Acoustic Species Identification

CSE 145, Prof. Kastner, Project Specification

PROJECT CHARTER

I. Project Overview

Accurate classification of animal vocalizations is crucial to ecological research and conservation efforts. Passive audio data collected from distributed sensors has the potential to provide a massive trove of information about species distributions and behaviors, but such data is time-consuming for experts to manually label. [1] This motivates bioacoustic classification efforts, which aim to build systems that can automatically identify species from vocalizations in their natural habitat.

Our project focuses on classifying bird vocalizations and extends previous work from Engineers for Exploration (E4E). Specifically, we will explore architectures **to overcome a challenging domain shift between labeled training data and unlabeled data from noisy soundscapes**. Our deliverables will be measured by their performance on the BirdCLEF 2024 Kaggle competition. [2]

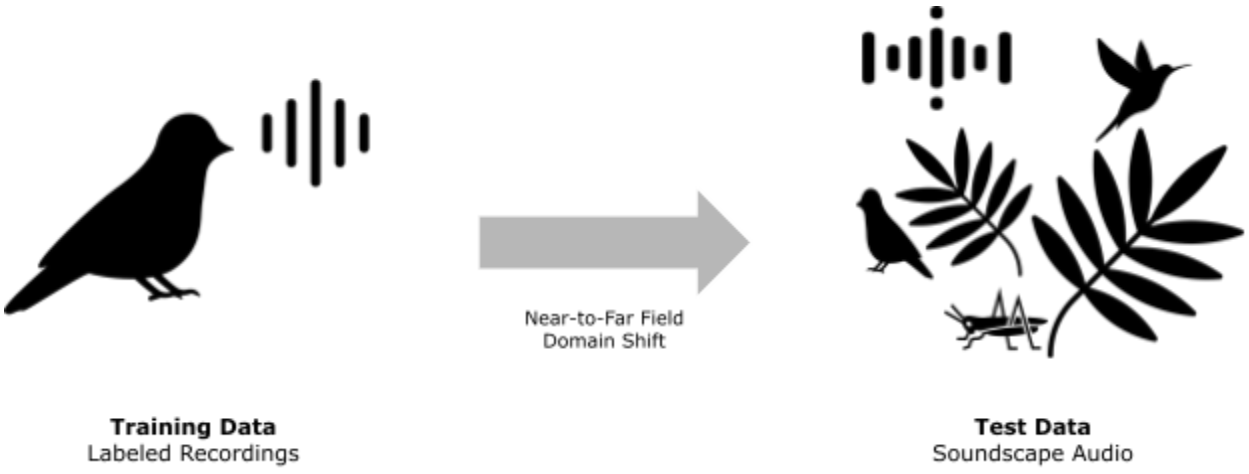


Figure 1: This project addresses the problem of *near-to-far field domain shift* in audio data for bird species classification.

II. Project Approach

We will explore two central directions to improve performance.

1. Modernizing classifier architectures.

Existing state-of-the-art approaches, as well as the existing PyHa pipeline, leverage convolutional neural networks (CNNs) on windows heuristically tuned spectrogram representations of data, treating recordings as images. While some modern works use vision transformers (ViTs) [3] and recurrent neural networks (RNNs), [4] these architectures have yet to gain traction in the field and are not available in PyHa.

Beyond standard EfficientNet baselines, we intend to explore the application of the Mamba selective-sequence-model architecture, which has shown promise across a number of audio-processing domains. Keeping the rest of our training pipeline stable, we hope to find improvement by training a more expressive model architecture.

2. Leveraging unlabeled soundscape data to overcome the train-test domain shift.

A central challenge to training avian classifiers is the strong domain shift between the labeled training distribution and unlabeled target distribution. Labeled data is often collected at zoos or in other relatively quiet environments, often with a single bird vocalizing at once; the most expansive collection of such recordings is the Xeno-Canto (XC) dataset. [5] However, the unlabeled soundscape recordings often have extensive background noise and may contain overlapping bird vocalizations, necessitating filtering and multi-way classification.

Existing approaches tend to train directly on Xeno-Canto data, and sometimes add synthetic noise to more closely approximate the soundscape distribution (see Section 2.3 of [6] for a modern survey of approaches).

One untapped approach to this problem is **task arithmetic**. [7] The finetuned weights of image classification and text generation models are found to compose together semantically when combined additively. More formally, finetuning a model with base weight $w_{pretrained}$ to obtain weights $w_{finetuned}$, we define the **task vector** as the difference $\tau = w_{finetuned} - w_{pretrained}$.

Adding two task vectors can create a multi-task model, and subtracting a task vector can reduce a learned behavior. Most relevant to our project is learning by analogy. For two tasks paired with two domains, one can construct the fourth model out of any of the other three (see below for a geometric intuition and [7] for more details). In our case, we can train both autoregressive and classification objectives in Xeno-Canto data, then train an autoregressive model in soundscape data to obtain the three task vectors necessary to approximate a classification model on soundscape data.

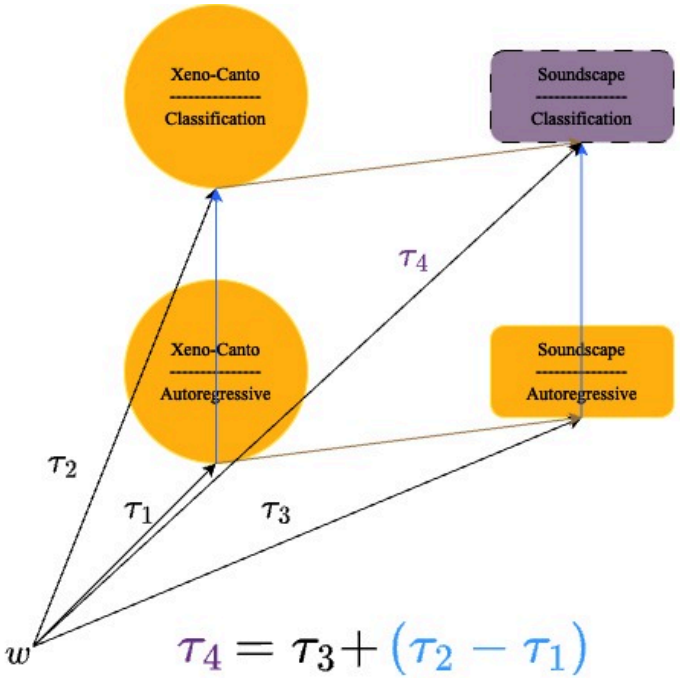


Figure 2: Task-vector approach to overcoming the Xeno-Canto to soundscape domain shift.

This idea is riskier than the others, as such behavior may not emerge at smaller scales, or on audio models that are not as saturated with data as language and image models are. As such, we hedge with more traditional development approaches on the architectural level. Still, if the task arithmetic works it would be novel and dispense with the need for heuristic noising functions to deal with the domain shift in these kinds of problems.

Minimum Viable Product

Our MVP is a machine learning model that can to a certain degree of accuracy detect and classify different bird species from bird vocalizations in 5 second segments of soundscapes. Specifically, we aim to develop a model trained on a variety of datasets including xeno-canto and then classify birds from soundscapes recorded in the Western Ghats of India. Our goal is to then submit the model to the yearly BirdCLEF competition hosted on Kaggle.

Constraints, Risks, Feasibility

One of the main risks of this project is that we might get too involved in the research and have no working MVP at the end. Hence, we decided to also in the meanwhile develop another model based on similar architectures and approaches as documented by previous BirdCLEF winners. We will also take into account any advice and recommendations from Sean, Sam, Jacob and the San Diego Zoo Collaborators to ensure we stay on the right track. Additionally, the submission deadline for the Kaggle competition

perfectly aligns with the quarter and will happen on June 10, 2024 which will be at the start of finals week. As pointed out earlier, our main problem will be the domain shift problem which we hope to overcome by using new approaches and conducting several experiments.

III. Group Management

Major Roles

There are three major roles: (a) project lead/E4E member, (b) researcher, and (c) engineer.

a. Project Lead/E4E Member: responsible for ensuring that the deliverables of this course project is relevant to stakeholders in the broader E4E project and the San Diego Zoo.
b. Researcher: understands literature, designs approaches, conducts experiments, reports findings.
c. Engineer: performs data engineering, implements evaluation pipeline.

The project lead is Ludwig von Schoenfeldt. Everyone will take on both roles as researcher and engineer as the needs of the project evolves throughout the quarter.

Decision Mechanism

Decisions should be made by consensus, unanimously if possible. Should there be serious disagreements, we can re-evaluate what specific decision-making mechanism is required to address the issue.

Communication

We communicate through the class discord, especially for sharing documents and scheduling meetings. We also have a weekly in-person work meeting on Fridays, 4PM.

Time Management

Our weekly work meetings will help keep us accountable to each other. Furthermore, there is a weekly meeting with partners at the San Diego Zoo. These will give us biweekly markers so that we can continually re-evaluate our progress and determine if adjustments need to be made.

Responsibilities

Our team consists of Ludwig von Schoenfeldt, Sean O'Brien, Geelon So, and Vibhuti Rajpurohit. We detail each individual's responsibilities below in the Project Milestones section. These assignments aim to leverage each members' strengths: Ludwig has ongoing work and domain knowledge in this project. Sean has expertise in deep learning. Geelon works in machine learning theory. Vibhuti is experienced in programming in Python and using its related libraries.

IV. Project Development

Development Roles

Our project can mostly be divided into 3 separate categories: Research, Development and Evaluation. During the initial stage of this project we will be mostly conducting research into several model architectures and approaches to the problem with a special focus on newer non computer vision approaches. While this research component will most likely stretch until week 6 we will develop a more safe computer vision based approach from previous years falling into the Development category. Towards the second half of the project we will then focus on evaluation of the newer models comparing it to already existing computer vision based approaches to create a fully functional architecture which we will then submit on Kaggle for BirdCLEF 2024.

Role Assignments

Ludwig: Lead, Research, Evaluation, Development
Geelon: Research, Development, Evaluation
Sean: Research, Development, Evaluation
Vibhuti: Development, Evaluation

Supplementary Materials Required

In terms of software, we will be using PyHa, a multi-class training pipeline – all of which are already setup for us. We will also be utilizing google cloud instances already set up for us for model training.

Testing

Our testing environment will be similar to the ones we use for training purposes. Additionally, we would be using padded cmAP from scikit- learn. The basic metrics would be precision and recall also from the scikit-learn.

Documentation

As in the previous year we will be using the background documentation and information that is provided in the official E4E_Passive_Acoustic_Biodiversity_Modelling google drive. We also will create several documents, outlines, videos and presentations for this project, which will be uploaded to the CSE 145/237D subfolder on the E4E shared drive. Any additional research papers that we will be utilizing for newer approaches will also be uploaded to the shared E4E folder.

V. Project Milestones and Schedule

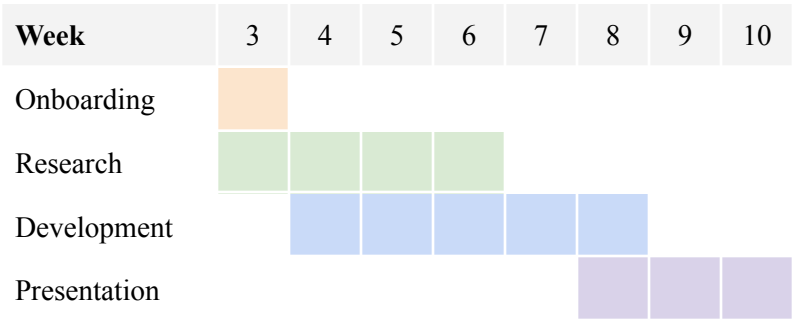
This research project explores the *near-to-far-field* domain shift problem in acoustics classification. We concretely evaluate our findings by participating in the BirdCLEF 2024 Kaggle competition, which is centered precisely around this domain shift question.

Our **final deliverables** are threefold:

1. Deployed model: the final model for the Kaggle submission.
2. Working paper: the accompanying paper for the Kaggle submission.
3. Project repository: this contains all models, experiments/pipeline, and notes required for reproducibility and downstream research beyond the Kaggle competition.

Within the context of the broader Acoustics Species Identification project, the deployed model and working paper are short-term contributions—they consolidate our findings and serve as a proof of concept on the specific competition data. In contrast, the repository is a long-term contribution meant to facilitate further research and the ongoing project with the San Diego Zoo.

We break down the project into smaller, more actionable **milestones**, which we broadly categorize into four phases: onboarding, research, development, and presentation.



Milestone #1: Build a model that does not work	
Aim	Set up cloud compute infrastructure. Understand Kaggle competition. Develop basic technical knowledge for sound data.
Requirements	1. Obtain access to the Google Cloud Platform. 2. Create/join project repository. 3. Individually build a model that could be submitted to Kaggle (correct "type signature"; can perform terribly).
Deadline	Week 3
Assignees	Everyone

Milestone #2: Write very short research proposals/design doc	
Aim	Gain familiarity with existing literature. Enable meaningful research discussions. Provide technical specifications.
Requirements	1. Perform literature review on approaches to the domain shift problem (e.g. foundation models, representation learning, models/architectures). 2. (Individually) write up short proposals/design docs, and discuss with the team to refine. 3. The design docs should be sufficiently detailed to answer relatively low-level implementation questions.
Deadline	Week 4
Assignees	Ludwig, Sean, Geelon

Milestone #3: Establish baseline by ignoring domain shift	
Aim	Leverage prior year's work. Build a very minimal MVP. Understand evaluation metrics.
Requirements	1. Set up experimental pipeline for evaluation. 2. Train/submit existing model using Kaggle data. 3. Evaluate baseline model using Kaggle's metrics.
Deadline	Week 4
Assignees	Vibhuti, Ludwig

Milestone #4: Experiment and compare with baseline	
Aim	Efficiently iterate on methodology for domain shift.
Requirements	1. Implement proposed approaches (see Milestone #2) 2. Evaluate approaches (see Milestone #3) 3. Record results; share findings in Oral Project Update (Week 6, 5/7) and Milestone Report (Week 6, 5/12) 4. (Fail fast and repeat Milestones #2 and #4)
Deadline	Week 6
Assignees	Everyone

Milestone #5: Specify and train Kaggle model	
Aim	Construct specialized model for deployment.
Requirements	1. Write design doc for deployment model (if new). 2. Implement and carefully train model. (Research phase: fail fast for more exploration. Development phase: spend more time exploiting).
Deadline	Week 8
Assignees	Everyone

Milestone #6: Submit model and working paper to Kaggle	
Aim	Consolidate findings. Evaluate approach in broader pool of competitors.
Requirements	1. Complete BirdCLEF submission (Week 10, 6/10) and working paper (also for the course Final Report, Finals week, 6/13).
Deadline	Week 10
Assignees	Everyone

Milestone #7: Document project for E4E Acoustics project	
Aim	Ensure longevity/reproducibility of contributions.
Requirements	1. Write documentation for approaches. This extends the original proposals, includes evaluation, and some discussion/lessons learned. 2. Complete requirements for Final Oral Presentation (Week 9, 5/30), Project Web Presence (Week 10, 6/4), and Final Project Video (Finals week, 6/10)
Deadline	Week 10
Assignees	Everyone

Citations

[1] Julia Shonfield and Erin M. Bayne. Autonomous recording units in avian ecological research: Current use and future applications. Avian Conservation and Ecology, 12(1), 2017. doi: 10.5751/ACE-00974-120114.

[2] HCL-Rantig, Holger Klinck, Maggie, Sohier Dane, Stefan Kahl, Tom Denton, Vijay Ramesh. (2024). BirdCLEF 2024. Kaggle. https://kaggle.com/competitions/birdclef-2024

[3] Dosovitskiy, Alexey et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” ArXiv abs/2010.11929 (2020): n. pag.

[4] Graves, Alex. “Connectionist Temporal Classification.” (2012).

[5] Vellinga W (2024). Xeno-canto - Bird sounds from around the world. Xeno-canto Foundation for Nature Sounds. Occurrence dataset https://doi.org/10.15468/qv0ksn accessed via GBIF.org on 2024-04-19.

[6] Rauch, Lukas et al. BirdSet: A Multi-Task Benchmark For Classification In Computational Avian Bioacoustics. arXiv, April 8 2024. https://arxiv.org/abs/2403.10380

[7] Ilharco, Gabriel et al. “Editing Models with Task Arithmetic.” ArXiv abs/2212.04089 (2022): n. pag.