Spotting the Mockingjay: Learning to Detect Bird Sounds Amidst Background Noise

CS 172B: Neural Networks and Deep Learning | 19 May 2020

Group 1

- Taneisha Arora (arorat@uci.edu)
- Thanasi Bakis (abakis@uci.edu)
- Theja Krishna (takrishn@uci.edu)
- Bryon Tjanaka (btjanaka@uci.edu)

Overview

Task

- Detect bird sounds in audio

Applications

- Wildlife monitoring
  - Chernobyl
- Audio library management
  - Xeno-Canto
- Pre-filtering for other tasks
  - Classification

Bird Audio Detection (BAD) Challenge

History

- DCASE 2018 (Detection and Classification of Acoustic Scenes and Events)
- 2016 IEEE Challenge (Machine Listening Lab, QMUL) (Stowell et al. 2016)

Description

- Train on 3 datasets, test on 3 separate datasets

Evaluation

- Output P (bird | audio), final score: AUC
- SOTA: 95% AUC on test datasets (Lasseck 2018)

Data Collections

Training

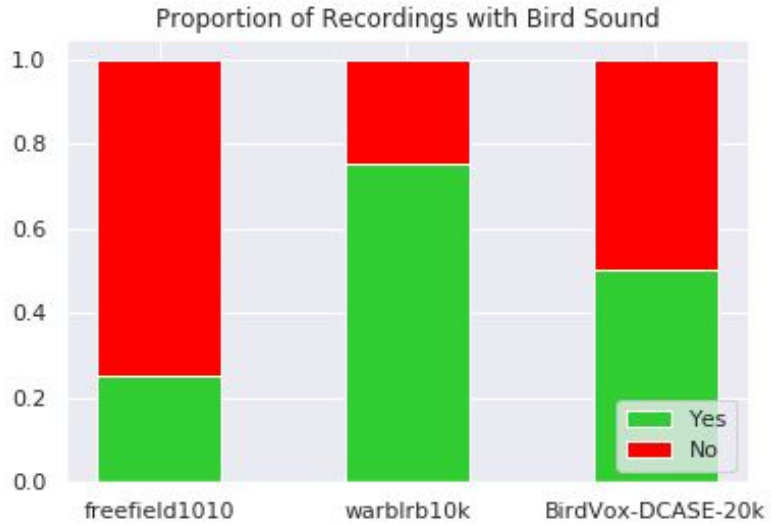
- freefield1010
  - 7,690 recordings from around the world
- warblrb10k
  - 10,000 recordings crowdsourced from around the UK
- BirdVox-DCASE-20k
  - 20,000 recordings from remote monitoring units around Ithaca, NY

Testing

- warblrb10k, Chernobyl, PolandNFC

Data Analysis

- Audio
  - 16-bit 44.1 kHz WAV files
  - ~10 s
- Proportion
  - Differs



Feature Extraction

- MFCC
- Spectral Contrast
- Chromagram

MFCC (Mel-Frequency Cepstral Coefficients)

What is it?

The rate of change of the frequencies over time:

Signal → Fourier transform → Mel scale frequencies → log magnitude → cosine transform

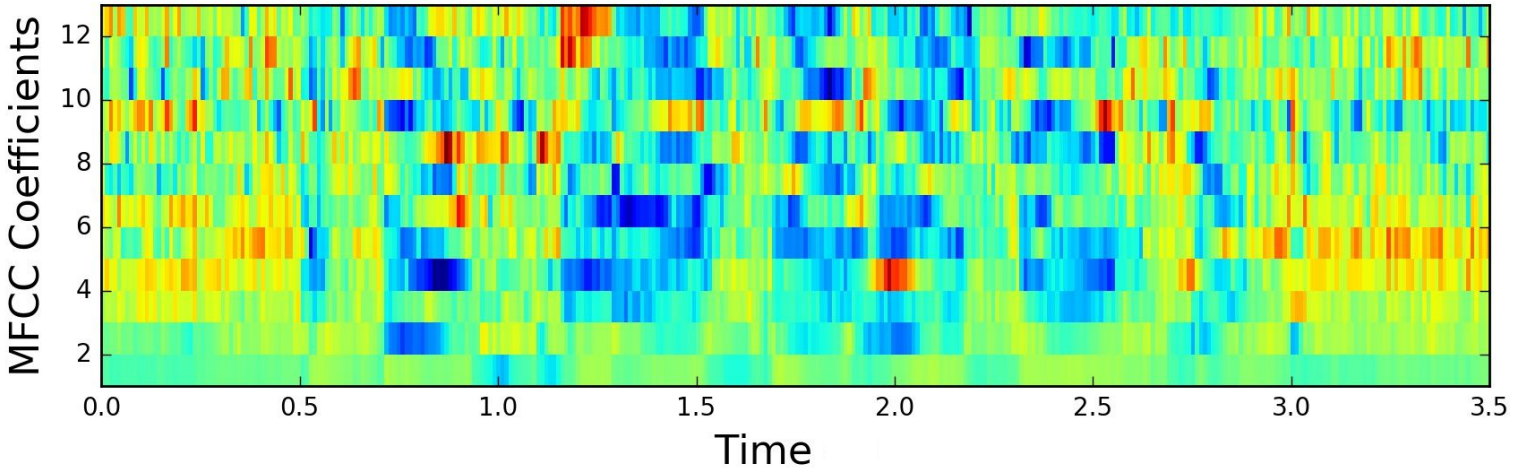
What does it tell us?

The timbre of the sound, or “what the sound sounds like”, at any given point in time

* Timbre is the characteristics of the sound; it’s what makes an instrument/voice sound like itself.

MFCC (Mel-Frequency Cepstral Coefficients)

What does it look like?



Posted by u/jewdai 4 years ago

10

ELI5: MFCC (Mel-frequency cepstral coefficients)

4 Comments Give Award Share Save Hide Report 100% Upvoted

This thread is archived

New comments cannot be posted and votes cannot be cast

SORT BY **BEST**

[deleted] 3 points · 4 years ago

I don't think a 5 year old is going to get this, unless he's somehow in grad school.

Share Report Save

Spectral Contrast

What is it?

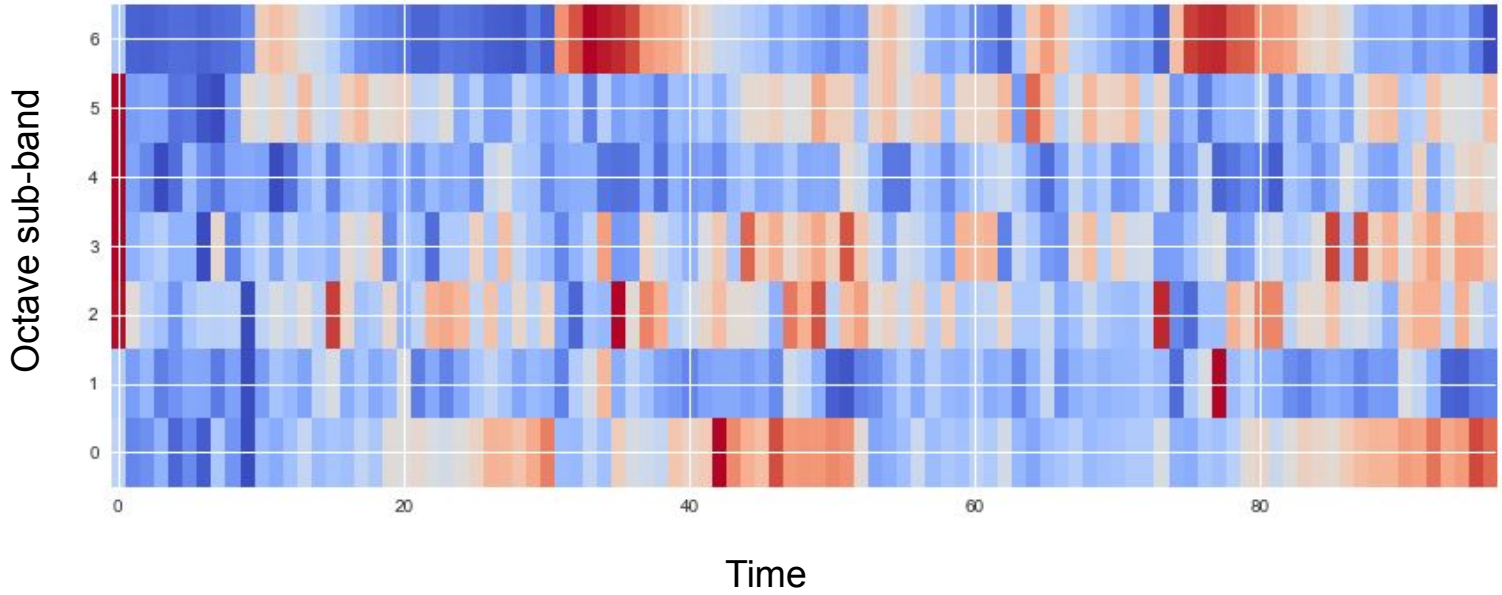
The difference in intensities of the spectral peak and valley (within octave sub-bands) over time

What does it tell us?

“The variety of sounds” or “noisiness” in the audio at any given point in time

Spectral Contrast

What does it look like?



Chromagram

What is it?

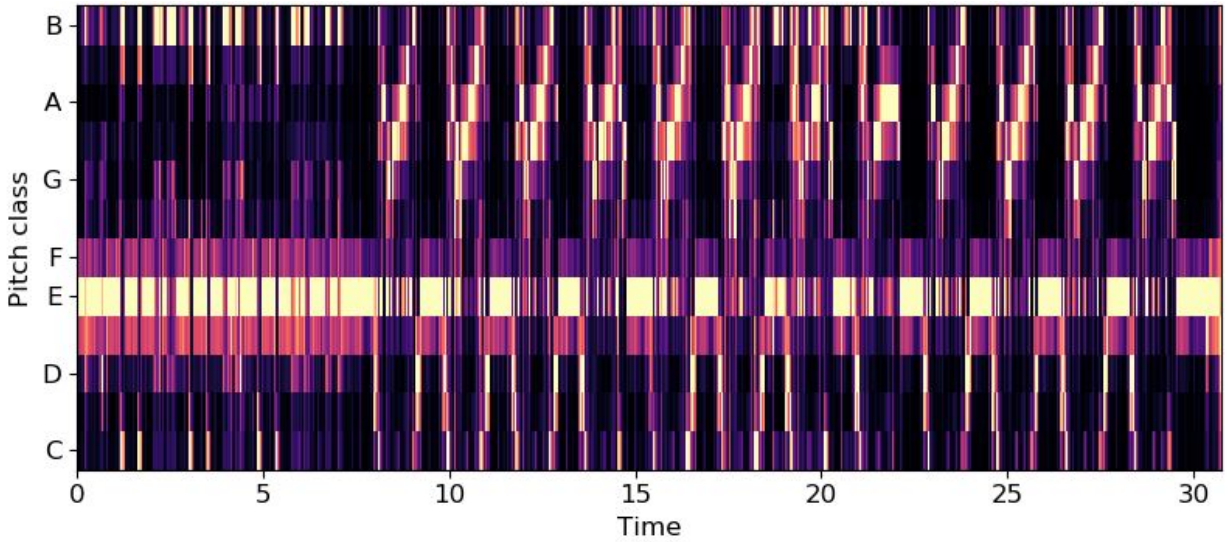
The intensities of the 12 distinct semitones (across all octaves) over time

What does it tell us?

“How much each note is present” in the audio (C, C#, D, D#, ...) at any given point in time

Chromagram

What does it look like?



Data Cleaning

1. Transform all audio files into a feature representation
    a. An N x T matrix, where N = # of filters and T = time
2. Clip longer features along time axis to be (N, 450)
3. Pad shorter features along time axis to be (N, 450)

Basic Model

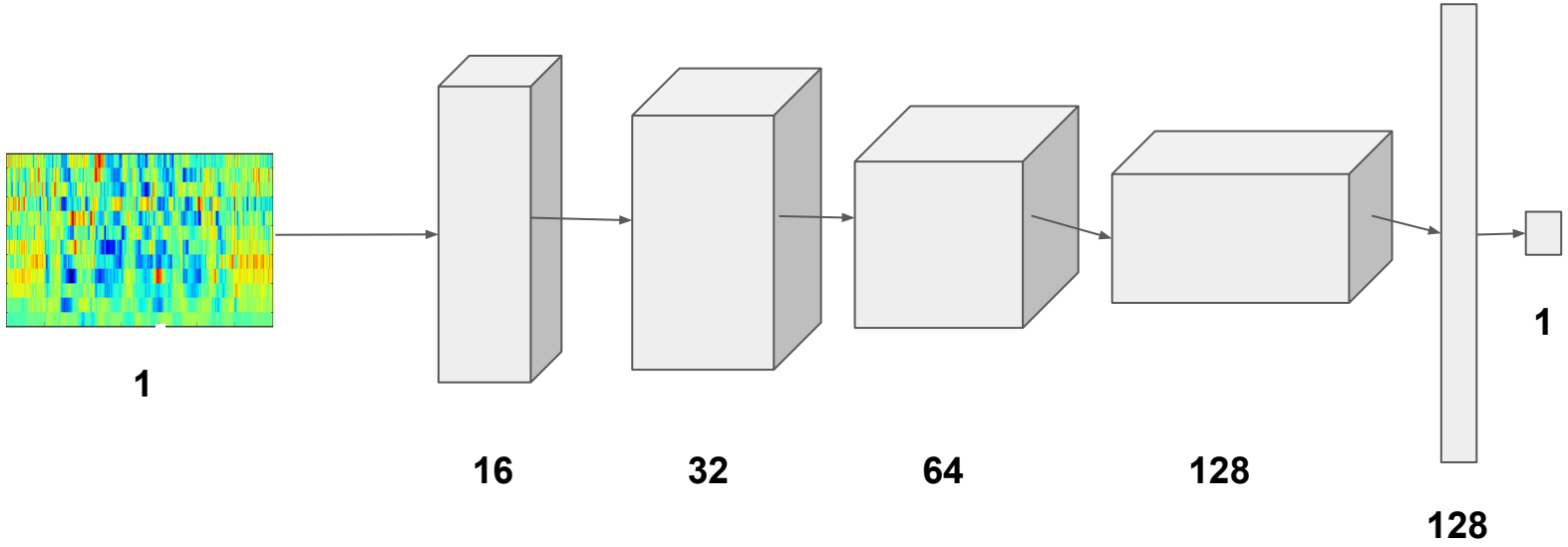
Inspiration: Urban Sound Classification example on Medium

Motivation: Get something to work, understand *how* it works later

Inputs: MFCC representations (as described earlier)

Output: P (bird | audio)

Basic Model Architecture



Multi-Feature Model

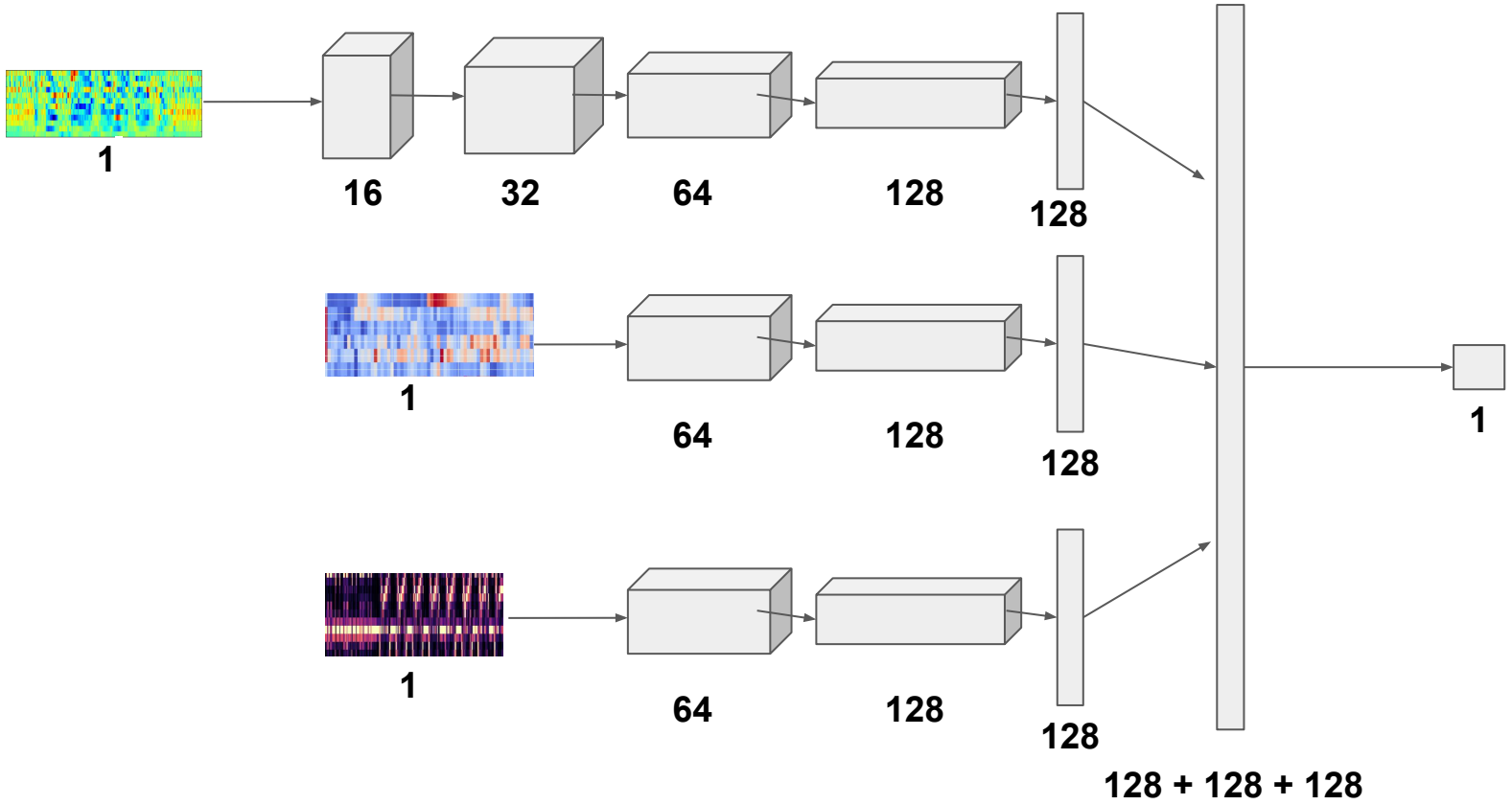
Inspiration: Su et al. 2019

Motivation: Combined representations -> greater accuracy?

Inputs: MFCC, Spectral Contrast, Chromagram

Output: P (bird | audio)

Multi-Feature Architecture



Training

Parameters

- 100 epochs
- Batch size 16
- Adam optimizer
- Binary Cross-Entropy loss

Hardware

- CUDA-enabled NVIDIA GTX 1650

Training Time

- 30-60 minutes per model

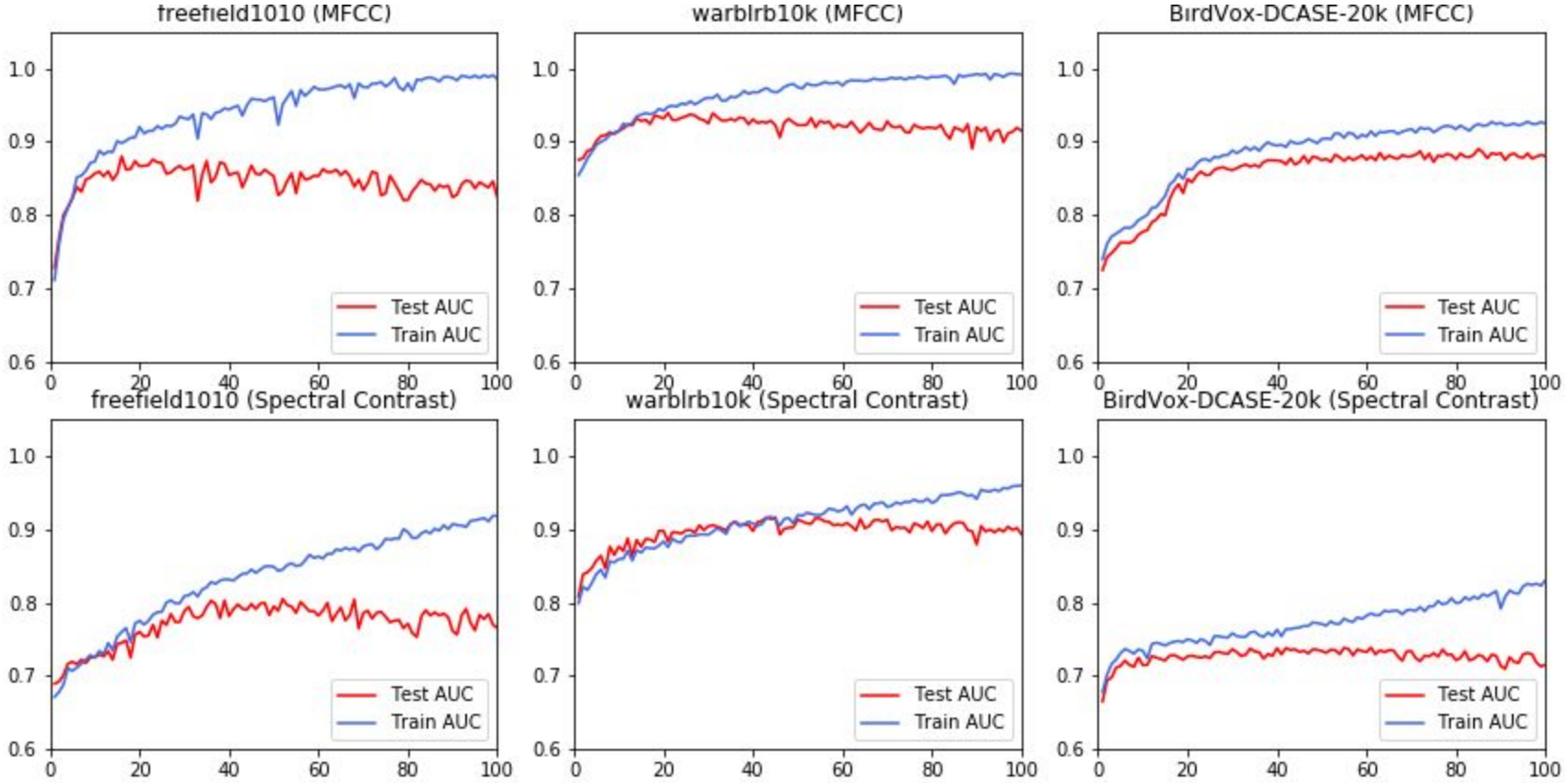
Data

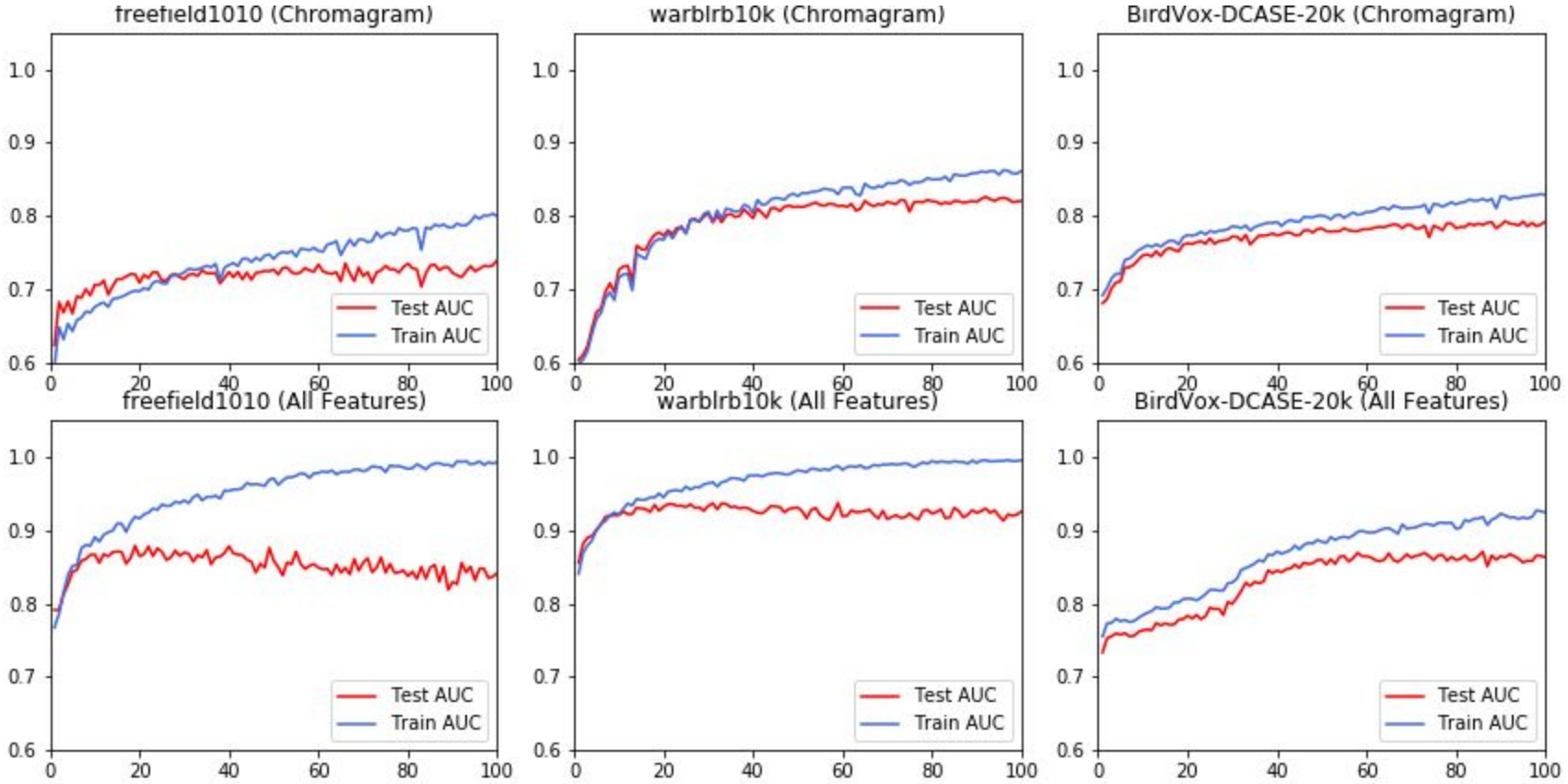
- 80-20 train-test split

Software

Results

Features	Test AUC		
	freefield1010	warblrb10k	BirdVox-DCASE-20k
MFCC	0.8236	0.9148	0.8800
Spectral Contrast	0.7653	0.8938	0.7156
Chromagram	0.7399	0.8203	0.7924
All	0.8425	0.9263	0.8635


freefield1010 (MFCC)
warblrb10k (MFCC)
BirdVox-DCASE-20k (MFCC)
freefield1010 (Spectral Contrast)
warblrb10k (Spectral Contrast)
BirdVox-DCASE-20k (Spectral Contrast)
Test AUC
Train AUC
1.0
0.9
0.8
0.7
0.6
0
20
40
60
80
100


freefield1010 (Chromagram)
warblrb10k (Chromagram)
BirdVox-DCASE-20k (Chromagram)
freefield1010 (All Features)
warblrb10k (All Features)
BirdVox-DCASE-20k (All Features)
Test AUC
Train AUC
1.0
0.9
0.8
0.7
0.6
0
20
40
60
80
100

Discussion

Observations

- All > Individual
- MFCC > Spectral Contrast (SC), Chroma

Explanation

- All provides more information
- MFCC provides more relevant information compared to SC and Chroma

Future Work

- Models
  - RNNs / LSTMs / GRUs
  - Different CNNs
  - Ensembles
- Data
  - Add small Gaussian noise
  - Add chunks from other files
  - Cross-validation on different datasets
  - Access test set
- Computation
  - Google Cloud

References

1. Y. Su, K. Zhang, J. Wang, and K. Madani, “Environment sound classification using a two-stream cnn basedon decision-level fusion,”Sensors, vol. 19, p. 1733, 04 2019.
2. M. Lasseck, “Acoustic bird detection with deep convolutional neural networks,” tech. rep., DCASE2018Challenge, September 2018.
3. T. Grill and J. Schlüter, “Two convolutional neural networks for bird detection in audio signals,” in201725th European Signal Processing Conference (EUSIPCO), pp. 1764–1768, 2017.
4. D. Stowell, M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin, “Automatic acoustic detection of birdsthrough deep learning: The first bird audio detection challenge,”Methods in Ecology and Evolution, vol. 10,no. 3, pp. 368–380, 2019.
5. D. Stowell, M. Wood, Y. Stylianou, and H. Glotin, “Bird detection in audio: a survey and a challenge,”CoRR, vol. abs/1608.03417, 2016.

Thanks!