Capstone Project Weekly Progress Report

Semester	Fall 2022
Course Code	AML 2404
Section	Section 2
Project Title	Generating Synthetic Voice
Group Name	Group A: Goal Diggers
Student names/Student IDs	1. Abhimanyu Sharma (C0859053) 2. April Alcantara (C0858528) 3. Harsharan Singh Raina (C0857931) 4. Krati Rastogi (C0858753)
Reporting Week	Week 4
Faculty Supervisor	William Pourmajidi

1. **Tasks Outlined in Previous Weekly Progress Report** (Provide detailed information on the tasks to be completed in this week)

These are the tasks outlined to be completed this week.

- Krati and April as mentioned analyzed the data further, but due to the large size of the dataset, it was taking a lot of time. So, the team decided to download the dataset in the local disk and then work on it.
- Harsh and Abhimanyu studied about conformer model and how it can be integrated in the project along with other models.

2. **Progress Made in Reporting Week** **(Provide detailed information on the progress that you made in the reporting week. Limit your write-up to no more than two page)**

This week's Task:

Professor William mentioned that integrating ML Ops would advance our project technology-wise. The team decided to dedicate two hours of the class to take up the LinkedIn learning course 'MLOps Essentials: Model Development and Integration' to familiarize ourselves with the concept of the purpose of applying it in our project.

The findings by the team are mentioned below:

Machine Learning Activities covers:

- Build focuses on creating and testing models and its core activities are feature engineering, training, testing, and packaging.
- It also covers deploying, executing, and implementing the model where the core activities include model deployment and inference.

Machine Learning lifecycle continuously refines the new training data, new user requirements and model decay.

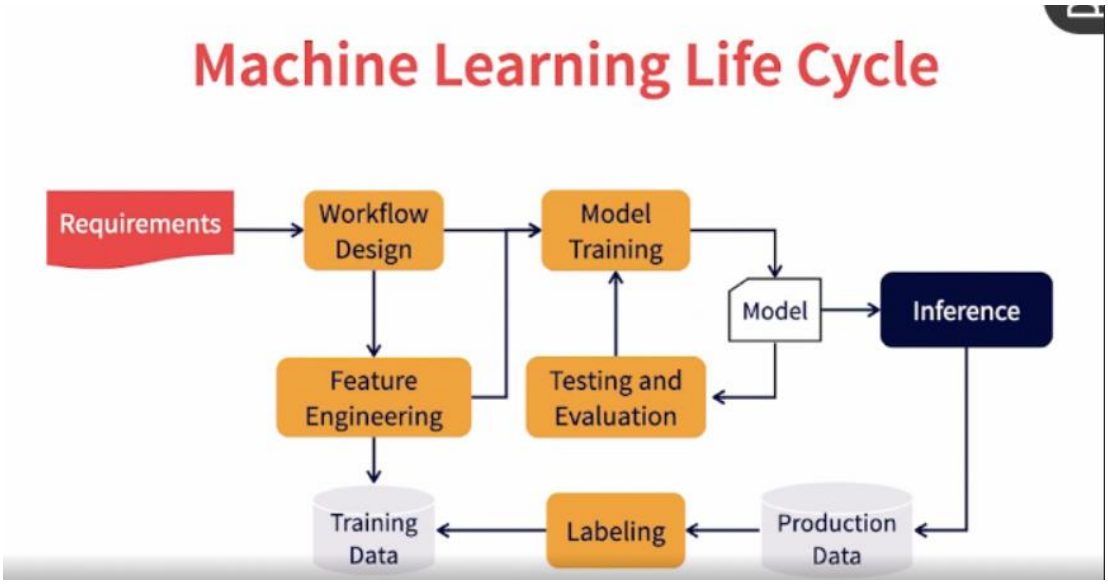

Figure 1: Machine Learning Lifecycle

Machine Learning Operations is a set of best practices that helps manage the creation and use of ML artifacts through efficient workflows, collaboration, and tracking.

Machine Learning Operation elements are:

- Extend DevOps to manage machine learning problems
- Integrate data engineering and model development into software engineering lifecycle
- Enable continuous model development and integration
- Serve ML models and provide feedback
- Manage ML process effectively through automation and tools

April and **Abhimanyu** worked on the Google Colab notebook.

- Downloaded and loaded the conformer encoder model from Kaggle

```
od.download('https://www.kaggle.com/datasets/tuannguyenvananh/conformer-ctc-model')
```

```
Please provide your Kaggle credentials to download this dataset. Learn more: http://bit.ly/kaggle-creds
Your Kaggle username: aalcan
Your Kaggle Key: ··········
Downloading conformer-ctc-model.zip to ./conformer-ctc-model
100%|██████████| 572M/572M [00:06<00:00, 87.9MB/s]
```

```
file =('conformer-ctc-model/\
conformer_ctc_subsampling.state')
file
```

```
'conformer-ctc-model/conformer_ctc_subsampling.state'
```

```
checkpoint = torch.load(file, map_location={'cuda:0': 'cpu'})
```

```
checkpoint
```

```
{'epoch': 14,
 'model_state_dict': OrderedDict([('encoder.encoder.conv_subsample.sequential.0.weight',
              tensor([[[[-0.0563, -0.3823,  0.0219],
                        [ 0.2353,  0.3854,  0.2552],
                        [-0.1907,  0.0550, -0.0325]]],


                      [[[-0.1255, -0.3426, -0.4033],
                        [-0.0517,  0.3258, -0.1315],
                        [-0.0560,  0.3776,  0.0674]]],
```

Figure 2: Download and loading of the conformer model

- Downloaded and loaded the Librispeech dataset from Kaggle rather than using the dataset on cloud

```
od.download('https:://www.kaggle.com/datasets/victorling/librispeech-clean')
```

```
Please provide your Kaggle credentials to download this dataset. Learn more: http://bit.ly/kaggle-creds
Your Kaggle username: aalcan
Your Kaggle Key: ··········
Downloading librispeech-clean.zip to ./librispeech-clean
100%|██████████| 28.1G/28.1G [03:29<00:00, 144MB/s]
```

Figure 3: Download LibriSpeech Dataset

- Started with the data exploration
  - Length of the train set, dev set, and test set are 132,553, 2,703 and 2,620, respectively.
  - The datasets are of torch.utils.data.Dataset type

```
[33] train_set = LibriSpeechDataset('train')
     dev_set = LibriSpeechDataset('dev')
     test_set = LibriSpeechDataset('test')
     len(train_set), len(dev_set), len(test_set)

     (132553, 2703, 2620)

     type(train_set), type(dev_set), type(test_set)

     (__main__.LibriSpeechDataset,
      __main__.LibriSpeechDataset,
      __main__.LibriSpeechDataset)
```

Figure 4: Librispeech dataset

Krati and **Harsharan** read up on the conformer encoder model. We have learned that:

- This specific model is a **Transformer and Convolution neural network (CNN)** based models which has shown promising results in **Automatic Speech Recognition (ASR),** as compared to **Recurrent neural networks (RNNs)** in order to see how it can be applicable to our project.
- Conformer is a mixture of convolution-augmented transformer
- It gives better accuracy than CNN and Transformers
- The conformer consists of two feed-forward layers that resemble macarons, with half step residual connections separating the convolution and multi-headed self-attention modules.

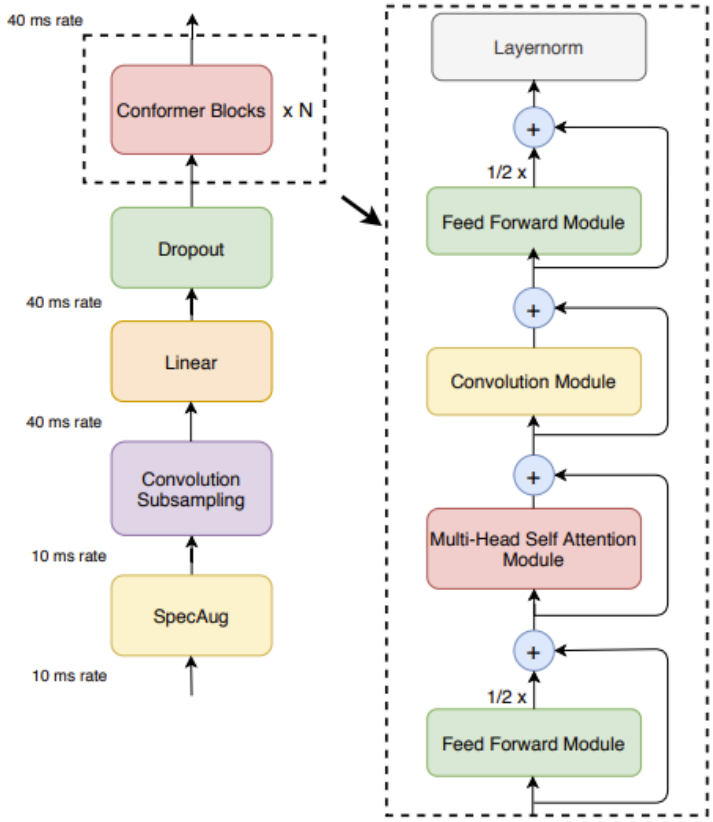

Figure 5: Conformer encoder model architecture

3. **Difficulties Encountered in Reporting Week** (Provide detailed information on the difficulties and issues that you encountered in the reporting week. Limit your write-up to no more than one page)

The difficulties we encountered this week are:

- Faced runtime issues when loading the conformer model

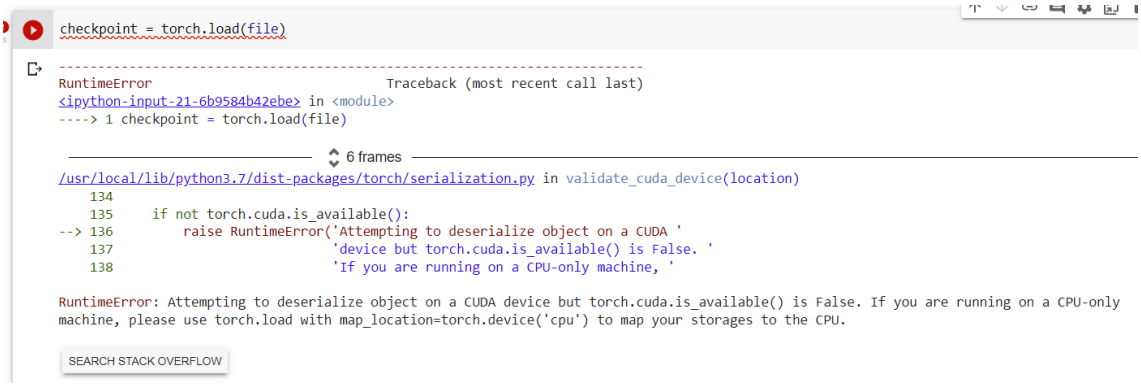

Figure 6: Error in trying to load the conformer model via torch.load() method

To fix this error, we remap the Tensor location at load time using the map_location argument at torch_load() as suggested from Stack Overflow.

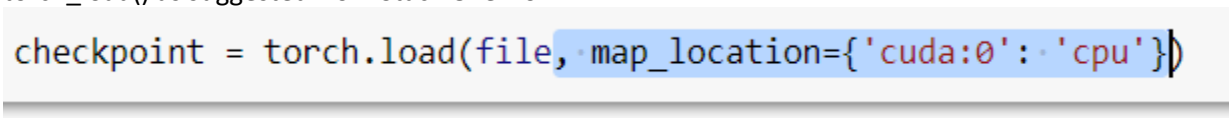

Figure 7: Code change to remap the Tensor location

2. **Tasks to Be Completed in Next Week** (Outline the tasks to be completed in the following week)
   - Study how to integrate ML Ops in our project
   - Explore if there is a way to download the Kaggle dataset in Google Colab without having to supply the Kaggle login token information
   - Research about the models which we will be using and how to integrate them in the project.

References:

Conformer: Convolution-augmented Transformer for Speech Recognition (2020, May 16).

https://arxiv.org/pdf/2005.08100.pdf

RuntimeError: Attempting to deserialize object on a CUDA device (2019, May 30). Stack Overflow. Retrieved October 9, 2022 from

https://stackoverflow.com/questions/56369030/runtimeerror-attempting-to-deserialize-object-on-a-cuda-device

conformer-ctc-librispeech. (2022, March 9). Kaggle. Retrieved October 2, 2022, from https://www.kaggle.com/code/tuannguyenvananh/conformer-ctc-librispeech/data

Starter: LibriSpeech 8da5c067-f. (2019, June 13). Kaggle. Retrieved October 2, 2022, from https://www.kaggle.com/code/kerneler/starter-librispeech-8da5c067-f/data

English Multi-speaker Corpus for Voice Cloning. (2019, December 17). Kaggle. Retrieved October 2, 2022, from https://www.kaggle.com/datasets/mfekadu/english-multispeaker-corpus-for-voice-cloning