



**CEA List**  
**Service d'Intelligence Artificielle pour le Langage et la Vision**  
Centre de Saclay 91191 Gif-sur-Yvette France  
<http://www.kalisteo.eu>

Contact Aleksandr Setkov  
Tél +33 (0)1 69 08 07 50  
E-mail [aleksandr.setkov@cea.fr](mailto:aleksandr.setkov@cea.fr)

**STAGE 2021**

Réf : LVA-21-S7

## Détection d'évènements anormaux dans les vidéos

### 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)
- Annotation intelligente (annotation à grande échelle de données visuelles 2D/3D de manière semi-automatique)
- Perception et décision (processus de décision markovien, navigation)

### Description du stage

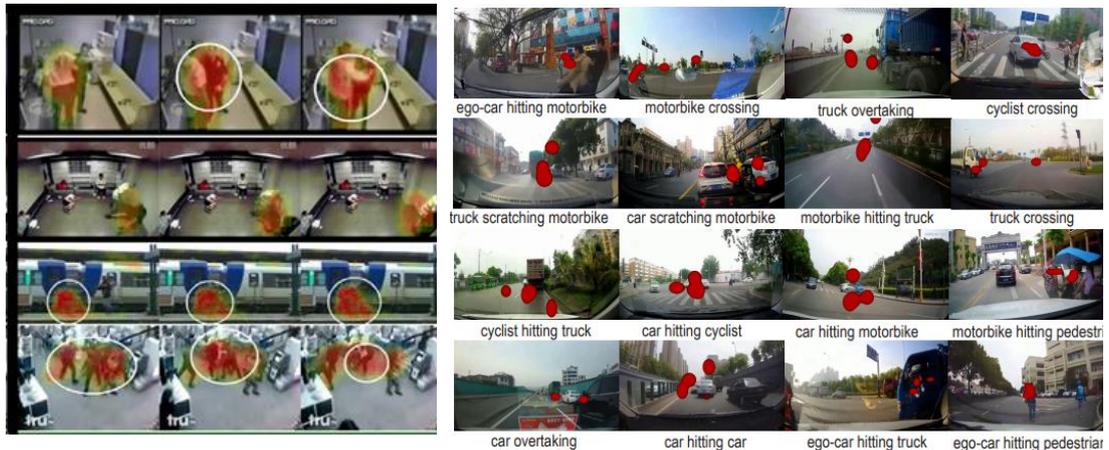
La reconnaissance d'évènements anormaux dans des vidéos par apprentissage profond est une fonctionnalité cruciale pour les applications de vidéo surveillance, de sécurité routière ou encore de conduite autonome. Il existe cependant encore beaucoup de verrous techniques et scientifiques. Les principaux sont:

- L'hétérogénéité des évènements anormaux : comment décrire et caractériser un évènement anormal ? (bagarre, vol, violence, accident routier, comportement dangereux, etc.) ;
- La rareté des évènements anormaux (souvent les données ne sont pas disponibles ou disponibles en petites quantités) ;
- Les différents niveaux sémantiques nécessaires pour la modélisation des évènements d'intérêt (de l'analyse bas niveau de motifs et de mouvements, à la détection des personnes et d'objets en présence, l'estimation de leur densité et de leurs trajectoires, jusqu'à la compréhension haut niveau d'évènements comme un vol, une agression, un incendie ou une manœuvre dangereuse sur la route) ;
- La complexité de la scène (densité d'objets présents, arrière-plan dynamique, mouvement égocentrique) ;
- Le temps de traitement et la latence (une réponse temps réel est requise dans certaines applications comme la conduite autonome)

Les méthodes par apprentissage supervisé nécessitent des annotations en grande quantité, ce qui dans ce contexte est compliqué à obtenir. Les données anormales sont rares par rapport aux données normales et l'annotation de vidéos est par ailleurs une tâche fastidieuse. Ces méthodes ne peuvent donc pas être directement appliquées pour la détection d'anomalies dans les vidéos. Une stratégie, dite « one-class », consiste à ne superviser et modéliser que la classe normale pour en déduire les évènements anormaux qui sont trop différents du modèle de normalité [1]. Mais cette modélisation se heurte souvent à l'hétérogénéité des données normales comme à celle des données anormales.

L'objectif du stage est, dans un premier temps, d'étudier plusieurs méthodes de l'état de l'art, notamment celles qui nécessitent peu d'exemples (« Few-Shot Learning » [2]) et/ou qui sont faiblement ou partiellement supervisées [3,4]. Le candidat devra les évaluer sur la tâche de détection

et localisation des évènements anormaux dans les vidéos afin d'apprécier leurs avantages et leurs limitations. Dans un deuxième temps, le candidat devra proposer des améliorations pour pallier un ou plusieurs problèmes identifiés. Ces travaux seront évalués quantitativement et qualitativement sur divers jeux de données [3, 5,6,7] et pourront faire l'objet de publications scientifiques. Ce stage ouvre la possibilité de poursuite en thèse sur cette problématique.



Exemple des résultats de la détection d'anomalie (agressions) dans les vidéos [1] et exemples des scénarios d'accidents dans les vidéos routières égocentriques [5]

## Références

- [1] "RIMOC, a feature to discriminate unstructured motions: Application to violence detection for video-surveillance", Pedro Canotilho Ribeiro, Romaric Audigier, Quoc Cuong Pham, CVIU 2016
- [2] "Any-Shot Sequential Anomaly Detection in Surveillance Videos", Doshi, Keval and Yilmaz, Yasin, CVPR Workshops 2020
- [3] "Real-world anomaly detection in surveillance videos", Waqas Sultani, Chen Chen, Mubarak Shah, CVPR 2018
- [4] "Weakly-supervised Deep Anomaly Detection with Pairwise Relation Learning", Guansong Pang, Anton van den Hengel, Chunhua Shen, CoRR 2019
- [5] "Dada: A largescale benchmark and model for driver attention prediction in accidental scenarios", Jianwu Fang, Dingxin Yan, Jiahuan Qiao, and Jianru Xue, CoRR 2019.
- [6] "When, where, and what? a new dataset for anomaly detection in driving videos", Yu Yao, Xizi Wang, Mingze Xu, Zelin Pu, Ella Atkins, and David Crandall, CoRR 2020.
- [7] "Future frame prediction for anomaly detection – a new baseline.", Liu, W., Luo, W., Lian, D., Gao, S., CVPR 2018

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