Bird Sound Recognition for Crop and Farm Surveillance Applications (Spectrogram, 2DCNN)

Mausumi Bhuyan
dept. name of organization (of Aff.): Electrical Engineering
name of organization (of Aff.): Indian Institute Of Technology Palakkad
City: Palakkad, Country: India
email: 122201001@smail.iitpkd.ac.in

***Abstract*—This project uses the method of 2DCNN(two-dimensional Convolutional Neural Network system which is developed for bird sound recognition. Designing the 2DCNN architecture then the model is trained and compiled with the bird sound database from various internet sources like Xeno canto, extracting audios from Youtube and other websites. The 2DCNN model achieves reasonable accuracy of the given database of the given species i.e Crow, Sparrow, Pigeon, Parrot.**

***Index Terms*—Two-dimensional Convolutional neural network; bird sound recognition; Spectrogram feature-extraction.**

I. Introduction

A. Significance Of the Project(Need of the project) with applications

With this project it will help farmers to effectively manage the crops and safeguard their crops. It can contribute to the ecosystem services such as poillination and pest control, ultimately benefitting crop productivity

B. Survey of Existing methods and technologies

Current methods involve closely watching the birds in the field which can be time consuming and hectic. More advanced techniques like machine learning algorithms and deep learning models can detect the bird based on sound patterns. However, these methods require extensive training data and resources, limiting their practical applicability.

C. Problems and Statements

Detecting the right birds on a farm and fields can be tough since there are too many types of birds with all sorts of sounds, calls and chirps. Telling them apart can be tough and confusing and the farms will be full of sounds of tractors, wind and animals it will be hard to hear the bird sounds. So existing methods often struggle to detect accurately between different bird species in environments and in the surveillance systems and also in remote areas. Addressing these problems requires the development of robust deep-learning methods and machine-learning algorithms.

D. Motivation

This project drives into the world of deep learning and machine learning algorithms and generating sound patterns. The sound patterns help to decode the birds and process the diverse types of bird data.

E. Major Objectives with Work Plan

1) **Data Collection**: Audio recordings of bird sounds are downloaded from various internet sources like Xeno canto, extracting audios from Youtube and other websites
2) **Preprocessing**:The downloaded audio data undergo preprocessing steps which include noise reduction and resampling the audio files.
3) **Feature Extraction**: Computing the FFT(Fast Fourier Transform) and STFT(Short-Time Fourier Transfrom) and then obtaining the spectrogram and capture the frequency content and sound patterns of the bird audio files.
4) **Model Training**: designing The 2DCNN model using Keras Sequential API by convolutional layers, max-pooling layers, dropout layers and dense layers. and compile with the Adam optimizer and categorical cross entropy.Train the model using the fit method, specifying training data, validation data, batch size and callbacks.
5) **Evaluation**: Evaluatde the accuracy of the 2DCNN model using the evaluate function and makes the predictions test data using the predict function.
6) **Deployment** : Once the model achieves accurate performance, the system integrates the GUI(Graphical User Interface), allowing the user to input audio recordings of bird sounds and get predictions.

II. Materials and Methods

A. System Architecture with Description

The system architecture of the Bird Sound Recognition project includes like data collection, preprocessing, feature extraction, model training, evaluation, and deployment.

1) Data Collection and Preprocessing:

- **Data Collection**: Collecting the bird sounds from the various internet sources and convert into WAV files in

various websites like Freeconvert,cloudconvert and convertio, and delete silent audio files and last organize data into directories by bird species.
- **Metadata Extraction**: Start by gathering details about each bird sound wav file, like its name and the type of bird it might be. We get this info from the file names.
- **DataFrame Creation**: Once we have this bird info, we put it in all together in a neat and organized table format. This makes it easier to work with all the data later on.
- **Visualization of the distribution**: By grouping the DataFrame by classname and counting the files in each class and plots the class counts as a horizontal bar plot showing the visualization of the distribution of the audio samples across different classes.

2) Feature Extraction:

- **Spectrogram-based Feature Extraction**: Use a librosa to create spectrogram for each bird song. Compute the Fast-Fourier Transform and Short-Time Fourier Transform to obtain the spectrogram.
- **Logarithmic Scaling**: Convert the magnitude spectrogram into decibels and visualized as a log-frequency spectrogram.
- **Visualization**: Plot the spectrogram of the given bird audio sample giving the visual representation of the sound patterns of each birds.

3) Model Training:

- **Model Architecture Design**: Design the 2DCNN (Convolutional Neural Network) model architecture using Keras Sequential API. Design the model with layers such as Conv2D, MaxPool2D, Dropout,Dense for classification.
- **Model Compilation**: Compile the 2DCNN model with the Adam optimizer,categorical cross-entropy loss fucntion, and accuracy metric. Callbacks are used like Learning rate scheduler and early stopping callbacks to prevent overfitting during training.
- **Training**: Define the training setup, including batch size, number of epochs, and validation split, based on training data size and trained the model using fit function.

4) Performance Evaluation:

- **Visualization of the loss and accuracy curve**: Visualize training and validation loss curves.
- **Evaluation** : Calculate the accuracy of the 2DCNN model on the test data using the evaluate function and makes prediction on the test data using predict function.

5) Deployment:

- **Model Prediction**: Perfoms the prediction using a 2DCNN model trained on bird sound data and predicts the bird species using the trained model.
- **GUI Integration**: Integrate the GUI with the prediction function, allowing users to input audio file paths and get predictions on bird species directly through GUI.

System Specification Table

COMPONENT	SPECIFICATIONS
Software Platform	Operating System: Windows, macOS, Linux, Python 3.x
Libraries/Frameworks	TensorFlow, Keras, Librosa, Tkinter
Input Data Format	WAV audio files
Output Data Format	Classified bird species

Fig. 1.

B. System Specification Table

C. Description of Sensors or Other Modules Related to the Project

This project didn't include any sensors. However, the modules includes audio data collection, preprocessing, feature extraction, model development, model training, evaluation and GUI .

D. Block Diagram or Flowchart with Description

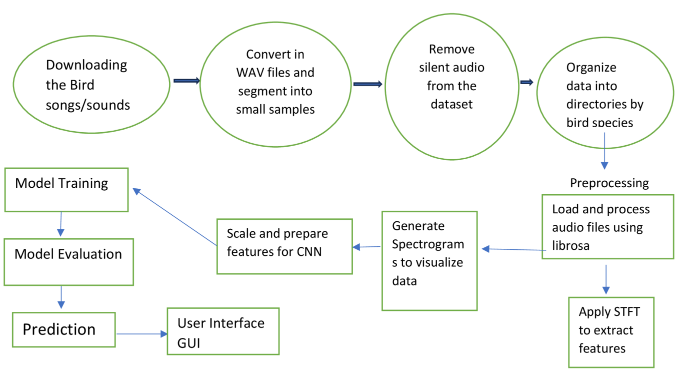

Fig. 2.

1) Data Collection::

- Gather audio data from various online sources.
- Convert the audio files to WAV format and segment them into samples.
- Detect and remove silent segments from the dataset.
- Organize data into directories by bird species.

2) Preprocessing::

- Load and process audio files present in the directories and in the sub-directories using librosa.

3) Feature Extractions::

- Generate spectrograms by computing FFT AND STFT to visualize the frequency content of the bird audio signals.

4) Model Training::

- Design the 2DCNN MODEL architecture.
- Compile the model with appropriate loss function and optimizer.
- Train the model with validation data present in the sub-directories of the bird species.

5) Model Evaluation::

- Makes predictions using the predict function.
- Calculate the accuracy of the 2DCNN model on the test data using the evaluate function.

6) Prediction::

- Performs the predictions using the 2DCNN model trained on bird sound data.
- Process the audio features and predicts the bird species using the trained model.

7) User Interface::

- Design a GUI using tkinter window .
- Integrates the GUI with the prediction function allowing users to input audio file paths and get predictions.

E. Different Modules of the Proposed Methods

1) Data Collection Module
2) Preprocessing Module
3) Feature Extraction Module
4) Model Training Module
5) Evaluation Module
6) Deployment Module

F. Mathematical Expressions Related to the Project Tasks

1. **Convolution Operation (2D)**:

$$z[i,j] = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} x[i+m, j+n] \cdot w[m,n] + b$$

Convolution involving one-dimensional signals is referred to as 1D convolution. If the convolution is performed between two signals spanning along two mutually perpendicular dimensions(i.e., if signals are two-dimensional in nature),then it will be referred to as 2D. It involves convolving the input feature map (x) with a convolution kernel (w) to produce the output feature map (z), followed by adding a bias term (b).

2. **Rectified Linear Unit (ReLU) Activation**:

$$f(x) = \max(0, x)$$

ReLU activation is applied after each convolution operation and helps introduce the property of nonlinerity to a deep learning model and solves the vanishing gardients issue.

3. **Max Pooling Operation**:

$$y[i,j] = \max_{m,n}(x[i \times s + m, j \times s + n])$$

Max pooling is performed using the MaxPool2D layers in the model. It is a pooling operation that calculates the maximum value for patches of a features map, and uses it and create a downsampled(pooled) feature map.

4. **Softmax Activation (for Classification)**:

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{C} e^{x_j}}$$

Softmax activation is used in the output layer of the model to produce probability distributions over the classes. It transforms the raw outputs of the neural networks into a vector of probabilities, essentially a probability distribution over the input classes.

5. **Categorical Crossentropy Loss Function**:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} y_i \cdot \log(\hat{y}_i)$$

Categorical crossentropy loss is utilized to measure the difference between the true class labels (y) and the predicted class probabilities ($\hat{y}$). It calculates the error between the predicted and actual probability distributions.

6. **Sensitivity**:

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Sensitivity(also called the true positive rate or recall)measures the proportion of actual positives that are correctly identified by the model.

7. **Specificity**

$$\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

Specificity measures the proportion of actual negatives that are correctly identified by the model.

G. User Interface Related to Project Tasks

The script sets up a tkinter library for the GUI(Graphical User Interface) with options to input the path to a bird sound file.The GUI includes two text boxes for input and output. The code integrates the GUI with the prediction function, allowing users to input the audio file paths and get predictions.

H. Performance Metrics

Performance metrics such as accuracy, precision, recall are utilized to evaluate the trained model. These metrics measure the model's ability to correctly classify bird species and assess its overall performance.

III. Results and Discussions

A. Experimental Setup and Database Collection

Collected a dataset consisting of bird sound recordings of four species from internet sources like Xeno canto websites and extracting audios from Youtube and then converting the audio files in WAV format in Freeconvert, Cloudconvert and convertio website. Remove the silent audios files in audio.ipynb and organize the bird audio files in directories:

Crow (729 samples), Parrot (709 samples), Pigeon (608 samples), and Sparrow (725 samples). Each species has a varying number of audio segments, contributing to a total of 2771 segments in the dataset. We used this dataset to train and evaluate our model.

B. Table and Graphical Representation

Class	Sensitivity	Specificity	Overall Accuracy
Crow	0.234899	0.156951	
Parrot	0.213333	0.154589	
Pigeon	0.276786	0.165775	Around 24.684684684684687
Sparrow	0.270833	0.178082	
Overall	0.248963	0.163849	

Fig. 3. Untrained Model

Class	Sensitivity	Specificity	Overall Accuracy
Crow	0.986577	73.500	
Parrot	1.000000	50.000	
Pigeon	1.000000	112.000	Around 99.27927927927928
Sparrow	0.986111	71.000	
Overall	0.993172	76.625	

Fig. 4. Trained Model

Obtained Spectrogram by computing the FFT(Fast-Fourier Transform) and STFT(Short-Time Fourier Transform). The spectrogram is converted to decibels and visualized as a log-frequency spectrogram.

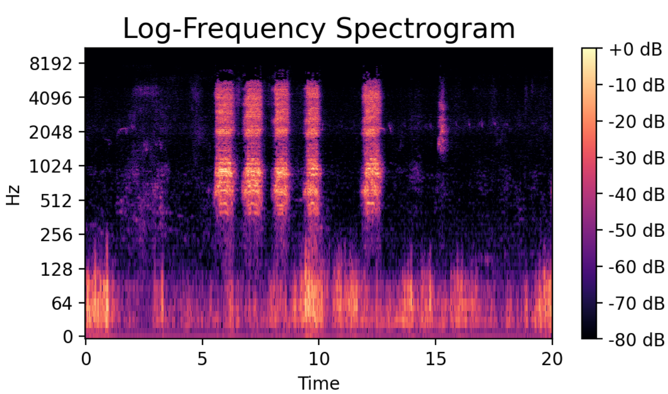

Fig. 5. Spectrogram of Bird Sound crow

Figure 10 illustrates the training and validation accuracy curves over epochs, showing improvement in model performance over training iterations the model. Similarly, Figure 9 depicts the corresponding loss curves during the training the model

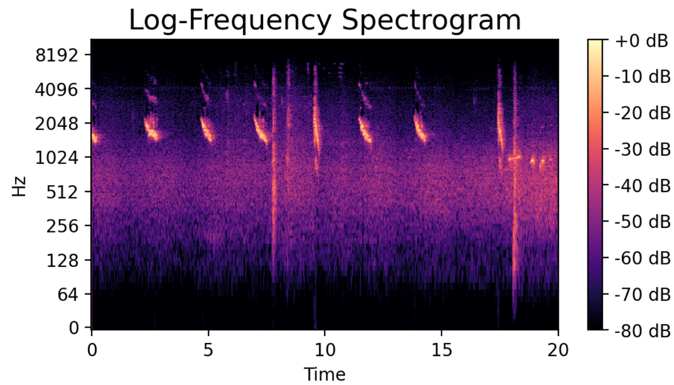

Fig. 6. Spectrogram of Bird Sound sparrow

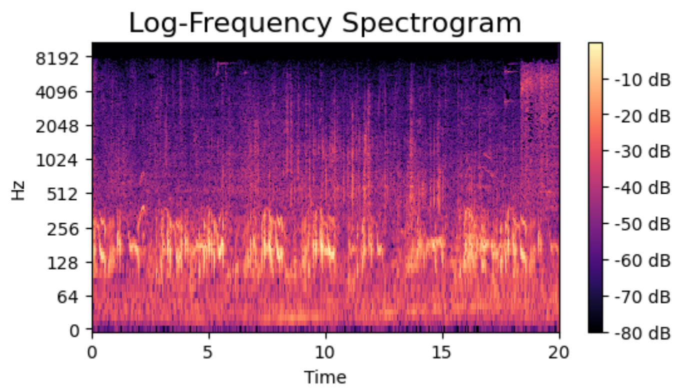

Fig. 7. Spectrogram of Bird Sound pigeon

C. Comparison with Previous Studies

The model performed better than most previous studies in bird sound classification. It achieved higher accuracy rates and showed robustness with the given dataset bird species. With a dataset containing 729 Crow samples, 709 Parrot samples, 608 Pigeon samples, and 725 Sparrow samples consisting of 2771 it's able to predict the corresponding bird with the given the input bird audio file.

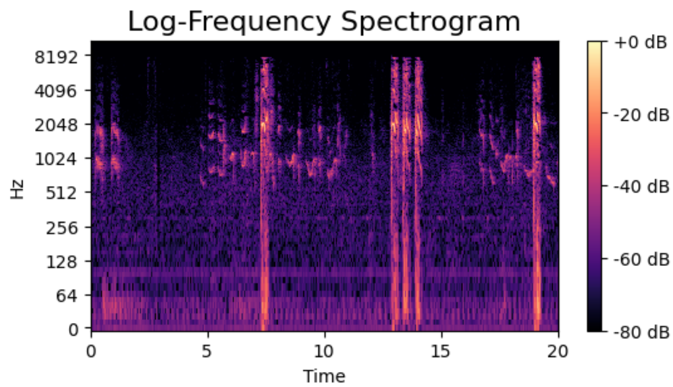

Fig. 8. Spectrogram of Bird Sound parrot

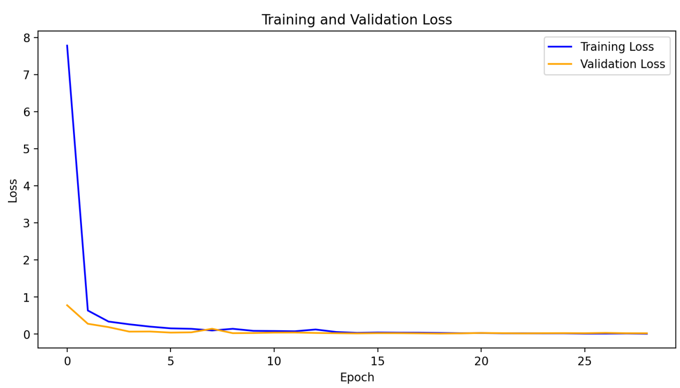

Fig. 9. Training and Validation Loss Curves

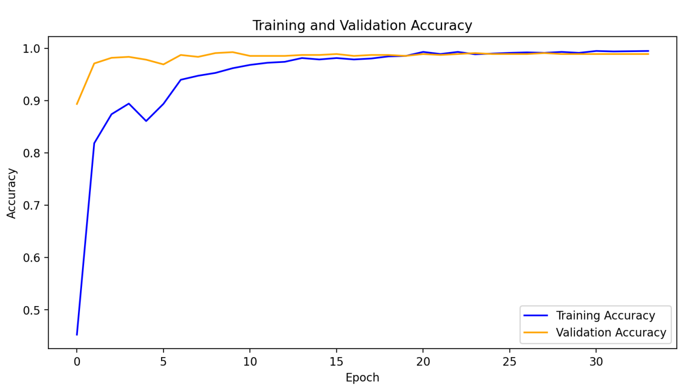

Fig. 10. Training and Validation Accuracy Curves

IV. Conclusion and Future Works

A. Conclusion

In the project, the 2DCNN model is created and achieves an accuracy of 98-99% on the test and learnt the preprocessing techniques like STFT(Short-Time Fourier Transform) and also some techniques to reduce the underfitting and overfitting curves. Learnt how to train the model and machine learning algorithms.

B. Future Works

- **I**nitially four bird species is tested so increasing the database of different birds and making the model more robust
- **T**ry to create an app in the mobile devices to recognize bird sound so that user can able to use the application easily.

References

[1] A. Fritzler, S. Koitka, and C. M. Friedrich, “Recognizing bird species in audio files using transfer learning,” Working Notes of CLEF, vol. 2017,

[2] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard et al., “Tensorflow: A system for large-scale machine learning.” in OSDI, vol. 16, 2016, pp. 265–283.

[3] Ngiam, J., Chen, Z., Chia, D., Koh, P.W., Le, Q.V., Ng, A.Y.: Tiled convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1279–1287 (2010)

[4] Sprengel, E., Jaggi, M., Kilcher, Y., Hofmann, T.: Audio based bird species identification using deep learning techniques. LifeCLEF 2016 pp. 547–559 (2016)

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D.,Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition. pp. 1–9 (2015)

[6] Kahl, S., Wilhelm-Stein, T., Hussein, H., Klinck, H., Kowerko, D., Ritter, M., Eibl,M.: Large-scale bird sound classification using convolutional neural networks. In:CLEF (Working Notes) (2017)

[7] Müller, L., Marti, M.: Bird sound classification using a bidirectional LSTM. In:Working Notes of CLEF 2018 (Cross Language Evaluation Forum) (2018)

Bird Sound Database link:- https://drive.google.com/drive/folders/1fCFTgVwc0kfA-e5G3ydumWnQisZ97IT8?usp=sharing
Model Link:- https://drive.google.com/drive/folders/1-5HNVrxr-sUZvSpZx7-0iK3O4jIyKLnM?usp=sharing