**EIE558 Speech Processing and Recognition**

**Lab 2: DNN-Based Speech Recognition**

**A. Objectives and Outcomes**

After finishing this lab, you should be able to perform the following:

* Use Google Colab to perform speech recognition.
* Use PyTorch to implement a 1D-CNN for recognizing spoken digits.
* Understand the role of convolution in CNN for speech recognition.
* Use Python speech recognition library to recognize continuous speech.

**B. Google Colab**

Google Colab is a free cloud service with GPU support. You may use Colab to develop deep learning applications based on popular libraries such as Keras, TensorFlow, PyTorch, and OpenCV.

**C. Submission**

Write a report, convert it to PDF, and submit it to Blackboard before the deadline. Your report may contain the following:

1. Discussions on your observations, e.g., the correlation between network depth and performance, the role played by the CNN with different kernel size, etc.
2. Waveforms and spectrograms of clean, noisy, and denoised speech.
3. The effect of reducing the number of feature maps in each convolutional layer
4. The difference between statistics pooling and PyTorch’s adaptive average pooling.
5. Performance of different pooling methods.
6. The effect of some other modifications that you can think of.

**D. Procedures**

***D.1 Prepare Colab Environment***

1. Colab runs on browsers. You need a Google account to use Colab. If you do not have one, visit https://support.google.com/mail/answer/56256?hl=en.[[1]](#footnote-1)
2. Display the Google Drive page (https://drive.google.com/drive/my-drive) in your browser. Use the “+ New” button on the left panel to create the following directory structure in your Google Drive:

My Drive/Learning/EIE558/

After creating the folders, you should see something like this:

Graphical user interface, text, application, chat or text message

Description automatically generated

Directory structure in Google Drive

Later, you may mount your folders in Google Drive to the working environment of Colab.

1. Go to <http://bioinfo.eie.polyu.edu.hk/download/EIE558/asr> and download the Colab file “digitrec.ipynb”. Remove the .txt extension if your browser adds one for you. Upload the file to your Google Drive under “My Drive/Learning/EIE558/asr”. Repeat the same procedure for “sphrec.ipynb”.
2. Right click the file “digitrec.ipynb” on your Google Drive folder and then select “Open with”. If you see “Google Colaboratory”, select it. If your Google account has not been connected with Colaboratory yet, select “Connect more apps” and then select Colaboratory. Then, click “Install”. Then, right click the file again and select “Google Colaboratory”.

Graphical user interface, text, application, chat or text message

Description automatically generated

1. Configure the Colab to use GPU by clicking Edit 🡪 Notebook settings. Select “GPU” in the pop-up window:

Graphical user interface, text, application, chat or text message

Description automatically generated

1. Follow the procedures in “digitrec.ipynb”. Report and explain your observations. Note that you may use the pre-trained models in the models/ folder to avoid the lengthy training process. Alternatively, you may use the trimmed dataset, as shown in the Colab file.
2. Follow the procedures in “sphrec.ipynb”. Report and explain your observations.

**References:**

1. <https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d>
2. H.S. Choi et al., “The algorithm and code are based on the paper "Phase-Aware Speech Enhancement with Deep Complex U-Net" (2019)

-- END --

1. If you use Anaconda environment, you must start an anaconda environment with PyTorch installed. Launch Jupyter notebook from that environment. For example:

   $ conda activate myenv

   $ juypter notebook

   You will see Jupyter on your browser and you are ready to go [↑](#footnote-ref-1)