Classification Of Birds Sounds And Predicting The Specification

MINI PROJECT REPORT

Submitted by

Jeetendra Prasad S (RCAS2021MDB047)

in partial fulfillment for the award of the degree of

MASTER OF SCIENCE
SPECIALIZATION IN
DATA SCIENCE AND BUSINESS ANALYSIS

DEPARTMENT OF COMPUTER SCIENCE

RATHINAM COLLEGE OF ARTS AND SCIENCE

(AUTONOMOUS)

COIMBATORE - 641021 (INDIA)

DECEMBER-2022

RATHINAM COLLEGE OF ARTS AND SCIENCE
(AUTONOMOUS)
COIMBATORE - 641021

BONAFIDE CERTIFICATE

This is to certify that the report entitled **Classification Of Birds Sounds And Predicting The Specification** submitted by **Jeetendra Prasad S (RCAS2021MDB047)** for the award of the Degree of Master in Computer Science specialization in **"DATA SCIENCE AND BUSINESS ANALYSIS"** is a bonafide record of the work carried out by him under my guidance and supervision at Rathinam College of Arts and Science, Coimbatore

Mr.S.Ravisankar,M.E.,(Ph.D).,
Supervisor

Mr.P.Sivaprakash, M.Tech., (Ph.D).,
Mentor

Submitted for the University Examination held on 01.12.2022

INTERNAL EXAMINER

EXTERNAL EXAMINER

RATHINAM COLLEGE OF ARTS AND SCIENCE

(AUTONOMOUS)

COIMBATORE - 641021

DECLARATION

I, **Jeetendra Prasad.S**, hereby declare that this report entitled **Classification Of Birds Sounds And Predicting The Specification** is the record of the original work done by me under the guidance of **Mr.S.Ravisankar,M.E,(Ph.D),.**, Faculty Rathinam college of arts and science, Coimbatore. To the best of my knowledge this work has not formed the basis for the award of similar award to any candidate in any University.

Signature of the Student

Jeetendra Prasad S

Place: Coimbatore

Date: 01.12.2022

COUNTERSIGNED

Mr.S.Ravisankar,M.E,(Ph.D),.

Supervisor

Contents

Acknowledgement

On successful completion for project, I look back to thank who made in possible. First and foremost, I thank **"THE ALMIGHTY"** for this blessing on me without which I could have not successfully my project.I am extremely grateful to **Dr.Madan.A. Sendhil, M.S., Ph.D.,** Chairman, Rathinam Group of Institutions, Coimbatore and **Dr. R.Manickam MCA., M.Phil., Ph.D.,** Secretary, Rathinam Group of Institutions, Coimbatore for giving me an opportunity to study in this college.I am extremely grateful to **Dr.R.Muralidharan, M.Sc., M.Phil., M.C.A., Ph.D.,** Principal Rathinam College of Arts and Science(Autonomous), Coimbatore. I Extend my deep sense of valuation to **Mr.A.Uthiramoorthy, M.C.A., M.Phil., (Ph.D),** Rathinam College of Arts and Science (Autonomous) who has permitted to undergo the project.

Unequally I am thank **Mr.P.Sivaprakash, M.Tech., (Ph.D).,** Mentor and **Dr.Mohamed Mallick, M.E., Ph.D.,** Project Coordinator, and all the Faculty members of the Department - iNurture Education Solutions pvt ltd for their constructive suggestions, advice during the course of study.I convey special thanks, to my supervisor **Mr.s.Ravisakar,M.E,(Ph.D),.** who offered their inestimable support, guidance, valuable suggestion, motivations, helps given for the completion of the project.

I dedicated sincere respect to my parents for their moral motivation in completing the project.

List of Figures

List of Abbreviations

CNN	Convolution Neural Network
MFCC	Mel Frequency Cepstrum Coefficients
R-CNN	Region-Based Convolutional Neural Network
ROI	Region of Interest

Abstract

In today's world, many new technologies are created for the identification of various species and resources. As many researchers, academicians, geologists and ornithologists are facing difficulties to identify and classify the endemic birds species in India. Species that are known to exist naturally exclusively in a single nation are called endemics. Birds are utilized to provide provisioning, regulating, and supporting functions.

Due to its rare sightings the identification and understanding of the bird is hard. The proposed model is aimed to bridge the gap for the identification of the species. The model takes the input as the Birds sound and then processes it with the log Mel-spectrogram and Mask R-CNN by integrating these two techniques for identification.

Bird sounds initially processing the log Mel-Spectrogram of the audio files using Mask R-CNN, a cutting-edge algorithm for detecting and recognizing in images. R-CNN Mask analyses audio sequences that might

By creating bounding boxes around the execution of an action in the time-frequency domain, will contain it. Then, we analyse each individual frame in the candidate segments suggested by Mask R-CNN then the identification of the bird is given out as the result. This proposed model would be of a huge change for the researchers.

Ornithologists and other various people who are trying to identify a particular bird is an endemic species or not without the sight of it.

Chapter 1

Introduction

In Nowadays Invention of technologies is more efficient ways in both software and hardware for identification of various things in the way of identify and classify the endemic bird's species in India. Is a challenging task for identify the birds and the specifications for the researchers and Ornithologists to find the particular bird in endemic species.This model will helps to identify it. Birds are good environmental indicators and may reveal if an ecosystem is sustainable.

In The Field of Machine learning (ml) and Artificial intelligence (AI) increased interest in a variety of classification issues, particularly those involving data in the form of photos, videos, and audio as advancements in the field of Deep learning(DL). Identifying categories for sounds and classifying them is one of the main classification issues. In Deep learning is used to analysis the all the image, audio and video data.

The proposed model will use to analysis the audio data by using Librosa it helps to read and visualize the sound signals and also do the feature extractions in it using different signal processing techniques. Then Librosa helps to convert the audio data into wave plot. and cepstral features in research on bird sound classification is the

Mel frequency cepstral coefficients(MFCC). Based on how people hear sounds like that frequency domain representations of the original bird sounds are given as input to the Mel-scale filter bank to produce mel-spectrum, To identify audio segments that relate to various sorts of events, computer vision experts use object detection models and mel-spectrograms of audio as images.

Implemented a region-based strategy while utilising a slight different version of R-FCN to find uncommon sound frequencies. Fully convolutional neural networks, or R-FCN, are used to detect objects. In order to create tight borders around the event while taking the background noise into consideration, region-based models can be helpful. These models look for patterns in log mel-spectrograms to identify events. Using both region-based and frame-based techniques, sound event detection Using Mask R-CNN mode, we first employ a region-based strategy to extract event-regions from the audio. A cutting-edge object recognition and segmentation model is Mask RCNN. We can identify candidate audio events with tight, pixel-precise bounding boxes because to its remarkable ability to construct tight and very accurate boundaries around the target objects.

After identifying probable event regions, we examine each of these regions by examining the little frames that make up each one of them, using a frame-level classifier made up of a number of convolutional and recurrent layers. This method successfully detects audio events in log-mel spectrograms that have very varied and irregular forms. The benefit of Mask R-CNN, a cutting-edge object detection model in computer vision, to suggest audio regions as potential event regions with patterns comparable to the

target events. Convolutional and recurrent layers are combined to create a frame-level classifier that we use to examine each candidate segment's frames and separate the actual event regions from those that the Mask R-CNN model has suggested. In this we can identify the particular bird.

1.1 Objective of the project

This project goal is to develop a web application that will support in classifying and recognizing the indigenous bird species of India. Researchers, geologists, and ornithologists are having trouble in identifying the endemic bird species that exist in India. It is challenging to recognise and comprehend the bird's taxonomy, season, and region. The species of birds are difficult to identify because they only occasionally appear. To fill in the gaps necessary for species identification.

The proposed methodology expects to address the challenges faced by researchers, geologists, and ornithologists when trying to determine which type of bird. The proposed work for the model analysis in audio moves to transform audio into a wave plot and to utilize that wave plot's Extraction based on Mel frequency cepstral coefficients (MFCC) for different purposes. Mel-scaled filter banks for logs It is simple to recognise log-mel spectrograms from sound frequency and temporal characteristics. Subsequently suggested using Mask R-CNN to find exact bounding boxes around spectrogram regions that might relate to a target event. After classifying the data, training the model to forecast which kind of birds by sound.

Neural Network:

Neural network is a sequence of algorithms that try to build the relationships between the data and the human brain. Neural networks adapt the changing input so, the network generates the best possible output without needs to redesign again.

1.2 Scope of the Project

The creation of a web application to categorise and identify the indigenous bird species In India is the aim of this project. The model will provide information on the region, season, scientific and common names, taxonomy, and other characteristics of the birds. The system analyses the sound file using computer vision and audio wave plot techniques in order to extract the Mel frequency cepstral coefficients (MFCC). This feature of identifying the sound frequency and time characteristics of the audio wave plot is used to analyse the audio wave plot. Then, in the audio wave plot, log-Mel-scaled filter banks log-Mel spectrograms for sound identification appear. To gather information from log Mel-scaled filter banks in order to anticipate bird noises.

CNN Algorithm is one of the key elements needed to identify this model. The Mask R-CNN technique is chosen for the model that is being suggested since it has demonstrated excellent performance in object detection and segmentation in computer vision. The Mask R-CNN is a fully convolutional network. A class name and a small bounding box with pixel-to-pixel alignment of the image's actual objects are output by the algorithm. Then utilise the log-Mel spectrograms that were derived from the audio data to train a Mask R-CNN model for each event. To create bounding boxes around

spectrogram regions that closely resemble the target event, we use Mask R-CNN. For each target event, we have a list of potential event regions on the log-Mel spectrograms. The frame-level classifier is now used to separate the real events from them.This method is used to assess the model and identify the bird by listening to its sound and evaluating it.

1.3 Module Description

OS: The OS module in Python has functions for adding and deleting folders, retrieving their contents, changing the directory, locating the current directory, and more. Before you can communicate with the underlying operating system, you must import the os module.

Pandas: Pandas is a Python data analysis library. Pandas has developed into one of the most well-liked Python libraries and is a strong and versatile tool for quantitative research

Numpy: NumPy is a Python library that allows you to work with arrays. Additionally, it provides functions for working with matrices, the Fourier transform, and the area of linear algebra.

Matplotlib: Python's Matplotlib toolkit provides a complete tool for building static, animated, and interactive visualizations. Matplotlib makes difficult things possible and simple things easy. Produce plots fit for publication

Sklearn: Scikit-learn is a free machine learning library for Python. Numerous algo-

rithms, a train-test split, preprocessing, and other elements are included.

Librosa: A Python package for analysing music and audio is called librosa. It offers the components required to build music information retrieval systems. Librosa will gives sample rate to the audio signals and its able to give many channels to the audio

Soundfile: soundfile is a python library is used to read and write audio files. It views an audio file as a NumPy array that includes all of the audio's pitches. This module can both read and write audio files.

Spectrogram: A spectrogram is a graphic representation of frequencies plotted against time that reveals the strength of the signal at a specific moment. A spectrogram is essentially a visual representation of sound. A spectrogram is a visual representation of the "loudness" or signal strength over time at different frequencies contained in a specific waveform.

MFCC: MFCC Is a feature Extraction in python that allows to frequency distribution across the window size, so it is possible to analyse both the frequency and time characteristics of the sound. .

Keras: It is a deep learning API written in Python language, running on the top of the machine learning platform i.e., Tensor flow. It is used to create layers in Neural Network.

1.4 Existing System

Building a lightweight feature extraction and recognition network for Bird Sound using MobileNetV3 as the backbone A multi-scale feature fusion structure is designed, and the Pyramid Split Attention (PSA) module is added to improve the network's adaptability to scale extraction of spatial information and channel information. And in this model, ordinary convolutions are introduced into the Bneck module, causing the Bneck module to become the Bnecks module and the model to identify 264 different types of birds on the self-built data set.

And the model recognition and classification are finished. The total number of bird Mel spectrogram samples after feature extraction is 229164, with 183690 samples chosen as the training set and 45924 samples chosen as the test set. as the testing set Stochastic Gradient Descent (SGD) is used as the model optimizer, the loss function is the cross-entropy loss function, and the learning rate descent strategy is Cosine Annealing. There are fewer parameters and computations. The analysis of ablation experiments shows that the improvement proposed in this model can improve model classification accuracy and make the model more generally applicable.

Chapter 2

Literature Survey

Design of Bird Sound Recognition Model Based on Lightweight

Recognizing bird sounds is critical in bird conservation. With appropriate sound classification, research can predict the quality of life in the area automatically. Deep learning models are now used to classify bird sound data with high classification accuracy. The generalisation, however, most existing bird sound recognition models have limited capability, and a complicated algorithm is used to extract bird sound features. To address these issues, this paper constructs a large data set containing 264 different types of birds to improve the model's generalisation ability, and then proposes a lightweight bird sound recognition model to build a lightweight feature extraction and recognition network with MobileNetV3 as the backbone.

The model's recognition ability is improved by adjusting the depth wise separable convolution in the model. A multi-scale feature fusion structure is designed, and the Pyramid Split Attention (PSA) module is added to improve the network's adaptability to scale extraction of spatial information and channel information. To improve the model's refinement ability toward global information, the channel attention mechanism

and ordinary convolution are introduced into the Bneck module, transforming it into the Bnecks module.

The majority of people in related fields use Internet of things devices to remotely monitor bird populations online. Because the majority of bird protection habitats are in the wild, it is difficult for the online monitoring system to transmit bird sounds back to the server for data processing, recognition, and feedback under normal network conditions. Off-line monitoring in a bird reserve is not possible with low-cost embedded equipment because the high-complexity sound feature extraction algorithm and high-precision sound recognition algorithm are not supported. As a result, the goal of this paper is to create a lightweight bird voice recognition algorithm that not only achieves high accuracy by using simple and single features, but also keeps the model small enough to run in low-cost embedded devices.

A model of a lightweight bird song recognition algorithm is proposed. This model's classification accuracy can reach 95.12 percentage. The recognition rate of the model proposed in this paper is higher when compared to other lightweight networks. When compared to other depth models, the accuracy of the This paper's model is slightly different, and the number of parameters and computations has been reduced. The analysis of ablation experiments shows that the improvement proposed in this paper can improve model classification accuracy and make the model more generalizable.

Robust cepstral feature for bird sound classification

Birds are excellent environmental indicators that can indicate ecosystem sustainability; birds can be used to provide provisioning, regulating, and supporting services. As a result, birdlife conservation research is always given priority. Because birds are airborne and the tropical forest is dense, audio identification of birds may be a better solution than visual identification. The purpose of this research is to identify the most appropriate cepstral features that can be used to more accurately classify bird sounds. An automated energy-based algorithm was used to select and segment fifteen endemic Bornean bird sounds. There are three types of cepstral features extracted: linear prediction cepstrum coefficients (LPCC), mel frequency cepstrum coefficients (MFCC), and gammatone frequency cepstrum coefficients (GFCC) (GTCC), and separately used for classification purposes with a support vector machine (SVM).

Birds play an important role in ecosystem stability as pollinators and seed dispersers, as well as in maintaining a balanced population of predators and prey in the ecosystem. As a result, birdlife conservation and species preservation projects are critical for a healthy ecosystem. These types of projects are difficult to implement because they require manual labour, labour, and physically demanding processes. Modern and advanced techniques and technologies have made environmental and biodiversity monitoring research easier and more feasible. Many researchers have used signal processing and machine learning techniques to help with difficult and complex processes. Bioacoustics signal processing and pattern recognition algorithms have been used for detection and identification of bird species.

Within the field of bird sound classification research, MFCC is one of the most commonly used cepstral features. Based on human perception of sound, frequency domain representations of original bird sounds are fed into the mel-scale filter bank, which generates mel-spectrum, which is then converted to MFCC using cepstral analysis. Each MFCC band has a weighted sum that represents the spectral magnitude in the corresponding channel. Because there are no tuning parameters involved, calculating MFCC parameters is quick and easy. Lee et al. extract features for their work using both static and dynamic two-dimensional MFCCs. MFCCs have also been combined with other feature extraction methods. Kogan and Margoliash, for example.For feature extraction, both MFCC and linear predictive coding (LPC) are used, whereas Leng and Tran use a combination of three methods: binned frequency spectrum, MFCC, and LPC. GFCCs with first and second derivatives, as well as power normalised cepstral coefficients, are other cepstral-based methods (PNCC)

After data collection, the basic structure of any typical audio classification has several stages; the first stage is pre-processing, which is done on the audio signal for noise cancellation, silence reduction, and normalisation. In this study, the sounds of fifteen (15) endemic Bornean birds were collected and segmented using automatic energy-based segmentation to remove silence and noise from the recording. The segmented audio signal was then used to extract three (3) cepstral features, namely LPCC, MFCC, and GTCC, in the feature extraction stage. Finally, six hundred (600) samples were used for training and 150 samples were used for testing. In this work, the SVM classifier is used separately for each feature type, for both training and testing. It is clear that Despite

the fact that MFCC has been more widely used by many researchers over the years, GTCC feature-based classification outperforms the other two LPCC and MFCC-based classifications.

Combining the three cepstral features does not improve accuracy over using only the GTCC features. The outcome is significant because it demonstrates that using GTCC alone would provide reasonably high accuracy 93.3 percentage for bird sound classification. However, there is still room for improvement by investigating different properties of bird sounds, combining GTCC with other signal features, and implementing the technique in real-time on portable multimedia devices, which could lead to new directions for this proposed study.

Chapter 3

Methodology

3.1 Convolution Neural Network(CNN)

A CNN is a particular type of network design for deep learning algorithms that is utilized for tasks like image recognition and pixel data processing. Although there are different kinds of neural networks in deep learning, CNNs are the preferred network architecture for identifying and recognising objects.

Convolutional neural networks (CNNs) are a subclass of neural networks that are mostly employed in voice and image recognition applications. With no loss of information, its integrated convolutional layer lowers the large number of features of images. CNNs are very well suited for this use case. CNN's input layer, an output layer, and hidden layers are all used to process and categorize images.

The hidden layers comprise convolutional layers, ReLU layers, pooling layers, and fully connected layers, all of which play a crucial role.they can recognise intricate elements in image data, which makes them useful for face detection and recognition

Convolutional Layer:The convolutional layer, the central component of a CNN, is

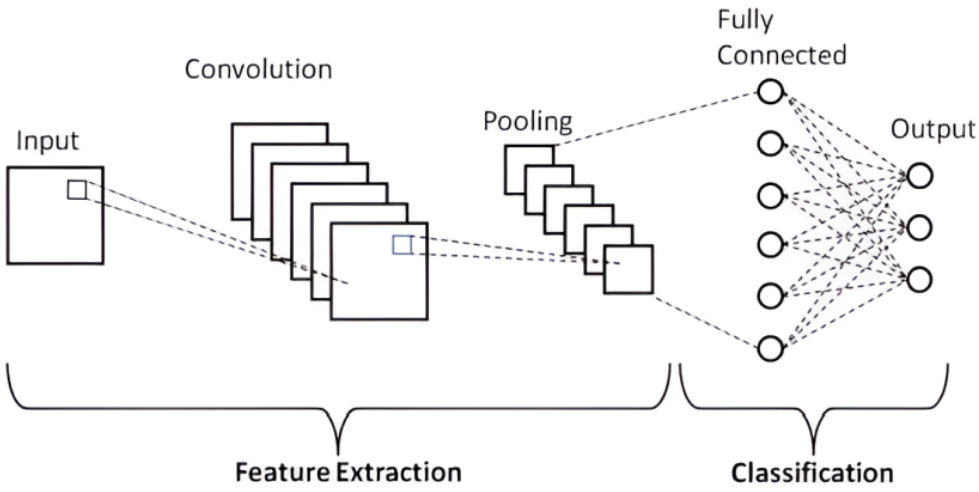

Figure 3.1: Convolution Neural Network(CNN)

where most computations take place. The first convolutional layer may be followed by a subsequent convolutional layer. A kernel or filter inside this layer moves over the image's receptive fields during the convolution process to determine whether a feature is present.

The kernel traverses the entire image over a number of iterations. A dot product between the input image and the filter is calculated at the end of each iteration. A convolution layer or convolved feature is the result of the dots being connected in a certain pattern. In this layer, the image is ultimately transformed into numerical values that the CNN can understand and extract pertinent patterns from.

pooling Layer:collecting layer The pooling layer similarly to the convolutional layer sweeps a kernel or filter across the input image. Compared to the convolutional layer, the pooling layer has fewer input parameters but also causes some information to be lost. Positively, this layer simplifies the CNN and increases its effectiveness.

Fully Connected Layer:whole layer connectivity Based on the features extracted in the preceding layers, picture categorization in the CNN takes place in the FC layer. Fully connected here refers to the connection of each activation unit or node of the following layer to each input or node of the previous layer.

3.2 Mask R-CNN

Modern in terms of picture and instance segmentation, Mask R-CNN is a Convolutional Neural Network (CNN). Faster R-CNN, a region-based convolutional neural network, served as the foundation for the development of Mask R-CNN.Faster R-expansion, CNN's Mask R-CNN, operates by simultaneously adding a branch for predicting an object mask (Region of Interest) and the branch already in place for bounding box recognition.

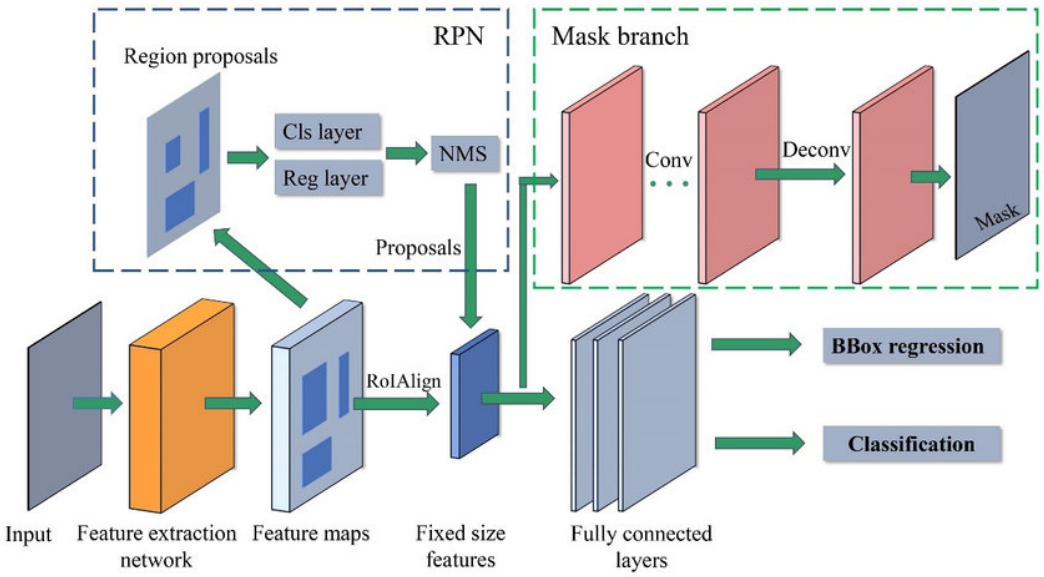

Figure 3.2: Mask R-CNN Frame Work

In Mask R-CNN the computer vision task Image segmentation is the process of dividing a digital image into various pieces (sets of pixels, also known as image objects). To locate objects and borders, this segmentation is used (lines, curves, etc.).

Backbone Network:

Mask R-CNN was tested on two of backbone networks. The first is the standard ResNet architecture (ResNet-C4), and the second is the ResNet with feature pyramid network. The standard ResNet architecture was similar to Faster R-CNN, but ResNet-FPN proposed some changes.

This is a multi-layer RoI generation process. This multilayer feature pyramid network generates RoI at various scales, which improves the accuracy of the previous ResNet architecture.

Region Proposal Network:

The previous layer's generated convolution feature map is passed through a 3*3 convolution layer. This is then fed into two parallel branches that compute the objectness score and regress the bounding box coordinates. For this feature pyramid, we only use one anchor stride and three anchor ratios (because we already have feature maps of different sizes to check for objects of different size).

Mask Representation:

This ConvNet accepts a RoI as input and returns the m*m mask representation as output. We also upscale this mask for inference on the input image and use 1*1 convolution to reduce the channels to 256. We use RoIAlign to generate input for this

fully connected network that predicts mask.

RoIAlign's purpose is to convert a variable-size feature map generated by a region proposal network into a fixed-size feature map. The Mask R-CNN paper proposed two architecture variants. The input of the mask generation CNN is passed after RoIAlign is applied in one variant (ResNet C4), but it is passed just before the fully connected layer in another (FPN Network).

RoI Align:

RoI align serves the same purpose as RoI pool in that it generates fixed-size regions of interest from region proposals. Given the previous Convolution layer's feature map of size h*w, divide it into M * N grids of equal size (we will NOT just take integer value). The mask R-CNN inference speed is around 2 fps, which is good considering the architecture's addition of a segmentation branch.

There are 2 main types of image segmentation that fall under Mask R-CNN:

1. Semantic Segmentation
2. Instance Segmentation

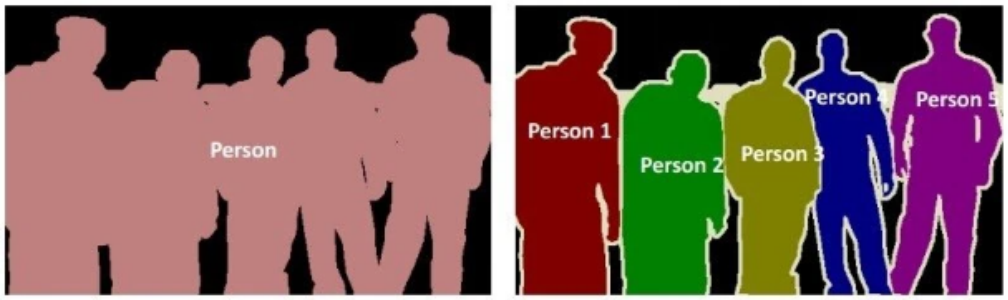

Figure 3.3: Semantic Segmentation and Instance Segmentation

Semantic segmentation is the process of identifying and categorising related ob-

jects into a single class at the pixel level. And every object was categorised as a single object. Because it separates the image's subjects from the background, semantic segmentation is also known as background segmentation.

Instance segmentation is concerned with accurately identifying every object in an image as well as correctly segmenting each instance. As a result, it combines object location, object classification, and object detection. Each object is separated as a single entity throughout this segmentation procedure. Additionally, it highlights the main subjects in the image instead of the background.

3.3 Advantages

This model will be useful to all ornithologists and the general public in identifying the endemic bird species of a specific region and in providing information about the specific family of birds that each bird belongs to.

3.4 System Design

Initially gathering information from the dataset, which includes a collection of bird sounds. These procedures are part of this model: Making a spectrogram, extracting MFCC features, creating a Mel log spectrogram, Make observations, processing, displaying, and classification of data sets Our first stage concerned the gathering of data while bearing in mind the general standard machine learning technique.

In the learning phase, we classify the model using the state-of-the-art technique for detecting and recognising in images, Mask R-CNN. R-CNN Mask analyses sound files

that could the execution of an operation in the time-frequency domain will be contained by the creation of bounding boxes around it. The model will be prepared to forecast the outcome once it has been constructed.

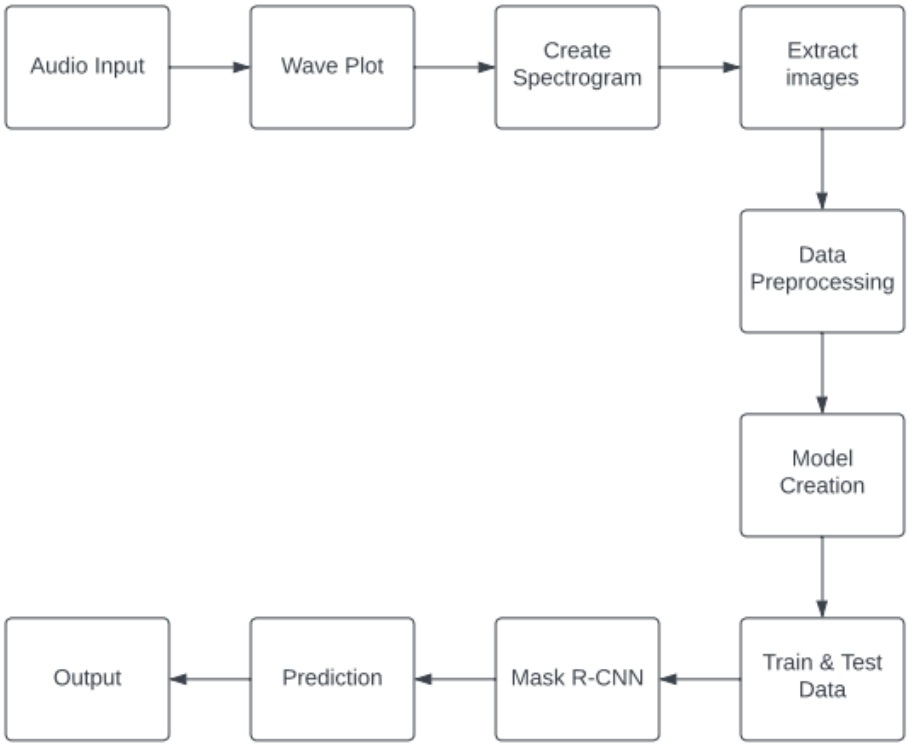

Figure 3.4: The Proposed Framework

3.5 Sequential Model

Make a sequential model. It is the simplest way to create a model in Keras. It allows you to build a model layer by layer (starting from input layer to output layer). The following are some key concepts related to the sequential moThe layer used is dense. In a dense layer, all nodes in the previous layer are connected to nodes in the current layer.

The activation function enables the model to represent non-linear relationships. The activation function is the Region Proposal Network (RPN), which is the first stage that produces an output approximate bounding boxes for potential objects in the image.

The network is slid over this feature map to generate a set of region proposals as well as an objectless score for each region. The second stage is in charge of predicting object class labels, tightening bounding boxes, and creating segmentation masks for objects. Mask R-CNN employs a method known as RoIAlign to generate fixed-size feature maps for each region while ensuring that the values on the feature map correspond to the regions on the actual image.

3.6 Keras

Keras API is a deep-learning library with methods for loading, preparing, and processing Images, running on top of the machine-learning platform Tensor Flow. Keras helps to Segregate the image layer by layer. It is simple, flexible, and powerful. The high-level API of Tensor Flow 2 is called Keras; it is a user-friendly, highly effective interface for Resolving machine learning issues with a focus on contemporary deep learning. It offers Crucial building elements and abstractions for creating and delivering machine learning Solutions quickly. Keras have several layers, which is Conv2D, MaxPooling2D, Flatten, Dense, Dropout, Batch Normalization.

1. **Conv2D** In order to generate a tensor of outputs, this layer generates a convolution Kernel, which is convolved with the layer input. The number of filters that the convolutional layers will learn from is the required Conv2D parameter. It is an integer number That also establishes how many output filters will be used in the convolution.

2. **MaxPooling2D** Max pooling for 2D spatial data. The input is down samples along Its spatial dimensions (height and width) by taking the maximum value for each input Channel over an input window with a size determined by pool size. Each dimension of The window is moved one step at a time.

3. **Flatten**The input is flattened using flatten. For instance, the layer's output shape Will be (batch size, 4) if flatten is applied to a layer with an input shape of (batch size, 2,2). The argument for flatten is as follows: keras. Layers. Flatten (data format = None)

4. **Dense** Dense layer is the regular deeply connected neural network layer. It is the Most common and frequently used layer. The dense layer does the below operation on The input and returns the output. Output = activation (dot (input, kernel) + bias)

5. **Dropout**In order to avoid over fitting, the Dropout layer randomly sets input units To 0 with a frequency of rate at each step during training. The sum of all inputs is Maintained by scaling up non-zero inputs by 1/ (1 - rate).

6. **Batch Normalization**, Batch Normalization applies a transformation that keeps the Output mean and standard deviation close to 0 and 1, respectively. Significantly, batch Normalization behaves differently during inference than it does during training

Chapter 4

Experimental Setup

4.1 Audio visualisation

To understand the audio for computer vision use, plot the audio in graphical representation in a waveform using Librosa, a Python library package. It is useful for reading and visualising sound signals, as well as extracting features from them using various signal processing techniques. Librosa will assign a sample rate to the audio signals. The sample rate is useful for creating a wave plot.

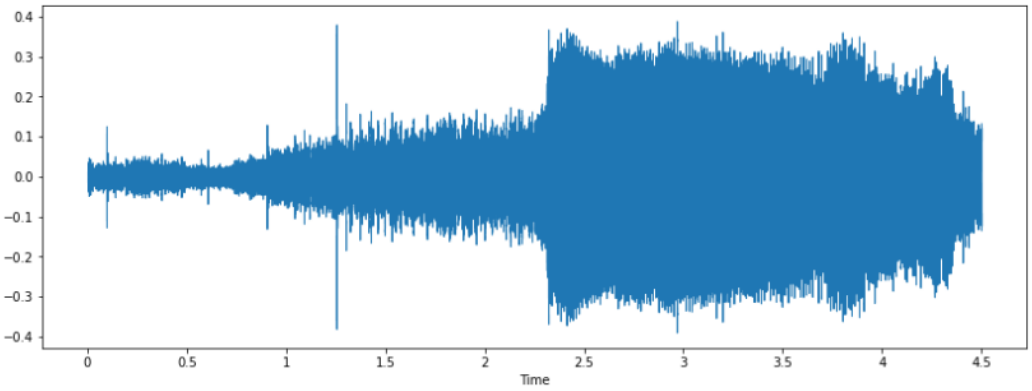

Figure 4.1: Audio Visualisation

4.2 Spectrograms images

To display a spectrogram like an image of a signal. And it graphs frequencies in the X-axis and time in the Y-axis. The colours of a spectrogram can also be used to represent the signal strength; the brighter the hue, the stronger the signal. Then, each frequency in the transmission is distributed according to signal strength.

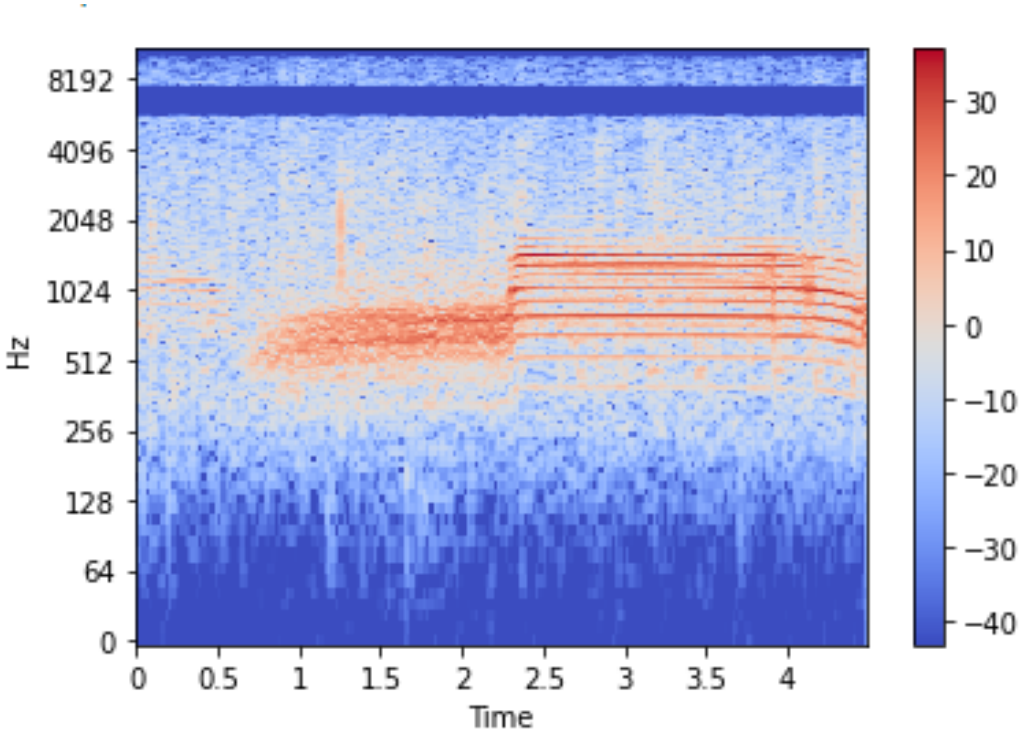

Figure 4.2: Spectrograms of Bird Sound

4.3 Data Preperation

Load the audio file, then extract the necessary information. The words variable contains all of the training data that has been tokenized, the classes variable contains all of the target labels and tags that correspond to each training data, and the documents variable

contains both training data (frequency of sample audio that has been tokenized using the tokenize(pattern)) and labels that correspond to each training data. This is a fundamental summary of the data pre-processing model. There will be two sets of data used for data classification: training data and testing data. Once the model is built, the output will be derived from it. the result is whether the Bird is recognised.

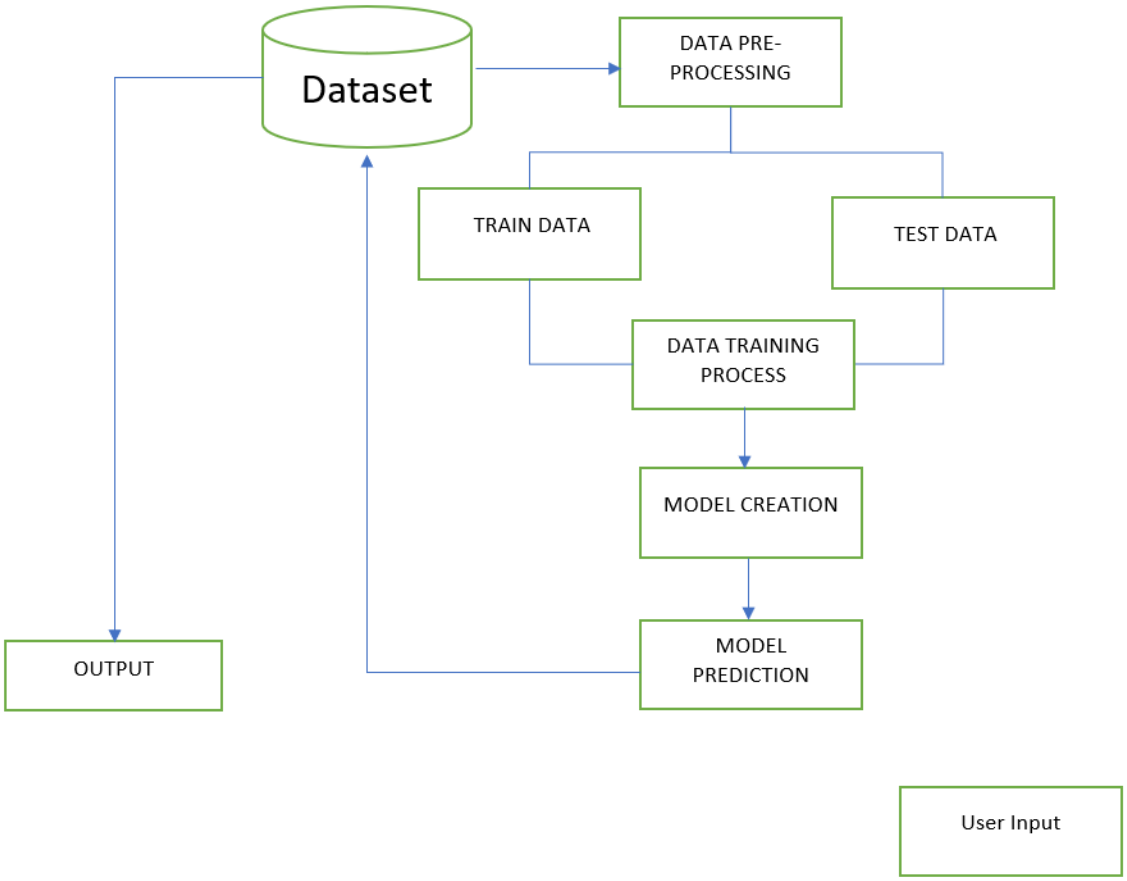

Figure 4.3: Dataset flow work

4.4 Model Training

Define Neural Network architecture for proposed model and for that use the Sequential model class of Keras. Keras is an open source, high level library for developing neural network model. The steps for creating a model as follow:

Step 1: First, to import the all needed packages, that is audio wave, OS,Sklearn, keras, from TensorFlow, matplotlib, librosa,Spectroogram etc.audio used for store the metadata. sklearn used to split the dataset into train and test. keras used to segregate the images layer by layer. librosa is used to visualize the audio files.

Step 2: Then create a spectrogram like an image of a signal. And it graphs frequencies in the X-axis and time in the Y-axis. then convert into frame level classifier because the computer did not understand the audios.

Step 3: And then spliting the dataset into train and test using train test split function from sklearn.model selection library. The train and test size will be 70

Step 4: Now, model is trained, an input can be used to make prediction

4.4.1 User Input

To get the input as a audio file from the user, this model will automatically analyze those above all steps.

4.4.2 Predict Output

After the all pre-processing of the dataset, the model will work on those data and the better result about the Birds identification from these prediction.

Chapter 5

Result and Discussions

We described a project that is still in progress to collect bird sounds from the dataset and researchers, academicians, geologists and ornithologists and analyse how can benefit from this information. These outcomes just hint at the possibility of this kind of study. Despite the fact that our findings are identification of bird species. Due to its rare sightings the identification and understanding of the bird is hard. This model will help to easily identify the bird species. We have temporarily restricted our use to a subset of the data. We also lack information on birds sound conditions, We're expanding our research to include audio samples that we've already gathered. Audio will provide helpful additional information for classification. Our dataset also contains birds sound with various species, and we hope to examine this further in order to investigate how identification is in this respect. Although we have only presented a small investigation of the birds sounds in the model.

Chapter 6

Conclusion

We have discussed our efforts to develop a identification of Birds that is soundbased. The tool was developed based on earlier research that demonstrated good accuracy for identifying other specification of birds. We emphasise the justification of selecting various database birds sound. The development of data collecting is characterised in terms of meta data statistics. We also draw attention to the birds complementary character. The next step in the development of the audio analysis tool will involve using machine learning algorithms to categorise various classificaton and identification look for sound-based birds sound.

6.1 Limitations

At the present model we will predict birds species identifiction using bird sound. In future there more possible for different kind of Classification of birds species. so, this model will update for that classification and the prediction.

6.2 Future Works

The future work of this paper is the concept of trust factor used for researchers and ornithologists. The next step in the development of the identification tool will involves using algorithms for machine learning to categorise various birds sound and look for sound-based database for identify birds species. Main future work is predict the bird species using sounds.

References

1. Design of Bird Sound Recognition Model Based on Lightweight Fan Yang; Ying Jiang; Yue Xu IEEE Access Year: 2022 — Volume: 10 — Journal Article — Publisher: IEEE

2. A Robust cepstral feature for bird sound classification M Ramashini, PE Abas, K Mohanchandra... - Int. J. Electr. Comput ..., 2022 - researchgate.net

3. I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," Applied Acoustics, vol. 80, pp. 1–9, Jun. 2014, doi: 10.1016/j.apacoust.2014.01.001.

4. S. Nowicki and P. Marler, "How do birds sing?," Music Perception, vol. 5, no. 4, pp. 391–426, Jul. 1988, doi: 10.2307/40285408

. N. Larsen and F. Gollerf, "Role of syringeal vibrations in bird vocalizations," Proceedings of the Royal Society of London. Series B: Biological Sciences, vol. 266, no. 1429, pp. 1609–1615, Aug. 1999, doi: 10.1098/rspb.1999.0822.

5. R. A. Suthers, "How birds sing and why it matters," in Nature's Music, Elsevier, pp. 272–295, 2004.

6. L. R. Hernandez-Miranda and C. Birchmeier, "Mechanisms and neuronal control of vocalization in vertebrates," Opera Medica et Physiologica, vol. 4, no. 2, pp. 50–62, Dec. 2018, doi: 10.20388/omp2018.001.0059

7. M. J. Ryan and E. A. Brenowitz, "The role of body size, phylogeny, and ambient noise in the evolution of bird song," The American Naturalist, vol. 126, no. 1, pp. 87–100, Jul. 1985, doi: 10.1086/284398.

8. Park, D. H., H. K. Kim, I. Y. Choi, and J. K. Kim. 2012."A Literature Review and Classification of Recommender Systems Research." Expert Systems with Applications 39 (11): 10059–10072.

9. L. Su, "Between homomorphic signal processing and deep neural networks: Constructing deep algorithms for polyphonic music transcription," in 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Dec. 2017, pp. 884–891, doi: 10.1109/APSIPA.2017.8282170.

10. G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," Applied Acoustics, vol. 158, Jan. 2020, Art. no. 107020, doi: 10.1016/j.apacoust.2019.107020.