
Détection acoustique d'oiseaux

Responsable du projet : Xavier HINAUT

Contact : xavier.hinaut@inria.fr / www.xavierhinaut.com

Date / Lieu : Hackathon du HPC, 3 & 4 décembre 2018, Bordeaux, France

www.plafrim.fr/fr/evenements/

Mots-clés : oiseaux, sons, audio, détection, classification, réseaux de neurones artificiels, deep learning, réseaux convolutifs.

La détection des sons produits par des oiseaux est une tâche importante pour la surveillance automatique de la faune, ainsi que pour la science citoyenne et la gestion de la bibliothèque audio. La détection du son des oiseaux est une première étape très courante avant toute analyse ultérieure (par exemple, classification, comptage) et permet de travailler avec de grands ensembles de données (par exemple, une surveillance continue sur 24h) en filtrant les données dans des régions d'intérêt.

Ce sujet est basé sur une compétition qui a eu lieu cette année et dont la description complète est disponible ici : <http://dcase.community/challenge2018/task-bird-audio-detection>

La tâche de la compétition consistait à concevoir un système qui, après un court enregistrement audio, renvoie une décision binaire sur la présence / l'absence de son d'oiseau (son de toute sorte). La sortie peut n'être que "0" ou "1", mais les sorties pondérées / probabilités dans la plage continue $[0,1]$ sont encouragées à des fins de l'évaluation. Pour l'évaluation principale, la mesure utilisée est la performance de la classification "Surface sous la courbe ROC" (« Area under the curve », AUC en Anglais).

L'objectif du projet est de choisir une ou deux méthodes proposées par les participants à la compétition DCASE (<http://dcase.community/challenge2018/task-bird-audio-detection-results#technical-reports>) et de l'optimiser et la paralléliser afin de la faire tourner en un temps raisonnable sur les clusters à disposition. Cela permettra d'accélérer les phases d'apprentissage et de test sur un grand nombre de fichiers audio. Beaucoup de solutions proposées sont basées sur des réseaux de neurones artificiels, notamment les réseaux de neurones convolutifs (https://fr.wikipedia.org/wiki/R%C3%A9seau_neuronal_convolutif). Ce type de réseaux font partie de la mouvance du Deep Learning (apprentissage profond) et sont plus connus pour leur utilisation en reconnaissance visuelle.

Un deuxième objectif de ce projet est la **reproductibilité scientifique** : c'est-à-dire vérifier que les méthodes et les codes sources proposées par d'autres permettent d'obtenir des performances similaires lorsque ces méthodes sont réimplémentées. C'est un problème majeur actuellement en science et notamment dans le domaine du Deep Learning où beaucoup de méthodes sont déposés sur des sites d'archivage tel arXiv.org sans être relu et vérifiés par des pairs.

Les méthodes proposées sont documentées et leur code source est souvent disponible. A vous de bien choisir les codes sur lesquels vous allez vous baser (modularité, faciliter de relecture, ...) avant de vous lancer.

Un des points intéressants de ce challenge est la diversité des jeux de données à laquelle doivent faire face les programmes développés. La liste des jeux de données est disponible ci-dessous.

Jeux de données d'apprentissage

- Enregistrements de terrain, dans le monde entier ("freefield1010") - une collection de 7 690 extraits d'enregistrements de terrain du monde entier, rassemblés par le projet FreeSound, puis normalisés pour la recherche. Cette collection est très diversifiée en termes d'emplacement et d'environnement. Pour le BAD Challenge, nous l'avons annotée sur la présence / l'absence d'oiseaux.
- Ensemble de données « Crowdsourced », Royaume-Uni ("warblrb10k") - 8 000 enregistrements audio sur un smartphone provenant des quatre coins du Royaume-Uni, externalisés auprès des utilisateurs de Warblr, l'application de reconnaissance des oiseaux. L'audio couvre une large distribution de lieux et d'environnements britanniques, et comprend le bruit causé par les intempéries, le bruit de la circulation, la parole humaine et même des imitations d'oiseaux humains.
- Appels de vols de télésurveillance, États-Unis ("BirdVox-DCASE-20k") - 20 000 clips audio provenant d'unités de télésurveillance placées près de Ithaca, dans l'État de New York, aux États-Unis, à l'automne 2015, dans le cadre du projet BirdVox. Plus d'info sur BirdVox-DCASE-20k.

Jeux de données d'évaluation

- Ensemble de données externalisé, Royaume-Uni ("warblrb10k") - Ensemble conservé de 2 000 enregistrements provenant des mêmes conditions que l'ensemble de données de développement de Warblr.
- Ensemble de données de surveillance à distance, Tchernobyl ("Tchernobyl") - 6 620 clips audio ont été collectés à partir d'un équipement de surveillance à distance sans surveillance dans la zone d'exclusion de Tchernobyl (CEZ). Ces données ont été recueillies dans le cadre du projet de recherche TREE (Transfert-Exposition-Effets) sur les effets à long terme de l'accident de Tchernobyl sur l'écologie locale. L'audio couvre un large éventail d'oiseaux et inclut des échantillons de bruits météorologiques, de grands mammifères et d'insectes échantillonnés dans divers environnements de la ZCE, y compris des villages abandonnés, des prairies et des zones forestières.
- Suivi des appels de vol de nuit à distance, Pologne ("PolandNFC") - 4 000 enregistrements du projet de thèse de Hanna Pamuła sur le suivi de la migration nocturne des oiseaux en automne. Les enregistrements ont été collectés toutes les nuits, de septembre à novembre 2016, sur la côte de la mer Baltique, en Pologne, à l'aide d'unités Song Meter SM2 avec microphones montés sur des poteaux de 3 à 5 m. Pour ce défi, nous utilisons un sous-ensemble dérivé de 15 nuits avec différentes conditions météorologiques et différents bruits de fond, notamment le vent, la pluie, le bruit de la mer, les appels d'insectes, la voix humaine et les appels de chevreuils.

La compétition DCASE est terminée pour l'année 2018, par contre étudier et réimplémenter les méthodes proposées par les participants sont un bon entraînement pour ceux qui souhaitent relever le défi pour les prochaines années.