TCU College of Science & Engineering

A Convolutional Neural Network to Classify Zebra Finch Song

Kevin Bien, Samantha Shah, Dimitri McLain, Emily Sugg, Diana Villalta Palenica, Shayla Smith, Maria Solis, Andrew Magee, Soseh Asadoorian, Brenton G. Cooper
Psychology Department, Texas Christian University, Fort Worth, TX 76129

Introduction

- The zebra finch (*Taeniopygia guttata*) is widely studied by neuroscientists as a model of animal vocal learning because of its singing behavior.
- Classification algorithms are used to separate zebra finch songs from others noises so that singing can be studied.
- Current classification algorithms are slow and difficult to use, but convolutional neural network (CNN) models offer a promising alternative.
- CNNs are a class of deep neural networks that apply learned filters to extract high-level information (features) from input data and are often used for image classification.
- **In this project, we developed a CNN based algorithm and compared its performance to a baseline template-matching algorithm that is currently in use.**

Methods

- Audio recordings can be represented as images, which can then be processed by a CNN for classification.

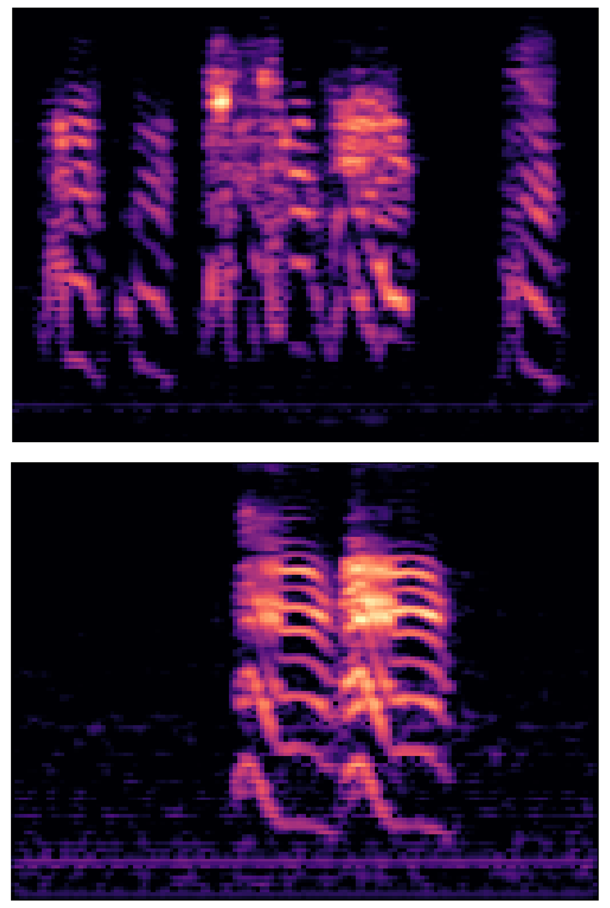


Figure 1. Spectrogram representations of a segment of zebra finch song (top) and other vocalizations (bottom).

- The CNN developed in this project was trained using recordings from 5 zebra finches.
- These recordings were processed into 0.5s files containing either song or other vocalizations and noise.

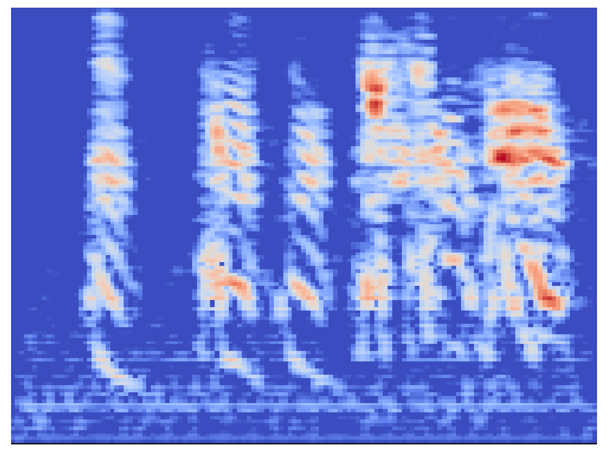


Figure 2. An example of a preprocessed spectrogram used to train the CNN.

Methods (Continued)

- The network contains 3 convolutional layers, which filter data from the previous layer of the network to generate feature maps, and 2 fully connected layers, which use the extracted features to compute the probability that the input is song.

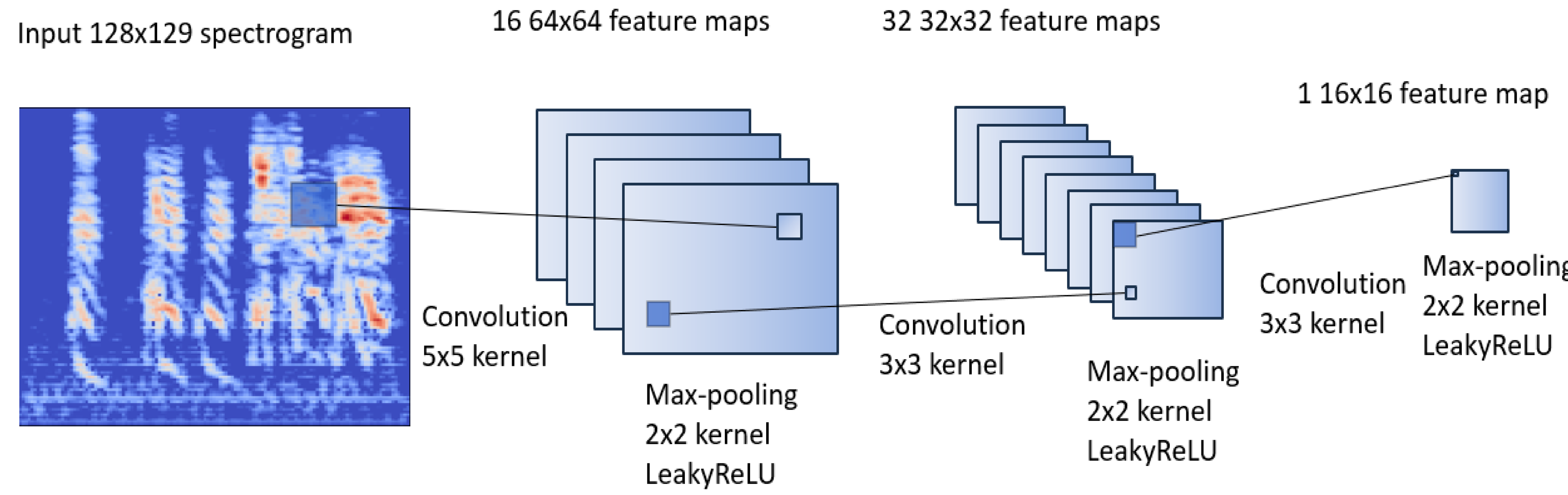


Figure 3. Overview of the convolutional layers of the CNN. Text on the top of the figure specifies the dimensions of data as it passes through the network, while the text on the bottom specifies which functions are being applied to the data.

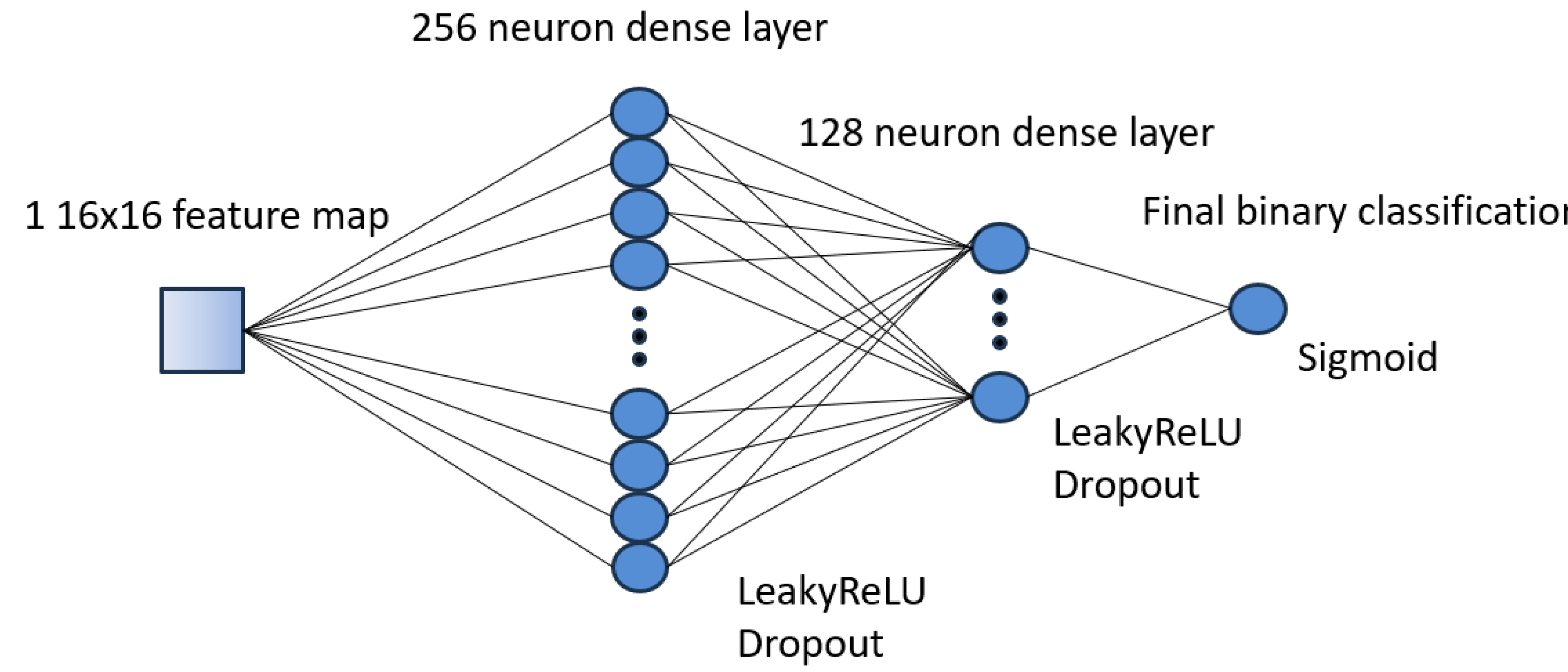


Figure 4. Overview of the dense layers of the CNN.

- The CNN was trained using the Adam algorithm, which is based on gradient descent.
- The trained CNN was compared to an existing laboratory standard template-matching algorithm to compare processing speed for birds outside of the original training set.
- Data for one of these birds was incorporated into the training set for a tuned model to test whether the performance improved with additional training data.
- To compare the performance of both algorithms, the number of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) classifications were measured based on visual classification of songs.
- Using the classification metrics, the accuracy (Acc) and Cohen's kappa (κ) values for each classifier were computed.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\kappa = \frac{Acc - p_e}{1 - p_e}, \text{with } p_e = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}$$

Analysis and Results

		True Class	
	Baseline	Song	Noise
Predicted Class	Song	170	54
	Noise	63	1867

		True Class	
	CNN	Song	Noise
Predicted Class	Song	160	52
	Noise	73	1869

Figure 4. Confusion matrices for baseline (left) and CNN (right) algorithms.

- For the baseline algorithm, $Acc = 0.946$ and $\kappa = 0.714$, while for the CNN-based algorithm, $Acc = 0.942$ and $\kappa = 0.687$.
- These data indicate that the baseline algorithm slightly outperforms the CNN-based algorithm on this dataset.

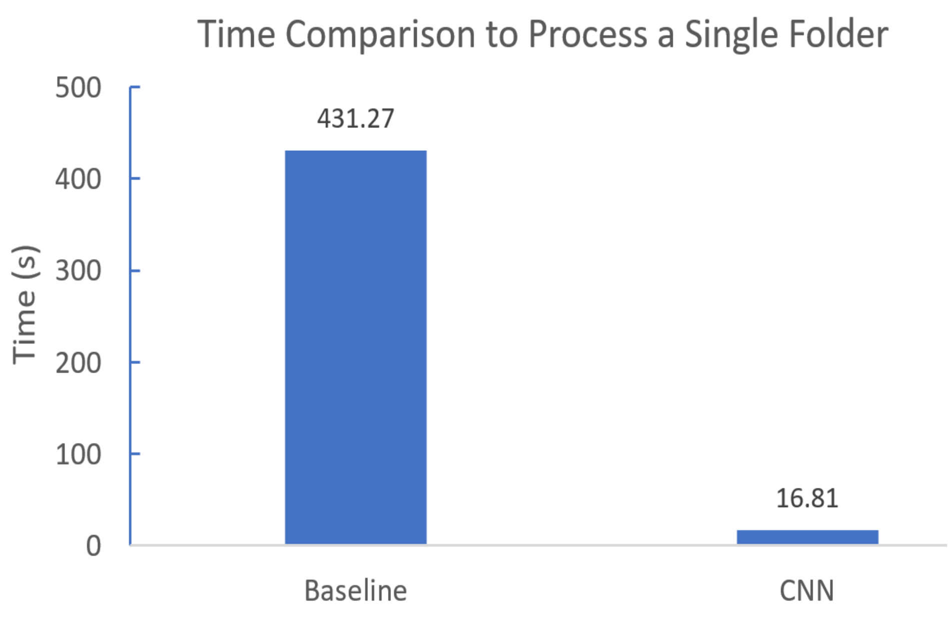
		True Class	
	Tuned CNN	Song	Noise
Predicted Class	Song	169	45
	Noise	64	1876

Figure 5. Confusion matrix for the tuned CNN algorithm.

- The algorithm based on the tuned CNN outperformed the other algorithms, with $Acc = 0.949$ and $\kappa = 0.7248$.

- The processing speed of both algorithms was compared for a single folder with 33 files.
- The CNN-based algorithm required only 3.9% of the time the baseline required.

Figure 6. Time comparison of the algorithms.



Conclusions

- CNN based audio classifications algorithms can achieve comparable performance to current algorithms for song identification.
- The CNN was able to recognize singing in birds that were outside of the original training set, suggesting that the model can generalize across zebra finches.
- CNN based methods are dramatically faster than current algorithms, potentially enabling real time song detection.
- Expanding the training data set improved model performance.

Acknowledgements: We thank Lindy Bledsue for expert animal care.