A Convolutional Approach to Birdsong Classification and Isolation

Gordon Doore & Ghailan Fadah

Department of Computer Science, Colby College, Waterville, ME

Background:

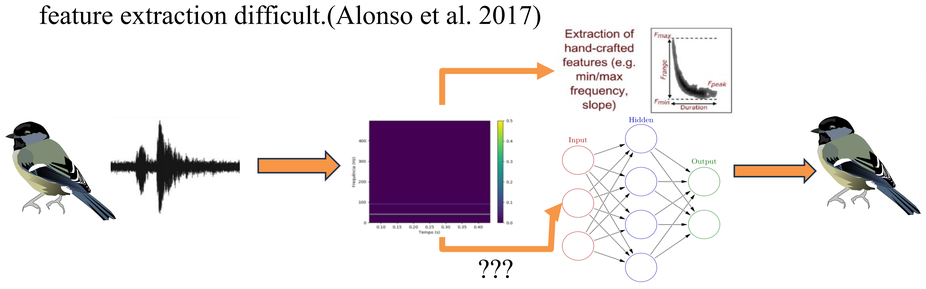
With the recent rapid decline in global biodiversity passive acoustic monitoring has become one of the standard methods to monitor ecosystems known as bioacoustics (Ella B. et al. 2017).

- non-invasive
- Extended deployment time
- Cost-effective and scalable

Allows for

- Estimation of species occupancy abundance, population density and community composition.
- monitoring illegal activities in high-risk areas.
- monitoring spatial and temporal trends in animal behavior

Current barriers in the field include noisy or uninterpretable data making feature extraction difficult.(Alonso et al. 2017)



Purpose: Leverage a neural network with encoder-decoder architecture to replace the traditional method of feature extraction.

RQ1: *To what extent can we classify bird species using a 3 second recording of their bird song?*

RQ2: *Can we isolate a single birdsong from a superposition of multiple?*

RQ3: *Once Isolated, how does this reconstruction behave as a birdsong? Can our classification model detect its class?*

Data:

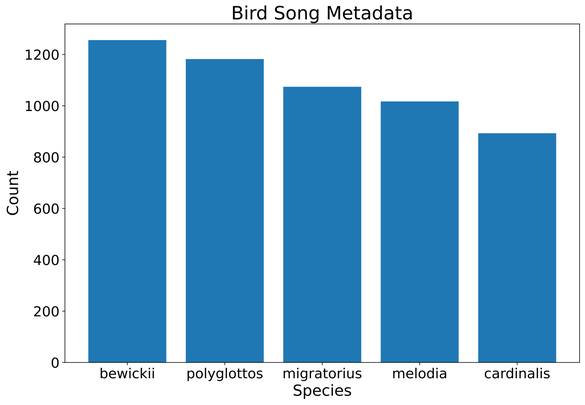
Dataset includes 5422 samples from xeno-canto (Naturalis biodiversity center, 2005). Each sample is a 3s recording. We construct superpositions of samples by adding waveforms of different species together.

Five species, Encoding

bewickii, 0
polyglottos, 1
migratorius 2
melodia 3
cardinalis, 4

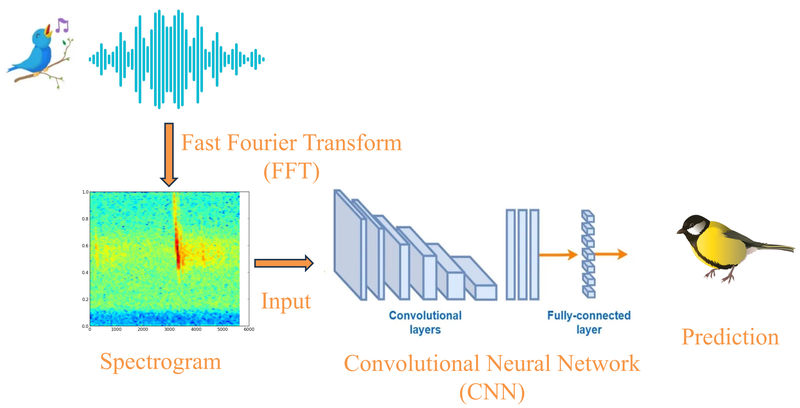
Split

5000 for training
300 for testing
122 for validation

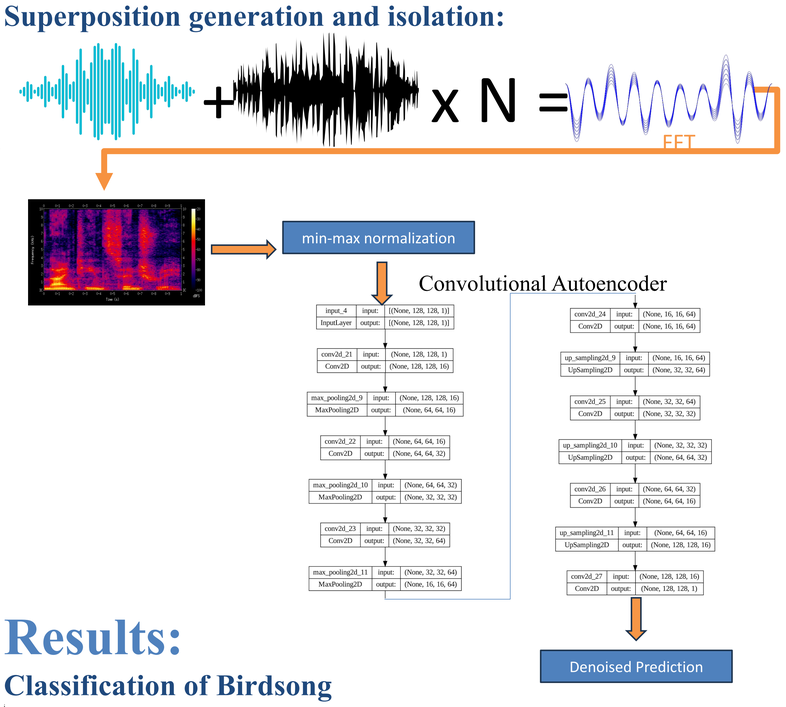


Methodology:

Bird classifying:



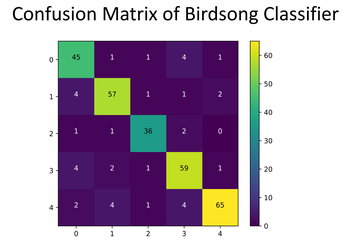
Superposition generation and isolation:

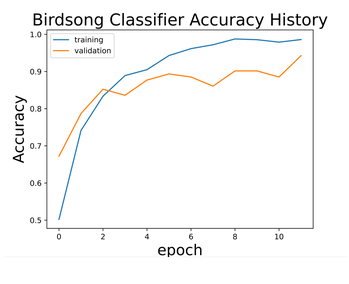


Results:

Classification of Birdsong

- Obtain an accuracy of 87% with CNN.
- The model makes mistakes at a similar rate across all samples (matrix).
- Classifier's accuracy increasing with diminishing improvements through the training session (accuracy history plot).

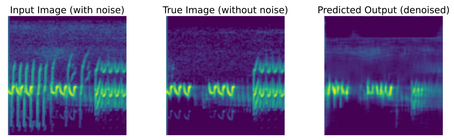




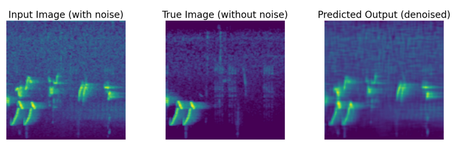
A Single Model or Many?

- Better performance obtained for specialized model.
- Bewickii model obtains a Structural Similarity Index (SSIM) of 0.408 and a mean squared error (MSE) of 0.0101
- General modelhas similar performance across our performance metrics (SSIM = 0.397, MSE = 0.0111), but tends to regenerate the input rather than denoise it.

Specialized Model



General Model



Classifying Denoised/Birdsong Isolation Examples

We chose to use the bewickii specialized model as it attains superior performance to the general model. We used the generated spectrograms from the test set of the data used to train the cardinal specialized model as an input to our classifier. We see below that the model significantly mistakes our denoised data between bewickii: 0, and polyglottos: 1.

Confusion Matrix of Specialized Reconstruction of Superpositions of Bewickii

Discussion:

Isolating superpositions of birdsong data can be done as demonstrated when building a specialized model, but fails under the current complexity of our single, general model. Implementing a general model or a variety of specialized models could allow for a more powerful tool for monitoring spatial and temporal trends in animal behavior.

Furthermore, the dataset used for these experiments is a small excerpt of a larger animal sound database, xeno-canto. It is conceivable to expand a model of this type to be able to more broadly classify or isolate different animal waveforms and superpositions.

We also see that there is room for further investigation into the dynamics of the interpretation of the classifications provided for the reconstructed bewickii data and the effects of fine-tuning using the reconstructed data are unstudied.

Conclusion:

- Birdsong is distinguishable by a CNN among the studied species
- The dominant birdsong can be isolated from a superposition. A specialized network performs more meaningful results, but achieves similar performance across our evaluation metrics.
- Denoised reconstructions are much harder to classify using the existing weights for the birdsong classification network.

Acknowledgements:

We'd like to thank the Colby College Computer Science Department, Professor Layton, and the CS443 Class Community for their help throughout this project