IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automated Bird Species Identification using Audio Signal Processing and Neural Networks

1 Avinash Tatar, 2 Bhushan Chavan, 3 Kashyap Bhamare, 4 Snehal Shirode, 5 Abhay Gaidhani

1, 2, 3, 4, 5 Department of Computer Engineering,

1, 2, 3, 4, 5 SITRC, Nashik, India.

Abstract: In this paper, techniques for identifying birds have been researched and an automatic system for recognizing their species has been devised. Significant study on taxonomy and other subfields of ornithology has been a challenging and difficult attempt for automatic identification of bird sounds without physical interaction. A two-stage identification approach is used in this work. The first step was creating an ideal dataset that included all of the bird species' sound recordings. The sound snippets were then put through a variety of sound pre-processing procedures, including pre-emphasis, framing, silence removal, and reconstruction. For each reconstructed sound clip, spectrograms were produced. In the subsequent stage, a neural network was set up and given the spectrograms as input. Using the input characteristics, the sound sample is categorized by Convolutional Neural Network (CNN), which also determines the type of bird. For the system mentioned above, a real-time implementation model was also created and put into practice.

Keywords - Bird species identification, bird sound, sound pre-processing techniques, Convolutional Neural Network, Spectrograms.

I. INTRODUCTION

Birds help us effectively identify different organisms in our climate (like the creepy reptiles they eat) by reacting quickly to ecological changes However, collating and collecting bird data is very labor intensive as it becomes a more expensive strategy. In this context, a solid framework is needed to provide a wide range of bird data processing and will serve as an important tool for experts, legislative offices, etc. Evidence of identifiable bird species therefore plays an important role in distinguishing a particular bird image from which species has a location.

The method of distinguishing and proving bird species uses an image to provide a location and type of bird species. Identifiable evidence must be obtainable by means of images, sounds or videos. Sound processors can be distinguished by capturing the vocal signals of birds. However, processing this data is made more complicated by the mix of sounds in the climate from insects, real-world objects, etc. Often people find that pictures are more successful than sound or recordings.

Automatic identification of bird calls from continuous recordings gathered from the environment would be significant addition to the research methodology in ornithology and biology, in general. Often these recordings are clipped or contain noise due to which reliable methods of automated techniques have to be used instead of manual conventional methods. Manual inspections of the spectrograms are often error prone and usually the techniques are esoteric in nature and involves multiple experts which makes it unreliable.

A technique has been proposed which involves the use of sound processing and convolutional neural networks to automate the entire process of bird sound identification. The first stage involves the creation of a database consisting of all the sound recordings. Subsequently, these recordings are subjected to sound pre-processing techniques like reemphasis, framing, silence removal and reconstruction. Spectrograms are generated for the sound clips and these spectrograms were given as input to a CNN which was trained on a GPU.

A real time implementation model was designed for the trained CNN model. A Graphical User Interface (GUI) was also designed for the system. An application can be designed in the future and deployed for mobile devices, which would enable users to use their smartphones as handheld devices for prediction and analysis of bird sounds.

II. PROBLEM STATEMENT

Identifying a bird can be a challenge, even for experienced birders. If you're new to using a field guide, it can be daunting to know where to start looking through hundreds of species pages. Birds can be classified based on characteristics such as size, shape, and color. Using CNN, we can classify birds.

III. PROPOSED METHODOLOGY

The methodology followed in the paper is summarized in the flowchart given in Fig.1.

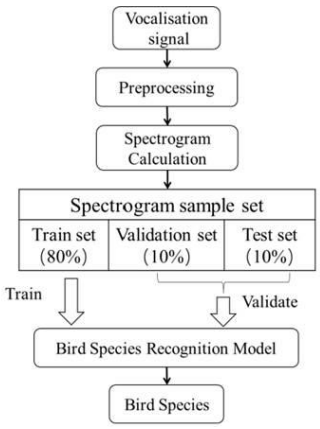

Fig.1 Flowchart of the Methodology

A. Data Acquisition

Dataset is of paramount importance and is a critical deciding factor for any machine learning based approach to a problem. A dataset should have certain crucial characteristics for it qualify as a good dataset. It should be accurate and precise, devoid of flawed elements and misleading information. It should be reliable and consistent, there shouldn't be contradictions in the dataset between different data elements irrespective of the source they have been collected from. The dataset must be complete, comprehensive and relevant, because fragmented data provides an inaccurate overall picture. Dataset should be granular and unique to prevent confusions which arise due to aggregated, summarized and manipulated data.

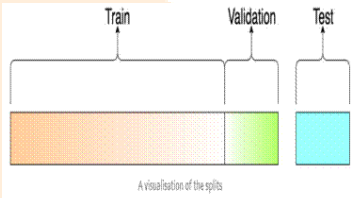

Fig.2 Division of the Dataset

B. *Framing and Silence removal and Reconstruction*

An audio signal is not a stationary signal. It consists of various statistical properties that can vary over the duration the time. It is necessary to first divide the recorded audio signal into a number of frames based on its length and then extract the signal which is devoid of any unwanted silence period. The frame length is decided based taking into consideration the total length of the signal and also the sampling period used. In this paper 2.5 % of the entire length of the audio clip is taken as the length of one single frame. After framing is completed, the silence removal is carried out using a thresholding function. The threshold function is chosen such that the audio signal above it is of our interest and the signal which falls below the threshold is considered as silence period or background noise. This silence removal is repeated for all the frames and dynamically changing the threshold value to 7% of the maximum amplitude present in that frame.

Reconstruction is the process of combining or concatenating all the frames that were obtained after the process of framing and silence removal. The result of this process is generation of a signal that is void of any noticeable silence period and still includes most of the information that is of our interest. The final step is to now procure the best sample in the pre-processed audio signal by considering 1 second of the clip that contains the highest amplitude in the total duration of the reconstructed signal.

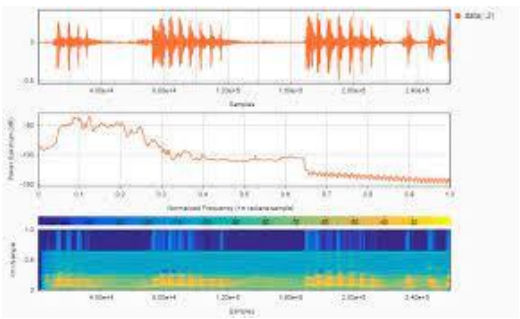

Fig.3 shows the result of reconstruction the signal after silence removal

C. Spectrogram generation

A spectrogram is a graphical representation of the range of frequencies and how they vary with time. It usually consists of time on x-axis, range of frequency on y-axis and the colour of the representation depicts the power/intensity of that particular frequency. A Spectrogram can be generated by first converting the signal in time domain to frequency domain using Fourier Transforms and then plotting the frequencies [4]. In this paper, the spectrogram is generated using an inbuilt MATLAB function. The spectrogram is generated only for the data that was obtained after reconstruction and not the entire signal. This process is repeated for all the audio-clips in the dataset and the respective spectrograms are stored in the labelled folders.

Each of the spectrograms are unique and have their own characteristics. For example, a spectrogram of sparrow's chirp will usually contain relatively low frequencies of high intensity whereas the spectrogram of a crow's chirp will have high intensities over a range of frequencies. These differences are easily picked up by a Neural Network when it is trained. In this paper, AlexNet is chosen as the Neural Network because of its high accuracy and easy implementation in the MATLAB. The spectrograms generated are scaled down to the resolution of 227 by 227 by 3 which can then be used to train the AlexNet Neural Network. Fig.4 shows an example of a spectrogram generated for an audio signal consisting of the chirps of a crow.

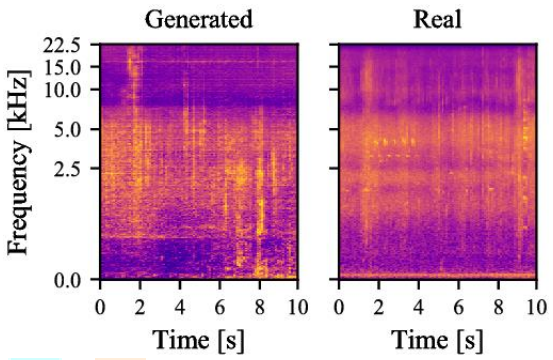

Fig.4 Example of a generated Spectrogram

IV. Real-Time Implementation

Now that we have trained a convolutional neural network (AlexNet), we can predict the bird species from an input audio recording. But these input audio recordings were collected in ideal environments where noise is almost non-existent, and in some cases the noise was removed by certain pre-processing techniques. Therefore, a network trained on a dataset can predict a given species when the input data is free from any disturbance or noise.

Unfortunately, this is not the case in live recordings due to ambient noise. Noise can be caused by many factors, such as vehicle noise, overlapping human voices, and natural phenomena. It is imperative to ensure that the network works as expected and performs at the same level as before in a simulated environment, rather than predicting bird species on a noiseless dataset. To ensure that the work then proceeds in real time, the CNN must be re-trained on a dataset containing the ideal dataset as well as sound samples taken from the surrounding environment.

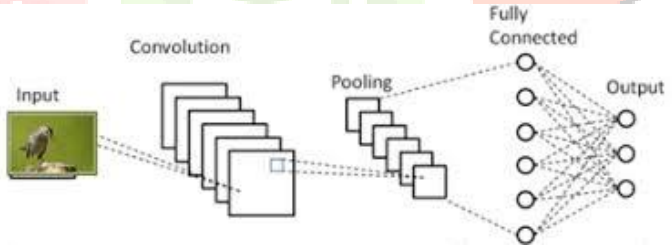

Figure 5 Real-Time Implementation

Audio clips of different birds in the dataset are randomly selected and down-converted to a fixed sample rate of 44100Hz or 48000Hz to maintain diversity and avoid overfitting. Bitrates are also set to 128kbps and 320kbps as these are the standard bitstreams used in audio applications as they ensure clear audio recordings with smaller file sizes. After converting all audio tracks to the desired sample rate and bitstream, a spectrogram will be generated for each audio track. These spectrograms are used to retrain the neural network. Once transfer learning is complete, the model can be saved and reused to classify audio signals in real time by converting them into spectrograms.

It is recommended to set the microphone to have a sample rate of 44100 Hz and a bit rate of 128 kbps or 320 kbps. The system was tested in a real-time environment and produced classification results with 91% accuracy. A Graphical User Interface (GUI) is designed to operate the above systems, involving all the processes to be performed, from recording sounds in a real-time environment to processing data and displaying results.

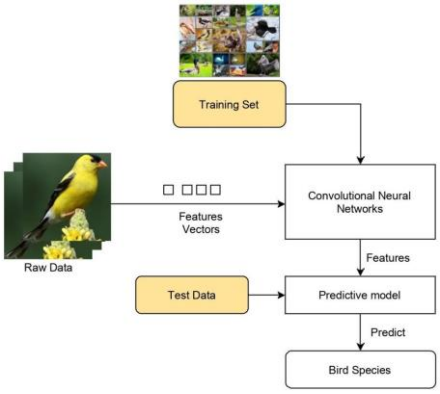

Figure 5 Real-Time Implementation

V. RESULT

Automated bird species identification using audio signal processing and neural networks is a promising area of research in the field of bioacoustics. By analyzing the audio recordings of bird sounds, it is possible to accurately identify the species of the bird and study their behavior and ecology.

The performance of automated bird species identification systems can be measured using various metrics, including accuracy, precision, recall, and F1-score. The results of these metrics depend on various factors such as the size and quality of the training data, the choice of features and neural network architecture, and the complexity of the bird species identification task.

Some studies have reported high accuracy rates in identifying bird species using audio signals, with reported accuracies ranging from 80% to over 90%. However, the accuracy may vary depending on the specific bird species, recording conditions, and the presence of background noise.

Number of species	Data Split	Epoch	Accuracy
2	80:20	20	92%
2	70:30	20	90%
4	80:20	20	88%
4	70:30	20	85.25%
4	80:20	35	97%
4	70:30	35	94%

Table 1 Training Results

VI. ADVANTAGES

1. Improved efficiency: Traditional manual methods of bird identification involve physically observing birds and identifying them by sight or sound. This can be time-consuming and labor-intensive, especially when monitoring large populations of birds. Automated systems can quickly process large amounts of audio data and identify bird species accurately and efficiently.
2. Increased accuracy: Bird species identification by sound can be challenging, even for experienced birders. Automated systems use advanced signal processing techniques and machine learning algorithms to analyze bird vocalizations and identify species with a high degree of accuracy.
3. Non-invasive monitoring: Automated bird species identification using audio signal processing and neural networks does not require capturing or handling birds, making it a non-invasive method of monitoring bird populations. This is particularly important for sensitive or endangered bird species, where minimizing disturbance is crucial.
4. Scalability: Automated systems can be scaled up to monitor large geographic areas or to process vast amounts of data collected over long periods. This makes it possible to monitor changes in bird populations over time and identify trends or patterns that may not be apparent through manual methods.

VII. LIMITATION

1. Limited accuracy: Despite significant advancements in the field, automated bird species identification using audio signal processing and neural networks still has some limitations in accuracy. The accuracy rates of these systems can vary depending on the specific dataset and the complexity of the bird vocalizations.
2. Limited dataset: One major challenge in developing accurate bird species identification systems is the limited availability of large and diverse datasets of bird vocalizations. This can limit the ability of these systems to generalize to new and unseen bird vocalizations.
3. Environmental noise: The presence of environmental noise, such as wind, traffic, and other bird vocalizations, can interfere with the accuracy of automated bird species identification systems. These systems need to be able to distinguish between the target bird species and other sources of environmental noise.
4. Variation in bird vocalizations: Bird vocalizations can vary significantly between individuals of the same species, as well as between different populations and regions. This can make it challenging to develop accurate and reliable bird species identification systems that can handle this variation.

VIII. CONCLUSION

In this paper, four different species of birds have been identified. The method involves pre-processing bird sounds and then generating their spectrograms, which are used to train a classification model. The data used for training consisted of real bird sounds recorded in their natural habitat, among all other sounds. Observe the output for different values of learning rate, number of epochs, and data distribution. The system was able to classify birds based on spectrogram images generated from bird sounds with 97% accuracy. This accuracy is achieved by taking into account human voices and bird calls. Accuracy can be further improved by fine-tuning the performance parameters.

IX. REFERENCE

[1] www.iucn.org/theme/species/our-work/birds

[2] www.xeno-canto.org/

[3] Sujoy Debnath, Partha Protim Roy , Amin Ahsan Ali, M Ashraful Amin" Identification of Bird Species from Their Singing", 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016

[4] Rong Sun, Yihenew Wondie Marye, Hua-An-Zhao "FFT Based Automatic Species Identification Improvement with 4- layer Neural Network", 2013 13th International Symposium on Communications and Information Technologies (ISCIT)

[5] www.mathworks.com/help/deeplearning/ref/alexnet.html

X. FUTURE SCOPE

This undertaking has an exponential scope for upgrades within the destiny in terms of economic as well as clinical opportunities. A software can be designed and deployed for mobile gadgets, which could enable users to apply their smartphones as hand-held devices for prediction and evaluation of chicken sounds. The fowl sound may be recorded by the consumer, the app then procedures the recording at the device and returns the result together with the photos, description and populace distribution of the chicken. The recording can also be despatched to a cloud server walking an advanced CNN for meticulous evaluation and evaluation to gain extra correct outcomes. The CNN can also be deployed on a hardware setup like a Neural Compute Stick or a Raspberry Pi. these hardware setups can be installed in ecological parks, conservation parks and chicken sanctuaries. The statistics obtained can both be saved domestically or at the cloud. The facts accordingly acquired can be of massive significance in research referring to hen migration patterns, population distribution, biodiversity and hen demographics in a given region.