Offre n°2024-08375

Internship - Diffusion-based Unsupervised Joint Speech Enhancement, Dereverberation, and Separation

Le descriptif de l'offre ci-dessous est en Anglais

Type de contrat : Convention de stage

Niveau de diplôme exigé : Bac + 4 ou équivalent

Fonction : Stagiaire de la recherche

Contexte et atouts du poste

This master internship is part of the**REAVISE** project: "Robust and Efficient Deep Learning based Audiovisual Speech Enhancement" (2023-2026) funded by the French National Research Agency (ANR). The general objective of REAVISE is to develop a unified audio-visual speech enhancement (AVSE) framework that leverages recent methodological breakthroughs in statistical signal processing, machine learning, and deep neural networks to design a robust and efficient AVSE framework.

The intern will be supervised byMostafa Sadeghi (researcher, Inria), Romain Serizel (associate professor, University of Lorraine), as members of the MULTISPEECH team, and Xavier Alameda-Pineda (Inria Grenoble), member of the RobotLearn team. The intern will benefit from the research environment, expertise, and powerful computational resources (GPUs & CPUs) of the team.

Mission confiée

Speech restoration regroups several downstream tasks that share a common goal of recovering a ground-truth speech signal that has been affected by one or many deformations. These deformations can be for example due to: a) a noise or a concurrent speech that adds up to the original speech signal, b) reflection of the speech signal by the walls in a room, c) limited dynamic range of a recording system that clips the speech waveform's amplitudes exceeding a certain threshold, d) packet loss occurring in transmission in telecommunication systems. Each of these degradations, has been most of the time studied separately in the literature leading to respective techniques like a) speech enhancement or speech separation b) dereverbation, c) declipping, and d) inpainting. Recently, the interest increased in learning universal models able to tackle simultaneously two or more tasks of speech restoration [1]. This is motivated by the fact that in real-life applications, a speech signal is likely tainted by several degradations at once. Various approaches have been proposed. They can be generative based [2, 3] or not [4], but they are mostly implemented in a supervised way leading to the requirement of pairs of training data, where each pair is composed of a degraded speech and the corresponding clean speech target. Better generalization for such a model is achievable at the cost of important training data. Particularly, speech denoising (or enhancement) is known to be vulnerable to mismatches since its standard approaches heavily rely on paired clean/noisy speech data to achieve strong performance.

Principales activités

Recent advances in generative modeling have seen the emergence of diffusion models as strong data distribution learners. It consists of gradually turning clean data into noise and learning a deep neural network to reverse this process so that one can generate samples of the clean data distribution starting from a pure Gaussian noise for example. This generative modeling has been successfully applied in an unsupervised way individually for the task of speech enhancement [10], and speech dereverberation [5, 6]. That is, the training no longer requires paired data as opposed to the supervised; only the clean speech data is required in training, and the enhancement or dereverberation task is performed in inference with some statistical modeling. Promising results for generalization have been found for these approaches. Utilizing diffusion models [7] also attempts to solve individually in an unsupervised way, various speech restoration but with little success particularly for the speech separation task while in a supervised way and still with diffusion model [8, 9] achieve competitive performance.

The goal of this internship will be threefold:

- Address joint speech enhancement and dereverberation tasks rather than separately with diffusion models in an unsupervised way,
- Address speech separation, with a diffusion model learned in an unsupervised way (i.e. learned only on
  clean speech),
- Address the three tasks (enhancement, dereverberation, separation) with a single unsupervised

framework.

References

[1] M. Maciejewski, G. Wichern, E. McQuinn, and J. Le Roux, (2020, May) WHAMR ! : Noisy and reverberant single-channel speech separation In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020.

[2] D. Yang, J. Tian, X. Tan, R. Huang, S. Liu, X. Chang, and H. Meng, Uniaudio: An audio foundation model toward universal audio generation arXiv preprint arXiv :2310.00704 2023.

[3] J. Serrà, S. Pascual, J. Pons, R. O. Araz, and D. Scaini. Universal speech enhancement with score-based diffusion arXiv preprint arXiv :2206.03065 2022.

[4] C. Quan, and X. Li, SpatialNet : Extensively learning spatial information for multichannel joint speech separation, denoising and dereverberation IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 1310–1323, 2024.

[5] J.-M. Lemercier, E. Moliner, S. Welker, V. Välimäki, and T. Gerkmann Unsupervised blind joint dereverberation and room acoustics estimation with diffusion models arXiv preprint arXiv :2408.07472 2024.

[6] J.-M.Lemercier, S. Welker, and T. Gerkmann, Diffusion posterior sampling for informed single-channel dereverberation In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) 2023.

[7] A. Iashchenko, P. Andreev, I. Shchekotov, N. Babaev and D. Vetrov, UnDiff: Unsupervised Voice Restoration with Unconditional Diffusion Model arXiv preprint arXiv :2306.00721 2023.

[8] B. Chen, C. Wu, and W. Zhao Sepdiff: Speech separation based on denoising diffusion model In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

[9] R. Scheibler, Y. Ji, S. W. Chung, J. Byun, S. Choe, and M. S. Choi, Diffusion-based generative speech source separation In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

[10] B. Nortier, M. Sadeghi, and R. Serizel, Unsupervised speech enhancement with diffusion-based generative models In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024.

[11] X. Bie, S. Leglaive, X. Alameda-Pineda, and L. Girin, Unsupervised speech enhancement using dynamical variational autoencoders IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 2993-3007, 2022.

[12] M. Sadeghi, and R. Serizel, Posterior sampling algorithms for unsupervised speech enhancement with recurrent variational autoencoder In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024.

[13] J. Richter, S. Welker, J.-M. Lemercier, B. Lay, and T. Gerkmann, Speech enhancement and dereverberation with diffusion-based generative models IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 2351-2364, 2023.

Compétences

Preferred qualifications for candidates include a strong foundation in statistical (speech) signal processing, and computer vision, as well as expertise in machine learning and proficiency with deep learning frameworks, particularly PyTorch.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

€ 4.35/hour

Informations générales

- **Thème/Domaine :** Langue, parole et audio

Calcul Scientifique (BAP E)
- **Ville :** Villers lès Nancy
- **Centre Inria :** Centre Inria de l'Université de Lorraine
- **Date de prise de fonction souhaitée :** 2025-04-01
- **Durée de contrat :** 6 mois
- **Date limite pour postuler :** 2024-12-15

Contacts

- **Équipe Inria :** MULTISPEECH
- **Recruteur :**
Sadeghi Mostafa / mostafa.sadeghi@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

Prospective applicants are invited to submit their academic transcripts, a detailed curriculum vitae (CV), and, if they choose, a cover letter. The cover letter should highlight the reasons for their enthusiasm and interest in this specific project.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.