BTP Project Update-1

High level Progress Overview

We have started by implementing automatic speech recognition trained on a very large clean speech dataset of raw audio file and corresponding transcript. Dataset used for this particular purpose is the Corpus Librispeech dataset.

Our speech recognition model is an end to end deep neural network model. We have performed audio preprocessing and implemented the architecture of the model till now.

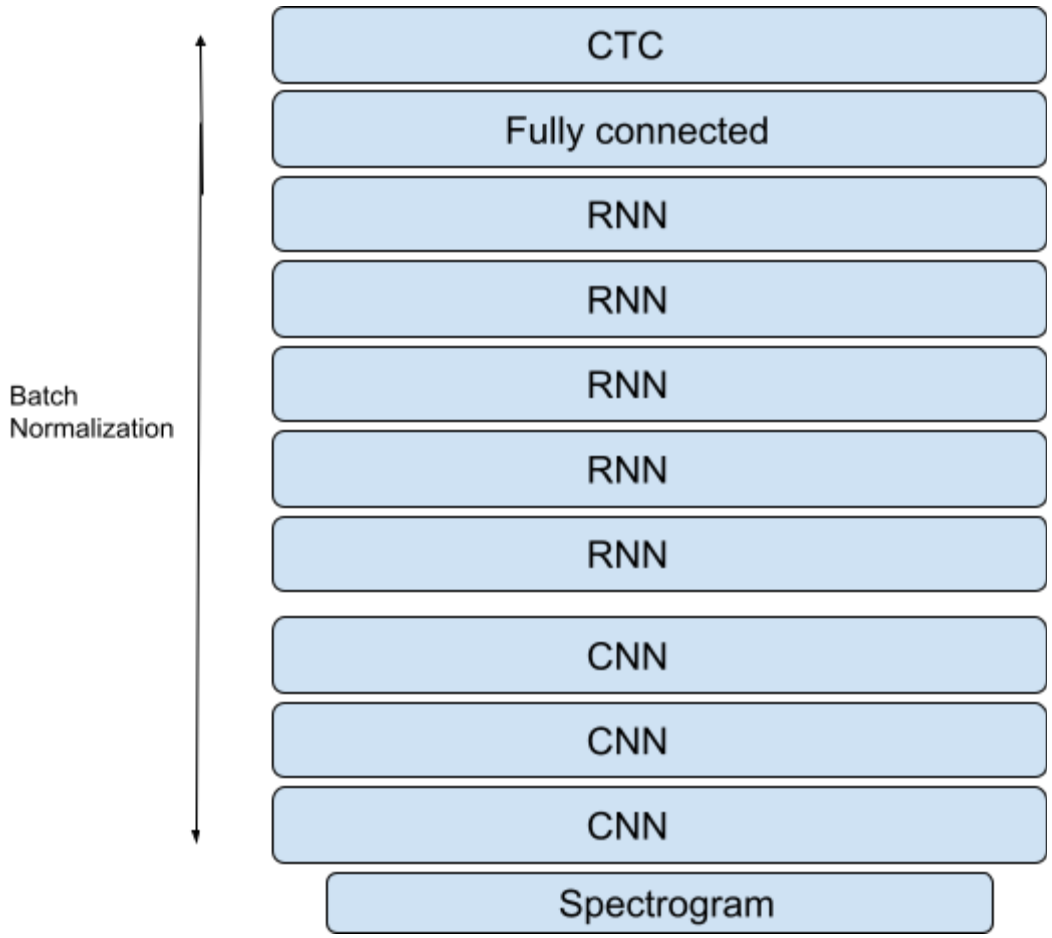
Low level Progress Overview

Steps performed till now:

- Data preprocessing
- Raw audio data augmentation
- Generation of Mel spectrograms
- MFCC
- Neural network architecture building

Architecture of our model:

The architecture of our model will use Convolutional neural network(CNN) as well as Recurrent neural network(RNN) packed with CTC(Connectionist temporal classification) function that will calculate the WER(Word error rate) for our model. Below is the architecture of our model before transfer learning.



Next goals to perform:

1. Training the model
2. Evaluating the model performance by calculating the WER(Word error rate) and CER(character error rate)
3. Working on gathering impaired speech dataset
4. Applying transfer learning paradigm
5. Building text to speech model

Future plan of action:

After we are done with end to end ASR model training, we will start working on how to gather impaired speech dataset so that we can apply transfer learning.

References:

1. https://arxiv.org/pdf/1512.02595v1.pdf
2. https://arxiv.org/pdf/1412.5567.pdf
3. https://www.biometricupdate.com/201906/google-building-impaired-speech-dataset-for-speech-recognition-inclusivity
4. https://ai.googleblog.com/2019/08/project-euphonias-personalized-speech.html