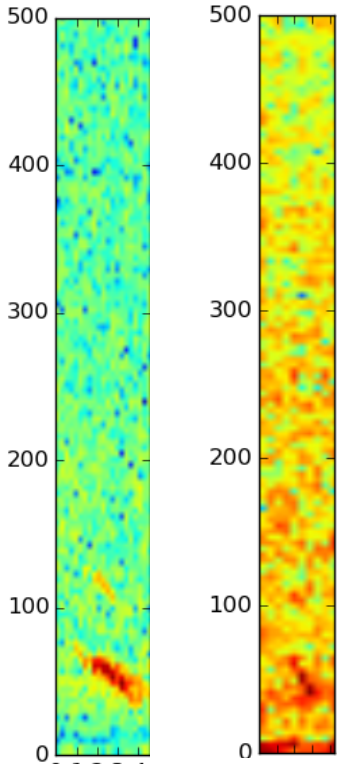
Weirathmueller , Wilcock , Soule (DCLDE 2011)

"Finally, ambiguity could arise in distinguishing blue whale D calls from fin whale 40-Hz calls in an LTSA even though D calls have a distinctly broader bandwidth (Oleson et al. 2007)"

Širović, Williams, Kerosky, Wiggins, and Hildebrand (2013)

Fin Whales and Blue Whales

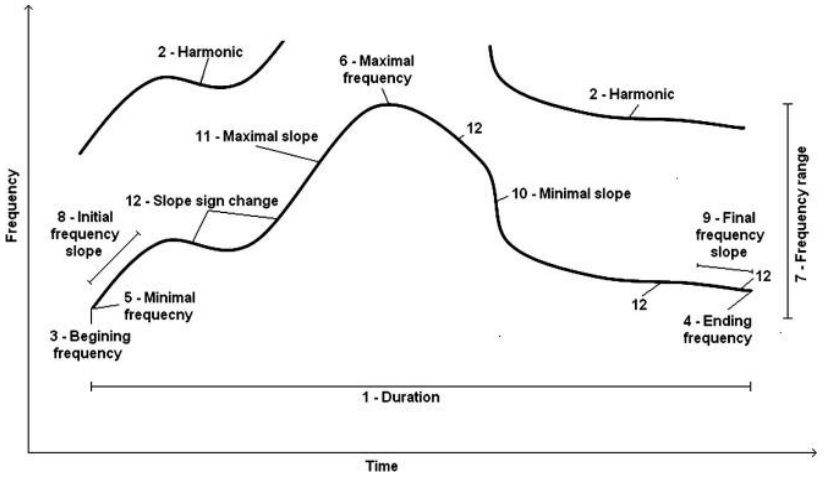




- Closely related species, so call production may be similar
- Evidence suggests that the fin whale 40-Hz call may be feeding call, similar to the blue whale D call

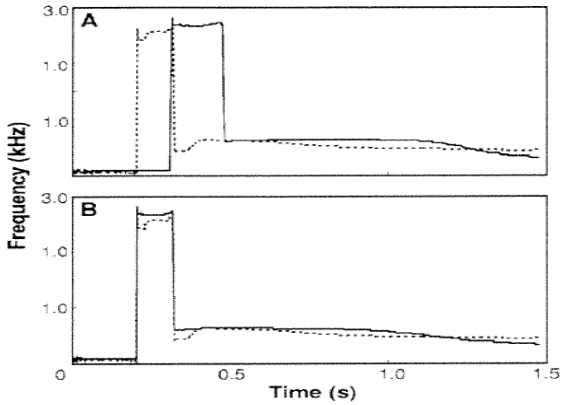
(Watkins (1981); Sirovic, Williams, Kerosky, Wiggins, and Hildebrand (2013))

Whistle Classification



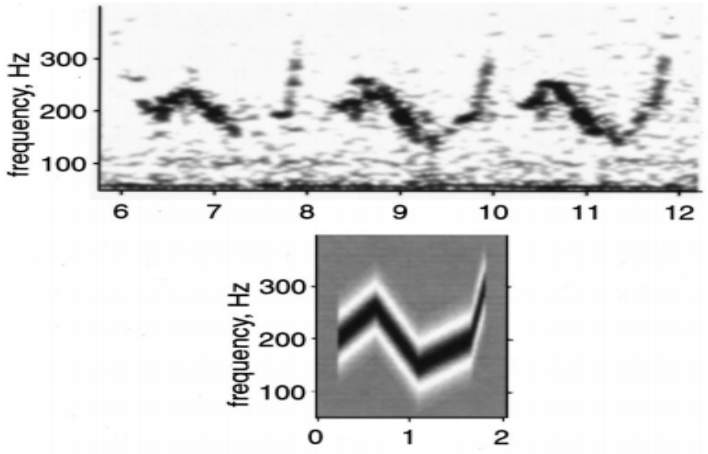
Feature Selection

(Gannier et al., 2000)



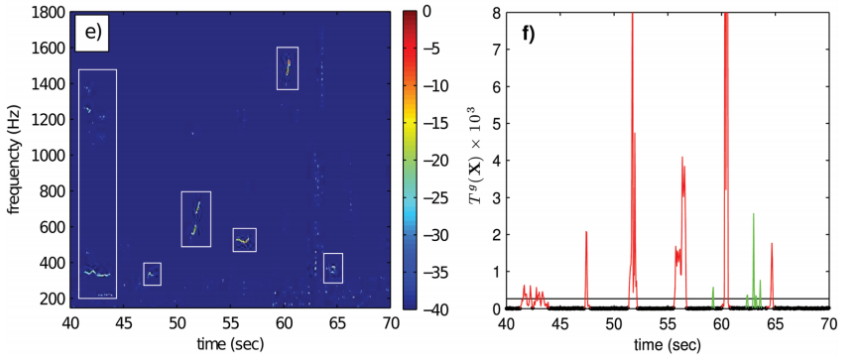
Dynamic Time Warping

(Deecke & Janik, 2006)



Spectrogram Correlation

(Mellinger & Clark, 2000)

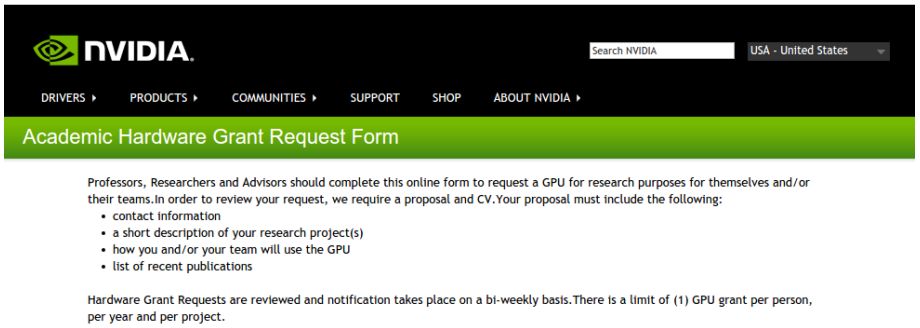


Generalized Power-Law

(Helble, Ierley, D'Spain, Roch, Hildebrand, 2012)

GPU and Deep Learning Packages

- Shallow Network:
  - Recognizing transient low-frequency whale sounds by spectrogram correlation (Mellinger, Clark 2000)
- Deep Network:
  - Practical deep neural nets for detecting marine mammals (Nouri, DCLDE 2013)



Academic Hardware Grant Request Form

BVLC Caffe (Setup)

theano

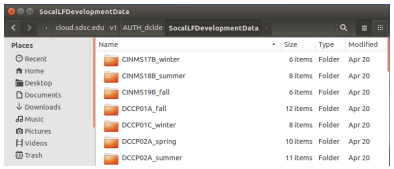
torch

Data Collection

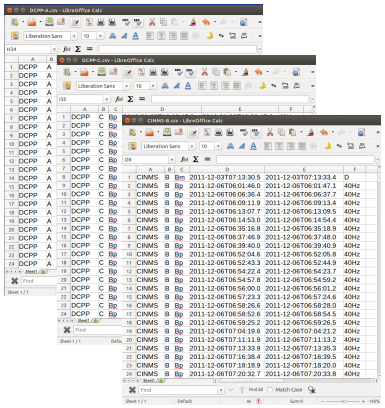
Over 1387 hours of audio recorded between 2009-2013 off the coast of Southern California

Dataset Creation

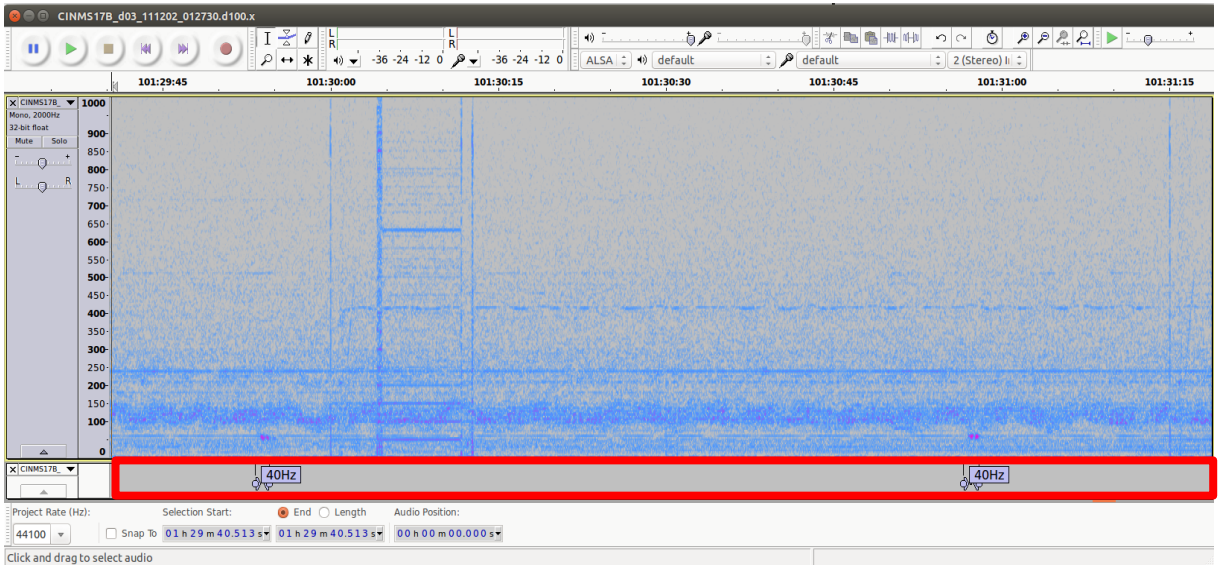
Creating annotation files for each audio file for visual inspection



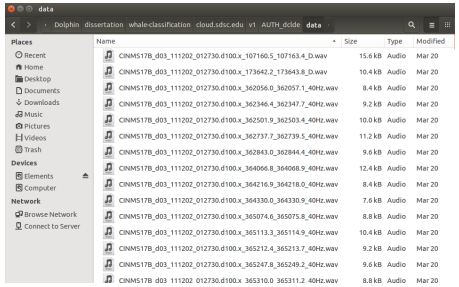
Audio files



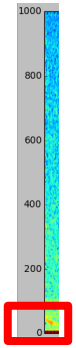
Annotations

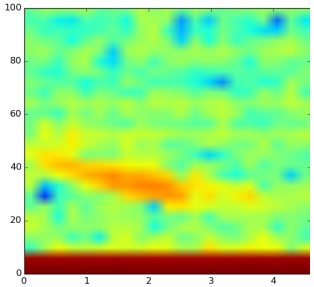


Creating audio dataset and spectrogram dataset

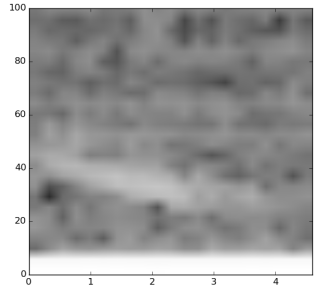


All vocalizations





Spectrograms (<100 Hz)

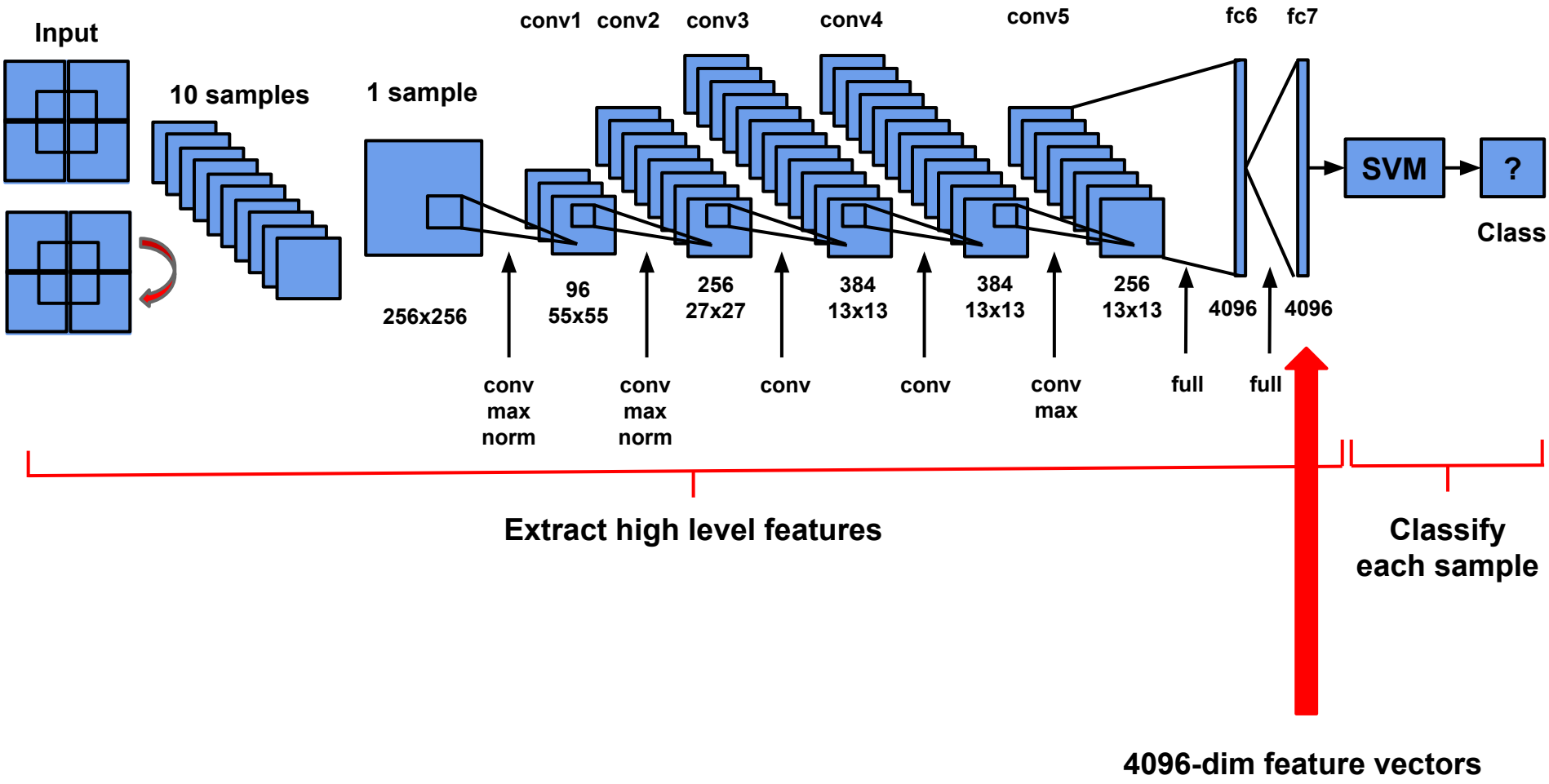


Scaled (0-255) as image

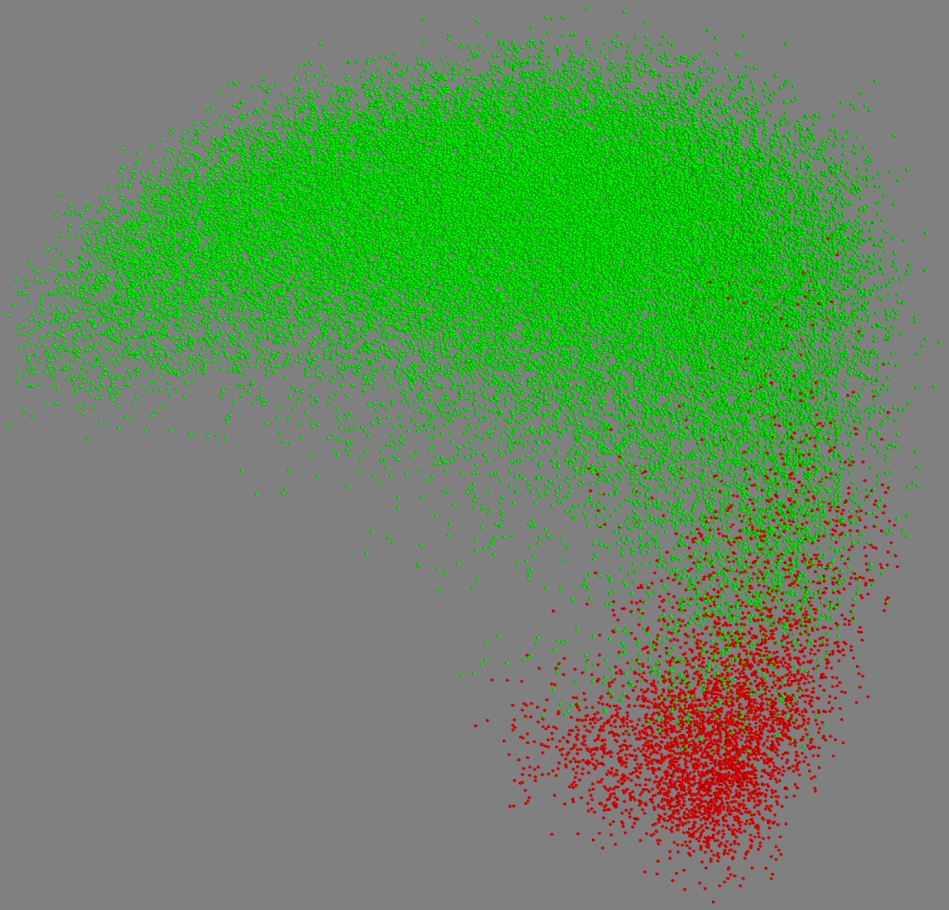
4796 D calls

415 40-Hz calls

AlexNet + SVM


Input
10 samples
1 sample
256x256
conv1
conv2
conv3
conv4
conv5
fc6
fc7
96
55x55
256
27x27
384
13x13
384
13x13
256
13x13
4096
4096
SVM
?
Class
conv
max
norm
conv
max
norm
conv
conv
conv
max
full
full
Extract high level features
Classify
each sample
4096-dim feature vectors

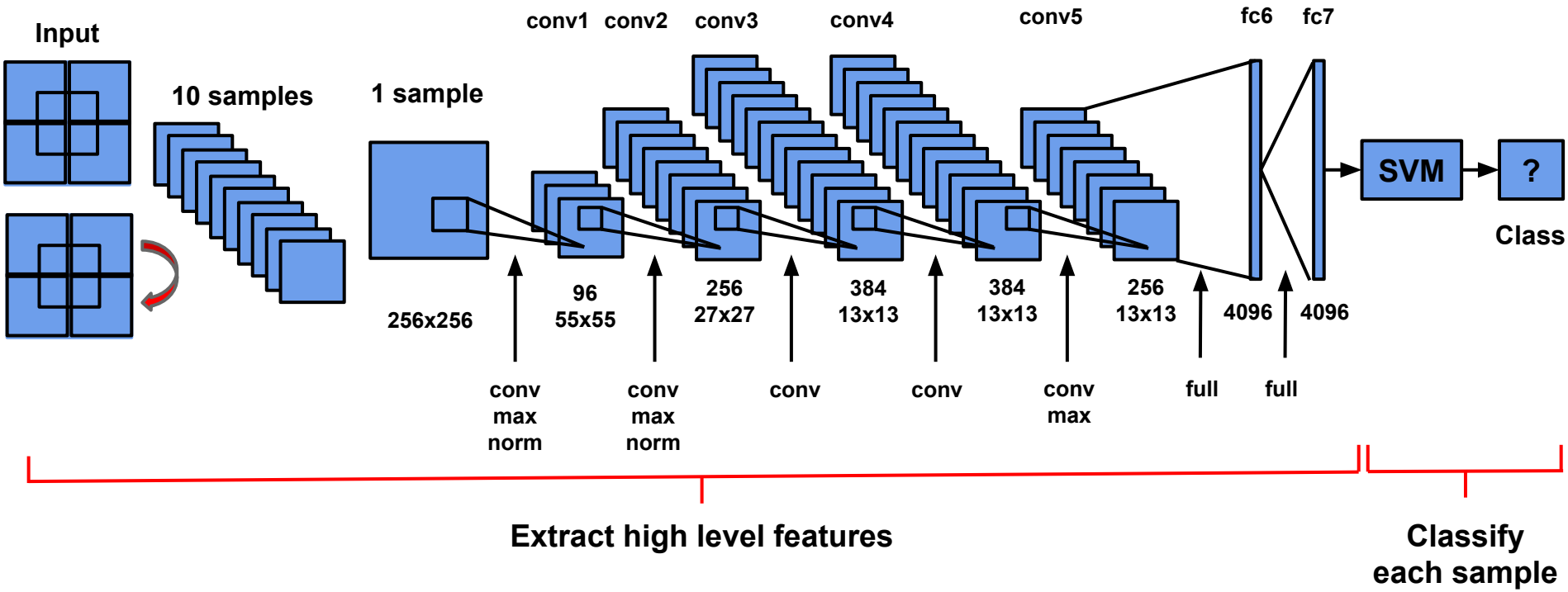
fc7 Feature Vectors in 2D (PCA)



blue whale
D calls

fin whale
40-Hz calls

AlexNet + SVM



- Linear SVM (C=0.0625)
- 10-fold cross-validation
- For each image, classify each sample as 0 (fin) or 1 (blue)
  - Take average of 10 samples and label call blue if > 0.5

Results

Confusion Matrix

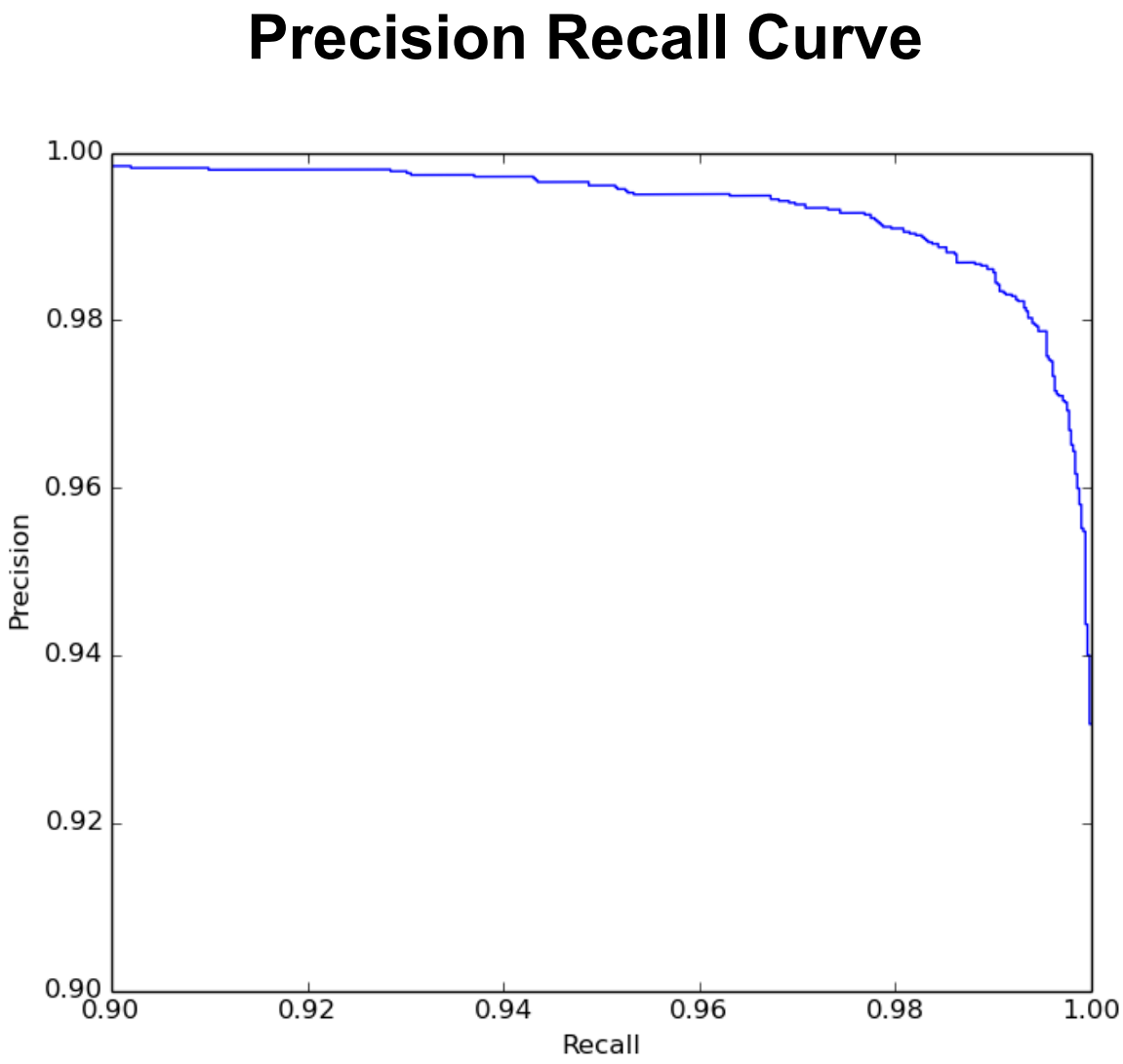
True \ Predicted	Blue whale	Fin whale
Blue whale	**4738**	**58**
Fin whale	**66**	**349**

97.62% Accuracy

For blue whales: 98.63% Precision

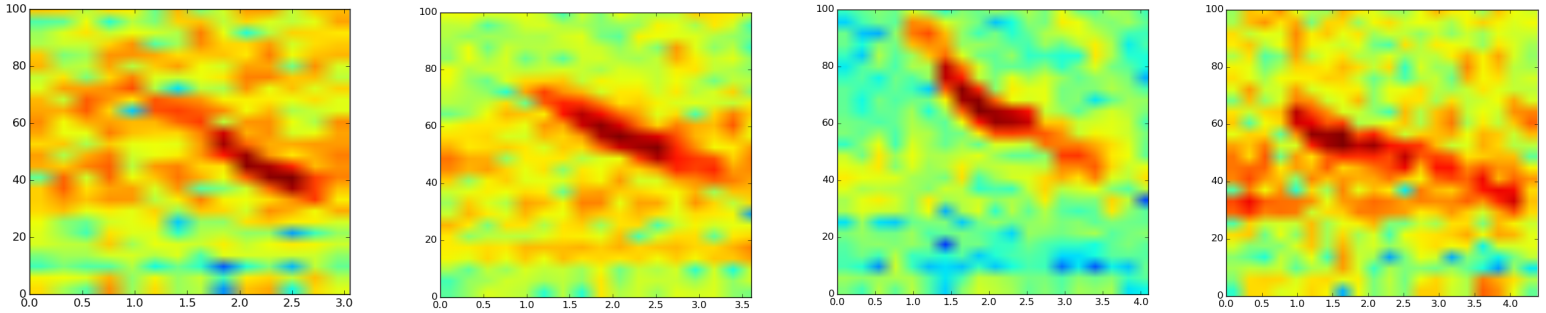
98.79% Recall

Results

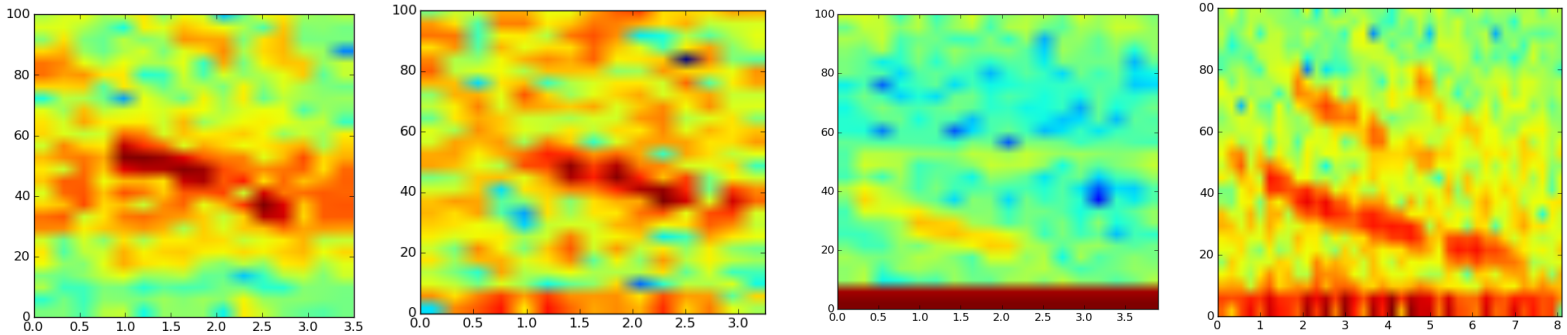

Precision Recall Curve
1.00
0.98
0.96
0.94
0.92
0.90
Precision
0.90
0.92
0.94
0.96
0.98
1.00
Recall

Results

Correctly classified blue whale D calls



Misclassified blue whale D calls



Lower frequency range? Different slope?

Harmonics? Different call?

Future Directions

- Compare this method with other classification methods
- Clean up noise in spectrogram before classification
- Obtain more data from noisier environments to make the detectors more robust
- Add in detection: Use GPL detector (Helble et al. 2012) and then classification on the found calls.
  - Determine the time savings for users

Contributions

- Created more targeted classification datasets
- Used deep learning methods for a novel whistle classification task
- Very good performance in accuracy, precision, and recall
- Easy to modify - researchers can add in additional categories: 50-Hz calls and other false tonal detections

Thanks!

(for all the fish)

Support

Elizabeth Vu

Tyler Helble

John Hildebrand

Edwin Hutchins

Christine Johnson

Funding

CARTA

Center for Academic Research & Training in Anthropogeny

"to explore and explain the origins of the human phenomenon"

THE KAVLI Institute for Brain and Mind at UCSD

NIMBioS

National Institute for Mathematical and Biological Synthesis

Temporal Dynamics of Learning Center

BROOKFIELD ZOO

Chicago Zoological Society

Inspiring Conservation Leadership