Topics in Speech Recognition on the Edge

Speech recognition is a challenging problem. Devices performing speech analysis of any sort usually rely on rather large models and large training datasets. Many of speech recognition tasks, however, do not need such large models, and many other complex speech-related tasks can be subdivided into sub-tasks that can be individually handled by relatively small models.

We focus on speech recognition "on the edge" – the sort performed by embedded systems or smart devices. Our aim is to develop small (about 20 to 100k parameters) neural models that can handle simple speech analysis tasks and intelligently call on a more potent shared resource when the model no longer suffices. Examples of our interests include speech enhancement, speech re-segmentation and canonisation, voice activity detection, wake-word detection, and parts of speech diarisation: unsupervised speaker profiling, profile activity detection.

About the thesis. You are going to be working very closely with DISCO group members, advancing the research of the group. In the first few weeks you will get familiar with the related work, and you will have the option to decide whether you want to pursue the topic further. Going forward, we will prepare a plan for the rest of your time with us while still leaving enough time for you to independently expand on the core objectives of your project.

Candidate Profile. A good candidate is a competent programmer in the language of his/her choice, has good knowledge of or solid experience with TensorFlow or (Py)Torch, has past experience with embedded system development or single-board computers, and is interested in one or more of the following fields: edge ML, speech recognition, deep learning for embedded systems, deep learning for time series.

Interested? Please contact us to learn more!

Contact

- Peter Belcak: belcak@ethz.ch, ETZ G61.3