Automatic bird sound detection in long range field recordings using Mel filter bank features

Suhas
Feb 2020

The Plan

- Motivation
- Previous works
- Dataset considered
- Features, methods and evaluation
- Some challenges involved
- Deliverables - How does the filterbank features stand against the current state of the art methods?

Why are we looking at this topic?

- Bird sound detection from far-field recordings is an important means for understanding the effect of urbanization on our environment.

- Can we substitute manual sightings?

The story so far

- Earliest works such as Tzanetakis [1] looked at applications : Speech vs music, Voices, Classical music recognition,etc. using DWTC, MFCC, STFTC

- M. Daniels [2] proposed a wavelet based method for percussion sound analysis. (In-house data) with db4, db5, and sym5 wavelets v/s MFCC features + SVM

[1] Tzanetakis, G., Essl, G., & Cook, P. (2001, September). Audio analysis using the discrete wavelet transform. In Proc. Conf. in Acoustics and Music Theory Applications (Vol. 66).

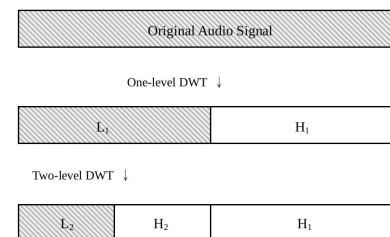[2] Daniels, M. (2010). Classification of Percussive Sounds Using Wavelet-Based. CCRMA, Stanford University thesis.

PennState
College of Engineering

ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE

The story so far

- Muscle Fish dataset, which consists of 410 sounds in 16 classes, is used to evaluate the performance. (Wavelets + FCC + SVM) [3]
- Hsieh [4] proposed a method of extracting wavelet features from audio - reduced the size of recordings. Followed by suprasegmental features.

Feature		Type of transforms	Number of features
Perceptual feature	Subband power P_j	Wavelet	3
	Pitch frequency f_p	Wavelet	1
	Brightness ω_c	Fourier	1
	Bandwidth B	Fourier	1
Frequency cepstral coefficient (FCC) c_n		Fourier	L

Original Audio Signal

One-level DWT ↓

L_1	H_1

Two-level DWT ↓

L_2	H_2	H_1

[3] Lin, C. C., Chen, S. H., Truong, T. K., & Chang, Y. (2005). Audio classification and categorization based on wavelets and support vector machine. IEEE Transactions on Speech and Audio Processing, 13(5), 644-651.

[4] Hsieh, S. L., & Wang, H. C. (2005). Feature Extraction for Audio Fingerprinting Using Wavelet Transform. In National Computer Conference.

PennState
College of Engineering

ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE

The story so far

clips containing one flight call (positive clips)   clips containing no flight call (negative clips)

BirdVox-70k dataset

convolutional neural network

presence vs. absence

- Mel Spectrogram + CNN using Urban-8K dataset and species classification from clips of avian flight calls (CLO-43SD dataset) [5]

  Looked at short term & long term recordings

- SoA : MFBE + CNN using Freefield1010 + Warblr (FF W 1) [6]

[5] Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., & Bello, J. P. (2019). Robust sound event detection in bioacoustic sensor networks. PloS one, 14(10).

[6] Pellegrini, T. (2017, August). Densely connected CNNs for bird audio detection. In 2017 25th European Signal Processing Conference (EUSIPCO) (pp. 1734-1738). IEEE.

PennState
College of Engineering

ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE
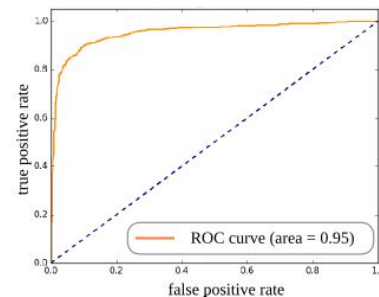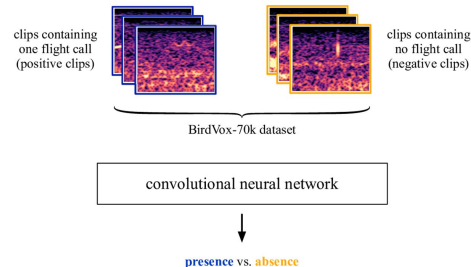
Dataset

- Field recordings (freefield1010)
- A collection of over 7,000 excerpts from field recordings around the world, gathered by the FreeSound project, and then standardised for research.
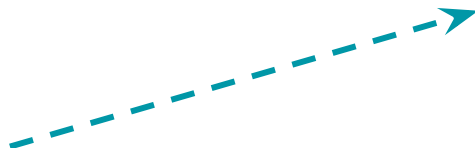- This collection is very diverse in location and environment.
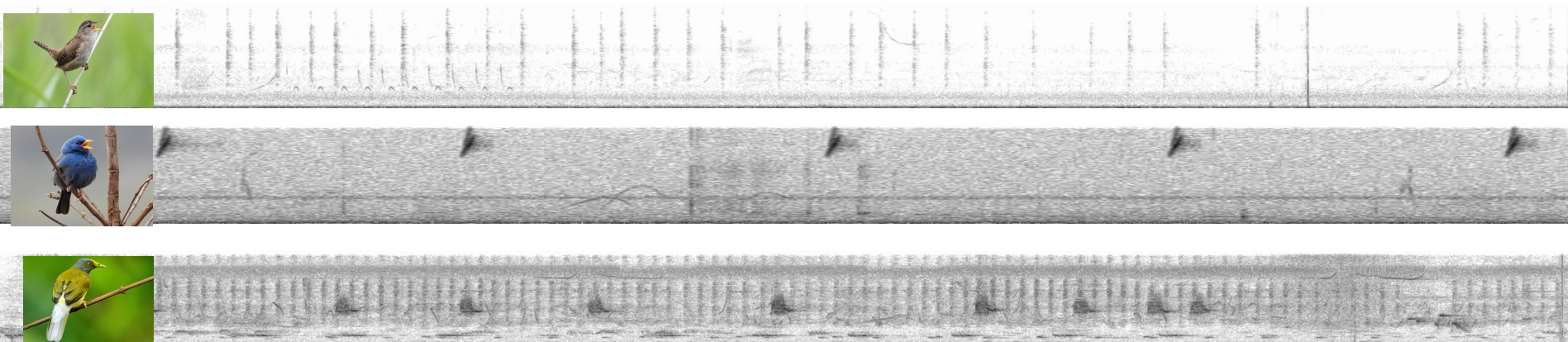
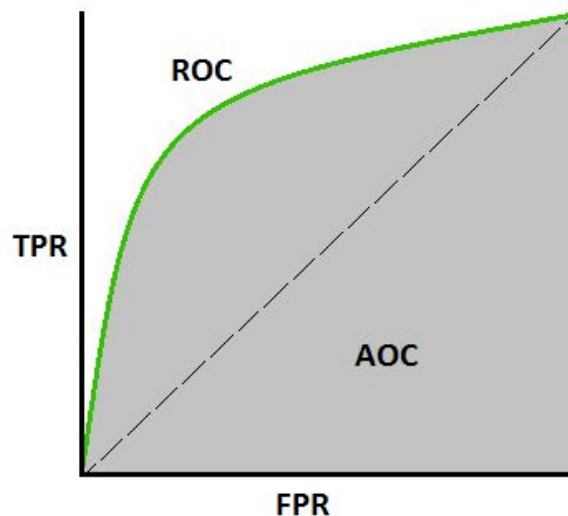# Features, methods and evaluation

- Mel frequency cepstral co-efficients
- Mel filter bank energies
- Two - level DWT
- Spectrogram features

- SVM
- CNN (SoA)

- AUC-ROC curves

SoA : State of the art methods

Deliverables - How does the filterbank features stand against the current state of the art methods?

Challenges involved

- Far field recordings pose some challenges - can we still achieve a good classification accuracy?

- Do these methods work for specific species of birds or is it generic?