Projets de fin d'étude // Master internships
Self-supervised audio representation learning

Keywords: audio and speech processing, deep learning, self-supervised learning, transfer learning

Self-supervised representation learning (SSRL) is the task of training a model (neural network) with labels obtained for "free". That is, part of the data is used as labels to be predicted by the model, using the rest of the information from these data. For instance, in speech, we can imagine a SSRL task where we ask a model to predict the next acoustic frames given the past ones.

Transfer learning is the task of adapting a model trained on a given task to our task at hand. We may retrain ("fine-tune") the whole model or part of it, depending on several factors (data size, task nature, etc.)

Self-supervised speech representation learning has recently witnessed a huge increase of interest from both the machine learning and audio (mostly speech) processing research communities: three benchmarks have been announced almost simultaneously by three independent research teams: *SUPERB*[1], *LeBenchmark*[2], and *HEAR*[3]. While *SUPERB* and *LeBenchmark* focus on the speech signal, *HEAR* aims at evaluating audio representation in other domains as well (such as music, environmental sounds, etc.).

In this internship, we propose to investigate transfer learning applied to self-supervised speech representation, with a focus on the French language:

1. Literature review on self-supervised speech representation learning
2. Comparison of existing pretrained models and their performance on transfer learning applied to *LeBenchmark* datasets
3. Training most promising approaches on a meta-dataset made of a large range of speech domains
4. Comparison of these models with existing pretrained models

In practice, those projects will be implemented in *pytorch* and experiments will make good use of *JeanZay* supercomputer[4].

Please send your application (including CV, grades and references)

- to thomas.pellegrini@irit.fr and herve.bredin@irit.fr
- with subject "Internship about self-supervised audio representation learning"

Location: France, Toulouse, University of Toulouse III Paul Sabatier, IRIT, SAMoVA team
Date and duration : 5 to 6 months, possibly leading to a PhD or research engineer position
Allowance: around 600 euros monthly

Qualifications: good programming skills, machine learning background, English proficiency

[1] superbbenchmark.org
[2] github.com/LeBenchmark
[3] neuralaudio.ai
[4] Google it: lots of GPUs!