Deep Learning for audio scene recognitions (mésocentre CRIANN)

Alain Rakotomamonjy[*1]

[1]Laboratoire d'informatique, de traitement de l'information et des systèmes (LITIS) – INSA - Institut National des Sciences Appliquées, Université de Rouen, Université du Havre – France

Résumé

Machines can now watch and interpret images, recognize speech and music genres, yet they are hardly capable of understanding daily life sound events e.g the sounds that occur in a kitchen at morning.

Today's researches dealing with audio scene understanding are mostly limited to the problem of categorization and localization of few tens of sound event classes and environmental contexts. While such tasks are useful, the ultimate goal of audio scene understanding goes far beyond the assignment of labels to few kinds of sound events. Instead, it aims at developing machines that fully understand audio input. However, before making sense from audio, it is necessary to be able to recognize these audio contexts.

In this talk, we present the results we achieved by performing representation learning (using deep learning) for classifying audio scenes.

---

[*]Intervenant