Proposal for a Bird Sound Identification System

Sarthak Sharma
ssharma19@earlham.edu
Earlham College
Richmond, Indiana, USA

ABSTRACT

My research study is about identifying birds sounds from recordings that include a mixture of sounds (birds, animals, insects, etc). The works build up on the research report from the Nobel Research Institute [14]. One of the ways my work can help social biologists determine if a particular species is present at a particular spot or not is by evaluating the condition (number of bird count) of particular bird species at a specific location [2]. Due to the wide range of bird cries and the difficulties in recognizing them, there is currently no approach giving a 100% accurate result for automating bird call recognition from audio recordings. I've described the procedures and methodology my aim will be to get above 70% result. In this paper, I will be using a machine learning approach. I will employ a deep convolutional neural network architecture to extract bird sounds from the mixed recordings of animal and birds sounds and after that to identify the bird species from the extracted recordings. The automatic bird call recognizer challenge have seen the highest progress with deep learning convolutional neural network (CNN) based designs. The LifeCELF Bird Competition (BirdCELF), an annual bench marking challenge to assess the state-of-the-art of audio based identification systems at scale, is one such bench marking challenge [15].

KEYWORDS

Audio Recordings, Graphs and Datasets,

ACM Reference Format:
Sarthak Sharma. 2022. Proposal for a Bird Sound Identification System. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 IMPORTANCE OF MY PROPOSAL

Bird Sound Identification is an important task that can be used for various purposes. One of the main applications of this skill is wildlife research. By identifying bird sounds, researchers are able to understand how different species interact with their environment and how they are affected by changes in climate or other factors. For example, recent research concerns how migrating birds respond when they encounter barriers to movement such as mountains or oceans [6]. In addition, bird sound identification can also be used for educational purposes. For example, teachers can use recordings of different birds to help students learn about these animals and their habitats. As a bird does say a lot more that just make sound - *Most children only consider one sound, but an observant child with more experience in bird sounds might ask, "What kind of bird?"* With the help of my research teachers can let student access sound of different bird species which would help answering the student questions. [12] The field of bird sound identification is broad and there are many unanswered questions that need to be explored. I am interested in this topic because it can provide solutions to pressing environmental concerns [7], such as deforestation, climate change, etc. [8]

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 PLANNED CONTRIBUTIONS

I will be using a publicly available data set of bird audio recordings found on Kaggle. The data set is provided by the Bird Celf 2021 competition and is available to all [3]. The data set includes over 1,000 species of birds from over 150 countries. I will convert the recordings into spectrograms. I'll be using software made available to students by "ravensoundsoftware," [4] which transforms audio recordings into spectrograms and wave patterns that CNN can access. I will use CNN to identify as many birds as I can available under folder 'train-short-audio'. This folder contains training data consisting of short recordings of individual bird calls from users of xenocanto.org. These audio files have been downsampled to 32 kHz and converted to the ogg format to match the test set audio. We anticipate that no additional training data is needed from xenocanto.org. Once my CNN has been successfully trained on the short audios I will move on to identifying the folder 'train-soundscapes' [3]. This folder contains audio files that are similar to the test set. They are all about 10 minutes long and in the ogg format. Using a CNN I created, I will determine the species of bird based on the more complicated patterns in the bird noises. By training the CNN on a large number of recordings (in this case 'train-short-audio'), the algorithm can learn to make accurate predictions on the type of bird based on their call. [16] This is an effective way to quickly identify bird species in large audio recordings (in this case the 'train-soundscapes' recordings will be used).

3 BACKGROUND RELATED / WORK SECTION

3.1 Recording

For my research study, I will be using recordings and data set of birds available publicly at Kaggle [3]. There is a long-standing practice of monitoring bird populations by conducting point count surveys [10]. At sampling locations, the observer will visually and aurally count every bird in a given time window (3 or 5 minutes). However, this process can be quite time consuming and requires expert knowledge in the identification of birds. With advances in technology however, there are now new ways to monitor bird

populations [10]. One of the main objectives for biologists and ecologists is to be able to collect and analyze data at large spatial scales to monitor the status, trends, distribution, and habitat use of wildlife species at various geographic locations.

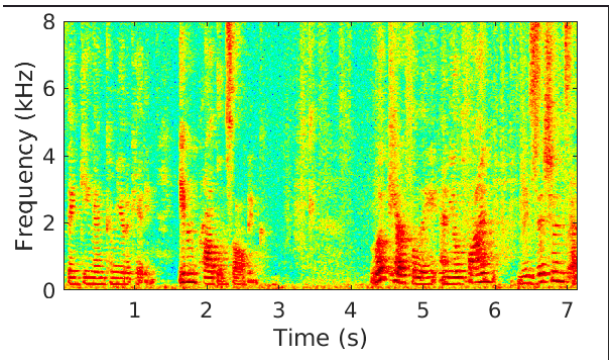

Figure 1: Audio recording converted to a spectrogram showing many different species over a period of 10 seconds
Doi - https://doi.org/10.1371/journal.pone.0179403.g007

3.2 Converting Audio to Visual Graphs

Converting the data set of short bird recordings into spectrogram is important. I will use Raven Sound Software to aid me with the process of doing my research project. The software allows users to listen to, view, and classify wildlife audio recordings as well as convert between different recording formats. An audio spectrogram is created from the audio files. The image Fig 1 displays it. The Y -axis displays the kHz, representing the recorded audio's frequency or loudness. The call length of birds are displayed on the X-axis, which may or may not be a single call. Spectrograms are important when identifying bird sounds because they provide a visual representation of a sound that can be used to distinguish different bird species. The visual representation of a sound is important as it helps to identify the unique characteristics of various bird sounds, such as frequency, amplitude, and duration. Additionally, spectrograms also help to identify and notice the sounds of birds that are too faint to be heard by the human ear [1].

3.3 Preprocessing

The next step is sorting, processing, and species identification from the transformed spectrograms. After we have successfully converted the audio into spectrograms we would be using CNN architecture to identify the sounds of the birds. The aim for my research would be introducing a new CNN structure that will aim in getting an accuracy above the already running softwares, for example the ResNet-50, a deep convolutional neural network architecture for automated bird call recognition. It has a 62 percent accuracy in identifying bird species [15].

3.4 Characteristics Of CNN

A convolutional neural network (CNN) is a special class of neural network that is built with the ability to extract unique features from image data. To train a CNN to help identify bird species from their bird sounds, the first step is to provide the CNN with a large data set of bird sounds and their corresponding species labels. For this we will be using data set provide by Bird Celf 2021 [3] named 'test-soundscapes'. To obtain accurate findings, we will employed CNN to classify the bird species. Once we successfully train the CNN to identify bird species we can collect more data, and further refine the model also if we have enough time in hand by the end of the next semester. CNNs can notice a variety of patterns in spectrograms. These patterns include frequency and amplitude, as well as rhythm and timing. Additionally, CNNs can also detect features such as formants, which are frequencies that appear in bird songs and calls. [17] Finally, CNNs can also detect harmonic structures, which are patterns of frequencies that occur in bird sounds.

3.5 How a CNN Operates

A CNN is a type of artificial neural network (ANN) that is used for image and sound recognition. The CNN is composed of multiple layers, with each layer processing the data in a different way. The layers are connected together, and they can be adjusted to improve the accuracy of the model. The CNN takes an input such as an image or sound recording, and it processes the data to identify patterns. CNNs have been used by researchers to identify bird sounds from audio recordings, with an accuracy of up to 70%. In a paper published in 2019 [18], researchers used a two-stage CNN-based model to identify bird species from audio recordings. The first stage used a CNN to extract features from a given audio recording, and the second stage used a CNN to classify the audio recording based on the extracted features. The researchers evaluated their model on two avian sound datasets, and achieved an accuracy of 70.4% and 70.7%, respectively. This demonstrates the potential of CNNs for identifying bird sounds from audio recordings. The outcome of the convolution procedure is known as an Activation Map or a Convolved Feature. We apply something called 'kernel' to reduce the size of the image. For example, given a 5 by 5 input, a kernel of 3 by 3 will output a 3 by 3 output feature map. For examples, See Fig 2.
This method will be similar to my CNN in that the input layer will be spectrograms of bird sounds, from which various layers will extract crucial information such as the harmonic structure, amplitude, frequency, and duration of the bird call, will help in identifying the class for each audio recording.

4 RELATED WORKS

ResNet-50, a deep convolutional neural network architecture for automated bird call recognition. The Res-Net 50 has a 62% accuracy in identifying bird species [15]. I will be using a similar approach and developing a CNN which would help identify the bird species. Although many other research papers discuss different techniques however, it has been determined that deep learning-based technique CNN with fully convolutional learning calls gets more accurate results because it eliminates the possible future modelling error caused by an imprecise knowledge of bird species [9]. I have never worked with CNN before and hence would like to start my research in this field and see how it follows. Similar technique, where passive acoustic monitors are used to record birds and animal sounds and then processed through bio acoustic analysis to get a spectrogram

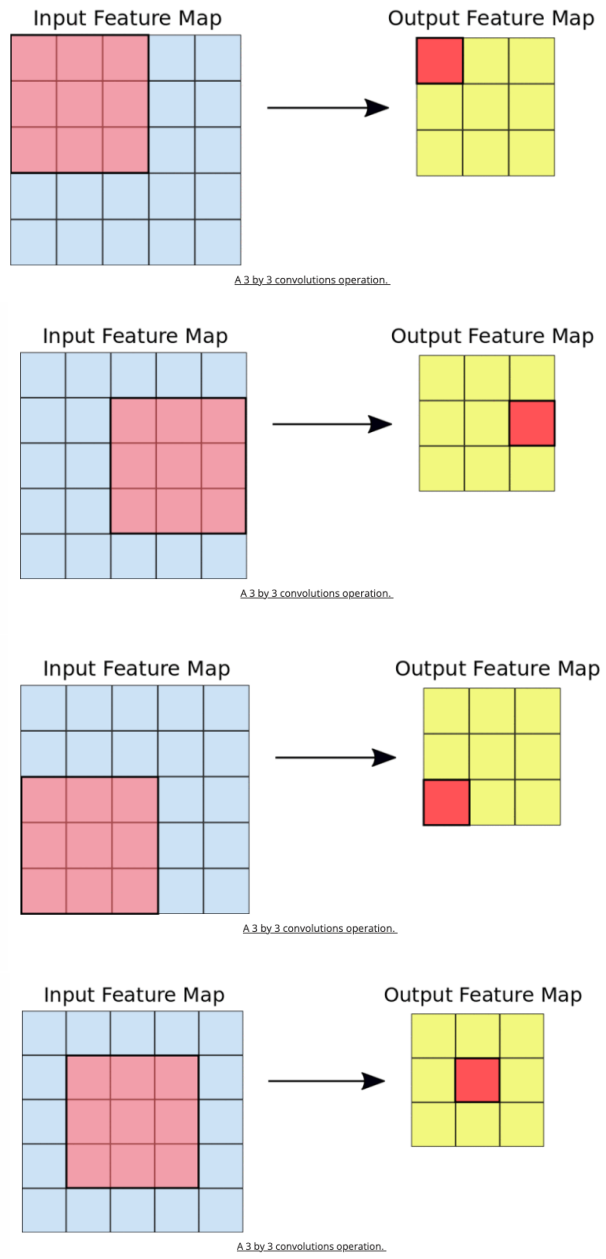

Figure 2: A 3 by 3 convolutions operation. [13]

output of the recordings so as to identify bird sounds are mentioned in the research paper [10]. The paper describes how the spectrograms are derived from the recordings, and the process of manually scanning, each bird's audio to classify it in its particular species using the Xeno-Canto Foundation's online bio-acoustic library[10]. The study also talks about how point count method was also used to know the frequency of birds flying by. The point count method is simply a process in which multiple individuals sit at different points, noting the frequency of birds and sounds. This way, it prevents overlapping the count of birds detected through ARUs [10].

5 ARCHITECTURAL DIAGRAM

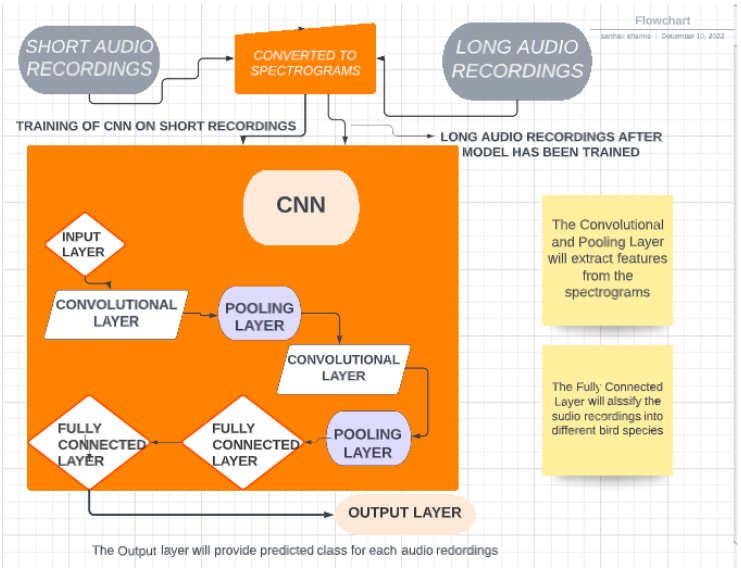

Figure 3: An Architectural Diagram of my plan

The input layer will be consist of both short audio recordings as well as the short sudio recordings. Once the CNN is trained on short 10 second videos we will switch the CNN to work on the 10 minute recordings. The Input Layer would consist of, Spectrograms of 10 minutes audio recordings.
Filtering will be done by Convolutional Layer: Filters with a kernel size of (3,3) with stride 2 (Stride is the number of pixels that the kernel moves when it slides across the input. It is used to control the output size of the convolutional layer.) and padding 'x' number. The Kernel size is the size of the filter used in the convolutional layer. It is usually a square or a rectangle and is usually set to (3,3). All the values I have taken right now have been decided after studying [13] research paper. In the upcoming semester with further research these values might or might not change.
Padding method discussed above is the number of pixels added to the input image before the convolutions are applied. This is done to ensure that the output size of the convolutional layer is the same as the input size.
Pooling Layer: Pooling with pool size (2,2) and stride 2
Convolutional Layer: 32 Filters with a kernel size of (3,3) with stride 2 and padding 'x' number.
Pooling Layer: Pooling with pool size (2,2) and stride 2. The Pooling layer is a subsampling layer that is used to reduce the spatial resolution of the input image. It is done by taking the maximum value of a certain region of the image and replacing it with the maximum value of the region.
Fully Connected Layer: 128 neurons
Fully Connected Layer: 64 neurons.
Output Layer: Bird Classification.

Loss Function: Cross Entropy Loss. In a CNN that identifies bird species from audio recordings, the Loss Function: Cross Entropy Loss describes how well the neural network is performing at predicting the correct bird species. Cross Entropy Loss is a calculation of how close the predicted output of the neural network is to the desired output. The lower the Cross Entropy Loss, the more accurate the neural network is at assigning the correct species [**?**].
The evaluation of the work will involve measuring the model's performance on a variety of metrics, such as accuracy, precision, recall, and F1-score. Accuracy is a measure of how often the model correctly identifies a bird species from audio recordings. Precision is a measure of how often the model identifies a bird species correctly, out of all the times it made a prediction. Recall is a measure of how often the model identifies a bird species out of all the times that species was actually present in the audio recordings. Finally, F1-score is a measure of the model's overall performance, which takes into account both precision and recall. The model will be evaluated based on its ability to generalize to unseen data, which is a measure of its robustness to changes in conditions such as noise and recording quality. Lastly, the model will also be evaluated based on its performance on other datasets, such as unseen images or sounds, which will give a better indication of its overall capabilities.This evaluation will help determine the model's effectiveness in recognizing different types of bird calls and accurately identifying bird species from audio recordings.

6 TIMELINE

- Week 1 - Gathering information for the paper.
- Week 2 - will get my First Draft Paper ready for over view and comments.
- Week 3 - Second draft of paper will be ready including everything (introduction, literature review, v1 of data architecture diagram, methods (data set(s), analysis plan)).
-Week 4 - A CNN trained on short audio recordings.
- Week 5 - Will make sure to get the first draft of CNN is due.
- Week 9 - After getting the necessary changes done the second draft of paper will be ready for submission(+initial results, +initial visualizations, and additional data if any).
- Week 10 - The First draft of demonstration video will be ready.
- Week 11 - First draft of poster will be ready.
- Week 12 - Third draft of paper.
- Week 13 - Second draft of demonstration video.
- Week 14 - Second draft of poster.
- Finals Week - final versions of paper, poster, and demonstration video.

7 CONCLUSION

An interesting application of bird sound identification is its potential use in solving various issues. By correctly identifying the sounds made by different species of birds, we can learn more about their behaviour and ecology [11]. This information can then be used to help us manage ecosystems better, prevent the spread of diseases among birds, and even mitigate the effects of climate change on bird populations. My research will contribute towards answering these many important questions. For my research I will be focusing on getting recordings from [3] and then, identifying as many birds as the CNN would allow me to. Numerous grassland bird species, like Northern Bobwhite Quail, are drastically declining. Therefore, it would be advantageous for society as a whole to have a tool that could be used to analyse the recordings in getting an accurate data on different bird species and the potential habitat destruction. The data can be used by biologists and academics to make projections about a species' population, size, and geographic dispersion. All of this data is essential for managing or monitoring a species. Bird Sound Identification has many uses and will solve issues. My research would help answering these problems. Unlike the common belief, bird sounds are not just beautiful music for people to appreciate. They actually have a lot of functions that can be very useful for humans too! For example, by identifying the different types of birds through their unique calls, we can get an understanding about what kind of environment they prefer and whether it is healthy or not. This knowledge can then be used to address some major environmental concerns that humanity is currently facing.

REFERENCES

[1] 2006. why are spectrograms important in identifying bird sounds. (2006), 1–4. https://www.audubon.org/news/why-are-spectrograms-important-identifying-bird-sounds
[2] 2022. ARTICLE Why Monitor Birds? *National Park Service* (2022), 4 pages. https://www.nps.gov/articles/000/why-monitor-birds.htm
[3] 2022. BirdCLEF 2021 - Birdcall Identification. (Dec. 2022), 1–4. https://www.kaggle.com/competitions/birdclef-2021/overview
[4] 2022. Raven Sound Analysis. (2022), e0211970. https://ravensoundsoftware.com/software/raven-lite/
[5]]CRROPY Jason Brownlee. [n. d.]. A Gentle Introduction to Cross-Entropy for Machine Learning. ([n. d.]).
[6] Northern Rocky Mountain Science Center. 2016. The ecology, behavior, and conservation of migratory birds. (2016), 1–4. https://www.usgs.gov/centers/norock/science/ecology-behavior-and-conservation-migratory-birds
[7] The Nature Conservancy. 2022. This Year's Big Wins in Land Conservation. (Dec 2022), e0211970. https://www.nature.org/en-us/
[8] Environment Defence Fund. 2022. We are Environmental Defense Fund, the organization that is all-in on climate — the greatest challenge of our time. Our game-changing solutions put people at the center of all we do. (Dec 2022), e0211970. https://www.edf.org/
[9] Hasan Abdullah Jasim, Saadaldeen R. Ahmed, Abdullahi Abdu Ibrahim, and Adil Deniz Duru. 2022. Classify Bird Species Audio by Augment Convolutional Neural Network. (2022), 1–6. https://doi.org/10.1109/HORA55278.2022.9799968
[10] Kinga Kułaga and Michał Budka. 2019. Bird species detection by an observer and an autonomous sound recorder in two different environments: Forest and farmland. *PLoS One* 14, 2 (2019), e0211970.
[11] Shelby Lawson and Mark E. Hauber. 2022. What Can We Learn from Bird Song? Recent Advances in Functional and Applied Avian Bioacoustic Research. *American Ornithological Society Journals* (2022), 8 pages. https://doi.org/DepartmentofAnimalBiology,SchoolofIntegrativeBiology, UniversityofIllinois,Urbana-Champaign,ILUSA
[12] Debbie Mondale. 2009. Exploring bird sounds with children. (August 2009), e0211970. https://musiconnx.wordpress.com/2009/08/30/exploring-bird-sounds-with-children/#:~:text=Fun%20vocal%20play%20with%20different,enough%20to%20imitate%20it%20effectively.
[13] Derrick Mwiti. 2022. Image Classification with Convolutional Neural Networks (CNNs). (May 2022), 4. https://www.kdnuggets.com/2022/05/image-classification-convolutional-neural-networks-cnns.html
[14] Mike Proctor and Stephen Webb. 2020. Capturing bird calls and other wildlife sounds with bioacoustics. *Capturing Bird Calls and Other Wildlife Sounds With Bioacoustics* 38, 12 (Dec. 2020), 1–4.
[15] Mangalam Sankupellay and Dmitry Konovalov. 2018. Bird call recognition using deep convolutional neural network, ResNet-50. 7, 9 (2018), 1–8.
[16] Jon Schwartz. 2020. Building a Convolutional Neural Network to Classify Birds. *Wildlife Acoustic* 5, 2 (2020). https://blog.jovian.ai/building-a-convolutional-neural-network-to-classify-birds-528794240fa1
[17] Mike Smales. 2019. Sound Classification using Deep Learning. (Feb. 2019), 1–4. https://mikesmales.medium.com/sound-classification-using-deep-learning-8bc2aa1990b7

[18] Stefano Ermon Volodymyr Kuleshov, S. Zayd Enam. 2017. Audio Super Resolution using Neural Networks. (Aug 2017). https://arxiv.org/abs/1708.00853 ,