ARS take-home task

Emotech

Instructions

On completion of the task please return a PDF file called report.pdf with the answers along with the repository containing all the code compressed either as a zip file or gzip tarball.

The focus of this task is to build and train a model for speech, music and noise detection. The architecture of interest is a SwishNet [1] neural network. Please familiarize yourself with the paper and complete the following set of subtasks:

1. Implement the SwishNet model in Tensorflow 1.15 (or >=1.10 if incompatible with your CUDA version). Feel free to use this open source Keras implementation as a reference [2]. Please highlight in the report an experiment showing that the outputs of your TF implementation and the Keras one match on the same input matrix.
2. Download the MUSAN dataset [3], prepare the preprocessing pipeline as described in [1] and obtain the preprocessed features of the dataset. Please highlight all of the steps you take in the report.
3. Prepare a training script for your Tensorflow SwishNet and train on MUSAN. Use [1] as a reference for relevant configurations.
4. Download the GTZAN dataset [4] and prepare it according to [1]. Benchmark your trained SwishNet according to Table II in [1]. Is the performance close to that reported (Undistilled version)? If not, can you think why? If the performance is very different train a Feedforward Neural Network (eg. the SNN in [1]) of similar size to your SwishNet and evaluate whether using SwishNet would still provide gains over a simple baseline at the relevant task.
5. The SwishNet paper was out in 2018. a) Summarize the main novelties of the model, b) have any new approaches improved over SwishNet at the same task? b) are there any novel approaches to similar classification tasks that you would try on this task to improve over SwishNet? Please highlight the main insights / novelties from the models you mention.

References

[1] Md. Shamim Hussain and Mohammad Ariful Haque. Swishnet: A fast convolutional neural network for speech, music and noise classification and segmentation. *CoRR*, abs/1812.00149, 2018.

[2] Swishnet keras implementation, `https://github.com/i7p9h9/swishnet`.

[3] David Snyder, Guoguo Chen, and Daniel Povey. MUSAN: A music, speech, and noise corpus. *CoRR*, abs/1510.08484, 2015.

[4] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.