BTP Project Update-4

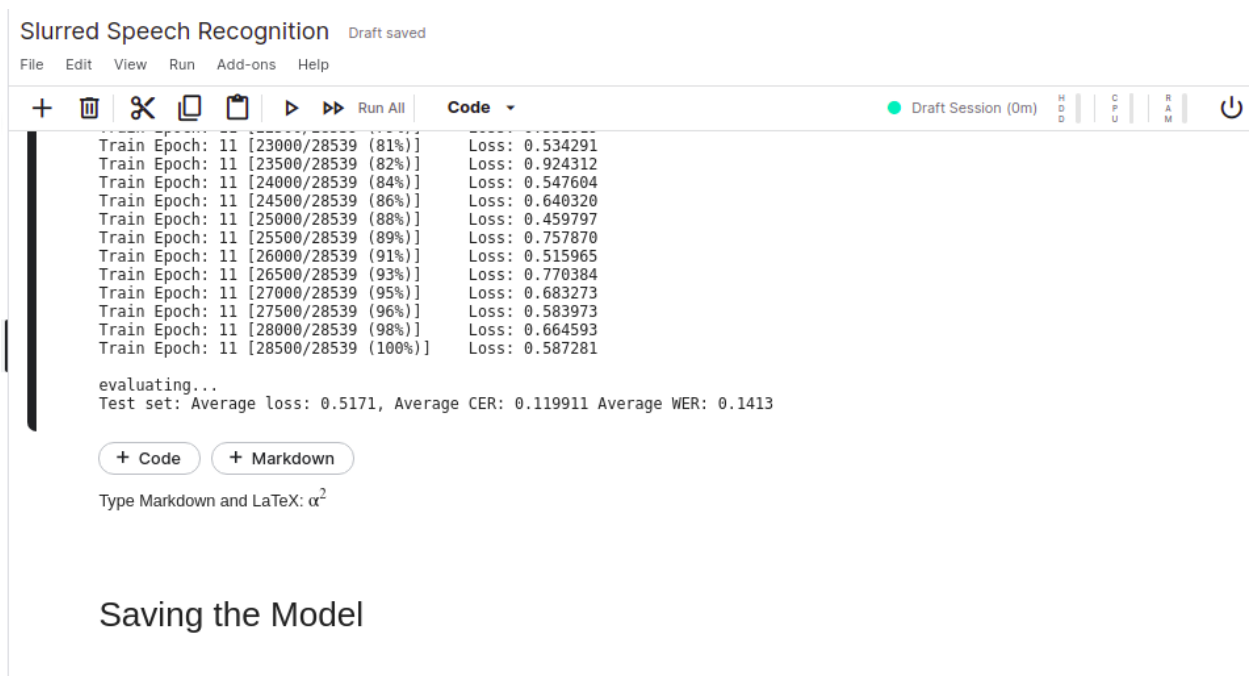
High-level Progress Overview

Last week, we completed the training of our end-to-end ASR deep learning model prior to transfer learning. We were also able to save the best fit model with a **Word error rate** of 0.17 and a **Character error rate** of 0.14 with epoch=5.

This week we performed mainly two tasks.

- Personalized Speech Dataset collection and preparation for fine-tuning our base ASR model.
- Improving the performance of our base model.

Our Base model WER now is 0.14 and CER now is 0.11.



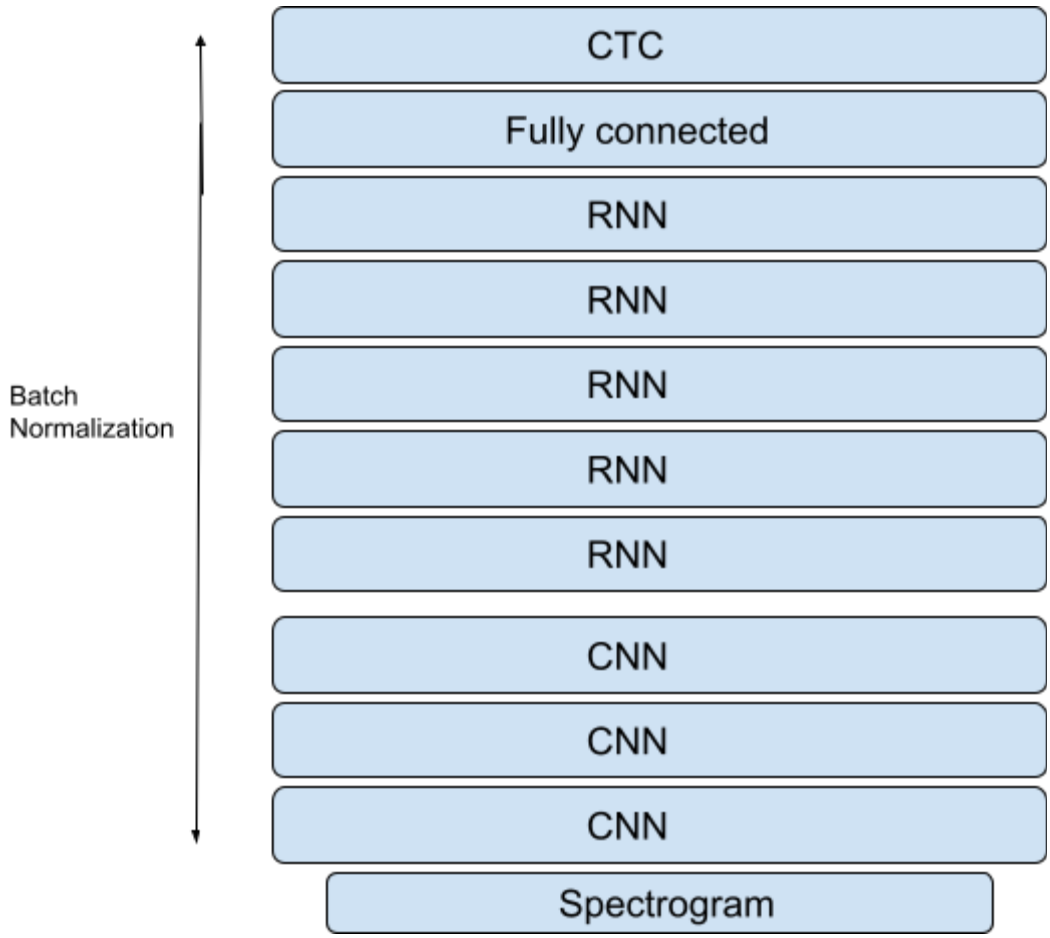
Low level Progress Overview

Steps performed till now:

- Data preprocessing
- Raw audio data augmentation
- Generation of Mel spectrograms
- MFCC
- Neural network architecture building
- Model Training
- Model Evaluation
- Saving Best fit Model
- Built Speech dataset Collection API using Django

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.



Model Performance:

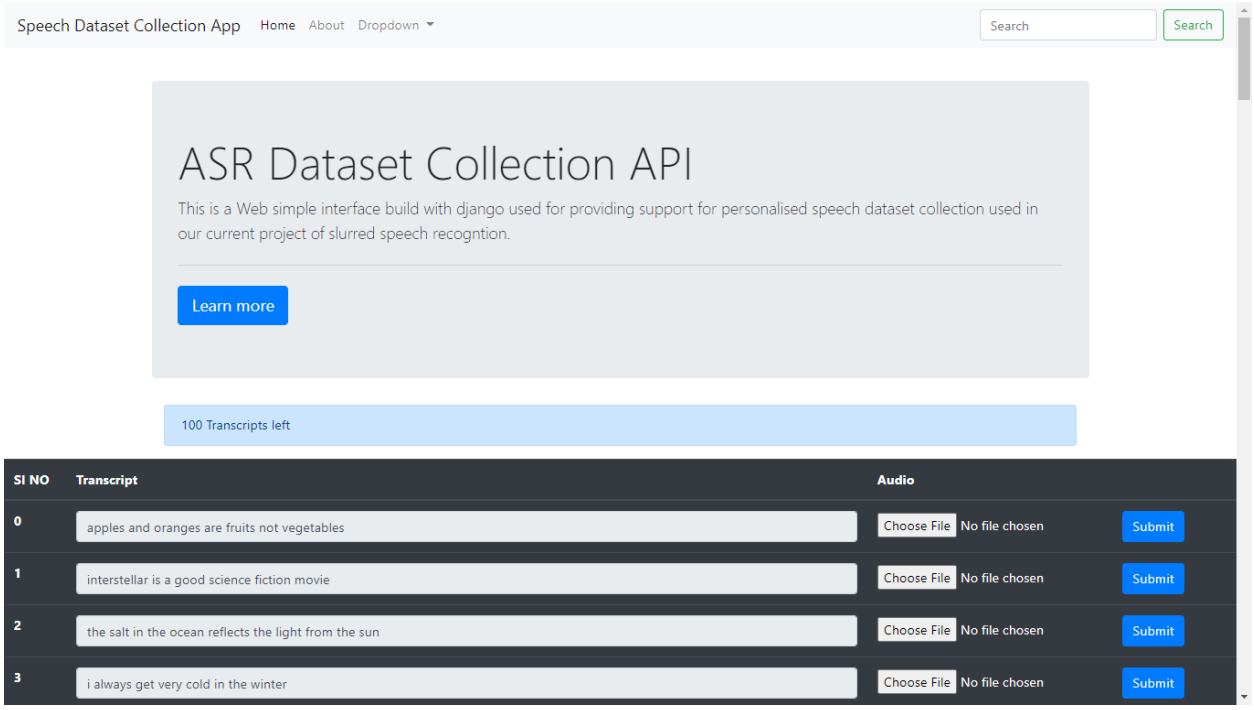
- Final Epoch Average Loss: 0.61
- Final Epoch Average CER: 0.11
- Final Epoch Average WER: 0.14

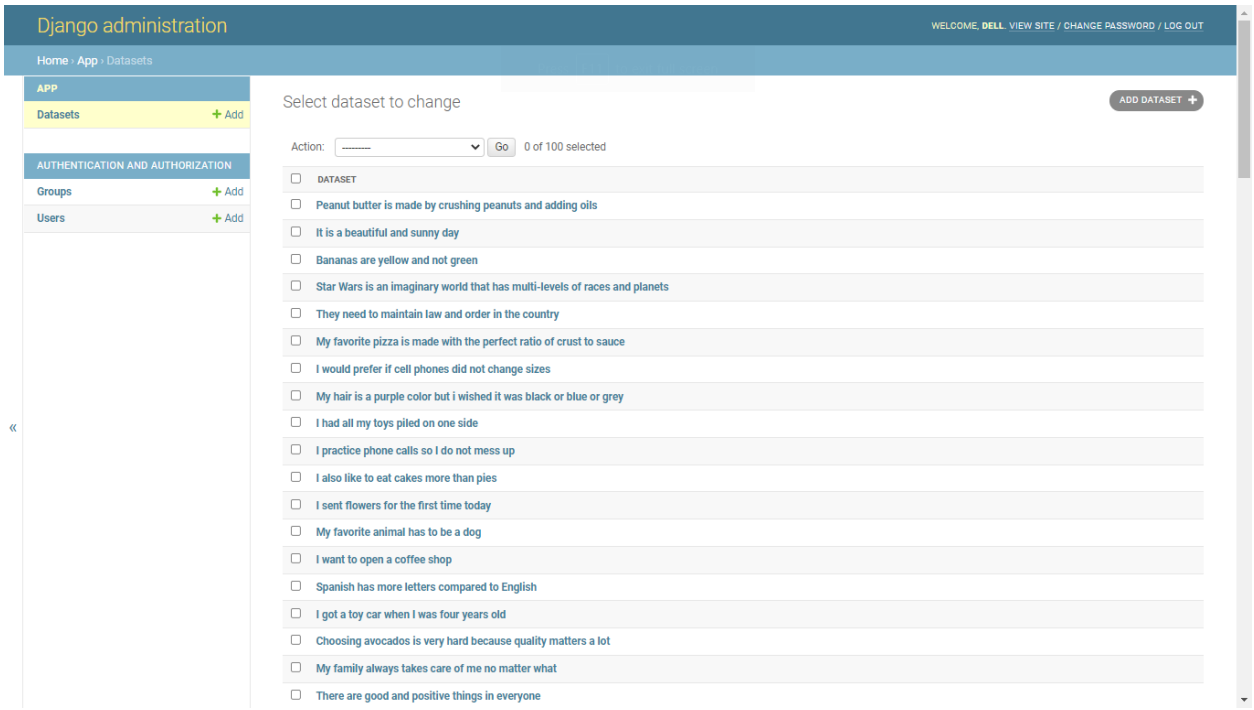
Testing our loaded base model

```
[ ] test(loaded_model, device, test_loader, criterion, 1, iter_meter)
```

```
evaluating...
Test set: Average loss: 0.5171, Average CER: 0.119907 Average WER: 0.1413
```

Snapshot of the Dataset collection Interface:





Link to web APP:

1. https://mmig.github.io/speech-to-flac/
2. http://speech-collection.herokuapp.com/index/

Next goals to perform:

1. Applying transfer learning paradigm
2. Building text to speech model

Future plan of action:

Since we are done with end to end ASR model training and built a solution for gathering impaired speech dataset , we can now focus on transfer learning. After that we will implement the next phase of our project i.e building text to speech model.

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