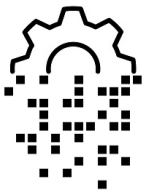
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

"I'm Done Talking" Automatic Profile Binding in Speech Recognition



The goal of this thesis is to implement a variant of the Mx convolutional neural network model [1] combined with a naive learning algorithm for detecting when the first speaker of an audio sample has finished speaking.

An additional constraint of our model is that it has to be suitable for machine learning on the edge – it needs to have at most 30-40k parameters and be ready to operate with raw waveforms in potentially noisy conditions, in real time.

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 our core objectives.

Candidate Profile. Generally speaking, 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, and is interested in one or more of the following fields: edge ML, speech recognition, deep learning for time series.

Interested? Please contact us to learn more!

Contact

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

References

[1] Wei Dai, Chia Dai, Shuhui Qu, Juncheng Li, and Samarjit Das. Very deep convolutional neural networks for raw waveforms. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 421–425. IEEE, 2017.