Reducing the Barriers of Acquiring Ground-truth from Biodiversity Rich Audio Datasets Using Intelligent Sampling Techniques

Jacob Ayers
Electrical and Computer Engineering
UC San Diego
La Jolla, CA 926093
jgayers@ucsd.edu

Sean Perry
Mathematics
UC San Diego
La Jolla, CA 926093
sperry@ucsd.edu

Vaibhav Tiwari
Computer Science and Engineering
UC San Diego
La Jolla, CA 926093
vktiwari@ucsd.edu

Mugen Blue
Computer Science and Software Engineering
Cal Poly San Luis Obispo
San Luis Obispo, CA 93407
mugen4college@gmail.com

Nishant Balaji
Electrical and Computer Engineering
UC San Diego
La Jolla, CA 926093
nibalaji@ucsd.edu

Curt Schurgers
Electrical and Computer Engineering
UC San Diego
La Jolla, CA 926093
cschurgers@eng.ucsd.edu

Ryan Kastner
Computer Science and Engineering
UC San Diego
La Jolla, CA 926093
kastner@eng.ucsd.edu

Mathias Tobler
San Diego Zoo Wildlife Alliance
Beckman Center for Conservation Research
Escondido, CA 92027
MTobler@sdzwa.org

Ian Ingram
San Diego Zoo Wildlife Alliance
Beckman Center for Conservation Research
Escondido, CA 92027
iingram@sdzwa.org

Abstract

The potential of passive acoustic monitoring (PAM) as a method to reveal the consequences of climate change on the biodiversity that make up natural soundscapes can be undermined by the discrepancy between the low barrier of entry to acquire large field audio datasets and the higher barrier of acquiring reliable species level training, validation, and test subsets from the field audio. These subsets from a deployment are often required to verify any machine learning models used to assist researchers in understanding the local biodiversity. Especially as many models convey promising results from various sources that may not translate to the collected field audio. Labeling such datasets is a resource intensive process due to the lack of experts capable of identifying bioacoustics at a species level as well as the overwhelming size of many PAM audiosets. To address this challenge, we

Tackling Climate Change with Machine Learning workshop at NeurIPS 2021

have tested different sampling techniques on an audio dataset collected over a two-week long August audio array deployment on the Scripps Coastal Reserve (SCR) Biodiversity located adjacent to sandstone cliffs and the Pacific Ocean in La Jolla, California. These sampling techniques involve creating four subsets using stratified random sampling, limiting samples to the daily bird vocalization peaks, and using a hybrid convolutional neural network (CNN) and recurrent neural network (RNN) trained for bird presence/absence audio classification. We found that a stratified random sample baseline only achieved a bird presence rate of 44% in contrast with a sample that randomly selected clips with high hybrid CNN-RNN predictions that were collected during bird activity peaks at dawn and dusk yielding a bird presence rate of 95%. The significantly higher bird presence rate demonstrates how intelligent, machine learning-assisted selection of audio data can significantly reduce the amount of time that domain experts listen to audio without vocalizations of interest while building a ground truth for machine learning models.

1 Introduction

Passive acoustic monitoring (PAM) is a method of garnering an understanding of various ecosystems that involves deploying a large amount of audio recorders that autonomously collect audio clips from natural soundscapes over time. [11] Combining PAM with machine learning techniques has created a niche to better understand the impacts of climate change on many noisy indicator species that are too small for large scale monitoring via traditional biodiversity surveying techniques such as trapping, monitoring feeding sites, and camera trap arrays [13, 9, 15, 16, 7, 2, 12, 14].

Due to the wider availability of low-cost hardware fit for PAM deployments [4] and open source pre-trained machine learning models [10, 5, 3] the barrier of entry for researchers breaking into the ecoacoustics [6] discipline is decreasing. However, it is very easy to be led on by promising results of pre-trained models on publicly available datasets that may not translate well when applied to noisier field audio recordings. [1] A crucial step towards understanding the biodiversity of an ecosystem using PAM and machine learning requires generating species level ground truth labels on a subset of audio recordings from the field for the purpose of testing and if necessary, validation and training of promising models. Creating ground truth requires a lot of time from the limited pool of experts capable of labeling audio data at a species level. For the sake of reducing the financial and temporal costs of the research process, it is in the best interest of researchers to develop methods that make sure that the audio being delivered to experts for labeling, have a higher probability of containing vocalizations of interest.

To meet this challenge, we have explored various methods to increase the probability of extracting bird vocalizations from a PAM field deployment we conducted on the California coast. These methods involve a baseline stratified random sampling technique, sampling with knowledge of diurnal bird vocalization trends, as well as using a neural network model designed for audio event detection with low resource training sets [8] that has been encapsulated in the Github repository Microfaune with a pre-trained model for binary bird classification. Microfaune can be decomposed into a convolutional neural network (CNN) layer that computes features of audio that have been converted into a mel spectrogram, a recurrent neural network (RNN) layer that computes features at each time step based on the neighboring time steps, and a final max-pooling layer that finds the highest prediction across an audio clip. The max-pooling layer for our purposes can be simplified down to Microfaune's prediction to the question: "what is the probability that at least one bird vocalization occurs in this clip according to this neural network".

These sampling methods were used to generate several subsets from our PAM deployment that were labeled for bird presence/absence and high/low activity to compare and contrast how effective the methods were in identifying audio recordings at the class level of the taxonomic tree to reduce the overhead costs of having experts label at the species level.

Figure 1: Coastal reserve AudioMoth (red dots) deployment region

Figure 2: AudioMoth in official housing on Lemonade berry bush

Figure 3: AudioMoth in Ziploc bag on California sagebrush

2 Methodology

2.1 Field data collection

We deployed 10 AudioMoths (version 1.2.0) on the Scripps Coastal Reserve (SCR) Biodiversity Trail, a private nature reserve in La Jolla, California (see figure 1). The reserve is managed by the UC San Diego Natural Reserve System and is home to over 150 bird species including the threatened California Gnatcatcher (*Polioptila californica*) according to the U.S. Fish and Wildlife Service. The devices were housed in either official Audiomoth cases or Ziploc bags and were attached to coastal sage scrubs such as Lemonade berry (*Rhus integrifolia*) bushes (see figure 2) and California sagebrush (see figure 3) (*Artemisia californica*) at a height of 30 to 150 cm from the ground across the soft chaparral environment.

The Audiomoths were set to record one minute every ten minutes at a 384 kilohertz sampling rate from August 10th to August 24th, 2021.

2.2 Subset creation

In order to test out different methods for extracting audio clips with bird vocalizations from the SCR dataset, we constructed four separate datasets of 240 audio clips. We constructed a baseline stratified random sample subset by selecting one audio clip from every hour of the day from each Audiomoth device. To test out the efficacy of neural network assisted sampling, we used the same stratified random sampling technique as the baseline, with the added parameter that each of the 240 clips sampled had a Microfaune prediction of 50% or more (see figure 5). The audio clips had to be downsampled from 384 to 44.1 kilohertz prior to being processed by the Microfaune prediction pipeline. The third dataset involved us randomly sampling 240 clips that were recorded at dawn and

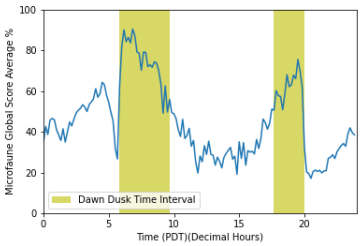

Figure 4: Scripps coastal reserve August bird vocalizations trends

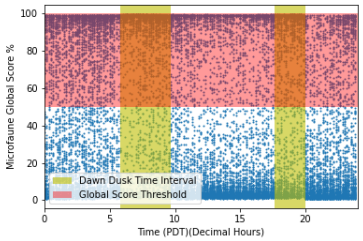

Figure 5: Intersection of Microfaune skew and dawn-dusk peak activity across all clips (blue dots)

Table 1: Bird presence (left) bird activity (right)

	Not Skewed	Skewed
Not Dawn-Dusk	.4375	.6583
Dawn-Dusk	.875	.95

	Not Skewed	Skewed
Not Dawn-Dusk	.2458	.3708
Dawn-Dusk	.6167	.7583

dusk intervals known to exhibit high bird activity (see figure 4). We averaged together the Microfaune predictions across each ten minute interval in the day to assist us in defining the dawn and dusk intervals.

The fourth and final dataset combined the techniques used in the second and third sets by randomly sampling clips within the dawn-dusk time intervals that were ranked as having a 50% or more chance of containing a bird vocalization by Microfaune (see figure 5).

2.3 Dataset labeling

To label the audio, the clips were uploaded to a web-based audio labeling system (Pyrenote). Volunteers familiar with the birds native to the SCR then labeled bird vocalizations directly on spectrograms of the audio. The annotations produced by the volunteers were post-processed down to whether or not they heard a bird vocalization (Bird Present/Absent) and whether or not there was more than one species in an audio clip (Heavy Activity/Low Activity) (see figure 6).

3 Results

At the end of the deployment, each device collected approximately 2000 audio clips amounting to about 336 hours of audio recorded across all of the devices. This means that each subset contains approximately 1% of the clips from the AudioMoth deployment. To compare and contrast the 4 subsets we divide the Bird Presence Count and the High Activity Count with respect to the Dataset Size (see table 1). That way we can see which sampling technique was the most effective at extracting bird vocalizations from the whole SCR audioset.

4 Conclusion

From the results, it appears that sampling from the dawn-dusk daily bird vocalization peaks was a larger factor in acquiring a higher rate of audio clips with bird presence and heavy bird activity compared to purely skewing the results based on Microfaune predictions.

We can see that combining the diurnal trends of bird vocalizations in our deployment region with binary neural network predictions in the process of sampling can assist in achieving a higher rate of audio clips with vocalizations of interest than each of the methods independently. Using these tools to focus on audio with a higher probability of relevant bioacoustics can greatly reduce the amount of time needed to acquire the necessary species focused training, validation, and testing sets that are required to confidently garner an understanding of how climate change impacts the biodiversity of a deployment region.

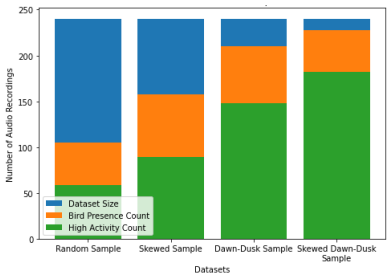

Figure 6: Labeling results across the four subsets

Broader Impact

Our hope is that the methods demonstrated in this paper can be used to help fellow researchers reduce the amount of time it takes to adequately process large audio datasets with critical biodiversity information. There is the possibility of our work not generalising to all audio datasets, and consequently working against the aforementioned goal. We chose the Scripps Coastal Reserve Biodiversity trail partially due to the fact that it is currently closed to the public, reducing the chance of privacy invasion. Our method of sampling from the dawn and dusk bird vocalization peaks reduces the chances of acquiring samples of nocturnal species such as the Common Poorwill (*Phalaenoptilus nuttallii*) native to the reserve. Furthermore, using a binary classifier such as Microfaune has the potential to skew results away from species of interest.

Additional Materials

Relevant Github Repositories

Microfaune Classifier: https://github.com/microfaune/microfaune

Pyrenote Audio Labeling System: https://github.com/UCSD-E4E/Pyrenote

Code for figures and data processing: https://github.com/UCSD-E4E/AID_NeurIPS_2021

Acknowledgements

This work was funded in part by the REU Site Engineers for Exploration, supported by the National Science Foundation (NSF) under grant number 1852403. Opportunities were extended to students on this project thanks in large part to the yearly Summer Research Internship Program (SRIP) run by the Electrical and Computer Engineering department of UC San Diego. Further opportunities were extended to students interested in this project thanks to the Summer URS Ledell Family Research Scholarship for Science and Engineering. Special thanks are in order for the Computer Science and Engineering (CSE) department of UC San Diego. We would like to thank Nathan Hui, the staff engineer of Engineers for Exploration for his help in guiding our team on our first field deployment. This work was performed in part at the University of California Natural Reserve System Scripps Coastal Reserve DOI: 10.21973/N3QH2R. We would like to thank the Reserve Manager, Isabelle Kay for her assistance and expertise that facilitated a smooth AudioMoth deployment. Sara Weitzel provided us with expertise in botany that guided us in describing the Biodiversity Trail's environment. We would like to thank Professor Sanjoy Dasgupta of UC San Diego's CSE department for providing us with many insights in the field of active learning that guided many of the sampling techniques used for this paper. This paper certainly could not have been made without Kathleen Dickey, Marty Hales, and Andy Rathbone of the Torrey Pines State Natural Reserve who volunteered their time and expertise by labeling hours of audio. Finally, we would also like to thank our friends and collaborators from the San Diego Zoo Wildlife Alliance that have contributed their expertise in ecology and conservation technologies to our team.

References

[1] Jacob G Ayers, Yaman Jandali, Yoo-Jin Hwang, Erika Joun, Gabriel Steinberg, Mathias Tobler, Ian Ingram, Ryan Kastner, and Curt Schurgers. Challenges in applying audio classification models to datasets containing crucial biodiversity information. 2021.

[2] Sérgio Henrique Borges. Bird assemblages in secondary forests developing after slash-and-burn agriculture in the brazilian amazon. *Journal of Tropical Ecology*, 23(4):469–477, 2007.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[4] Andrew P. Hill, Peter Prince, Evelyn Piña Covarrubias, C. Patrick Doncaster, Jake L. Snaddon, and Alex Rogers. Audiomoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, 9(5):1199–1211, 2018.

[5] Stefan Kahl, Connor M. Wood, Maximilian Eibl, and Holger Klinck. Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61:101236, 2021.

[6] Bernie Krause and Almo Farina. Using ecoacoustic methods to survey the impacts of climate change on biodiversity. *Biological conservation*, 195:245–254, 2016.

[7] Adria Lopez-Baucells, Ricardo Rocha, Paulo Bobrowiec, Enrico Bernard, Jorge Palmeirim, and Christoph Meyer. *Field Guide to Amazonian Bats*. 09 2016.

[8] Veronica Morfi and Dan Stowell. Deep learning for audio event detection and tagging on low-resource datasets. *Applied Sciences*, 8:1397, 08 2018.

[9] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25):E5716–E5725, 2018.

[10] Tessa Rhinehart, Barry E Moore, and Justin A Kitzes. Opensoundscape: Machine learning for scalable acoustic surveys. In *2019 ESA Annual Meeting (August 11–16)*. ESA, 2019.

[11] Len Thomas and Tiago A Marques. Passive acoustic monitoring for estimating animal density. *Acoustics Today*, 8(3):35–44, 2012.

[12] Mathias W. Tobler, Rony Garcia Anleu, Samia E. Carrillo-Percastegui, Gabriela Ponce Santizo, John Polisar, Alfonso Zuñiga Hartley, and Isaac Goldstein. Do responsibly managed logging concessions adequately protect jaguars and other large and medium-sized mammals? two case studies from guatemala and peru. *Biological Conservation*, 220:245 – 253, 2018.

[13] Michelle Ward, Ayesha IT Tulloch, James Q Radford, Brooke A Williams, April E Reside, Stewart L Macdonald, Helen J Mayfield, Martine Maron, Hugh P Possingham, Samantha J Vine, et al. Impact of 2019–2020 mega-fires on australian fauna habitat. *Nature Ecology & Evolution*, 4(10):1321–1326, 2020.

[14] Hartwell H. Welsh Jr. and Lisa M. Ollivier. Stream amphibians as indicators of ecosystem stress:a case study from california's redwoods. *Ecological Applications*, 8(4):1118–1132, 1998.

[15] Marco Willi, Ross T Pitman, Anabelle W Cardoso, Christina Locke, Alexandra Swanson, Amy Boyer, Marten Veldthuis, and Lucy Fortson. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 10(1):80–91, 2019.

[16] James E Woodford and Michael W Meyer. Impact of lakeshore development on green frog abundance. *Biological Conservation*, 110(2):277–284, 2003.