Polyphonic Avian Vocalizations Classification using Transfer Learning

1st Enrique Nueve
NAISE
enrique.iv@northwestern.edu

2nd Mia Wang
Purdue University
wang4070@purdue.edu

3rd Nicola J. Ferrier
Argonne National Laboratory
nferrier@anl.gov

Abstract—In the application of audio classification, audio signals have various noises overlapping the signal of interest. The signal may also have more than one sound of interest embedded. Thus, classification models of a single sound are, at times, unrealistic for application. Thus, polyphonic classification models can be used to determine if multiple sounds of interest are present within a given audio source. In this work, we propose a new polyphonic variant of the BirdCLEF 2021 data set. The mixing of avian vocalizations is dependent on regional relationships. We propose a novel architecture to create mixed audio embeddings that can be used as initial features to perform transfer learning on polyphonic classification methods in general. We show that through experimentation, our proposed mixed audio embedding method enhances the performance on the task of multi-classification on our proposed polyphonic variant of BirdCLEF 2021.

Index Terms—BirdCLEF 2021

I. To Do

A. Rick

1) Make mixed dataset
2) Train mixed embedder
3) Write up in Experiment and Results about how A Framework for the Robust Evaluation of Sound Event Detection (https://arxiv.org/abs/1910.08440) works and set up code examples with (https://github.com/audioanalytic/psds_eval)
4) Write lit review
5) Pick a conference and get that info
6) Learn everything on here (https://www.birds.cornell.edu/home/)

B. Mia

1) one

II. Introduction

Why do we need to classify avian vocalizations, why does it need to be polyphonic, what is the BirdCLEF 2021 challenge and dataset, we perform this task on the edge?

III. Literature Review

[1] proposed a way to address BirdCLEF 2021. Pick literature from the reference section of [1] on other papers about BirdCLEF 2021. Refer to work from https://www.birds.cornell.edu/home/ about why bird call classifcation matters. Refer to work from A Framework for the Robust Evaluation of Sound Event Detection (https://arxiv.org/abs/1910.08440) why polyphonic classification matters.

IV. Data

Our polyphonic variant of BirdCLEF 2021 was created using the tool Scraper [2].

V. Methodology

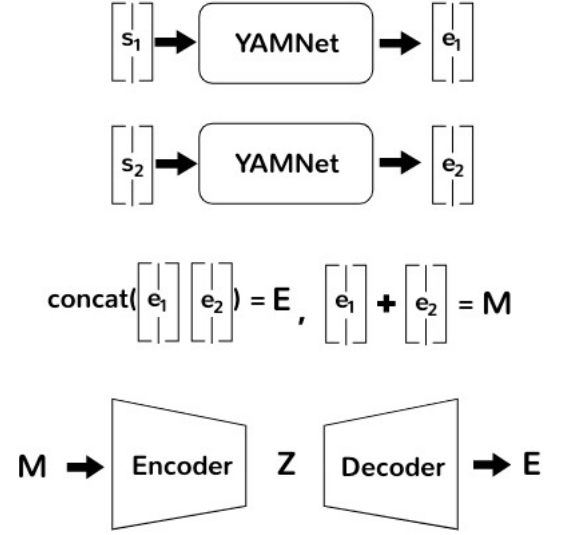### A. Polyphonic Transfer Learning Embedding

Fig. 1. Polyphonic Transfer Learning Embedding

VI. Experiment and Results

To score our methodology on our proposed data set, we used the proposed framework by [3]. This consist of the following metrics: metricA, metricB, etc.

A. Summarize the work of [3]

VII. Conclusion

REFERENCES

[1] M. V. Conde, K. Shubham, P. Agnihotri, N. D. Movva, and S. Bessenyei, "Weakly-supervised classification and detection of bird sounds in the wild. a birdclef 2021 solution," 2021.

[2] J. Salamon, D. MacConnell, M. Cartwright, P. Li, and J. P. Bello, "Scaper: A library for soundscape synthesis and augmentation," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2017, pp. 344–348.

[3] Ç. Bilen, G. Ferroni, F. Tuveri, J. Azcarreta, and S. Krstulović, "A framework for the robust evaluation of sound event detection," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 61–65.

[4] P. Beckman, C. Catlett, I. Altintas, E. Kelly, and S. Collis, "Mid-scale RI-1: Sage: A software-defined sensor network (NSF OAC 1935984)," 2019, https://sagecontinuum.org/.