



Offer #2024-07920

Doctorant F/H Conception d'un modèle de fondation pour la reconnaissance d'actions dans les flux vidéo de drones

The offer description below is in French

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

Le centre Inria d'Université Côte d'Azur regroupe 42 équipes de recherche et 9 services d'appui. Le personnel du centre (500 personnes environ) est composé de scientifiques de différentes nationalités, d'ingénieurs, de techniciens et d'administratifs. Les équipes sont principalement implantées sur les campus universitaires de Sophia Antipolis et Nice ainsi que Montpellier, en lien étroit avec les laboratoires et les établissements de recherche et d'enseignement supérieur (Université Côte d'Azur, CNRS, INRAE, INSERM ...), mais aussi avec les acteurs économiques du territoire.

Présent dans les domaines des neurosciences et biologie computationnelles, la science des données et la modélisation, le génie logiciel et la certification, ainsi que la robotique collaborative, le Centre Inria d'Université Côte d'Azur est un acteur majeur en termes d'excellence scientifique par les résultats obtenus et les collaborations tant au niveau européen qu'international.

Context

Inria, l'Institut national français d'informatique et de mathématiques appliquées, promeut "l'excellence scientifique au service du transfert de technologie et de la société".

Diplômés des plus grandes universités mondiales, les 2 700 collaborateurs d'Inria relèvent les défis des sciences numériques. Grâce à son modèle ouvert et agile, Inria est en mesure d'explorer des approches originales avec ses partenaires industriels et académiques et de répondre efficacement aux défis pluridisciplinaires et applicatifs de la transformation numérique. Inria est à l'origine de nombreuses innovations qui apportent de la valeur ajoutée et créent des emplois.

Equipe :

L'équipe de recherche STARS associe une théorie avancée à une pratique de pointe axée sur les systèmes de vision cognitive.

Site Web de l'équipe: <https://team.inria.fr/stars>

Assignment

1 - Objet de la thèse

Dans ce projet de thèse, nous visons à concevoir un nouveau modèle de fondation unifiant séquences de squelettes (i.e., postures des personnes) et actions RVB, pour cibler à la fois la classification et la segmentation des actions humaines à partir de flux vidéo acquis par des drones. Ce modèle de fondation traitera les séquences de squelettes humains ou les clips vidéo RVB en utilisant une architecture Transformer pour extraire des caractéristiques génériques.

2 - Descriptif (objectifs, aspects innovants)

Les problèmes posés par ce travail de thèse consistent principalement à concevoir un nouveau modèle de fondation unifiant les séquences de squelettes (i.e., posture 2D ou 3D) et les actions RVB sémantiques. Un premier défi consiste à rassembler suffisamment de données pour l'entraînement du modèle, rassemblant en même temps des actions RVB et des annotations de squelettes suffisamment précises.

Un autre défi est d'arriver à construire un modèle générique unifiant des données hétérogènes comme le sont les données squelettes par rapport aux actions RVB et aux données textes (i.e., sémantiques). Pour cela, il faudra concevoir des pré-tâches d'auto-apprentissage spécifiques permettant de construire un modèle générant des représentations génériques d'actions humaines à partir de flux vidéo acquis par des drones.

L'objectif général de la thèse est d'étendre et adapter les modèles de fondation Langage Visuel (e.g., CLIP) aux modèles pouvant être appliqués à des vidéos issues de caméras de vidéosurveillance et en particulier de flux vidéos provenant de drones. Il s'agira de construire une base de données vidéos (appelée Skeletics), spécifiques aux applications de vidéosurveillance et de vidéos issues de drones et de tester ces modèles aux cas d'analyse de vidéos issues de drones et de la reconnaissance d'actions centrées sur l'homme (c'est-à-dire la classification ou la segmentation d'actions). Comme cette détection sera générique et sémantique, elle pourra combiner des détections d'action à partir de drones (i.e., entre plusieurs drones), mais également avec des caméras de vidéosurveillance installées au sol.

Main activities

Résultats attendus

Les contributions de cette thèse de doctorat seront les suivantes :

1. Nous concevrons T-MOR, un nouveau modèle de fondation basé sur les mouvements de squelette transférable qui peut être généralisé à des tâches de reconnaissance d'actions centrées sur l'homme dans des vidéos réelles. Nous proposons d'améliorer la capacité de représentation des modèles vidéo de Langage Visuel récents en utilisant des données de mouvement humain et un apprentissage contrastif multimodal.
2. Nous construirons Skeletics, une nouvelle base de données vidéos d'action à grande échelle centrées sur le mouvement humain, comprenant des annotations de bonne qualité sur le squelette humain en 2D et en 3D, pour l'apprentissage de modèles génériques de mouvements de squelette.
3. Nous mènerons une étude de validation et montrerons que le pré-entraînement de T-MOR sur Skeletics et son transfert sur un ensemble de données vidéo cibles, représentent une méthodologie générique et efficace pour la reconnaissance d'action. Nous montrerons que le modèle de fondation d'action squelette-RGB-texte appris est suffisamment général pour améliorer de nombreuses tâches de reconnaissance d'actions (c'est-à-dire la classification ou la segmentation d'actions) à partir de vidéos de drones ou de vidéos au sol ou avec uniquement des descriptions textuelles, des clips vidéo RVB ou uniquement des séquences de squelettes.

L'objectif est d'avoir un modèle de fondation suffisamment générique pour être directement utilisable sur de nouvelles vidéos de drones. Pour valider ce modèle, on le testera sur de nouvelles vidéos qui n'auront pas été observées et n'appartenant pas à l'ensemble d'apprentissage. Ce modèle de fondation pourra être utilisé dans une large gamme d'application, par exemple comme la gestion de catastrophes naturelles. Ainsi, un drone envoyé sur un terrain d'observation pourra par lui-même automatiquement détecter des situations d'urgence, par exemple, une personne faisant des signes de détresse et envoyer une alerte directement au poste de commandement ou poste de secours.

Skills

Des connaissances dans les domaines suivants seront appréciées :

- Expérience dans les langages de programmation C++/Python,
- Apprentissage automatique, réseaux neuronaux profonds, PyTorch, TensorFlow
- Modèles graphiques probabilistes,
- Vision par ordinateur, et techniques d'optimisation (descente de gradient stochastique, Message-passing).

Benefits package

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés : 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Participation mutuelle (sous conditions)

Remuneration

2100€ brut mensuel (année 1 & 2) et 2190€ brut mensuel (année 3)

General Information

- **Theme/Domain** : Vision, perception and multimedia interpretation Statistics (Big data) (BAP E)
- **Town/city** : Sophia Antipolis
- **Inria Center** : [Centre Inria d'Université Côte d'Azur](#)
- **Starting date** : 2024-11-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2024-07-31

Contacts

- **Inria Team** : [STARS](#)
- **PhD Supervisor** :
Brémond François / Francois.Bremond@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

- Étudier les limites des algorithmes de reconnaissance d'activités existants.
- Annoter les vidéos en fonction des activités à reconnaître.
- Étude de la faisabilité et de l'adéquation du cadre dans des situations pratiques.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.