

STAGE 2020

Réf : LVA-2020-S5

Détection d'interactions entre personnes et objets dans une image par apprentissage profond

Présentation du laboratoire d'accueil

Basé à Paris-Saclay, le CEA List est l'un des quatre instituts de recherche technologique de CEA Tech, direction de la recherche technologique du CEA. Dédié aux systèmes numériques intelligents, il contribue au développement de la compétitivité des entreprises par le développement et le transfert de technologies.

L'expertise et les compétences développées par les 800 ingénieurs-chercheurs et techniciens du CEA List permettent à l'Institut d'accompagner chaque année plus de 200 entreprises françaises et étrangères sur des projets de recherche appliquée s'appuyant sur 4 programmes et 9 plateformes technologiques. 21 start-ups ont été créées depuis 2003.

Labellisé Institut Carnot depuis 2006, le CEA List est aujourd'hui l'institut Carnot Technologies Numériques

Le Laboratoire de Vision et Apprentissage pour l'analyse de scène (LVA) mène ses recherches dans le domaine de la Vision par Ordinateur (Computer Vision) selon quatre axes principaux :

- La reconnaissance visuelle (détection et/ou segmentation d'objets, de personnes, de patterns; détection d'anomalies; caractérisation)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, de comportements anormaux ou spécifiques pour des individus, un groupe, une foule)
- L'annotation intelligente (annotation à grande échelle de données visuelles 2D/3D de manière semi-automatique)
- La perception et la décision (processus de décision markovien, navigation)

Description du stage

Le LVA a développé un algorithme de reconnaissance d'interactions entre personnes et objets à partir d'une image [1]. A terme, cette brique technologique permettra d'analyser finement l'activité d'une personne. Cet algorithme, basé sur un réseau de neurones profond, est capable à l'étape d'inférence de retourner en une seule passe, toutes les interactions contenues dans une image à partir des détections des objets fournies par un détecteur d'objet externe.



Exemples d'interactions détectées par l'algorithme du LVA

Le premier objectif du stage est de développer une architecture de réseaux de neurones pour la détection d'objets et la reconnaissance d'interactions conjointes. Les publications [2], [3] et [4] proposent des méthodes incluant la détection des objets dans le réseau de reconnaissance d'interactions mais elles présentent l'inconvénient de décomposer les tâches. Dans un premier temps, leur réseau estime les boîtes d'objets puis dans un deuxième temps teste tous les couples d'interactions possibles entre les personnes et les objets, ce qui augmente notablement leur temps de calcul en fonction du nombre de sujets et d'objets.

Notre but est de proposer et développer un détecteur qui soit capable de retourner à la fois les boîtes d'objets et leurs interactions en passant l'image une seule fois dans le réseau. Les principaux challenges à relever sont donc

l'apprentissage de deux tâches distinctes au sein d'un même réseau et la gestion simultanée de bases d'images annotées différemment.

La plateforme Mobile Mii [5] du CEA List, est un véritable appartement permettant de réaliser des acquisitions et de mettre en place des démonstrations dans un environnement réel. Le second objectif du stage est d'intégrer l'algorithme de détection d'interactions à un démonstrateur en temps réel dans la plateforme. L'apprentissage du modèle sera réalisé sur le dataset DAHLIA [6] qui contient des activités filmées sur la plateforme.



Plateforme MobileMii - Acquisition de la base de données DAHLIA

Mots-clés :

Vision par ordinateur, apprentissage profond, détection d'objets, reconnaissance d'interactions.

Références

- [1] Chafik, S., Orcesi, A., Audigier, R., Luvison, B. , Classifying All Interacting Pairs in a Single Shot, ArXiv 2019 (à paraître)
- [2] S. Gupta and J. Malik. Visual semantic role labeling. arXiv preprint arXiv:1505.04474, 2015.
- [3] G. Gkioxari, R. Girshick, P. Dollár, and K. He. Detecting and recognizing human-object interactions. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8359–8367. IEEE, 2018.
- [4] Y.-W. Chao, Y. Liu, X. Liu, H. Zeng, and J. Deng. Learning to detect human-object interactions. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 2018.
- [5] Mobile Mii, <http://www-mobilemii.cea.fr/>
- [6] Vaquette, G., Orcesi, A., Lucat, L., & Achard, C. (2017, May). The DAily Home Life Activity Dataset: A High Semantic Activity Dataset for Online Recognition. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on* (pp. 497-504).

Profil du candidat/de la candidate

Niveau demandé :	Ingénieur, Master 2
Ce stage ouvre la possibilité de poursuite en thèse et ingénieur R&D dans notre laboratoire.	
Durée :	6 mois
Rémunération :	entre 700 € et 1300 € suivant la formation.
Compétences requises :	
<ul style="list-style-type: none"> - Vision par ordinateur - Apprentissage automatique (deep learning) - Reconnaissance de formes - C/C++, Python - La maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch) est un plus. 	