ILLEGAL LOGGING, POACHING AND ELEPHANT CALL DETECTION REPORT

PREPARED BY

Yash Sharma

EXECUTIVE SUMMARY

Excessive timber exploitation above the prescribed limits is known as illegal logging. Companies and people continue to intrude deeper into forests as worldwide demand for commodities such as paper, wood, and palm oil rises, exacerbating the illegal logging problem. Illegal logging destroys trees that serve as carbon sinks and climate regulators, contributing to global warming and climate change. Aside from raising global temperatures, the practise has resulted in a loss in total forest cover, exposing the majority of the world's territory to high temperatures and extreme weather.

Poaching is the greatest current threat to tigers, rhinos, elephants, gorillas and other African and Asian species. It’s a crime and it’s driving species to extinction. Tigers and rhinos are particularly vulnerable, their body parts being prized in traditional Asian medicine.

In this report we discuss a deep learning model which can detect and reduce the current rate of logging and poaching. The model also consists of a class namely ‘elephant_call’, which is used to detect the presence of an Elephant. Finally the last class is composed of noise data which is essentially required to chain another model which detects elephant rumbles with this one.

Brief Overview of Proposed Technique

The most widely used technique to classify between different sounds is using the Mel Frequency Cepstrum Coefficient (MFCC). MFCCs allow us to extract the linguistic content and discard all the other stuff which carries information like background noise. In more formal terms the job of MFCCs is to determine the shape of vocal tract which could give us an accurate representation of the phoneme being produced, this shape determines what sound comes out and is regarded as an important feature to distinguish between different sounds.

Model Overview and Results

In the upcoming section we'll discuss a Convolutional Neural Network architecture which uses 13 MFCC features to make predictions.

Data Extraction and pre-processing:

There are 4 classes in the model that need to be classified, chain_saw, gun_shot, elephant_call and lastly a noise class. The data for gun_shot audio collection was supported by NIJ grant 2016-DN-BX-0183 "Development of Computational Methods for the Audio Analysis of Gunshots.". Other classes' data was acquired using Google's Audio Set data collection.

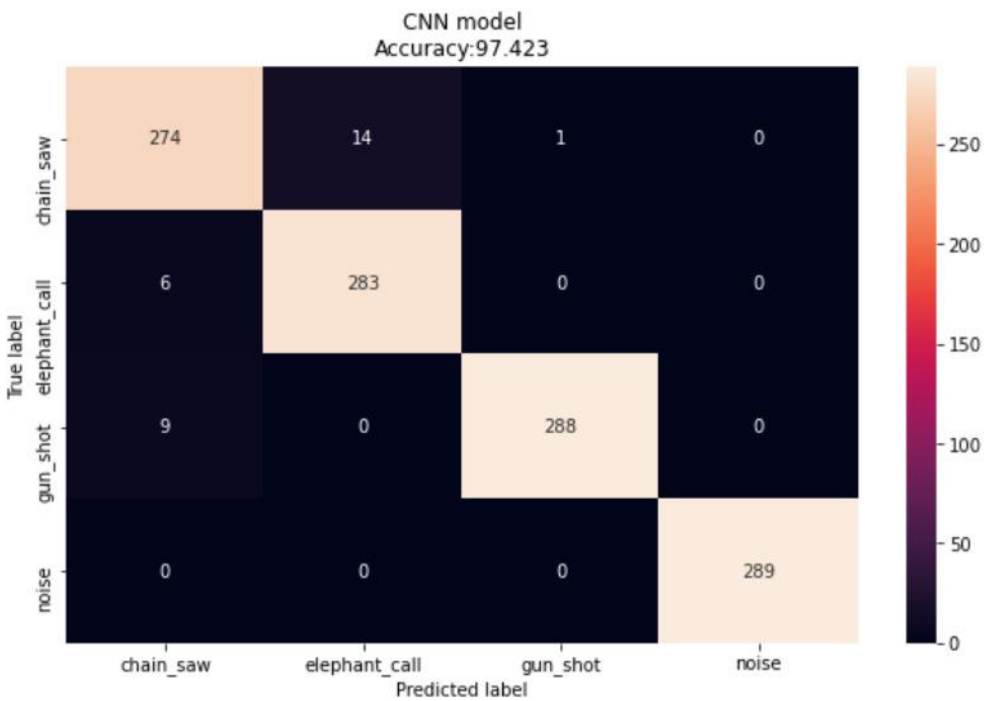
Since the length of the audio file for 'gun_shot' class ranged between 1 to 2 seconds. All other audio files were also trimmed down to 2 seconds each. The final sampling rate of the audio sample was set to 22.5 kHz.

Model Architecture:

The CNN model is based on the sequential API of the tensorflow library. First layer is Conv2D layer with a size of 64 output space in other words 64 filters with a kernel size of 3 followed by a batch normalisation layer, an activation layer with activation function as 'relu' and a max pooling layer of pool size 2. These 4 layers are again applied but this time the number of filters is reduced to 32. Lastly the model features are flattened and two batch normalised dense layers with 64 and 32 parameters are added. Finally to classify between the 4 classes a final dense layer with 4 neurons is added. Adam optimiser is used for optimization of gradient descent. To prevent model from overfitting, regularization methods such as batch normalisation and max pooling layer were employed.

Results:

1. Model Accuracy on training set: 99.33 %
2. Model Accuracy on validation set: 98.38%
3. Model Accuracy on test set: 97.42 %
4. Confusion Matrix:



.