Transfer Learning with Semi-Supervised Dataset Annotation for Birdcall Classification

BirdCLEF 2023, Team DS@GT

Anthony Miyaguchi
acmiyaguchi@gatech.edu

Georgia Institute of Technology

2023-09-11

DS@GT Competition Team

Recruitment

- Built a team of 4 people from DS@GT in Spring 2023
- 3 masters, 1 undergraduate

Technical Approach

- Use Bird-MixIT to sound separate audio
- Use BirdNET to generate embeddings and labels
- Generate pseudo-labels with heuristics
- Solve supervised classification problem

Figure 1: DS@GT: a student-run data science organization

Team DS@GT

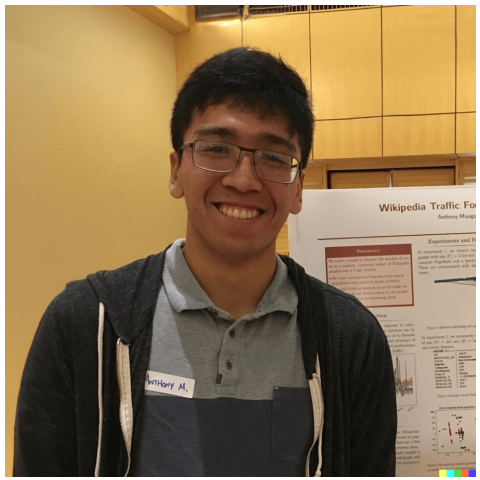


Figure 2: Anthony Miyaguchi

Figure 3: Nathan Zhong

Figure 4: Murilo Gustineli

Figure 5: Chris Hayduk

Why is audio classification challenging?

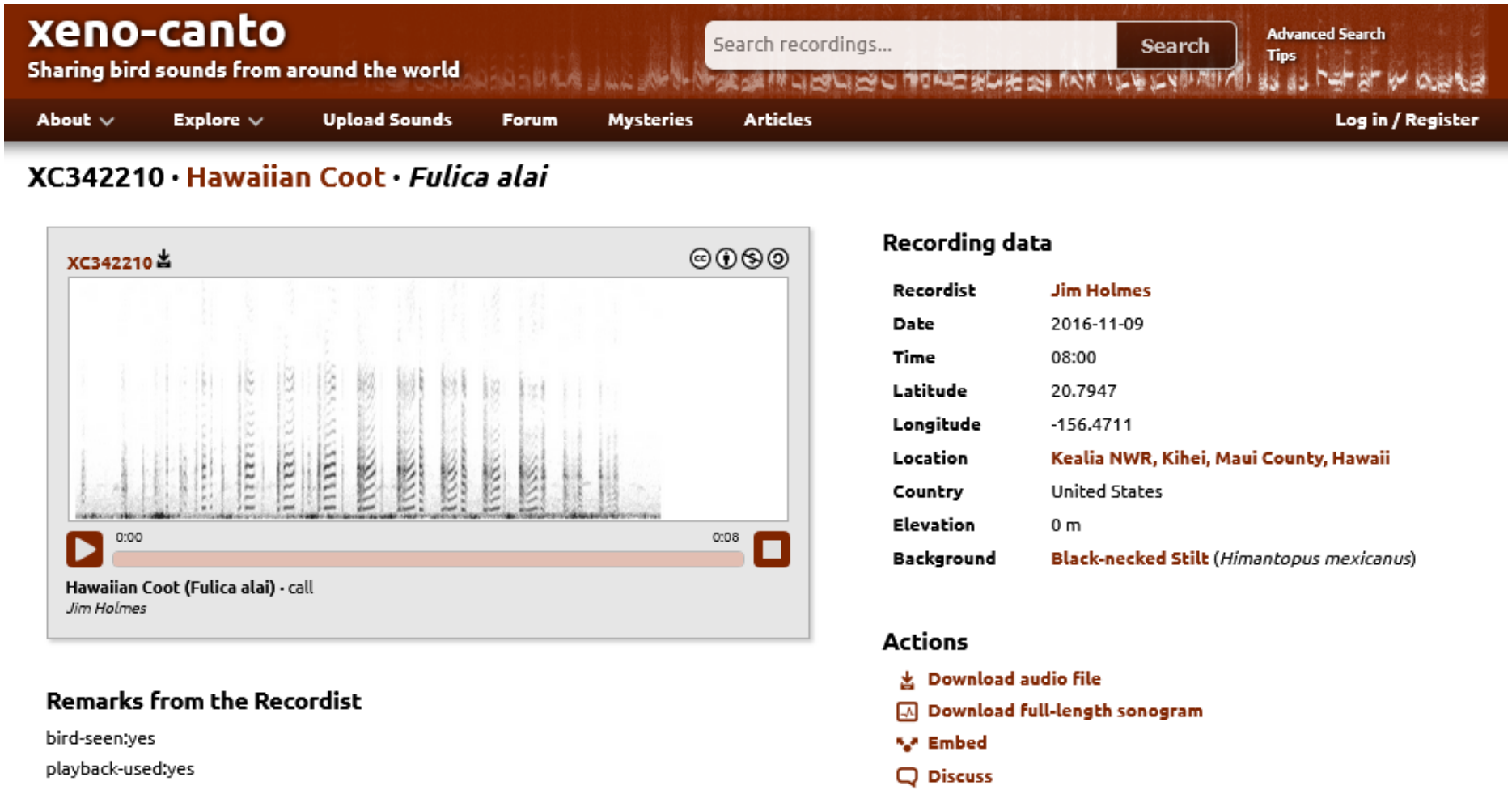


Figure 6: xeno-canto is a crowd sourced database of bird sounds.

Reading the literature

Domain specific deep learning model - BirdNET

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics, 61, 101236.

Sound separation - MixIT

Denton, T., Wisdom, S., & Hershey, J. R. (2022, May). Improving bird classification with unsupervised sound separation. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 636-640). IEEE.

Approach

Outline

- Use Bird-MixIT to sound separate audio
- Use BirdNET to generate embeddings and labels
- Generate pseudo-labels with heuristics
- Solve supervised classification problem

Sound Separation with MixIT

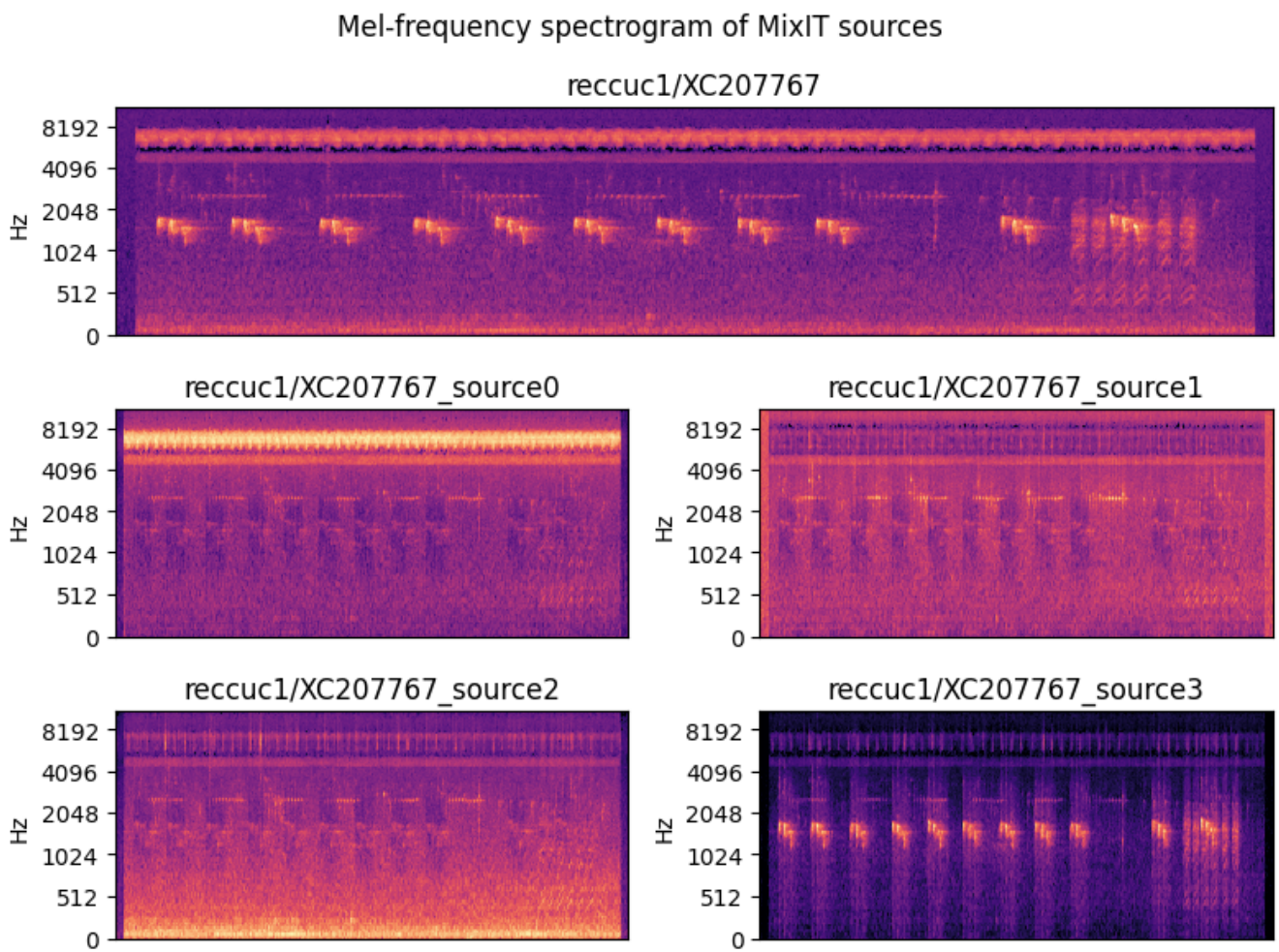


Figure 7: MixIT is a sound separation algorithm.

BirdNET embeddings

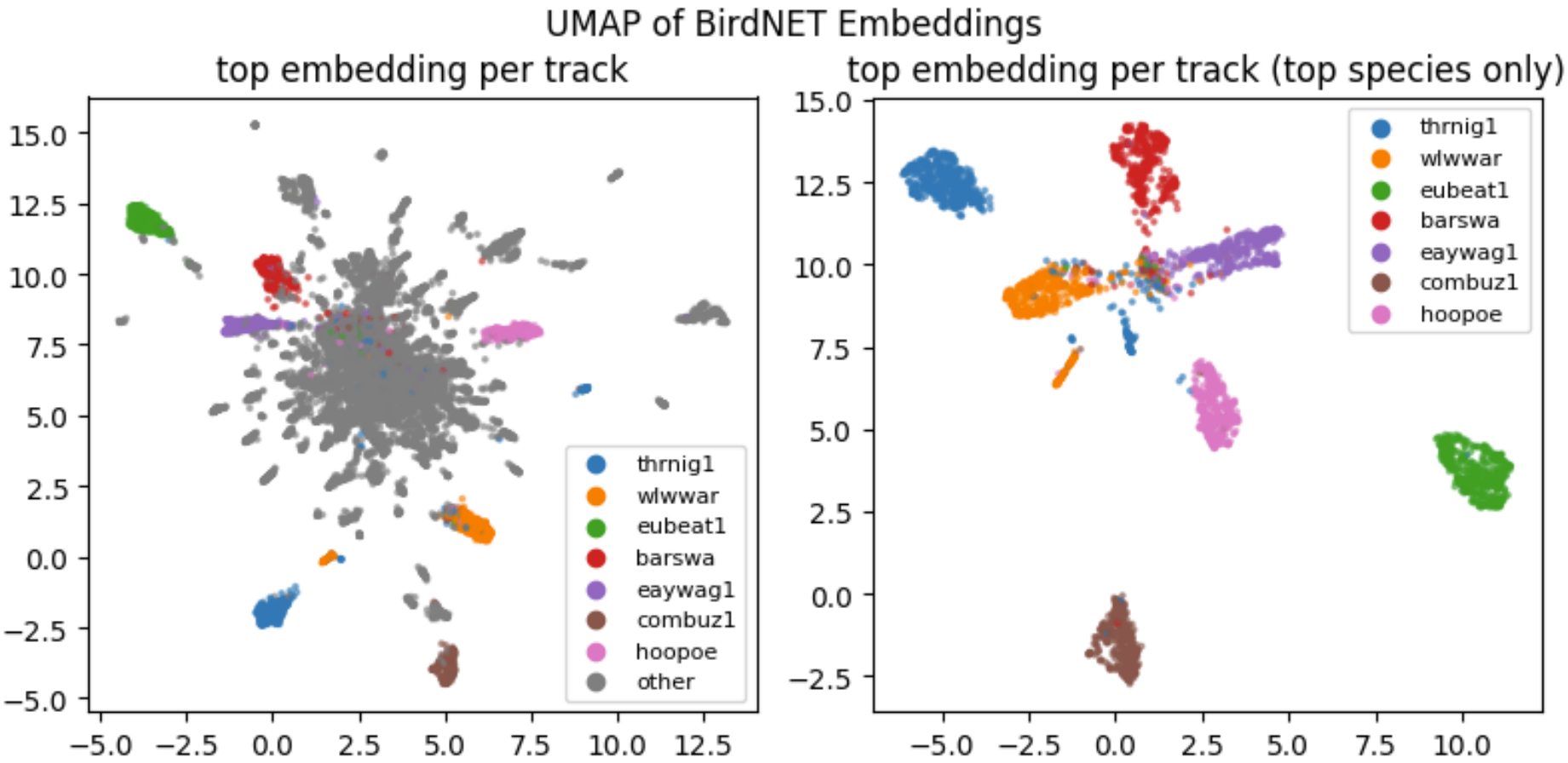


Figure 8: We can use the BirdNET embedding space for search and nearest neighbor queries.

BirdNET soft-labels as pseudo-labels

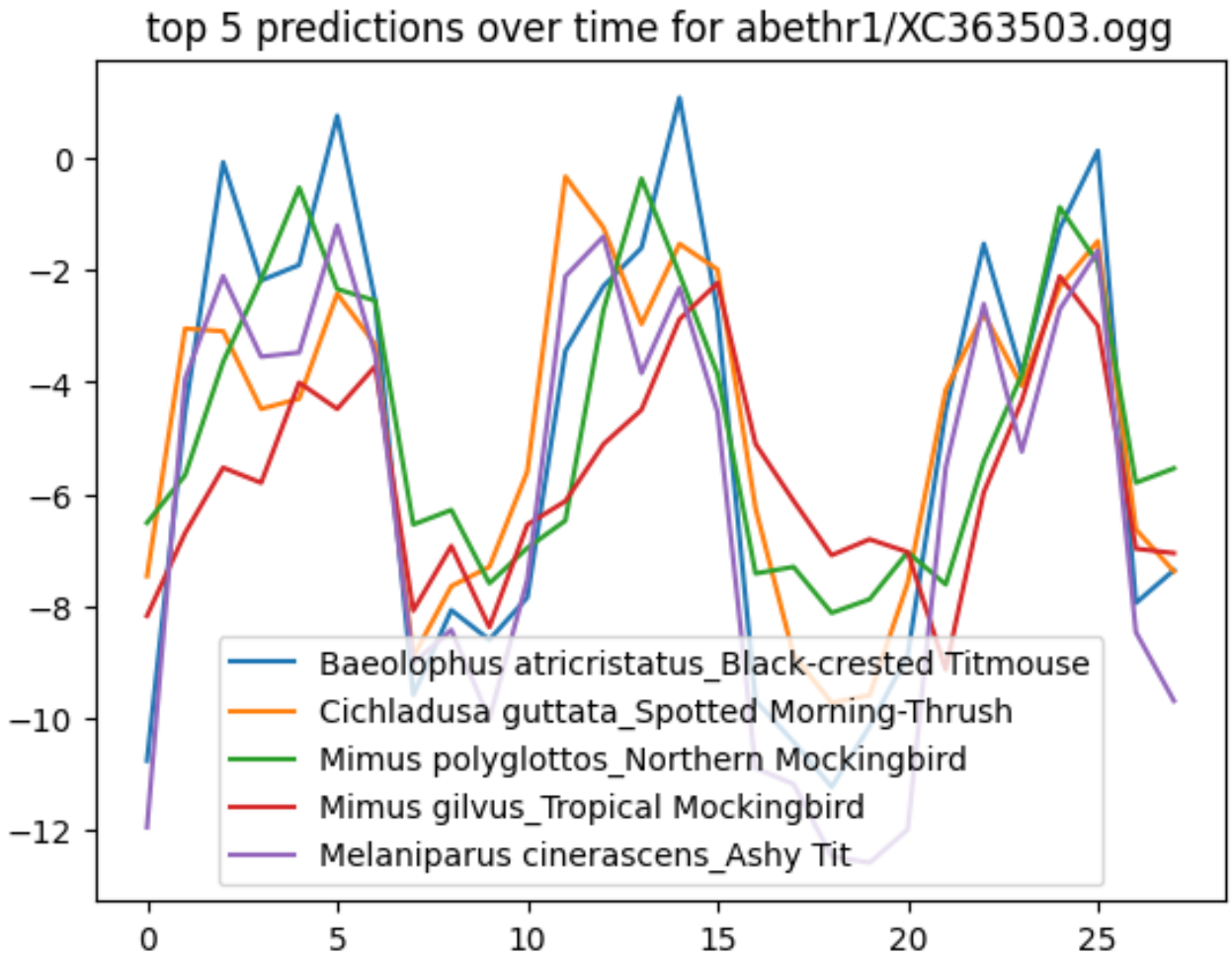
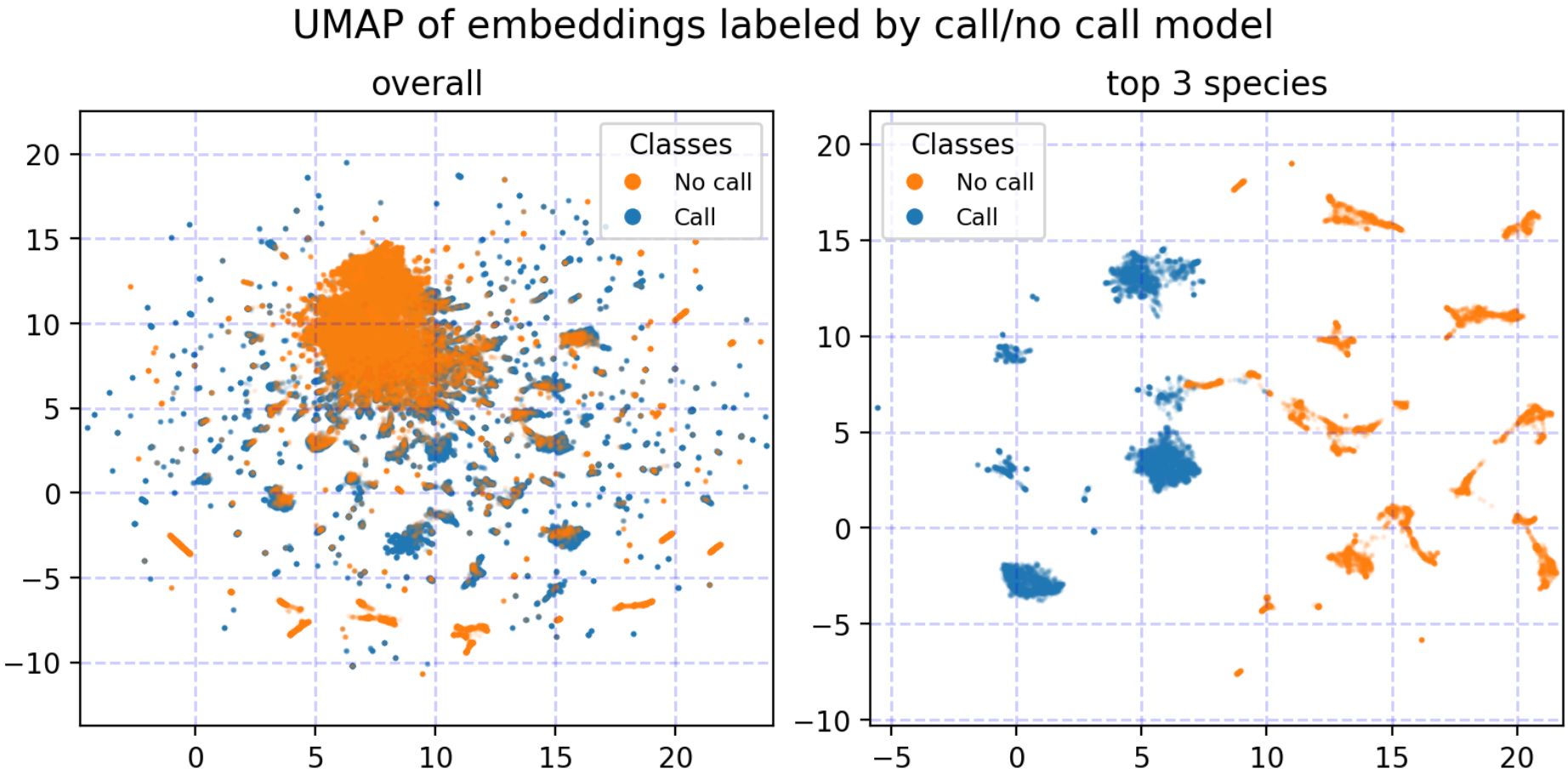


Figure 9: The BirdNET predictions can help with data annotation.

Supervised learning: classification



Engineering challenges

Outline

- Pre-processing audio data
- Time-resolution mismatches
- Classification implementations and performance

Preprocessing audio data: Luigi pipelines

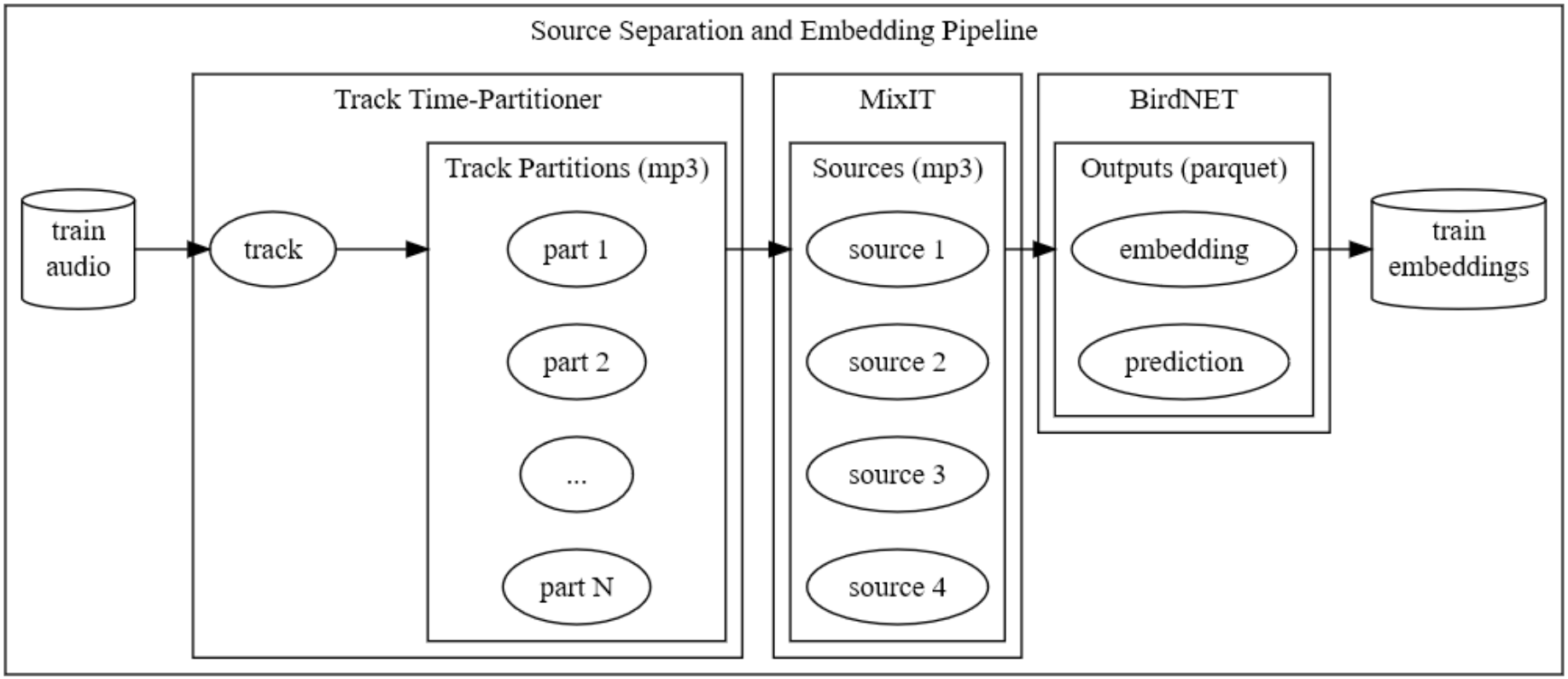


Figure 10: We use Luigi to coordinate a processing pipeline spanning days on an n2-standard-16 compute instance. We prevent processing skew across workers by recursively training audio. The audio is then source separated and embedded, resulting in a parquet file per audio chunk. We consolidate the parquet files into the final dataset.

Time-resolution mismatches

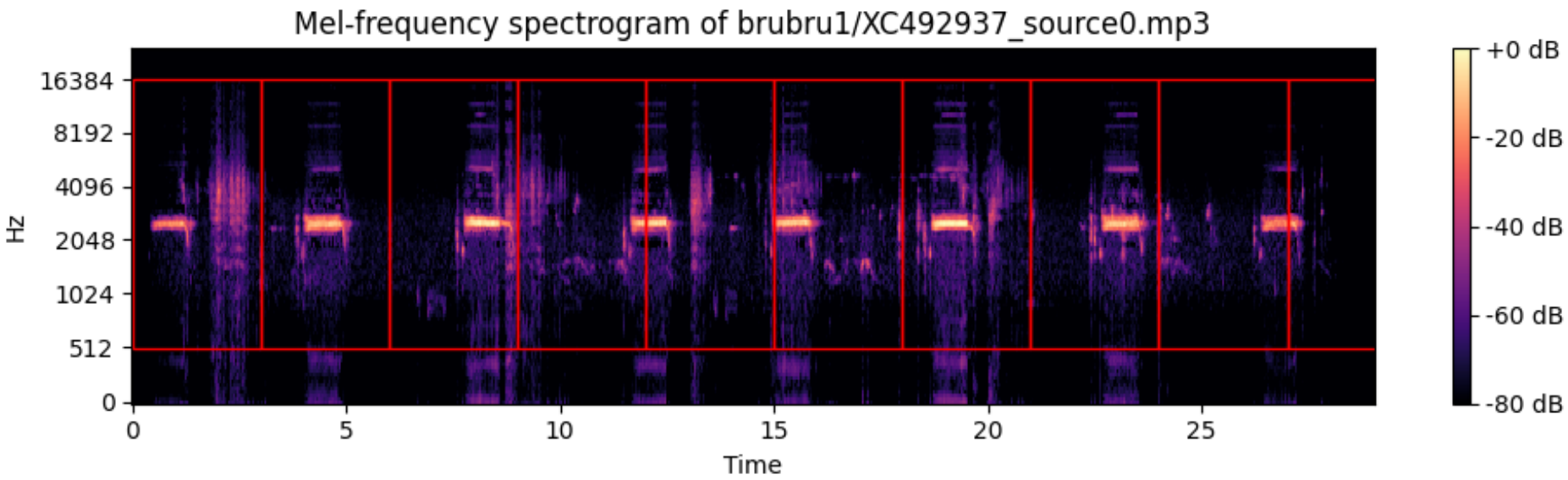


Figure 11: Chunked spectrogram of a bird call.

Classification implementations and performance

Table 1: A comparison between fit and predict the time for various models fit on a GCP n1-standard-4 compute instance with a Telsa T4 GPU. We fit the post-v7 dataset, which has 255,372 rows.

Model	GPU	Fit time	Predict time
Logistic Regression	No	59 min 17 s	1.5 s
SVC	No	90 min +	-
MLP	No	4 min 14 s	3.5 s
XGBoost (hist)	No	48 min 20 s	14.4 s
XGBoost (gpu_hist)	Yes	5 min	15.3 s
ComplementNB	No	4.2 s	1.5 s

Results

Outline

- Overall performance
- Experiments with psuedo-labeling and augmentation

Overall performance

Table 2: A summary of few models. Logistic regression is our simplest model. XGBoost is trained on a multi-label dataset.

Model	**Public Score**	**Private Score**
Logistic Regression	0.78541	0.68369
MLP	0.74014	0.62283
XGBoost	0.79068	0.68181

Experiments

Psuedo-labeling

- Use the primary class of the track as the label, when the confidence is above a threshold
- Multi-label classification

Data Augmentation

Embedding augmentation via concatenation and averaging.

$$\hat{y} \sim M_1(v_t) \tag{1}$$

$$\hat{y} \sim M_2(v_t \oplus v_{t+1}) \tag{2}$$

$$\hat{y} \sim M_3(v_t \oplus \sum_{i=0}^{n} v_i) \tag{3}$$

$$\hat{y} \sim M_4(v_t \oplus v_{t+1} \oplus \sum_{i=0}^{n} v_i) \tag{4}$$

Next time

- More variations of semi-supervision
  - A more rigorous approach to evaluation
- Comparisons of embedding models
  - Meta AudioGen, OpenAI Whisper, Mozilla DeepSpeech
- Embedding dynamics
  - Linear dynamics with forcing function, found via SVD?
  - Motif mining of the forcing function
  - Scale of the problem is best solved by Spark
- Sequence models
  - Learn a sequence model that best predicts the optimal set of classes
  - All data can be fed in one model start to end
    - Would rather this be in Torch, rather than Tensorflow, which leans toward the direction of AudioGen

And for the DS@GT Folks...

Figure 12: Bird conservation is a worthy cause and a great opportunity to learn.

Advice for myself two years ago

Building a team is worthwhile

- A strong team can help you achieve more than you could on your own. It's also an opportunity to connect with other students.

Be prepared to learn how to lead a team

- Effective communication and clear timelines are key to keeping the team on track
- Remember that everyone on the team is capable and valuable, and make an effort to recognize and appreciate their contributions

Reach out to OMSCS and OMSA early

- Working professionals have *a lot* to bring to the table.

Recruitment: Outreach

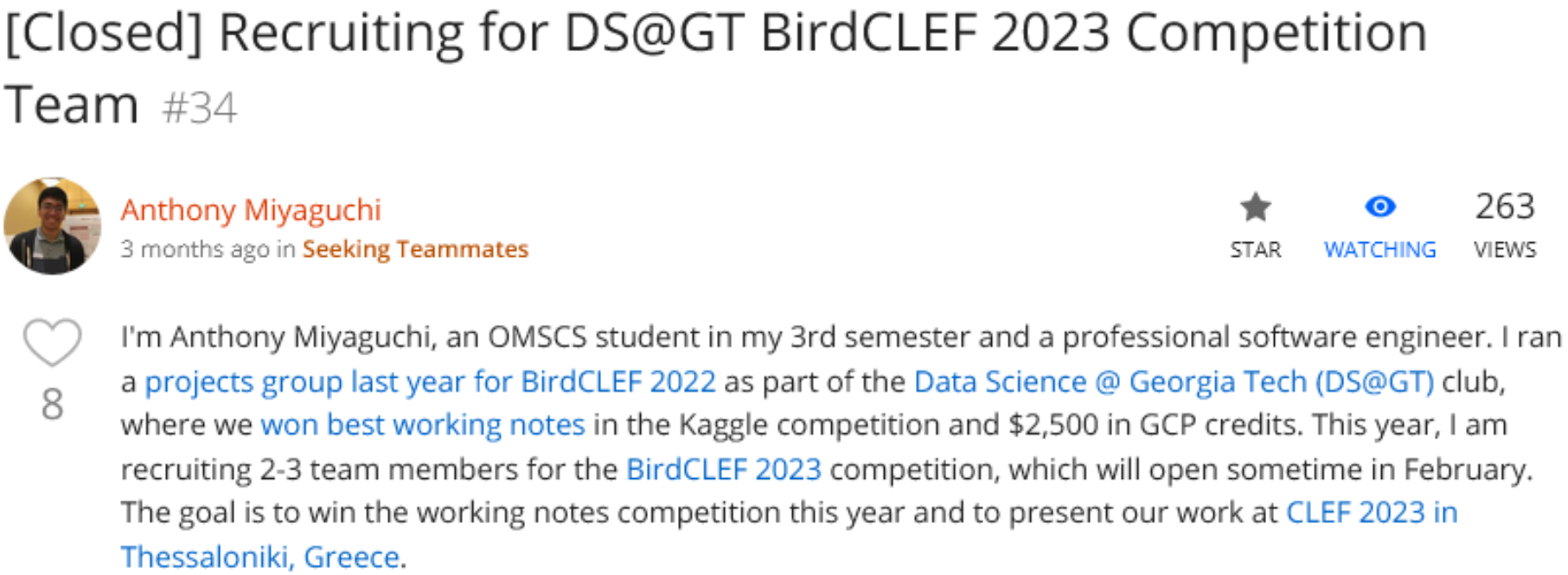


Figure 13: A post on the OMSCS Research EdStem board.

Be on the lookout for opportunities

Figure 14: Be on the lookout!

There's an abundance of opportunities for OMSCS students to collaborate with other students.

Thank you to everyone involved

DS@GT Leadership

- Pulak Agarwal
- Krishi Manek

BirdCLEF 2022

- Jiangyue Yu
- Bryan Cheungvivatpant
- Dakota Dudley
- Aniketh Swain

BirdCLEF F22 EDA

- Jinsong Zhen
- Kien Tran
- Siying Liu
- Muskaan Gupta
- Xinjin Li

BirdCLEF 2023

- Chris Hayduk
- Erin Middlemas
- Grant Williams
- Nathan Zhong
- Murilo Gustineli

Links and Resources

- Working Notes, "Motif Mining and Unsupervised Representation Learning for BirdCLEF 2022"
- DS@GT, Kaggle Competition Team Proposal, BirdCLEF 2022
- DS@GT, Project Group Proposal, BirdCLEF EDA Fall 2022
- DS@GT, Kaggle Competition Team Proposal, BirdCLEF 2023
- DS@GT, Assessment, BirdCLEF EDA Fall 2022
- DS@GT, Assessment, BirdCLEF 2023
- BirdCLEF Motif Viewer, Barn Owl, XC138041
- BirdCLEF 2023 MixIT Exploration, Red-chested Cuckoo, 2FXC207767

Thank you!

Time for Questions and Answers

Figure 15: Q&A