Acoustic fingerprints in nature: A pilot test for monitoring biodiversity at the edge

SAMIR RAJANI & DARIO DEMATTIES
Northwestern-Argonne Institute of Science and Engineering

Northwestern University
Argonne National Laboratory
Institute of Science and Engineering

Argonne NATIONAL LABORATORY

IPDPS 2023 • St. Petersburg, Florida USA

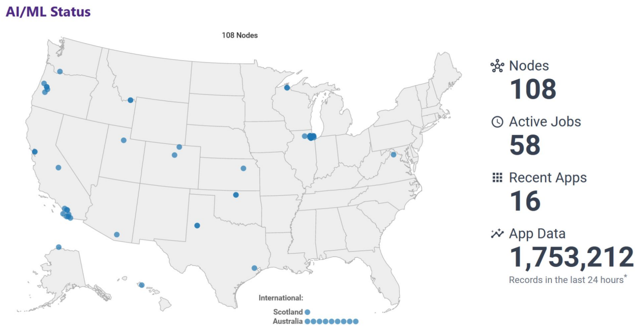
U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

INTRODUCTION

Why and How to Monitor Ecosystems

- Ecosystems play a key role in regulating the Earth's climate, nutrient cycling, and other critical processes that support life on our planet.
- Recently, our group developed *Sage: A Software-Defined Sensor Network*, deploying a national-scale cyberinfrastructure for linking advanced sensors with AI-enabled computation at the edge.
- We propose to use Sage as a starting point for advancing biodiversity characterization by means of continuous ecosystem monitoring.

SAGE
A Software-Defined Sensor Network
Cyberinfrastructure for AI at the Edge
www.sagecontinuum.org



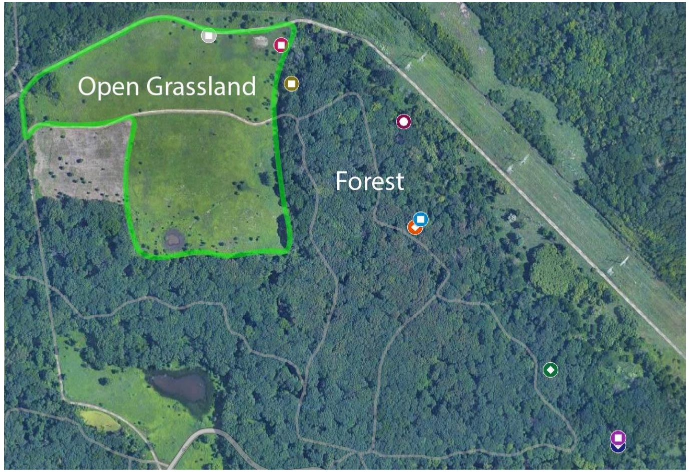
DATA COLLECTION

DATA COLLECTION

The Morton Arboretum

- Data collected from May 2021 - August 2021
- Waveforms sampled at 22,050 Hz
- Continuous (24 hrs. / day) recording

The Morton Arboretum, Lisle, IL
Image from mortonarb.org



Locations of Recording Nodes: Open Grassland and Forest

DATA COLLECTION

Recording Devices



Canopy Recorder (x7)

~2.1 TB of data
151 days of recordings



Open Grassland Recorder (x2)

~210 GB of data
29 days of recordings

DATA COLLECTION

Preprocessing

- Applied the following transformations to our data:
  - Mix down to a single channel
  - Split into nine-second clips
  - Apply Mel spectrogram

BACKGROUND: SELF-SUPERVISED LEARNING

SELF-SUPERVISED LEARNING

In Contrast with Supervised Learning

Supervised Learning

$$\theta_{t+1} = \theta_t - \eta \sum_i \nabla \ell(y_i, \hat{y}_i)$$

η (learning rate)

ℓ (loss function)

y_i (predicted label)

$\hat{y}_i$ (ground-truth label)

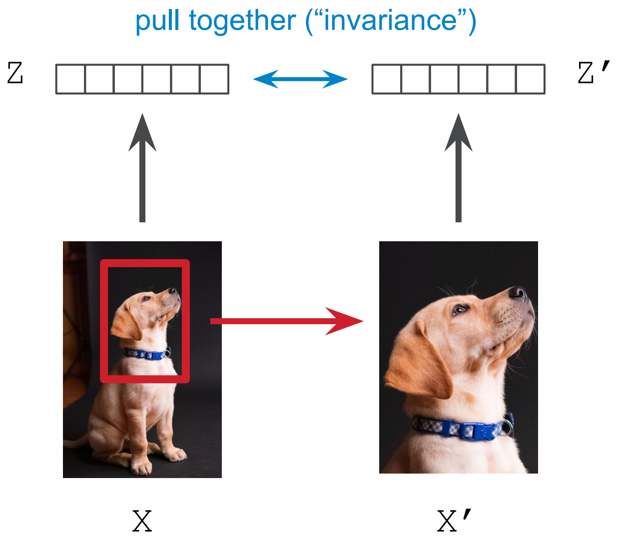
Self-Supervised Learning

- No ground-truth label
- Must devise some clever objective to facilitate feature extraction

SELF-SUPERVISED LEARNING

Augmentations

- A common theme in self-supervised learning: augment your data, and then devise a task making use of these augmentations.
- For us, these augmentations look like:
  - Random resized crops
  - Random horizontal flips
  - Gaussian blur



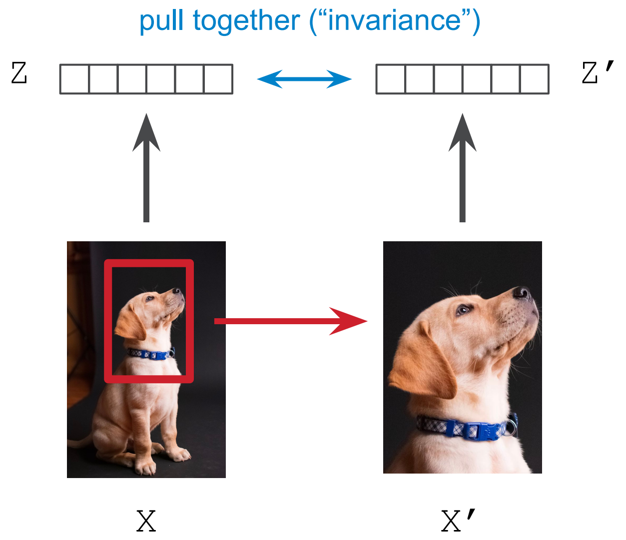
SELF-SUPERVISED LEARNING

Augmentations

- But… what if our loss function included only an invariance term?

$$\ell(Z, Z') = \frac{1}{n}\sum_i \|z_i - {z_i}'\|^2$$

- We now have a trivial solution: a constant mapping.
- This problem is known in the literature as **collapse**.

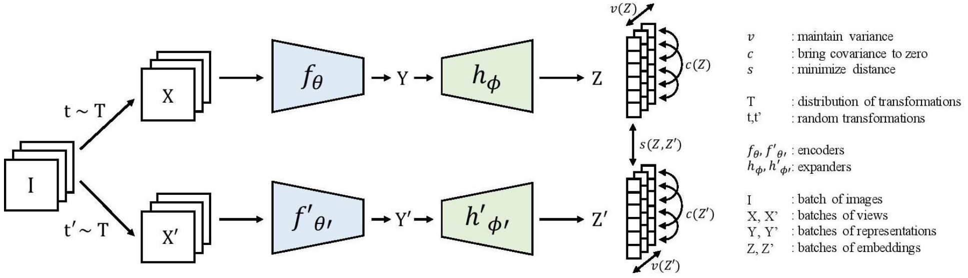


SELF-SUPERVISED LEARNING

VICReg (Bardes et al. 2021)

- Invariance: augmented views of the same image should produce the same embedding
- Variance: embedding vectors for different samples should be different
- Covariance: decorrelate features in latent representations

$$\ell(Z, Z') = \lambda s(Z, Z') + \mu[v(Z) + v(Z')] + \nu[c(Z) + c(Z')]$$



VICReg Architecture

Image from Bardes et al. (2021)

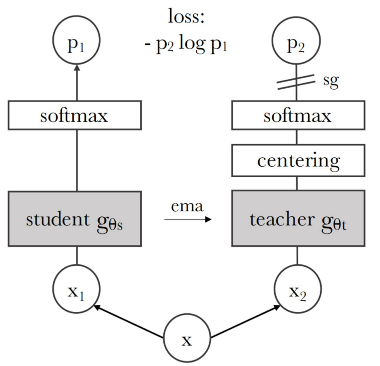
SELF-SUPERVISED LEARNING

DINO (Caron et al. 2021)

- Architecture of feature encoder is a vision transformer (ViT)
- Makes use of knowledge distillation via student-teacher architecture

Student receives:

- Several local crops (< 50% of image)
- Two global crops (> 50% of image)



DINO Architecture

Image from Caron et al. (2021)

Teacher receives:

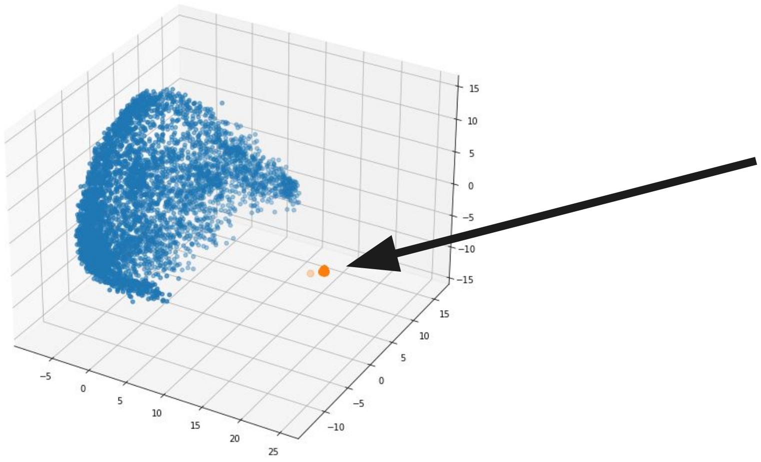
- Only two global crops

TRAINING AND ANALYSIS

TRAINING AND ANALYSIS

Training: First Pass

- Initial training: 45 epochs on all ~2 million spectrograms



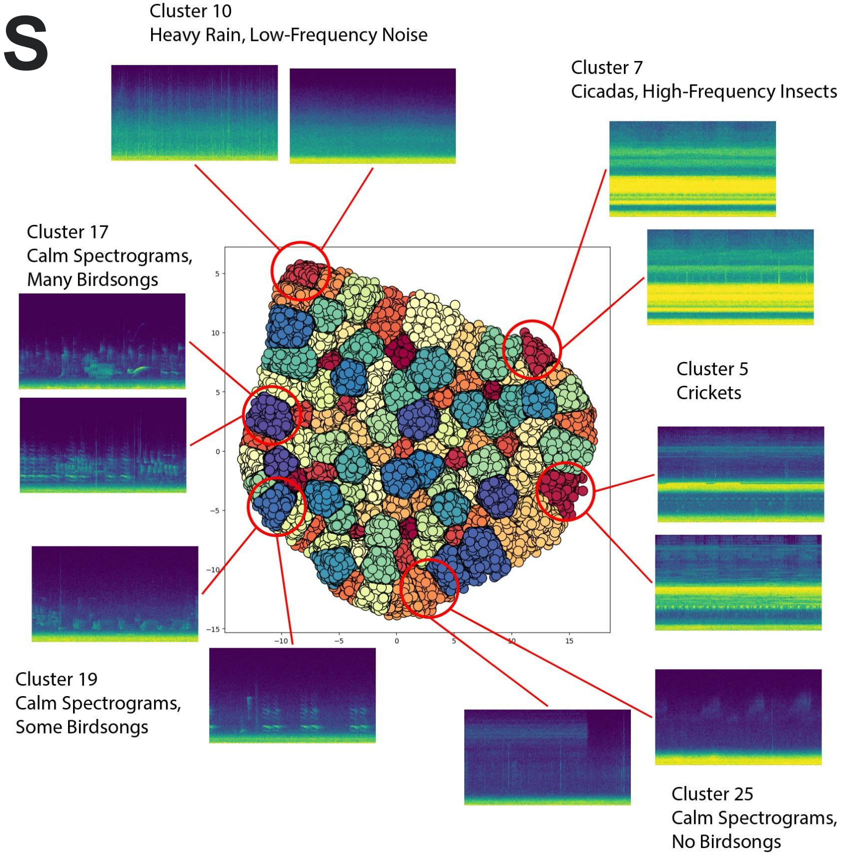
Identified and removed cluster of spectrograms found by BirdNET to have no detections. Confirmed this cluster to consist almost entirely of null (empty) spectrograms.

A subsample of feature vectors, visualized via PCA reduction to 3 dimensions.

TRAINING AND ANALYSIS

Training: Second Pass

- 45 epochs on all non-empty spectrograms
- k-means clustering on a subsample of spectrogram embeddings to identify cluster centers
- k-NN clustering on remaining spectrograms
- Why is this useful?
  - Analyzing the detections in a cluster by week, time of day, and location, allows us to monitor biological phenomena, no labels required!



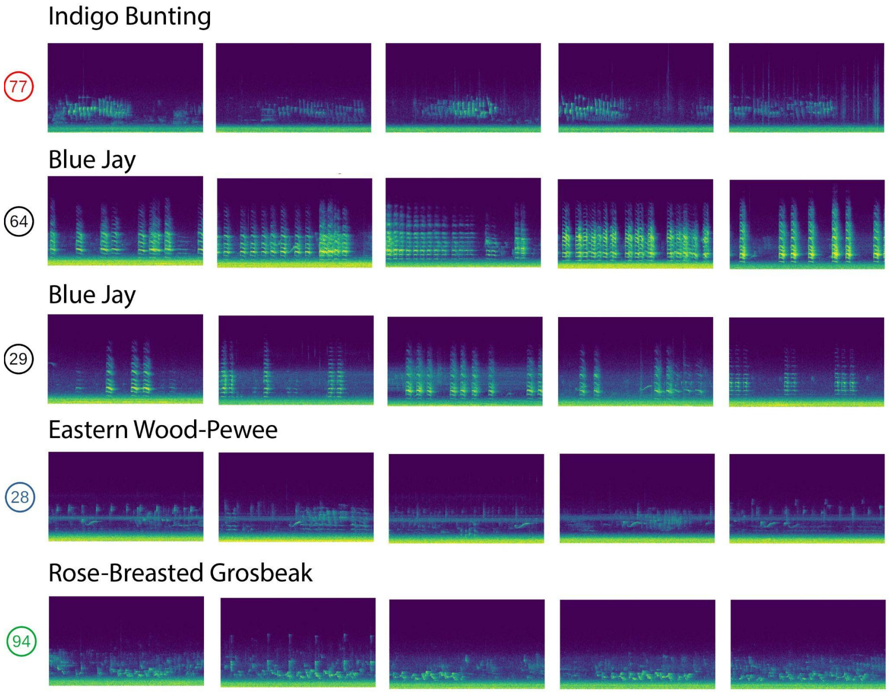
TRAINING AND ANALYSIS

Training: Third Pass

- Identified the ten clusters with the highest bird detection rates according to BirdNET, comprising ~200,000 spectrograms (~10% of the original dataset)
- Why not just use BirdNET directly on the spectrograms?
  - BirdNET fails to identify many species of birds present in our dataset
- Trained from scratch for 200 epochs; applied k-means and k-NN clustering with 100 clusters
- Many clusters exhibited a strong correlation with bird species
  - This correlation tended to be stronger in VICReg than in DINO

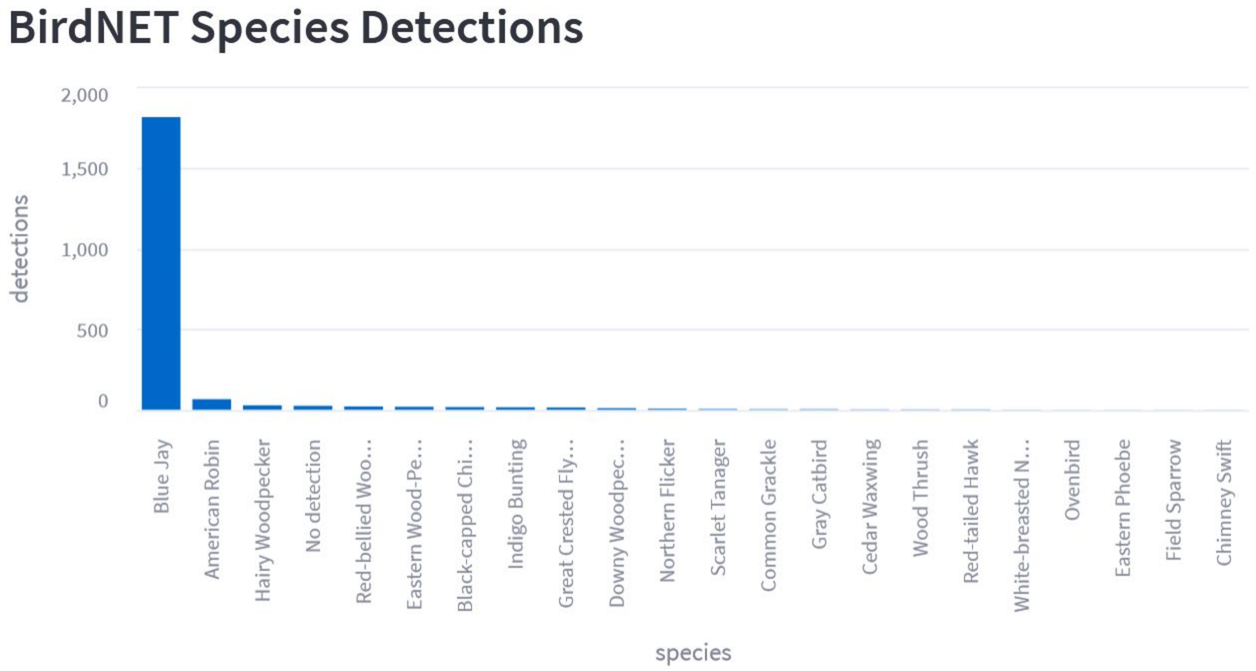
TRAINING AND ANALYSIS

Training: Clusters of Species



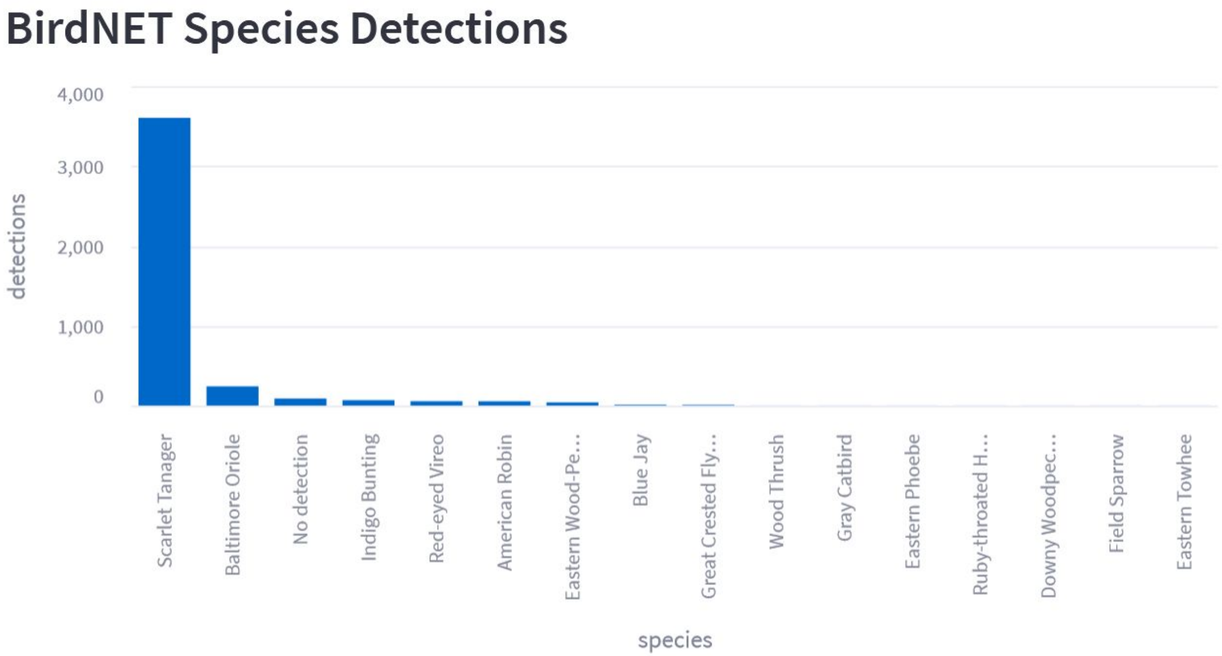
TRAINING AND ANALYSIS

Training: Third Pass - Cluster Analysis (VICReg, Cluster #50)



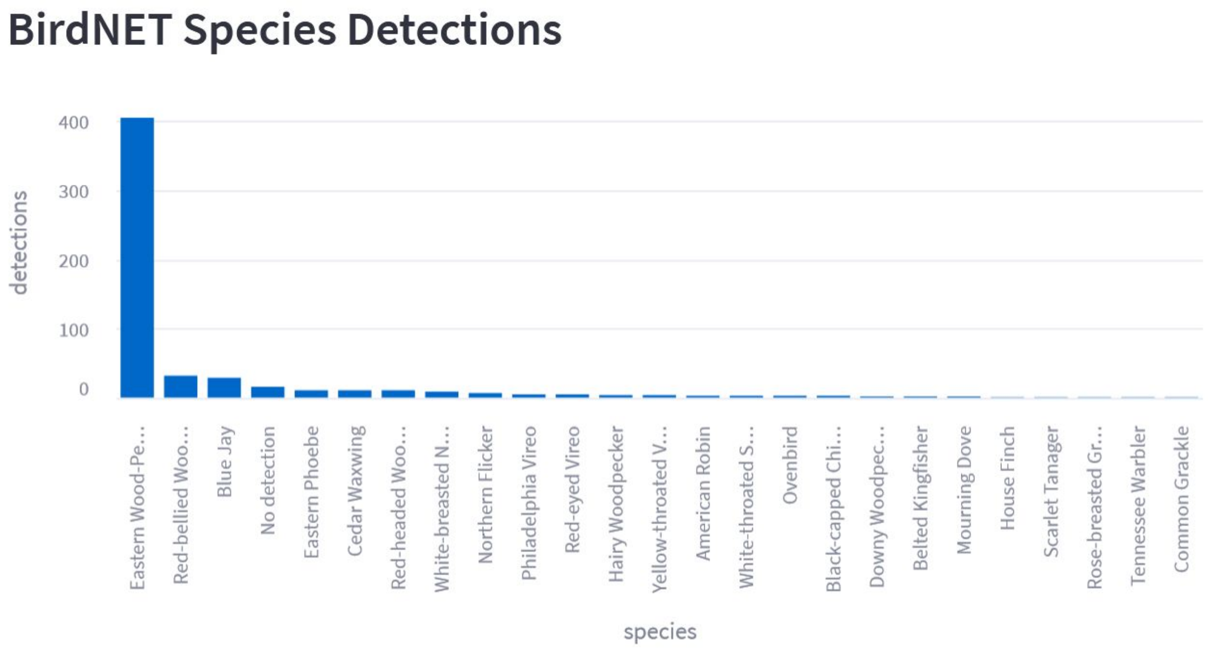
TRAINING AND ANALYSIS

Training: Third Pass - Cluster Analysis (VICReg, Cluster #38)



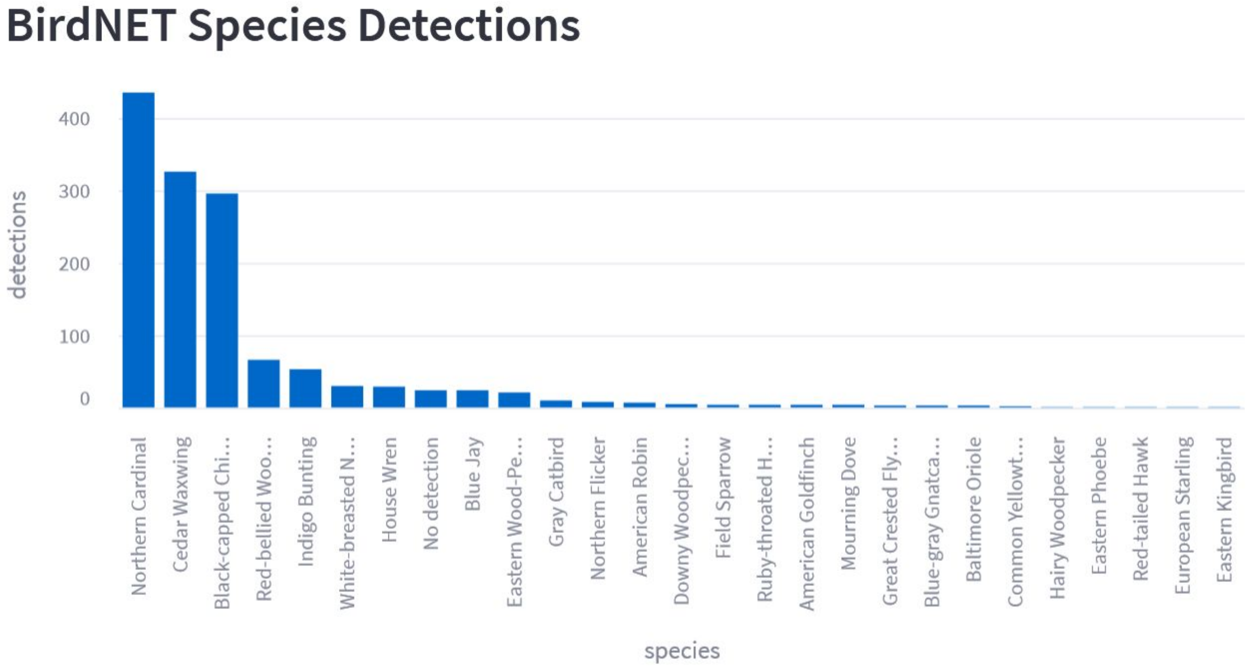
TRAINING AND ANALYSIS

Training: Third Pass - Cluster Analysis (VICReg, Cluster #83)



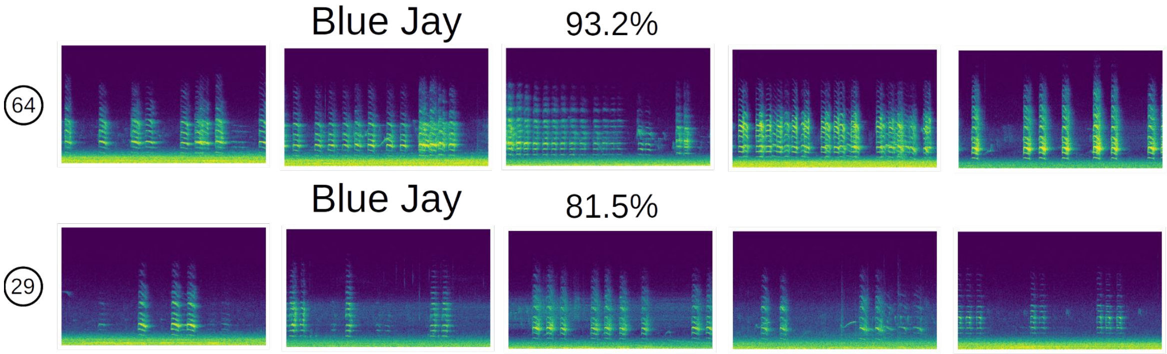
TRAINING AND ANALYSIS

Training: Third Pass - Cluster Analysis (VICReg, Cluster #85)



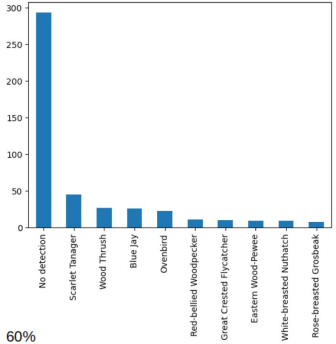
TRAINING AND ANALYSIS

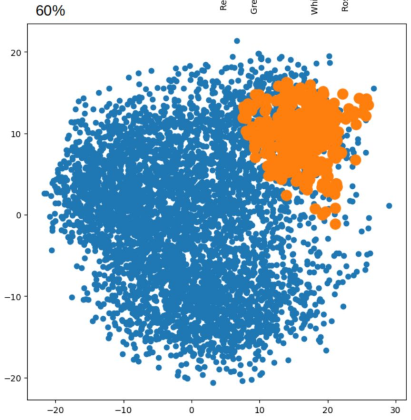
A Case Study: Blue Jays

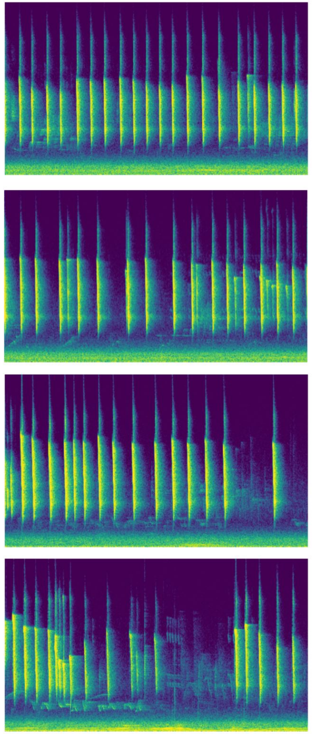


TRAINING AND ANALYSIS

A Case Study: Limitations of BirdNET

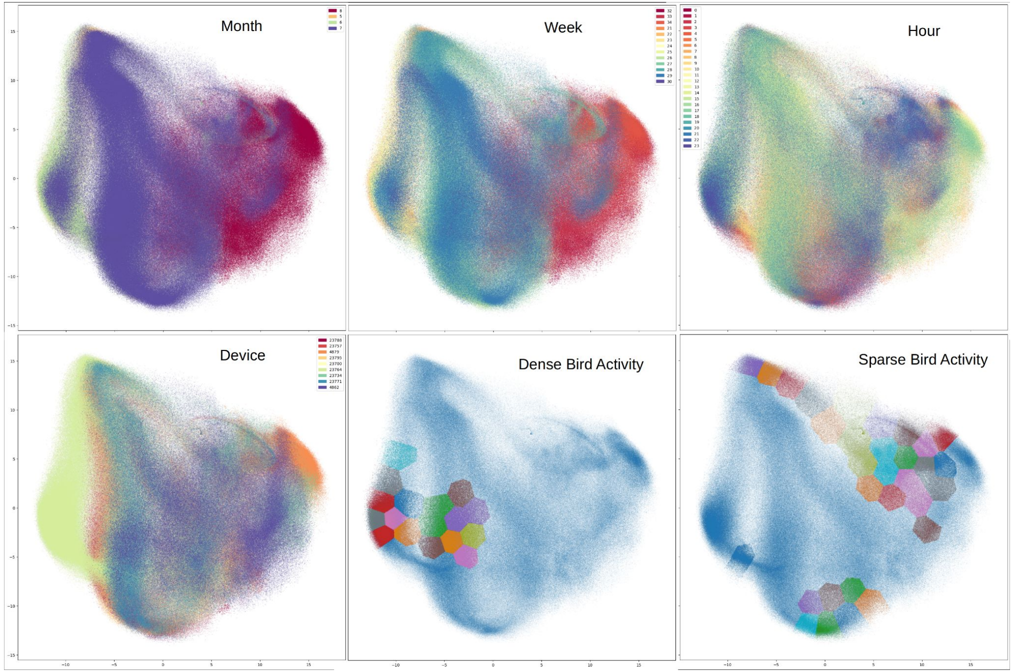






TRAINING AND ANALYSIS

Non-Species Clustering



LESSONS LEARNED AND FUTURE WORK

ON THE HORIZON

Analysis Tools

Avian Diversity Cluster Visualization

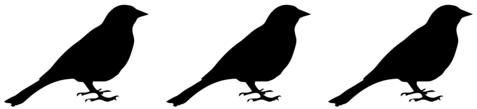
Select a cluster to analyze:

50 − +

Cluster Overview

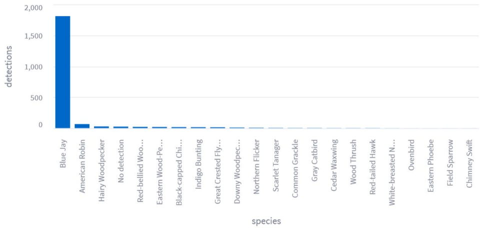
Cluster Number	Spectrograms	Detection Rate
50	2079	0.99

Top Bird Species

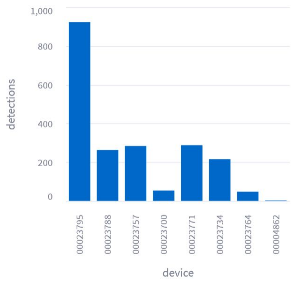


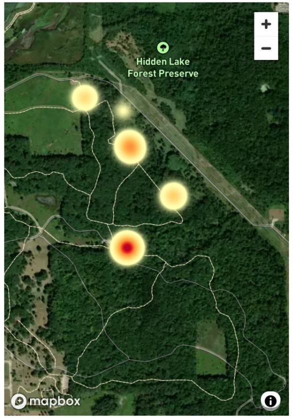
Blue Jay American Robin Hairy Woodpecker

BirdNET Species Detections

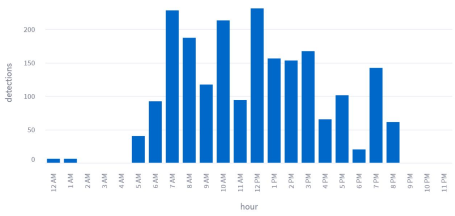


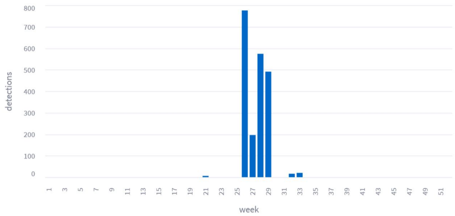
Detections by Location





Detections by Time





CONCLUSION: LESSONS LEARNED

Three Key Takeaways

- Self-supervised learning techniques are capable of making sense of large amounts of unlabeled data; allowing us to create powerful tools that can be used by ecologists.
- The application of self-supervised learning to unstructured data can lead to unpredictable results.
- Existing methods (e.g. BirdNET) are still limited; these limitations might be overcome through the clever application of SSL to large volumes of data collected by edge nodes, such as those deployed via Sage.

FUTURE WORK

Validation, Analysis, and New Techniques

- Further validation of features with supervised finetuning
- Applying unsupervised sound separation techniques (e.g. MixIT)
- Developing analysis tools
- Deploying to edge nodes for inference across many geographical locations

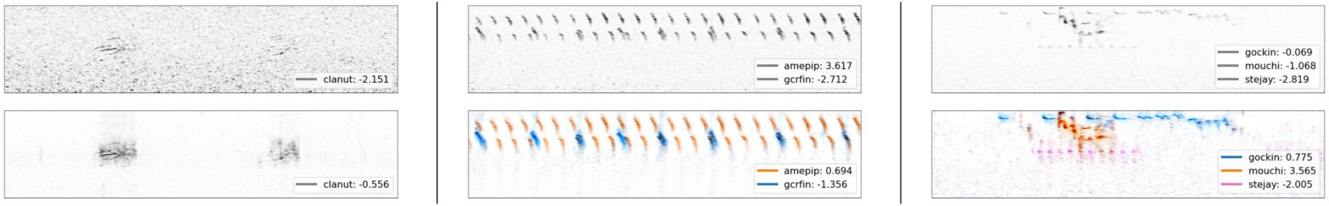


Image from "Improving Bird Classification with Unsupervised Sound Separation" (Denton et al. 2021)

Thanks for your time!

Questions?

Contact: rajanisamir@gmail.com