iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43309

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Bird Species Identification using Audio Signal Processing and Neural Networks

Dr. Amol Dhakne[1], Vaishnav M. Kuduvan[2], Aniket Palhade[3], Tarun Kanjwani[4], Rushikesh Kshirsagar[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, DYPIEMR, Akurdi, Pune*

Abstract: ***In this work, automatic bird species recognition systems were developed, and their identification methods were investigated. Automatically identifying bird calls without physical intervention has been a large and tedious endeavor for major studies in various subfields of taxonomy and other ornithology. This task uses a two-step identification process. In the first phase, an ideal dataset was created containing all recordings of different bird species. Next, the sound clip was subjected to various sound pre-processing techniques such as pre-emphasis, framing, silence removal, and reconstruction. A spectrogram was generated for each reconstructed sound clip. The second step used a neural network given with the spectrogram as input. A convolutional neural network (CNN) classifies sound clips and recognizes bird species based on input characteristics. In the above system, a real-time implementation model was also designed and executed.***
Keywords: Bird species identification, bird sound, sound pre-processing techniques, Convolutional Neural Network, Spectrograms

I. INTRODUCTION

According to the International Union for Conservation of Nature (IUCN), there are about 10,000 known species of birds scattered throughout ecosystems, from the rainforests of Brazil to the icy coasts of Antarctica [1]. These species show amazing diversity in behavior and morphology and are the essence of the normal functioning of ecosystems. However, this magnificent biodiversity has been threatened by recent human activity, from habitat invasion to complete extinction of habitats, coupled with natural phenomena such as global warming and climate change. Many species are being driven to extinction. About 1,370 species are endangered and are estimated to account for about 13% of the total bird population. Many bird species are common, but they are difficult for humans to identify. The purpose of this research is to develop automated techniques for identifying bird species based on sound.

Automatic identification of bird calls from continuous environmental records will be an important addition to the research methodology of ornithology and biology in general. Often, these recordings are truncated or noisy. Therefore, you should use a reliable automation method rather than the traditional manual method. Manual inspection of the spectrogram is often error-prone, and this technique is usually esoteric in nature and involves multiple experts, which makes the spectrogram unreliable and requires an automated system. Become. Bird watching is a popular hobby in many countries, and such systems have great commercial potential. Extensive international programs also stimulate activities in the fields of bioacoustic signal processing and pattern recognition.

In this paper, we proposed a method that uses sound processing and convolutional neural networks to automate the entire process of identifying bird calls. The first step is to create a database containing all the recordings. These recordings are then subjected to acoustic preprocessing techniques such as pre-emphasis, framing, silence removal, and reconstruction. Spectrograms were generated for audio clips and these spectrograms were provided as input to GPU-trained CNNs.

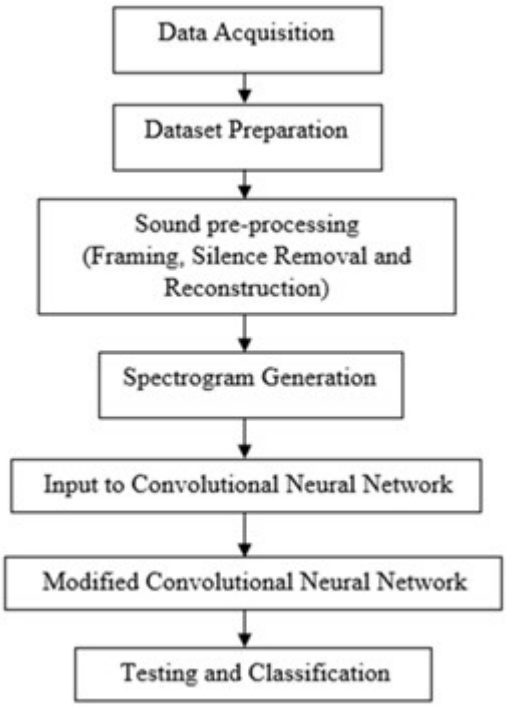
Data sets are the most important and important determinants of a machine learning-based approach to a problem. The dataset must have certain important properties in order to be certified as a good dataset. It must be accurate and accurate, free of false elements and misleading information. It is reliable, consistent, and there should be no inconsistencies in the dataset between different data items, regardless of the source. Fragmented data provides an inaccurate picture, so the dataset must be complete, comprehensive, and relevant. Data sets should be fine-grained and unique to avoid the confusion that results from aggregated, summarized, and manipulated data.

One of the main limitations of our research was the availability and accessibility of bird song datasets, as few studies were done in this area. So I manually created a dataset of bird calls from that area. We also downloaded a bird song recording from xenocanto.com, a dedicated website for sharing bird songs around the world [2]. The length of each sound clip ranged from 5 to 20 seconds. The dataset was created for four birds: cuckoo, sparrow, crow, and laughing dove. Each bird has 100 recording sample spaces, for a total of 400 bird sound recording samples.

©IJRASET: All Rights are Reserved

We also decided to create a dataset containing 100 samples of human audio clips and ambient sounds taken from the Google Audioset and LibriSpeechASRCorpus datasets. In real-time recordings, bird sound clips are usually interleaved with ambient noise, and human voice samples are used to build a reliable network, so include ambient noise and human voice samples. The decision is justified.

II. METHODOLOGY



A. Pre-Emphasis

The audio signal picked up by a microphone usually contains a wide range of frequencies. Audio signals and bird vocalizations are usually composed of higher frequency energies. In other words, predistortion amplifies or enhances high frequency components. Therefore, it is important to emphasize this high frequency energy that we are interested in and lower other frequencies. This is done using a simple first-order highpass filter.

All data points in the signal pass this filter given by the following formula. y [n] = x [n] S * x [n1]The audio signal picked up by a microphone usually contains a wide range of frequencies. Audio signals and bird vocalizations are usually composed of higher frequency energies.

In other words, predistortion amplifies or enhances high frequency components. Therefore, it is important to emphasize this high frequency energy that we are interested in and lower other frequencies. This is done using a simple first-order highpass filter. All data points in the signal pass this filter given by the following formula. y [n] = x [n] S * x [n1]Your paper must use a page size corresponding to A4 which is 210mm (8.27") wide and 297mm (11.69") long.

B. Framing and Silence Removal

The audio signal is not a stationary signal. It consists of various statistical properties that can change over time. You need to divide the first recorded audio signal into several frames based on its length, and then extract the signal without unwanted silence periods. The frame length is determined by considering the total length of the signal and the sampling period used. In this work, it is assumed that 2.5% of the total length of the audio clip is the length of a single frame. After framing is complete, the threshold feature is used to perform silence removal. The threshold function is selected so that you are interested in the above audio signals and signals below the threshold are considered a period of silence or background noise. This silence removal is repeated for every frame and dynamically changes the threshold to 7% of the maximum amplitude present in that frame.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Reconstruction

The audio signal is not a stationary signal. It consists of various statistical properties that can change over time. You need to divide the first recorded audio signal into several frames based on its length, and then extract the signal without unwanted silence periods. The frame length is determined by considering the total length of the signal and the sampling period used. In this work, it is assumed that 2.5% of the total length of the audio clip is the length of a single frame. After framing is complete, the threshold feature is used to perform silence removal.

The threshold function is selected so that you are interested in the above audio signals and signals below the threshold are considered a period of silence or background noise. This silence removal is repeated for every frame and dynamically changes the threshold to 7% of the maximum amplitude present in that frame.

D. Spectogram Generation

Reconstruction is the process of combining or concatenating all the frames captured after the framing and silence removal process. The result of this process is a signal that contains most of the information of interest without a noticeable period of silence. The final step is to get the best sample of the preprocessed audio signal by examining the 1 second of the clip with the highest amplitude of the total duration of the reconstructed signal. Figure 3 shows the result of signal reconstruction after removing silence.

The spectrogram is a graphical representation of the frequency range and how it changes over time. It usually consists of x-axis time and y-axis frequency range, and the colors in the graph indicate the power / intensity of that particular frequency.

The spectrogram can be generated by first transforming the time domain signal into the frequency domain using the Fourier transform, and then plotting the frequency.

In this document, the spectrogram is generated using the built-in MATLAB function. Spectrograms are created only for the data captured after reconstruction, not the entire signal. This process is repeated for all audio clips in the dataset and each spectrogram is saved in the specified folder.

Each spectrogram is unique and has its own characteristics. For example, a sparrow bark spectrogram usually contains relatively low frequencies and high intensities, while a crow call spectrogram has high intensities over a range of frequencies. These differences are easily captured by neural networks during training. For this article, we chose AlexNet as our neural network because it is highly accurate and easy to implement in MATLAB. The generated spectrogram is downscaled to a resolution of 227 x 227 x 3 and can be used for training AlexNet neural networks.

III.REAL TIME IMPLEMENTATION

We have trained a convolutional neural network (AlexNet) so that we can predict bird species for a particular input recording. However, these input recordings were collected in an ideal environment with virtually no noise, and in some cases noise has been removed by some pre-treatment techniques.

Therefore, a network trained with a dataset that can predict a particular species if the input data is free of glitches or noise. Unfortunately, this is not the case for real-time recording due to the presence of noise in the environment. Noise can be caused by many factors such as vehicle noise, human voice interference, and natural phenomena. You need to make sure that your network works as intended and at the same level that it worked in a previously simulated environment, such as when predicting bird species on a noise-free dataset.

For the network to work in real time, the CNN needs to be retrained with an ideal dataset and a dataset containing sound samples collected from the local environment. Randomly selected audio clips from different bird species in the dataset are converted to a fixed sample rate of 44100Hz or 48000Hz to maintain diversity and prevent overfitting. Also, the bit rates are set to 128kbps and 320kbps. These are standard bitstreams used in audio applications to ensure clear audio recording with small file sizes. After all audio clips have been converted to the desired sample rate and bitstream, a spectrogram of each audio clip is generated. These spectrograms are used to retrain the neural network. Once transmission training is complete, you can save and reuse the model and convert real-time audio signals into spectrograms for classification. We recommend setting the microphone to a sample rate of 44100Hz and a bitrate of 128kbps or 320kbps. The system was tested in a real-time environment with 91% accuracy in classification results. A graphical user interface (GUI) has been developed to operate the above system, including all the processes it performs, from recording in a real-time environment to processing data and displaying results.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



IV. FUTURE SCOPE

This project has an exponential range of future improvements in terms of economic and scientific opportunities. The application can be designed and deployed for mobile devices that allow users to use their smartphones as wearable devices to predict and analyze bird calls. The song of the bird can be recorded by the user. The app processes the recording on your device and returns the results along with a bird image, description, and population distribution. Records can also be sent to cloud servers running sophisticated CNNs for careful analysis and evaluation for more accurate results. CNN can also be deployed in hardware setups such as Neural Compute Stick and Raspberry Pi. These hardware settings can be installed in ecological parks, sanctuaries, and bird sanctuaries. The resulting data can be stored locally or in the cloud. The data thus obtained will be of great importance in the study of bird migration patterns, population distribution, biodiversity and bird demography in specific areas.

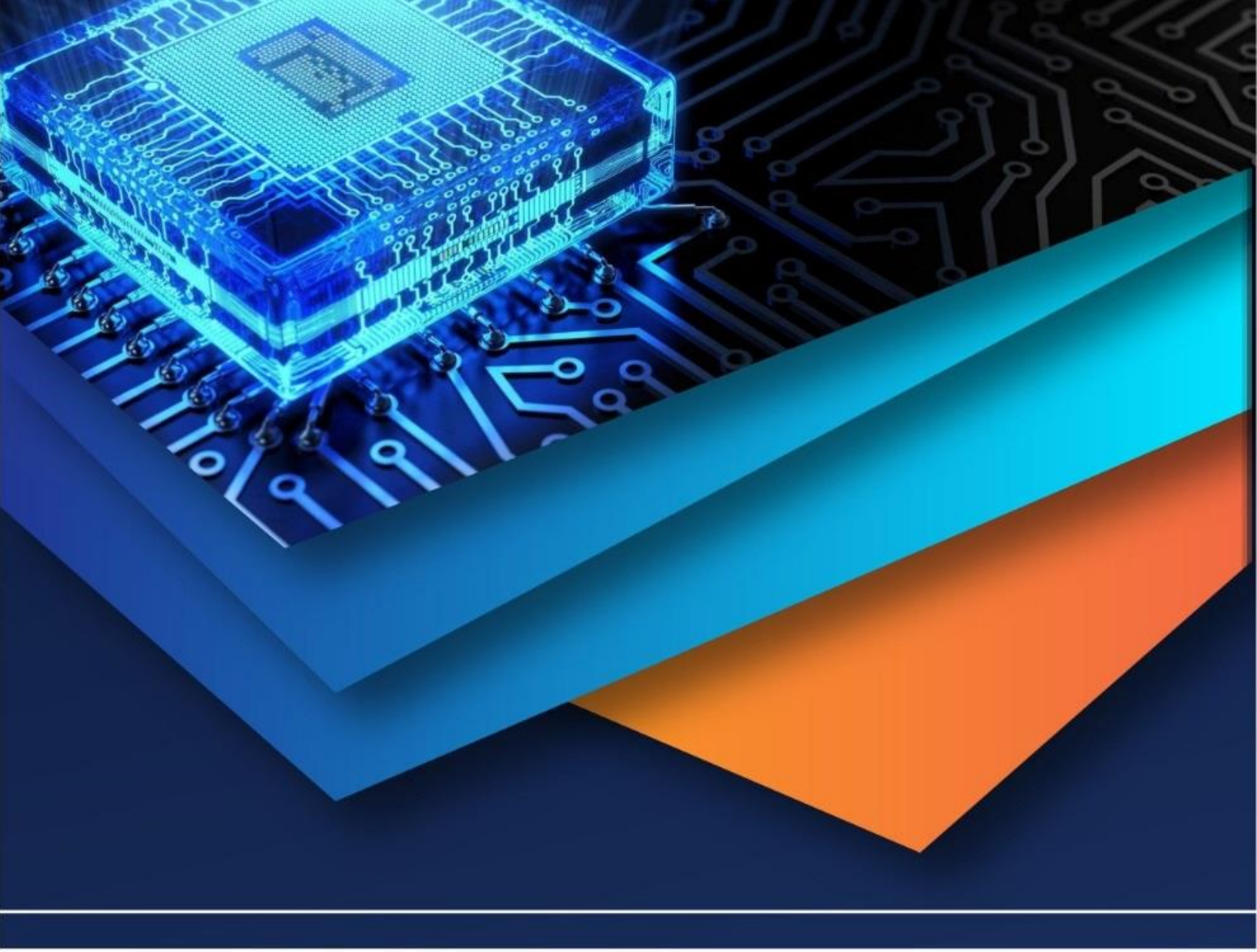
V. CONCLUSIONS

Four different bird species were identified in this project. In this approach, after preprocessing the bird's bark, we generated a spectrogram and used it to train the model for classification. The data used for training consisted of real bird sounds recorded in natural habitats, among all other sounds. Results were observed at various values of learning rate, number of epochs, and data partitioning. The system was able to classify bird species with 97% accuracy based on spectrogram images generated from bird sounds. This accuracy was achieved by considering the human voice as well as the bird's bark. You can further improve the accuracy by fine-tuning the performance parameters.

REFERENCES

[1] www.iucn.org/theme/species/our-work/birds
[2] Vemula Omkarini, G. Krishna Mohan. (2021). Automated Bird Species Identification Using Neural Networks. Annals of the Romanian Society for Cell Biology, 25(6), 5402–5407. Retrieved from https://www.annalsofrscb.ro/index.php/journal/article/view/8556
[3] Lasseck, Mario. (2018). Audio-based Bird Species Identification with Deep Convolutional Neural Networks..
[4] Stastny, Jiri & Munk, Michal & Juranek, Lubos. (2018). Automatic bird species recognition based on birds vocalization. EURASIP Journal on Audio, Speech, and Music Processing. 2018. 10.1186/s13636-018-0143-7.
[5] Madhavi, A. Sita and Rajni Pamnani. “Deep Learning Based Audio Classifier for Bird Species.” (2018).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)