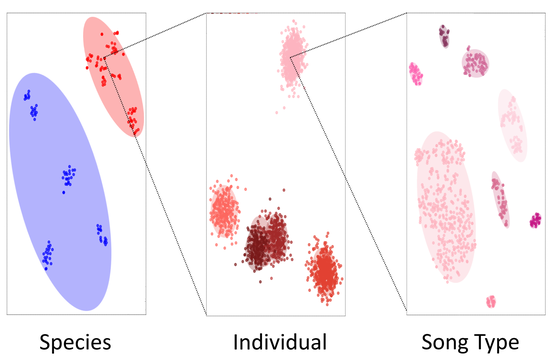
What do BirdNET Embeddings Capture: the Singer, the Song, or the Setting?

Akashram[1], Léa Bouffaut[2], Irina Tolkova[2]
1. Dwarkadas J. Sanghvi College of Engineering, akashram288@gmail.com
2. K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University

BirdNET is a widely used model for extracting deep learning embeddings in bioacoustics[1], but the factors shaping these embeddings remain poorly understood. In this study, we systematically altered song recordings from the Wytham Great Tit Song dataset[2] to assess how these changes influence the embedding structure and the performance of downstream tasks.



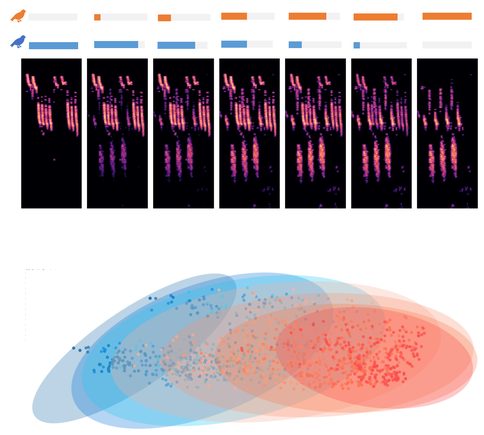


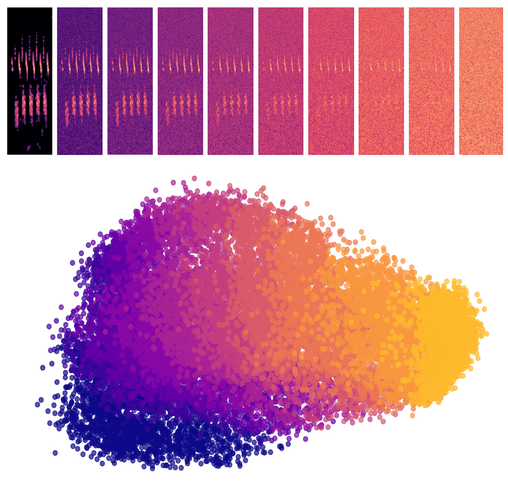
BirdNET embeddings reflect a hierarchical encoding of species, individual, and song-types

The figure on the left shows that BirdNET embeddings clearly separate Great Tit and Eurasian Blue Tit recordings, reflecting **species-level distinctions**. Linear Discriminant Analysis (LDA) projections further reveal **clustering by individual**, with **sub-clusters for different song types**. This suggests that the embeddings hierarchically encode species, individual identity, and vocal repertoire.

BirdNET embeddings capture dominant acoustic patterns

The figure on the right shows that as Great Tit and Eurasian Blue Tit vocalizations are mixed, their **embeddings shift smoothly between species clusters**. This nearly linear transition suggests BirdNET encodes dominant acoustic features proportionally, maintaining stable representations even when signals overlap.



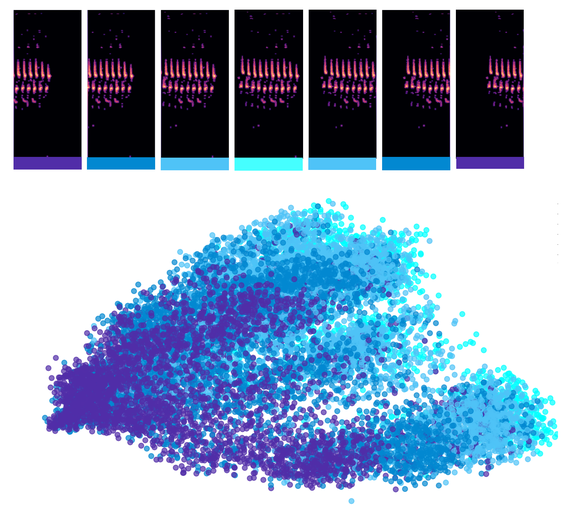


BirdNET embeddings are sensitive to SNR

Embeddings become **increasingly distorted** as the signal-to-noise ratio (SNR) decreases. The figure on the left illustrates a **near-linear shift** in embedding space as white noise is added at increasing levels. High levels of background noise push the embeddings further from their clean counterparts, confirming that BirdNET is sensitive to SNR **and may struggle to extract relevant features** when noise dominates the signal.

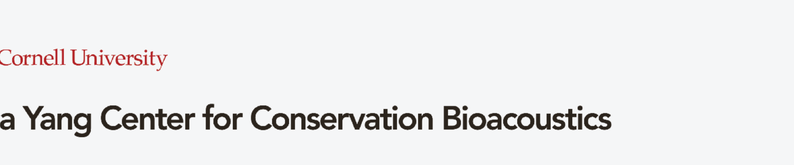
BirdNET embeddings are sensitive to truncation

Truncating the recordings causes a systematic shift in the embeddings, as shown in the figure on the left. The **more the signal is truncated, the further the embeddings drift** from their original position in a nearly linear fashion. This experiments highlights how missing or shortened signal segments can adversely affect BirdNET's performance.



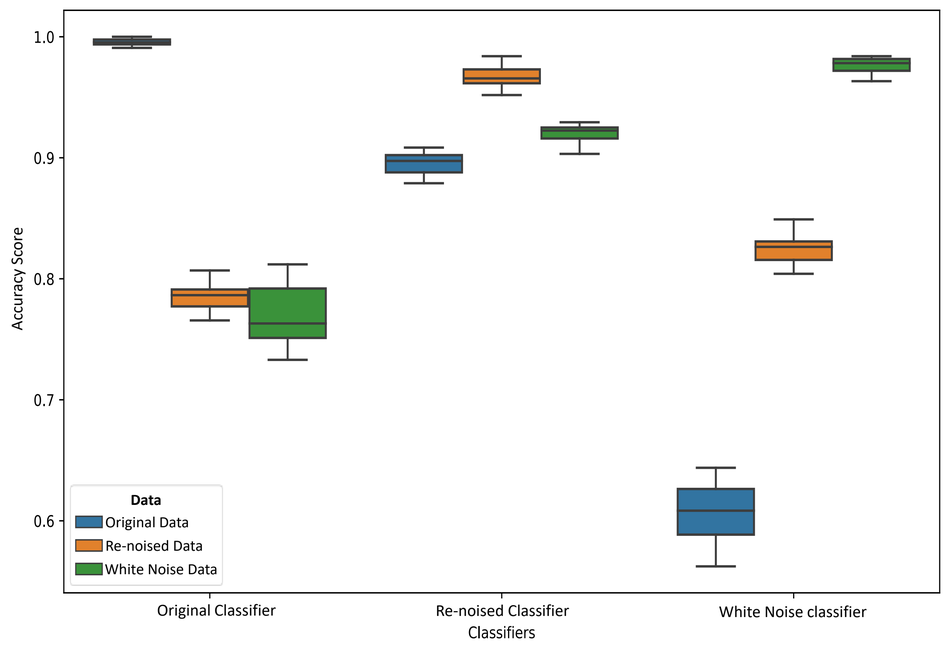
What is Re-noising ?

Re-noising involves first denoising a recording and then artificially adding background noise sourced from other individuals or environments. This **process disrupts individual-specific noise patterns**, encouraging models to focus on **meaningful vocal features** rather than background artifacts, thereby improving generalization.



Re-noising enhances generalization by reducing noise bias

The figure on the right shows that Support Vector Machines (SVMs) trained on recordings with **characteristic background noise perform poorly** on new or re-noised samples, indicating strong bias towards background noise. In contrast, **re-noising**—replacing background with that of other individuals—**improves generalization** by reducing this bias. All models were trained on BirdNET embeddings and compared to a baseline model trained on recordings with artificially added white noise.



References

[1] B. Ghani, T. Denton, S. Kahl, and H. Klinck, "Global birdsong embeddings enable superior transfer learning for bioacoustic classification," *Sci. Rep.*, vol. 13, no. 1, p. 22876, Dec. 2023, doi: 10.1038/s41598-023-49989-z.
[2] N. Merino Recalde, A. Estandía, L. Pichot, A. Vansse, E. F. Cole, and B. C. Sheldon, "A densely sampled and richly annotated acoustic data set from a wild bird population," *Anim. Behav.*, vol. 211, pp. 111–122, May 2024, doi: 10.1016/j.anbehav.2024.02.008.

Our analysis shows that BirdNET embeddings are influenced by multiple signal attributes, including background noise and vocal individuality. These insights highlight the need for caution when interpreting results from clustering, search, or similar analyses, and offer strategies for improving robustness in tasks like individual identification and species monitoring.