Speech Enhancement using Deep Learning

Introduction

Whenever we work with real time speech signals, we need to keep in mind about various types of noises that gets added to the original noise and hence resulting in corruption of noise. Therefore, in order to make a better sense of the signals, it is very much necessary to enhance the speech signals by removing the noises present in them.

Requirements:

- Tensorflow version 2.2.0
- Pyhton 3.7.6
- Librosa library

Dataset used:

The dataset used for this project is TCD-TIMIT speech corpus,a new Database and baseline for Noise-robust Audio-visual Speech Recognition.

Downloadable link for the dataset:

You can find the complete dataset here https://zenodo.org/record/260228

Clean dataset link: https://zenodo.org/record/260228/files/Cafe_5.tar.gz?download=1

Noisy dataset link: https://zenodo.org/record/260228/files/Cafe_-5.tar.gz?download=1

Approach followed:

Step 1: downloaded the dataset from the above link

Step 2: rename the dataset downloaded to Café_-5 for noisy and Café_5 for clean.

Step 3: unzip the tar file

Step 4: Found the mfcc for every file ,saved it in to pickle file.

Note :I have provided two files noisycafe.pickle and Cleancafe.pikle for noisy data and cleaned data in this repository. You can directly use these file for training the model and skip first 3 steps.

Architecture used:

I have used Autoencoder here:

```
2 import keras
3 from keras import layers
4 input_img = keras.Input(shape=(X_train.shape[1],X_train.shape[2], 1))
5 x = layers.Conv2D(64, (3, 3), activation='relu', padding='same',kernel_initializer='glorot_uniform')(input_img)
6 x = layers.Dropout(0.3)(x)
7 x = layers.MaxPooling2D((2, 2), padding='same')(x)
8
9 x = layers.Conv2D(64, (3, 3), activation='relu', padding='same',kernel_initializer='glorot_uniform')(x)
10 x = layers.Dropout(0.3)(x)
11 x= layers.Conv2D(32, (3, 3), activation='relu', padding='same',kernel_initializer='glorot_uniform')(x)
12 x = layers.Dropout(0.3)(x)
13 encoded = layers.MaxPooling2D((2, 2), padding='same')(x)
14
15
16 x = layers.Conv2D(32, (3, 3), activation='relu', padding='same',kernel_initializer='glorot_uniform')(encoded)
17 x = layers.Dropout(0.3)(x)
18 x = layers.UpSampling2D((2, 2))(x)
19 x = layers.Conv2D(64, (3, 3), activation='relu', padding='same',kernel_initializer='glorot_uniform')(x)
20 x = layers.Dropout(0.3)(x)
21 x = layers.UpSampling2D((2, 2))(x)
22 decoded = layers.Conv2D(1, (3, 3), activation='relu', padding='same',kernel_initializer='glorot_uniform')(x)
```

Reverences use:

1. https://github.com/achaitu/SpeechDenoisingDNN
2. https://www.youtube.com/watch?v=_xcFAiufwd0&list=PL-wATfeyAMNrtbkCNsLcpoAyBBRJZVlnf&index=13