Offre n°2024-07408

PhD Position F/M Explainable and frugal audio scene description

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

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

Contexte et atouts du poste

Inria Défense&Sécurité (Inria D&S) was created in 2020 to federate Inria's actions for the benefit of military forces. The PhD will be carried out within the audio processing research team of Inria D&S, under the supervision of Jean-François Bonastre and co-supervised by Raphaël Duroselle.

The automatic audio scene description task is to present operators with a summary of the information present in the scene, in the form of augmented text. This text provides a visual summary of the most important information, while efficiently structuring access to specific information. Here is an illustrative example of a summary: « This five-minute recording features three different speakers. Speaker A corresponds to a known identity in the database and speaks French with a strong Monawa accent, speakers B and C are unknown in the database and speak English in their interactions with A and use an unidentified language when talking to each other. The voices of B and C show strong similarities with speakers from the Eastern Quabar region. The main theme of the recording concerns a transfer of goods between the cities of Orienta and Flagrance. The date July 8, 2023 is mentioned three times.». Clicking on A gives the operator information about A and details of the voice identification performed. There will be direct access to the time segments during which A spoke and to their transcription. The transcription will highlight names of people, places or dates (named entities).

Mission confiée

Goal

The aim of this thesis is to propose a general framework for processing audio recordings for intelligence purposes. It consists in defining a high-level application adapted to the needs of end users, favouring the presentation of a recording in the form of a summary report to highlight its salient points.

Approach

This approach is inspired both by textual description of video scenes [1] and by dialogue systems based on audio-visual scenes [2]. The system will be based on the extraction of speech signal representations at different scales (frame, speech segment or sound event, complete recording), possibly dedicated to different tasks. The representations, useful for the various technological bricks of the system, will be embeddings extracted from deep neural networks, either generic [3] or dedicated to each task. The fusion between the different levels of information can be achieved with an architecture inspired by the multi-stream "Encoder-Decoder" scheme [4], with several encoders producing sequences of representations and one or more decoders performing the tasks or sub-tasks required by the system. One of these decoders will produce a textual summary of the scene.

Potential research directions, aiming to go beyond an audio scene description system by assembling existing bricks, can be discussed and refined with the candidate.

Principales activités

- Bibliography, development and evaluation of deep learning systems ;
- Definition of a new task, definition of a corpus and evaluation protocol ;
- Work on the alignment between self-supervised representations of the speech signal and large language models ;
- Weakly supervised system training ;
- System evaluation.

Compétences

Master level in computer science, mathematics or phonetics.

Strong interest in applied research.

Written and spoken English

Signal processing

Machine learning and deep learning

Experience with deep learning toolkits such as pytorch or keras

Speech processing experience, knowledge of open source toolkits such as kaldi or speechbrain.

References

[1] Aafaq, N., Mian, A., Liu, W., Gilani, S. Z., & Shah, M. (2019). Video description: A survey of methods, datasets, and evaluation metrics. *ACM Computing Surveys (CSUR)*, *52*(6), 1-37.

[2] Hori, Chiori, Huda Alamri, Jue Wang, Gordon Wichern, Takaaki Hori, Anoop Cherian, Tim K. Marks, et al. « End-to-End Audio Visual Scene-Aware Dialog Using Multimodal Attention-Based Video Features ». In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2352□56. Brighton, United Kingdom: IEEE, 2019. https://doi.org/10.1109/ICASSP.2019.8682583.

[3] Zhang, C., & Tian, Y. (2016, December). Automatic video description generation via lstm with joint two-stream encoding. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 2924-2929). IEEE.

[4] Pratap, Vineel, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky, et al. 2023. « Scaling Speech Technology to 1,000+ Languages ». arXiv. http://arxiv.org/abs/2305.13516.

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 and flexible organization of working hours,
- Professional equipment available (videoconferencing, loan of computer equipment, etc.),
- Social, cultural and sports events and activities,

Rémunération

- 1st and 2nd year : 2082 € bruts - gross /month
- 3rd year : 2190 € bruts - gross /month

Informations générales

- **Ville :** PARIS
- **Centre Inria :** Siège
- **Date de prise de fonction souhaitée :** 2024-05-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2024-08-31

Contacts

- **Équipe Inria :** MIS-DEFENSE (DIRECTION)
- **Directeur de thèse :**
  Maillet Florence / florence.maillet@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.

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.