# Brief

The py files are based on yeyuepiaoling’s VoiceprintRecoginition-Tensorflow project on GitHub and I modified them to fit in our project (parameters of audio files are the same in **RecordTrainingData.ipynb** and the data list will be similar format which contains file name and labels).

The audio is recorded and then the data are being processed using short-time Fourier transform for obtaining a 257\*257 spectrogram. They are trained using a modified ***resnet34*** model with input shape as [None, 1, 257, 257] to fit the spectrogram. The model is not trained to remember certain amount of specified people talking but to train to extract features from new audio and distinguish them. Therefore, for our project it is essential to not only train using our own voices but also with public data. With just 20 files for 5 people, it will surely go into overfitting.

**Train.py** is for model training. It loads the data list in format “file\_path label” and use buffer to split and shuffle per 1000 samples. **Eval.py** is used for evaluating the model. **Infer\_contrast.py** and **infer\_recognition.py** are two scripts for practical use, where the first one can be used to tell if two audio file belongs to the same person and the second one can be used to register new member into database and can be used to tell if the one currently speaking is registered to the database.

Things to notice:

1. **Create\_data** is not used as it is used to transfer mp3 files into wav files which is something we already did on RecordTrainingData.ipynb. We do need a py file to create a list containing data and label.

2. The time duration for a person’s speaking should be no less than **1.3 s**. This is because sample rate is 16k, window len 400 with separation 160. Input shape is 257, so the time is 1/16000\*(400+160\*256) = 2.585. When doing preprocess of the data the audio is flipped and concated so the original voice should be half which is 1.3 s.