ELEG5491 Project Initial Proposal

name: YU Si Hong

student ID: 1155141630

Introduction

As an undergraduate, I would like to choose a Kaggle challenge as my project, which is called "Multi-label Bird Species Classification - NIPS 2013" (https://www.kaggle.com/competitions/multilabel-bird-species-classification-nips2013/overview). It is a classification task about the bird songs, which of 87 classes of birds and amphibians are present into 1000 continuous wild sound recordings.

The dataset is relatively small, so it is easy to handle in personal computer. The task is quite old, so I want to see if nowaday's deep learning models can outperform the previous models.

Data

- Input:

The training and test set contain 687 and 1000 files respectively. The input data can be WAV format files or MFCC (Mel-frequency cepstral coefficients) files.

- Output:

The objective prediction is the predicted species for each records.

Proposed Method

It would be interesting to deploy deep learning models to solve this problem. As audio is difficult to process, in comparison to images.

As a sequential data, I think RNN maybe a good choice. I will try to use LSTM to process the audio data. To beat the baseline, fine-tuning the pre-trained model is a good choice. And I will try to optimize the parameters of the model. Data augmentation is promised to improve the performance of the model. But t is not easy to do data augmentation on audio data. I will also explore the ways to do data augmentation.