GONTIER – Félix

Analyse et synthèse de scènes sonores urbaines par approches d'apprentissage profond

Prédiction d'attributs perceptifs et resynthèse de signaux temporels à partir de spectres issus de réseaux de capteurs acoustiques à grande échelle

L'avènement de l'Internet des Objets (IoT) a permis le développement de réseaux de capteurs acoustiques à grande échelle, dans le but d'évaluer en continu les environnements sonores urbains. Dans l'approche de paysages sonores, les attributs perceptifs de qualité sonore sont liés à l'activité de sources, quantités d'importance pour mieux estimer la perception humaine des environnements sonores. Utilisées avec succès dans l'analyse de scènes sonores, les approches d'apprentissage profond sont particulièrement adaptées pour prédire ces quantités. Cependant, les annotations nécessaires au processus d'entraînement de modèles profonds ne peuvent pas être directement obtenues, en partie à cause des limitations dans l'information enregistrée par les capteurs nécessaires pour assurer le respect de la vie privée.

Pour répondre à ce problème, une méthode pour l'annotation automatique de l'activité des sources d'intérêt sur des scènes sonores simulées est proposée. Sur des données simulées, les modèles d'apprentissage profond développés atteignent des performances « état de l'art » pour l'estimation d'attributs perceptifs liés aux sources, ainsi que de l'agrément sonore. Des techniques d'apprentissage par transfert semi-supervisé sont alors étudiées pour favoriser l'adaptabilité des modèles appris, en exploitant l'information contenue dans les grandes quantités de données enregistrées par les capteurs. Les évaluations sur des enregistrements réalisés in situ et annotés montrent qu'apprendre des représentations latentes des signaux audio compense en partie les défauts de validité écologique des scènes sonores simulées.

Dans une seconde partie, l'utilisation de méthodes d'apprentissage profond est considérée pour la resynthèse de signaux temporels à partir de mesures capteur, sous contrainte de respect de la vie privée. Deux approches convolutionnelles sont développées et évaluées par rapport à des méthodes état de l'art pour la synthèse de parole.

Mots-clés : Paysages sonores, Réseaux de capteurs acoustiques, Perception de sources sonores, Synthèse sonore

Visa du Directeur de Thèse