

Détection de deepfakes audios en fonction d'une base de signatures



Mots-clés : audio deepfake detection, source tracing, base de signatures

Description

Au sein du Pôle Recherche de l'AMIAD (Agence Ministérielle pour l'IA de Défense), vous rejoindrez notre équipe, où vous développerez un système de détection de deepfakes audios reposant sur une base de signature des générateurs.

Les deepfakes notamment audio posent un problème de sécurité qui motive la production de systèmes de détection. La majorité des systèmes de détection existants reposent sur une classification binaire d'un audio (deepfake ou non). Afin de comprendre la nature de la menace, il peut être souhaitable d'aller plus loin en caractérisant le système de génération, on appelle cette tâche *source tracing* [Klein et al.,].

D'autre part, des nouveaux systèmes de génération étant produits très fréquemment, il est souhaitable de pouvoir maîtriser l'ensemble des générateurs reconnus par un système de détection, y compris en ajoutant des générateurs à partir d'exemples, sans besoin de réentraîner l'ensemble du système. Ce stage vise à découpler l'ensemble d'attaques pris en compte par le détecteur de l'entraînement du modèle. Une première preuve de concept a été réalisée par [Firc et al.,] avec une tâche de vérification : décider si deux enregistrements audios proviennent du même générateur, potentiellement inconnu.

Ce stage vise à passer de la tâche de vérification à l'identification du générateur à partir d'une base de signatures de générateurs connus. Pour cela, une première étape consistera à construire une base de générateurs, à partir de corpus en accès libres¹ et de générateurs *open source*. Une deuxième étape consistera à évaluer la qualité des signatures produites, à la fois en terme de quantité et de diversité des exemples, mais aussi de similarité avec d'autres signatures de la base.

Objectifs

- Création d'une base de signatures de deepfakes audios à partir des corpus publics existants ;
- Evaluation de la qualité de la signature d'un générateur en fonction du nombre d'exemples.

Profil recherché

Vous suivez une formation d'ingénieur ou de master en mathématiques ou informatique et vous disposez :

- de connaissances préalables dans les domaines de l'apprentissage automatique, de l'apprentissage profond, ou du traitement de la parole ;
- d'un intérêt prononcé pour la recherche et les nouvelles technologies ;
- de solides compétences en programmation *python* et *PyTorch*.

Informations générales

- Durée : 6 mois
- Niveau : Master 2 ou dernière année d'école d'ingénieur
- Début : dès que possible
- Lieu : AMIAD - X-Innovation Center, École Polytechnique, Palaiseau
- Ressources mises à disposition : l'AMIAD dispose de ressources computationnelles dédiées (Serveurs et GPU).
- Contacts : Thomas Thebaud et Raphaël Duroselle, CR AMIAD,
thomas.thebaud@polytechnique.edu, raphael.duroselle-fourcade@polytechnique.edu

Points de contact

Raphaël Duroselle : raphael.duroselle-fourcade@polytechnique.edu
Thomas Thebaud : thomas.thebaud@polytechnique.edu

Références

[Firc et al.,] Firc, A., Chhibber, M., Mishra, J., Singh, V. P., Kinnunen, T., and Malinka, K. Stopa : A database of systematic variation of deepfake audio for open-set source tracing and attribution. *Interspeech 2025*.

[Klein et al.,] Klein, N., Chen, T., Tak, H., Casal, R., and Khoury, E. Source tracing of audio deepfake systems. *Interspeech 2024*.

1. <https://huggingface.co/spaces/Speech-Arena-2025/Speech-DF-Arena>