IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11629

Bird Species Recognition Using Audio Signal processing and Convolutional Neural Network[1]

Hanumanthappa H [1],S M Mujeeb Ul Rehaman[2],Sandesh K C[3], Vinay R[4],Sagar Shapure[5]

Atria Institute of Technology, Computer Science & Engineering Department Bangalore 560024, India[1-5]

Abstract: In this research, a system for accurately identifying bird species was developed, and tactics for identifying them were studied. Automatic recognizing bird sounds with no physical interaction has proven to be a challenging and time-consuming task for significant research in ornithology's taxonomy and other subfields. Birds make a wide range of vocalisations, and different species of birds have distinct functions.
Manual annotation of each recording is used in current methods for processing big bioacoustic datasets. This necessitates specialised expertise and an inordinate amount of time. Recent advances in machine learning had made it easier to identify specific bird sounds for popular species with enough training data. However, developing such tools for rare and endangered species remains difficult.
This problem has been addressed in two stages : pre-processing and modelling (CNN model). The first step is to create a spectrogram from an audio input. The spectrograms were used as input in the second stage, which required establishing a neural network. Depending on the input properties, the Convolutional Neural Network identifies the sounds clip as well as separates the bird species.

Key-words: Bird Species Classification , Bird audio , method of pre-processing of bird sound, Convolutional Neural Network (CNN), Spectrogram

INTRODUCTION

As a result of technological improvements, a variety of audio recording equipment have grown more available, portable, and extensively used. The tracking devices were capable of capturing bird cries and songs in its native habitat in such an intrusive manner, with no need for direct interaction. As a normal pattern recognition problem, pre-processing, extraction of features, and categorization may all be used to bird species identification (Fagerlund, 2007). To our knowledge, the first publications on identification of bird species have been published within the nineties by the author (Anderson et., 1996). (Kogan & Margoliash, 1998). I improved upon those previous work aimed at performing bird species classification using spectrograms in this paper. Spectrograms have been proven to be effective in a variety of audio classification applications.
This time, we'll look at bird species classification in the context of a more difficult scenario involving a significantly larger number of classes. The tests were run on the same database that was used for the BirdCLEF identifying challenge. A total 40 species were used for each of the subgroups of the original data set.
In the future, a mobile device application could be developed that allows individuals to forecast and analyse bird sounds using their phones as handheld equipment.

DATA ACQUISITION

Every machine learning approach that tries to give problem-solving strategy demands a substantial quantity of data collection, which is a crucial decision factor. I used the kaggle dataset, which was gathered by Cornell Lab of Ornithology, in this paper.
The dataset contains audio signals from 40 different species, each with a sample space of 100 recordings, for a total of 10955 bird sound recordings samples.
Short recordings of individual bird sounds make up the majority of the training data; these files have been downsampled to 32 kHz to match the test set audio.

Data Pre-processing: The data collected from the source is in raw format and must be pre-processed and converted to a usable format before being fed into the Convolutional Neural Network (CNN).

1. At least 200 recordings are required for a species to be considered.
2. The recordings are then subjected to pre-processing techniques such as alteration, composing, and silence removal before reconstruction.
3. Using the librosa library, create a spectrogram from an audio file.

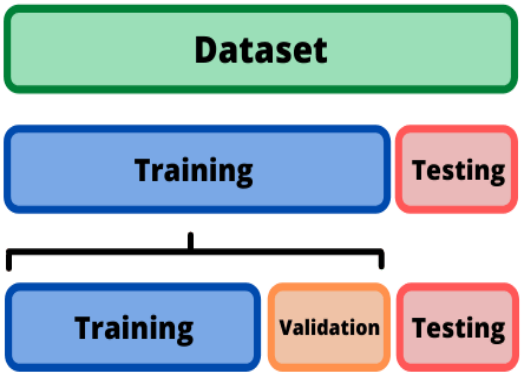
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11629

4. Divide the database into train-validation-test sets in the given ratio of 70:10:20.



Dataset challenges

• Audio recordings of the very same bird breed were obtained from several birds throughout the world;
• bird sounds were captured with the help of several individuals, which might or might not had shared the very same microphones and recording equipment.
•Voice signal had been collected from samples taken at different times of the year, and other sounds from the neighbourhood .
• The presence of a silent period with in audio transmission during where no bird sound can be heard.

METHODOLOGY

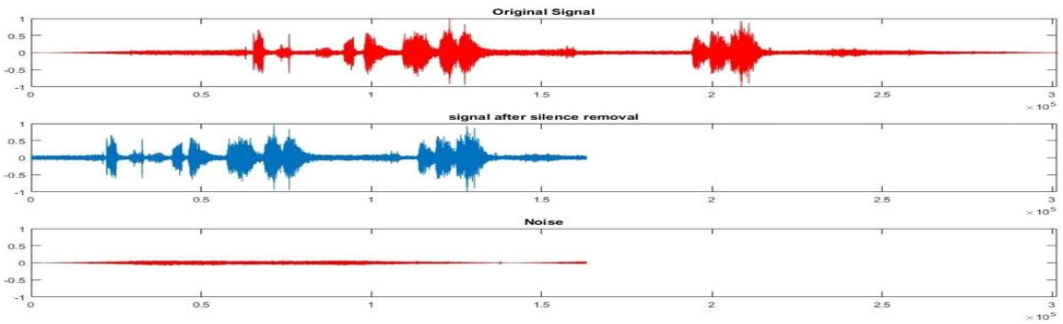
Pre-Emphasis

A wide range of frequencies can be found in a microphone-recorded audio stream. High frequencies energy is available in speech signals and bird chirps. In other words, pre-emphasis emphasizes or amplify component of high-frequency. As a result, it's vital to put more attention on the higher frequency energy that we care about while simultaneously diminishing other frequencies. A simple 1st order high band - pass filter is used to achieve this.

y[n] = x[n] - S*x[n-1] is the mathematical equation for this filter, which passes all of the data points in the signal. The value of S can be modified to change the level of attention required. In this study, the value of S is taken to be 0.95.

Silence removal and framing

A fixed signal can not be considered as an audio signal.They are made up of a variety of statistical characteristics that can change over time. While collecting the signal that is devoid of any undesired quiet interval,The duration of the recording audio signal must be separated together into number of frames.The frame length is determined by the total length of the signal and also the sampling time employed. One single frame represents 2.05 percent of the overall runtime of the voice recording in this work.Only after frame is complete, the silence is removed using a thresholding method.
The cutoff point function is set so that the sound wave above it is of relevant, but the audio wave below it is overlooked as background noise . The silence is removed in all frame, with the thresholding dynamically moving to 7 percent of the frame's peak amplitude.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11629

Reconstruction

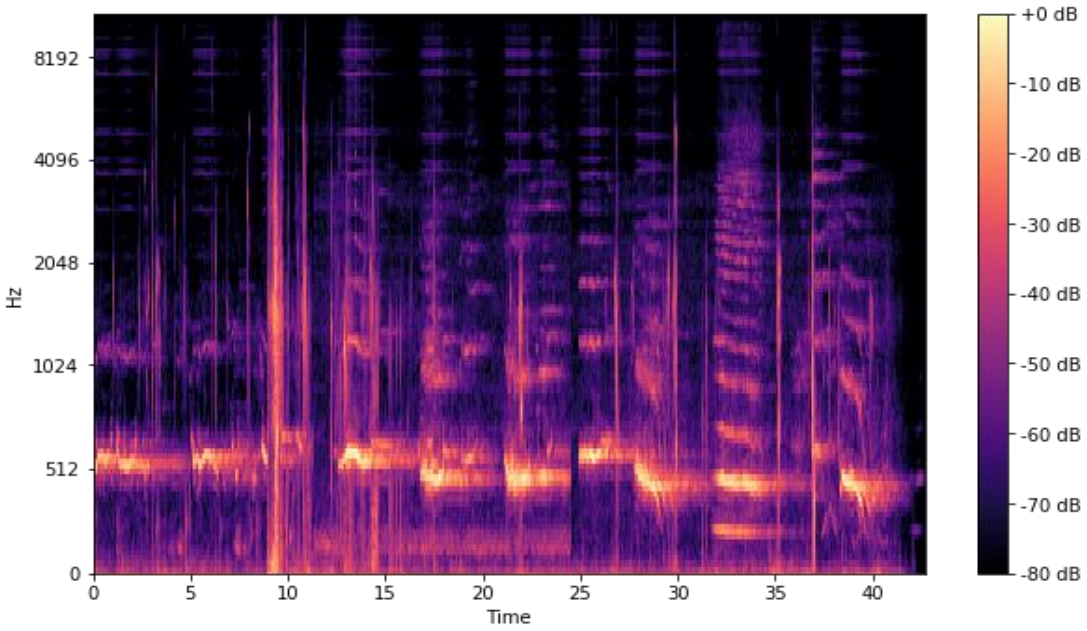
Reconstruction is the act of integrating or attaching all of the frames received after the silence removal operations. As an outcome of all this procedure, we also have a signals with no apparent silence intervals while still retaining the majority of the data we want. Finally, choosing the perfect sample with in pre-processed voice signal requires around a second of the clip with the highest amplitude throughout the whole length of the reconstructed signal.

Spectrogram generation

Sound visualisation is a strange idea. There are several fascinating methods for doing so, as well as some that are more mathematical. The Mel Scale is indeed the result of a non-linear modification of the scale based in mathematics. The Mel Scale is established such that sounds on the Mel that are appropriate distance apart "sound" the same though to humans. The gap around (500 to 1000) Hz is easily audible on the Hz scale, yet the difference among (7,500 & 8,000) Hz is scarcely perceptible.

It divides the Hz scale in to the bin and transforms every bin into a comparable bins in the Mel Scale using overlapped triangle filters.

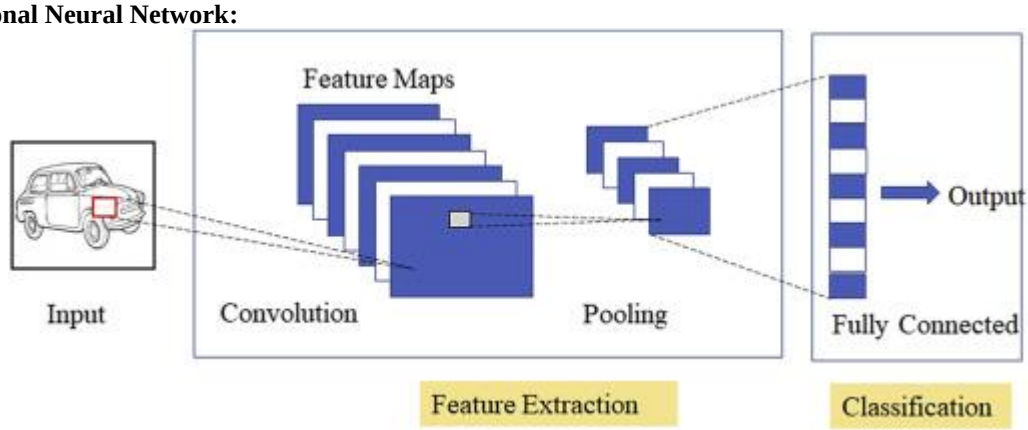
The Mel Spectrogram, unsurprisingly, can be said to be a spectogram where y-axis is Mel scale itself.



Modeling (Convolutional Neural Network)

The data has been pre-processed and translated to an image format using spectrogram, and now it's time to build a CNN model that will consume the spectrogram.

Convolutional Neural Network:



This is a type of image classification and required to process artificial neural network designed exclusively for pixels data processing.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11629

I used F1-score to evaluate the model's performance because it's commonly utilised when dealing with unbalanced datasets.
The F1- score takes the natural logarithm of a classifier's accuracy and recalls to generate a unique statistic. It's primarily used to evaluate 2 separate classifications' output.

$$F_1\text{-}score$$

$$\frac{1}{\frac{1}{2}\left(\frac{1}{\text{recall}}+\frac{1}{\text{precision}}\right)}$$

I trained two distinct CNN models and compared their F1-score performance:

Model 1 parameter:
5. Consists of Five convolution block , *each block has the sequence CONV --> RELU --> BNORM --> MAXPOOL.*
6. Using a global average pooling layer, a dense layer, and a classifier layer with softmax function, we converted a 2D kernel to a 1D vector.
7. *It has 414760 trainable parameters*

Model 2 parameter:
Everything is the same as model 1, except we only employed four layers of convolution blocks, which raises the model's parameter to 119784.
Compile the model with metrics, loss, and optimizer before training.

Parameters
Metrics : Accuracy, F1-score
Loss : Categorical_Crossentropy (Used for *multi-label task)*
Batch Size : 32
Validation Split : 20%
EPOCHS : 50

Callbacks used:
8. **ReduceLROnPlateau**: When learning becomes stagnant, reduces the learning rate by a factor.
9. **EarlyStopping**:This method starts with developing an unbounded large set of testing epochs and then end testing only when the model effectiveness is on a hold, validating database plateaus.
10. **ModelCheckpoint**: It can be utilized in combination of modelfit() training to get another models or weight at one predefined interval so that they may be downloaded later to restart training.

RESULT

The Model 1 that was used in this investigation proved to be the most successful. However, both the model and the training set were overfitted, resulting in a substantial disparity between validation and train accuracy.

Model 1:
Train-set Accuracy : 86.28 Validation-set Accuracy : 97.66
Train-set Loss : 0.24 Validation-set Loss : 1.31

Model 2:
Train-set Accuracy : 86.28 Validation-set Accuracy : 73.30
Train-set Loss : 0.60 Validation-set Loss : 1.32

According to the given statistics, model 1 outperformed model 2, however overfitting must be avoided by implementing an L1 or L2 regularisation in each layer of the convolution block; otherwise, the unbiased performance will suffer (test-set).

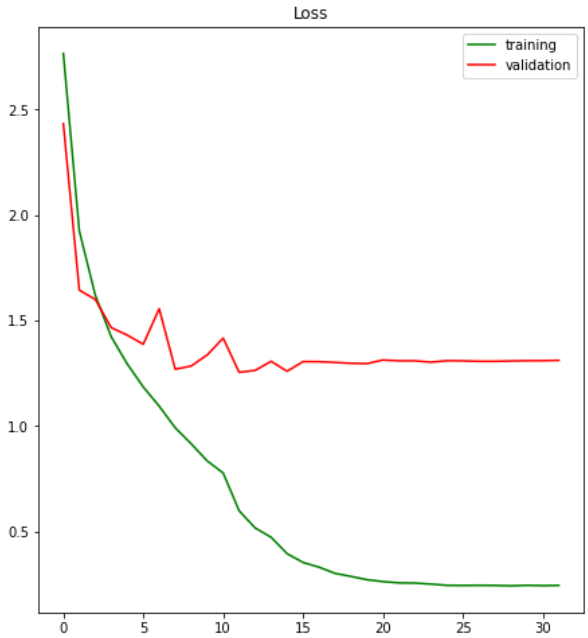
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

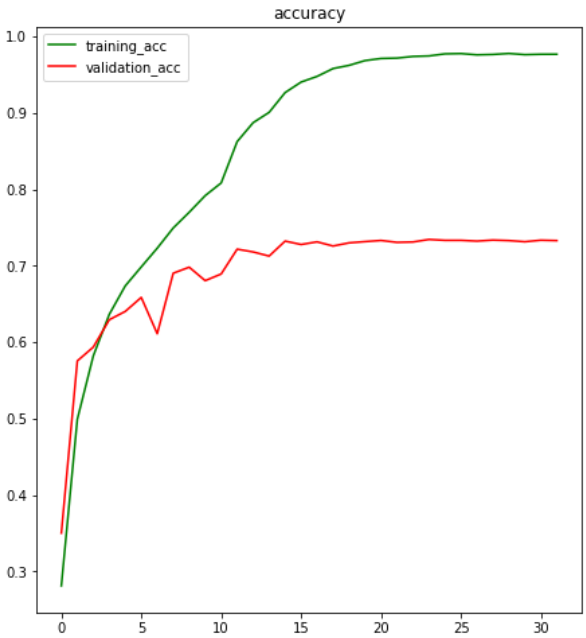
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

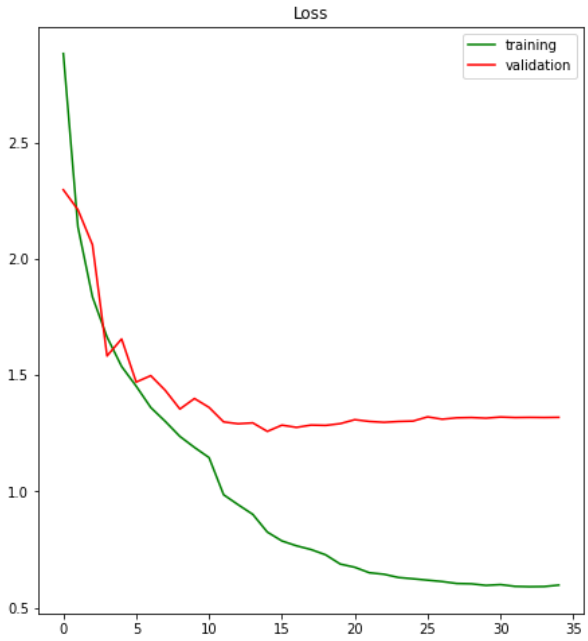
DOI: 10.17148/IJARCCE.2022.11629

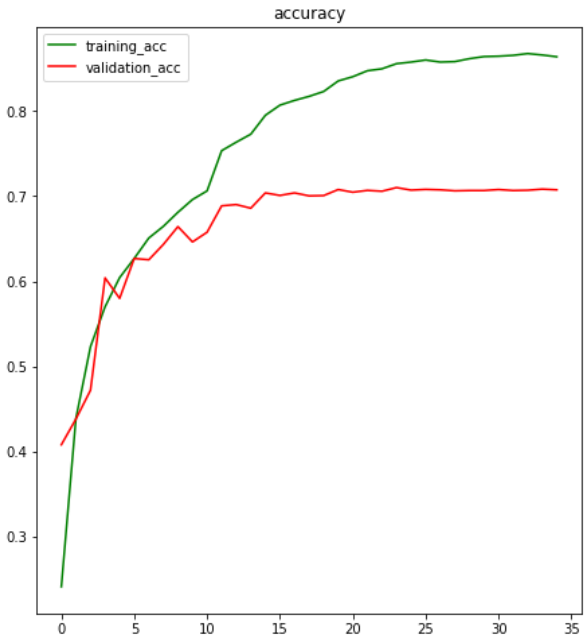
Model 1 performance





Model 2 performance :





FUTURE SCOPE

The work does have an enormous prospect for future development in economic and scientific improvements. It is possible to create and deploy a mobile application that allows people to foresee and analyse bird noises using their cellphones as portable equipment. The user can capture the sound of the bird, which is then processed by the software on the smartphone and delivered along with images, a description, and the bird's population distribution. The video may be transmitted to a cloud service with a powerful Cnn for further processing and evaluation, resulting in high precise results.

The CNN may also be operated on a hardware platform like a Neural Computing Stick or a Raspberry Pi . Ecology park, wildlife parks, and sanctuaries can benefit from these hardware installations. The information obtained can be kept locally or on the cloud. The information gathered will be invaluable in research of bird migration routes, demographic structure, diversity, and demographic in a specific region.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11629

CONCLUSION

This effort resulted in the identification of four different bird species. The tests included 40 species and 10955 training samples from 8 distinct subgroups of said BirdCLEF identification project challenge (audio files). The spectrogram has been selected also as input for obtaining the characteristic since this has been successfully employed in many previous audio classification projects. The findings make us believe that the suggested approach is one of the best ever devised, despite the fact that the values obtained can really be not directly matched to all other outcomes because of subgroups used in this study were evaluated for the very first time. We achieved a 97 percent identification rate in the most difficult scenario assessed, which included 40 classes. Fine-tuning and adding a regularization/dropout layer to the prototype can enhance its efficiency on the validating and testing dataset.

REFERENCES

- Bird species recognition by comparing the HMMs of the syllables. In: Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on, IEEE, pp. 143. https://doi.org/10.1109/ICICIC. 2007.199.
- Multiresolution gray-scale and ro-tation invariant texture classification with local binary patterns, Pattern Analysis and Machine Intelligence, IEEE Transactions on 24 (7) (2002) 971–987.
- Automatic segmentation of audio signal in bird species identification. In: Computer Science Society (SCCC), 2016 35th International Conference of the Chilean, IEEE, pp. 1–11.
- Use of spectrogram to classify North Atlantic right whale voice from audio recordings, in: Computer Science Society In: Computer Science Society (SCCC), 2016 35th International Conference of the Chilean, IEEE, 2016, pp. 1–6.
- Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP J. Appl. Signal Process. 2007 (1), 64