A Progress Synopsis

On

BIRD CALL RECOGNITION

carried out as part of the course: AI2270

Submitted by

Harsh Goyal

219311227

IV-AIML

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

In

Computer Science and Engineering(AIML)

Department of Artificial Intelligence and Machine Learning
School of Computer Science & Engineering
Manipal University Jaipur,
March 2022

BIRD CALL RECOGNITION

DESCRIPTION - Bird Sound Recognition using Deep Learning and Exploratory Data Analysis (EDA)

PROJECT OVERVIEW - This project aims to develop an efficient and accurate bird sound recognition system using advanced deep learning techniques, specifically artificial neural network (ANN), in conjunction with exploratory data analysis (EDA). Bird sound recognition has significant applications in ecology, ornithology, and biodiversity monitoring. By leveraging state-of-the-art technology, we seek to automate the identification of various bird species based on their unique vocalizations.

ROADMAP FOR THE IMPLIMENTATION

1. Data Collection and Preparation:

Gather a comprehensive dataset of bird sound recordings, encompassing diverse species and environmental conditions. Annotate the dataset with bird species labels.

Employ exploratory data analysis (EDA) to understand the dataset's characteristics, addressing data quality, class imbalances, and noise issues.

Sources: Kaggle, For bird call - https://xeno-canto.org/

2. Feature Extraction and Augmentation:

Transform audio data into suitable formats for deep learning, including Mel-frequency cepstral coefficients (MFCCs), spectrograms, and raw waveforms.

Implement data augmentation techniques to increase dataset diversity.

3. Neural Network Architecture Design:

Create a custom neural network architecture, integrating network layers for feature extraction and recurrent layers for capturing temporal patterns in the audio. Here, we're implementing artificial neural network (ANN).

Experiment with various architectures, hyperparameters, and optimization techniques to optimize model performance.

4. Dataset Description:

Train audio – The train data consists of short recordings of individual bird calls generously uploaded by the users of https://xeno-canto.org/

Test audio - The test_audio directory contains approximately 150 recordings in mp3 format, each roughly 10 minutes long. They will not all fit in a notebook's memory at the same time. They were divided into sited in 5 second increments and need matching predictions.

5. Model Training and Validation:

- **Data Feeding:** Feed the training data into the model. This involves passing audio features (e.g., MFCCs) through the network's input layer. The labels (bird species or sound categories) should also be associated with each input sample.
- **Loss Function Selection:** Choose a suitable loss function for your classification task. For multi-class classification, categorical cross-entropy is commonly used. The loss function quantifies the difference between the predicted outputs and the actual labels.

- **Model Compilation:** Compile the ANN model by specifying the optimization algorithm (e.g., Adam, SGD), the loss function, and any evaluation metrics you want to monitor during training (e.g., accuracy).
- **Hyperparameter Tuning:** Set hyperparameters such as the learning rate, batch size, and the number of training epochs. These parameters can significantly impact training outcomes. Hyperparameter tuning may require experimentation and validation set monitoring.
- **Training Loop:** Begin training the model by feeding batches of data into the network. The training loop involves the following steps:
    1. *Forward Pass*: The input data is passed through the network, resulting in predictions.
    2. *Loss Computation*: The difference between the predictions and actual labels is calculated using the chosen loss function.
    3. *Backpropagation*: The gradients of the loss with respect to the model's parameters are computed.
    4. *Weight Update*: The model's parameters (weights and biases) are updated using the optimization algorithm. This process aims to minimize the loss.

6. Testing and Evaluation:

Assess the trained model's real-world performance by evaluating its accuracy, precision, recall on a separate test dataset.

Implement techniques for model interpretability to understand its decision-making process.

7. Legal and Ethical Considerations:

Ensure compliance with legal and ethical considerations, including data usage rights and privacy concerns.

8. Documentation and Reporting:

Thoroughly document the project, including model architecture, hyperparameters, and evaluation results.

Share the findings and methodology to contribute to the field of bird sound recognition and deep learning applications.

This project endeavours to automate bird sound recognition, benefiting fields such as ornithology, wildlife conservation, and ecological research. By combining deep learning techniques with exploratory data analysis, it seeks to provide a robust and accurate solution for identifying bird species based on their vocalizations, ultimately aiding in the understanding and preservation of avian biodiversity.

9.Advantages:

1-High Accuracy: Machine learning models can achieve a high level of accuracy in identifying bird species based on their vocalizations, often surpassing human capabilities.

2-Continuous Monitoring: Automated bird sound recognition systems can continuously monitor audio recordings, allowing for long-term ecological studies without human intervention.

3-Efficiency: These systems process large datasets quickly, making them more efficient than manual bird sound analysis, which can be time-consuming and labor-intensive.

4-Non-intrusive: Bird sound recognition is a non-invasive method that doesn't disrupt bird behavior, making it an ethical choice for ecological research.

5-Remote Monitoring: Bird sound recognition technology enables remote monitoring in challenging or inaccessible environments, such as dense forests or remote wetlands.

10.Limitations:

There are many problems you can encounter:

1-background noise — especially while using data recorded in a city (e.g. city noises, churches, cars)

2-multi-label classification problem — when there are many species singing at the same time

different types of bird songs (as described earlier)

3-inter-species variance — there might be a difference in birdsong between the same species living in different regions or countries

4-data set issues — the data can be highly imbalanced due to bigger popularity of one species over another, there is a large number of different species and recordings can have different length, quality of recordings (volume, cleanliness)