AI based Birds Sound Classification Using Convolutional Neural Networks

Anand R¹, Shanthi T¹, Dinesh C¹, Karthikeyan S¹, Gowtham M¹, Veni S²

¹Department of ECE, Sona College of Technology, Salem.
²Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore.

E-mail: anand.r@sonatech.ac.in

Abstract. To classify and recognize the sounds which are produced by the birds which are used for the identification of species of the bird. The identification of bird species in captured audio files will be a transformational method for scholars, wildlife biologists and birders. In latest years, artificial neural networks have dramatically increased the detection efficiency of bird species recognition using machine learning systems. There is a lot of study in audio recognition using machine learning. This work may help for the easy identification of birds living in a locality and studying of birds’ migration.

1. Introduction

Birds are an important ecological species which are important in maintaining biodiversity. Most of the birds are vertebrate animals adapted for flight. Many can also run, swim, jump and dive. Some bird species like penguins, have lost their ability to fly but they have retained their wings. Birds are found worldwide and in all habitats. The largest is the ostrich and the smallest is the hummingbird. Everything about the anatomy of a bird reflects its ability to fly. Also, birds have their own unique sounds which are only meant to birds of same species. Birds are identified by their color of the feathers, sounds they make, body size and shapes. In our work we make use of the bird sounds for their identification. Every bird has their own sounds. Some bird sounds may consist of one or two-part length sounds and some bird’s sounds are three or more-part length sounds. Hence it makes the identification of bird species based on their sounds difficult. We make use of the uprising branch of technology i.e. Science-based artificial intelligence and deep learning algorithms for our categorization. Analytical AI has only similar features of cognitive intelligence; creating a cognitive image of the environment and utilizing information, based on knowledge, to guide possible decisions. Artificial intelligence that incorporates components of social and emotional AI, coupled with cognitive elements and considering them in their decision making. Humanized artificial intelligence displays both logical, mental, and social competencies, may be self-aware, and able to be conscious [1]. Artificial intelligence was established as an empirical field in 1956 and since then has undergone numerous periods of hope, accompanied by dissatisfaction and lack of funding (known as a “AI winter”), followed by new methods, growth and revived funding. While AI research has had its share of divisions, there has been opposition to interacting with each other. The subfields are focused on technological factors including the need for precise targets and specific uses. New methods or a new analysis, or simply new conceptual discrepancies. Subfields are often affected by socioeconomic and institutional influences (institutions or the work of re-searchers). AI [2],[3],[4] study focuses on classical and conventional problems (or goals), including thinking, information representation, preparation, learning, natural language processing, cognition, and the capacity to transfer and control artefacts. Intelligence enhancement is among the field’s long-term
priorities. Approaches include mathematical approaches, computational intelligence, and rule-based structures. Many methods used in AI involve programming languages and algorithms, neural networks, data analysis, economics, statistics [5],[6] and chance. AI encompasses several diverse disciplines from computer technology, software processing, mathematics, psychology, linguistics, and philosophy. The area was built on the premise that human intellect may be completely quantified and therefore scientifically defined. This poses crucial metaphysical questions concerning the existence of the soul, our handling of other people and intelligent robotics [7,8]. These topics have been discussed in a multitude of history, prose, and theory since antiquity. Some still think that AI could contribute to dangers for humanity if left unchecked. Some claim that artificial intelligence would displace human labor, contributing to mass unemployment. In recent decades, artificial intelligence techniques have undergone a revival following parallel advancements in computational capacity, massive volumes of data, and scientific understanding; and AI techniques have become an integral part of the technology industry. Artificial intelligence is the computer software that simulates human intelligence processes [8,9]. Weak AI, such as Siri, are a form of AI. Strong AI or artificial general intelligence is some type of AI that outperforms a person. When assigned a challenge that is new, a good AI framework will solve the problem on its own. The identification of bird’s species is mainly based on the sounds they produce. Some birds may look like each other, but they belong to different species. We can identify those types of birds only by the sounds they produce. Manually identifying all those birds are a tough job for the ornithologists. But the development of artificial intelligence and machine learning algorithms made the classification of audio sound easier. A bird sound consists of two part or more than two-part call sounds. We can make use the sounds produced by the birds for the classification of the bird species [10]. This work is about the identification of the bird species by the bird’s sound file given. We use the deep learning algorithm convolutional neural networks for the training of model and classification. This method can also be used for the classification of other animal species by training the model with the respective training model. Birds’ sounds are very important for the classification of the bird species, since they are unique to each species. It can used for the very precise identification of the bird species in real time environmental conditions. In this work the sounds recognition system is implemented using convolutional neural networks.

2. Literature Review

The identification of the bird sounds as such cannot be done. So, the bird sounds are converted into their respective Spectrograms. A Spectrogram is a graphic description of the wavelengths of a signal as it changes over time, or the numerical significance of the frequency. The spectrogram is used in music, geology, radar, sonar, and meteorology. Sound spectrograms may be used to generate audio that records the different noises produced by animals. It can be generated through techniques such as light spectroscopy, band-pass filters, Fourier transforms, or wavelet transformations.

2.1. Random forest classifier:[4]

A Random Forest (RF) classifier with an ECC classifier as the basis were used in the proposed classifier. The ensemble is an example of a machine learning algorithm named the Ensemble of Probabilistic Classifier Chains (EPCC) (Dembczynski et al., 2010). Therefore, odds can be aggregated from the SISL classifiers rather than a
vote of 0 or 1. The total odds are merged into one predictor. Algorithm 2 helps one to create class-score vectors using the ECC-RF algorithm efficiently.

2.2. Deep neural network: [9]
This suggested baseline structure is based upon eight weighted layers, no bottleneck links, and five layers with a shortcut relation. The new net has not yet completely converged and does not have tightly linked layers at the top. They have an implementation of every phase of the workflow with customized formats to render improvements simple. The model's implementation is based on Theano (Theano's Production Team, 2016) and Lasagne (Dieleman et al., 2015).

2.3. Support vector machines: [11]
The approach used in this machine learning research is the LIBSVM classifier which is applied using the one-against-one decomposition technique. The Gaussian radial base function is the kernel function (RBF). The classifier is trained using a five-fold cross-validation as part of the supervised quest to find the pair of regularization factor, C and RBF parameter, γ. The points of light in the picture taken for cross-validation.

2.4. Convolutional neural networks: [12]
In this method they decided to reduce the amount of the production team took into consideration existing best trends in CNN formats. Both layers (except the input and output layers) are initialised using the He initialization scheme and are a part of the Batch Normalization scheme. Our first convolutional layers were planned to have broad receptive fields, and this proved very successful for spectrograms during the experiments.

2.5. Hidden Markov model: [13]
HMM offers a basic and efficient framework for modeling time-varying spectral vector sequences. Any system designed to work reliably as a real-world application must be robust because of the modulations in humans and another species of the ecosystem. The goal of our work is to ease the way of identification of bird species using their sounds. We give our input as the sound file and we get the name of the bird.

Table 1: Comparison of different types of methods of classifiers.

| Reference paper | Method used | Performance of the classifier | Drawbacks |
|-----------------|-------------|-------------------------------|-----------|
| Evaluation of a Sparse Representation-Based Classifier for Bird Phrase Classification Under Limited Data Conditions.[11] | Support vector machine | Better performance when wide range of conditions | There are only four tokens per phrase in the training set. |
| Large-Scale Bird Sound Classification using Convolutional Neural Networks. [12] | CNN | Obtained a mean average precision of 0.645. | Less improvements for the sound scape domain. |
| Multi-Label Classifier Chains for Bird Sound. [4] | Random forest classifier | Poor performance of the model. | Worse in others compared to the MIML algorithms. |
| Our practice of using machine learning to recognize species by voice. [13] | Hidden Markov model | Highest accuracy obtained is 90% | The Viterbi algorithm is expensive. |
| Recognizing Birds from Sound The 2018 Bird CLEF Baseline System. [9] | Deep convolution | Obtained Mean Label Ranking Average | limited when compared to the scope of real- |
species as the output. In this work we used the classification of sounds for classifying five different types of bird species. We have used the extracted spectrograms of five species and trained the model. The 2018 Existence CLEF Bird Identification challenge (Joly et al., 2017) (Goeau et al., 2017) utilises two primary source recording methods. The training dataset exhibits 36,493 monophonic records for 1500 birds of South America.

Figure 1. Block Diagram for Bird Sound Classification

3. Methodology:

The first step involved is extracting of the spectrograms from the given wav file inputs are shown in figure 1. Then they are separated for the train and test datasets. The main thing involved after that is to use the best suited methodology for our classification. A typical architecture of a convolutional neural network contains an input layer, some convolutional layers, some fully connected layers, and an output layer. It has several layers with considering input and output. We can see the CNN architecture used in this work in Data modelling section.

3.1. Spectrogram extraction

Spectrograms are extracted for the given wav input files. Bird sounds consist of various foreground and background noise signals. Elimination of such noise is necessary in our work. So, we need to use the best suited filters for removing such noises. Also, many ways are there for the spectrogram extraction. Librosa is an inbuilt python package for audio and music signal processing. Spectrogram operations include the short-time Fourier transform (STFT), inverse STFT (ISTFT), and instantaneous frequency spectrogram. Using librosa the spectrograms are obtained which is shown in figure 2.

Fig 2. Spectrogram Extraction

CNN is designed with some modification on NaK Architecture. It has 6 layers without considering input and output. The architecture of the Convolution Neural Network [15,16] used in the work is shown in the following figure 3.

Fig 3. Proposed CNN Architecture for bird’s sound classification

4. Results and Discussion

Training of the dataset is crucial. We have used the 7:3 ratio of the dataset is used for training of the model and testing of the model, respectively. We have used only with a smaller number of class samples of bird
species. We searched at ways to render our preparation phase more effective but at the same time maintaining our success at an optimal stage. We analyzed the efficacy of various kinds of parameter configurations and selected the solutions that had the most impact. One of the most significant parameters when practicing Convnets is the learning pace. We used a learned model to evaluate 10 spectrogram images for each animal. We have used 80% of the data for preparation and the other 20% for research. We were able to get results from the 5 bird classification datasets with an accuracy of 90%, as is seen in figure 4. Even since species is a more precise classification than order or family, species classification may be used to classify birds into order or family for a multitask classification and output measure are seen in table 2. Figure 5 shows the uncertainty matrix of the classification findings, which display the recognition accuracy for each group.

Table 2: Performance Measure for AlexNet and ResNET

| Performance measure | AlexNet | ResNET |
|---------------------|---------|--------|
| Accuracy            | 84%     | 90%    |
| F-score             | 86.05%  | 91.28% |
| Recall              | 98.74%  | 98.76% |
| Sensitivity         | 83%     | 87.73% |
| Specificity         | 94.38%  | 92%    |

6. Conclusion

Birds species identification can be easily done with the help of the complex machine learning algorithms. It helps for the identification of the birds in any locality or environment. It helps the ornithologists for studying the migration of birds into any locality at different seasons. This bird sound classifier can also be made into simple mobile application which can help the people for knowing the species of the birds around them by simply providing the bird sound as input to the mobile application and they obtain the bird’s name as output.

REFERENCES:

[1] Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus).
[2] D. Stowell and M. D. Plumbley, (2014) “Audio-only bird classification using unsupervised feature learning,” Working Notes of CLEF,
[3] E. C¸akir, S. Adavanne, G. Parascandolo, K. Drossos, and T. Virtanen, (2017) “Convolutional recurrent neural networks for bird audio detection,” in Proceedings of the 25th European Signal Processing Conference (EUSIPCO). IEEE,

[4] Forrest Briggs, Xiaoli Z. Fern, Jed Irvine “Multi-Label Classifier Chains for Bird Sound”.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[6] Huang, G., Liu, Z., Weinberger, K. Q., & van der Maaten, L. (2016). Densely connected convolutional networks.

[7] J. Allen, (2009) “Short term spectral analysis, synthesis, and modification by discrete fourier transform,” IEEE Transactions on Acoustics, Speech, and Signal Processing.

[8] Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W., Lombardo, J., Planquè, R., Palazzo, S., & Müller, H. (2017). LifeCLEF Lab Overview: multimedia species identification challenges. In Proceedings of CLEF.

[9] K. J. Piczak, (2016) “Recognizing bird species in audio recordings using deep convolutional neural networks,” Working notes of CLEF.

[10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems.

[11] Lee Ngee Tan1, Kantapon Kaewtip1, Martin L. Cody2, Charles E. Taylor2, Abeer Alwan. “Evaluation of a Sparse Representation-Based Classifier For Bird Phrase Classification Under Limited Data Conditions”

[12] S. Kahl, T. Wilhelm-Stein, H. Hussein, H. Klinck, D. Kowerko, M. Ritter, and M. Eibl,(2017) “Large-scale bird sound classification using convolutional neural networks,” Working notes of CLEF,

[13] Sprengel, E., Martin Jaggi, Y. K., & Hofmann, T. (2016). Audio based bird species identification using deep learning techniques. Working notes of CLEF.

[14] Shanthi, T., Sabeenian, R.S. And Anand, R., 2020. Automatic Diagnosis Of Skin Diseases Using Convolution Neural Network. Microprocessors And Microsystems, 76, P.103074.

[15] Anand R, Shanthi T, Nithish Ms, Lakshman S. Face Recognition And Classification Using Googlenet Architecture. InSoft Computing For Problem Solving 2020 (Pp. 261-269). Springer, Singapore.

[16] R. Anand, S. Veni, P. Geetha, S. Rama Subramoniam, Extended morphological profiles analysis of airborne hyperspectral image classification using machine learning algorithms, International Journal of Intelligent Networks, Volume 2, 2021, Pages 1-6, ISSN 2666-6030 https://doi.org/10.1016/j.ijin.2020.12.006.