



Offres de stage 2023

Vision par ordinateur / Apprentissage automatique / Intelligence artificielle



- ▶ LVA-23-S1 : Modèles de perception 3D multicaméras agnostiques à la configuration des caméras
- ▶ LVA-23-S2 : Recherche d'une représentation robuste 3D d'une classe sémantique à partir d'une collection de modèles CAO
- ▶ LVA-23-S3 : Estimation de flux optique à partir de séquences vidéo de plus de trois images et d'architecture transformers
- ▶ LVA-23-S4 : Détection d'évènements anormaux dans les scènes dynamiques
- ▶ LVA-23-S5 : Stage de recherche scientifique : apprentissage par renforcement
- ▶ LVA-23-S6 : Reconstruction d'un objet 3D à partir d'une seule image 2D
- ▶ LVA-23-S7 : Détection et suivi du porteur de ballon dans des vidéos de matchs de sports collectifs
- ▶ LVA-23-S8 : Méthodes Non-Supervisées Appliquées aux Vision Transformers pour la localisation d'objets
- ▶ LVA-23-S9 : IA de confiance : vers une détection d'objets robuste face à l'inconnu



CEA List Laboratoire de Vision et Apprentissage pour l'analyse de scène Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu	Contact	Valentin Belissen, Florian Chabot
	Tél	+33 (0)1 69 08 33 63
	E-mail	valentin.belissen@cea.fr florian.chabot@cea.fr
Réf : LVA-23-S1		

Stage 2023

Modèles de perception 3D multicaméras agnostiques à la configuration des caméras

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).

Contexte

Une direction prise par les recherches en perception pour le véhicule autonome consiste à tenter de se passer de dispositifs coûteux comme le LiDAR, et réaliser les tâches de perception uniquement à partir des images captées par des caméras RGB embarquées.

Deux tâches classiques de perception sont notamment la segmentation de la route en différentes catégories d'intérêt et la détection de classes d'objets en 3D. La réalisation de ces tâches est désormais couramment menée par des modèles dont les cartes de sortie sont en vue aérienne (Birds-Eye-View ou BEV), dans un repère global associé au véhicule.

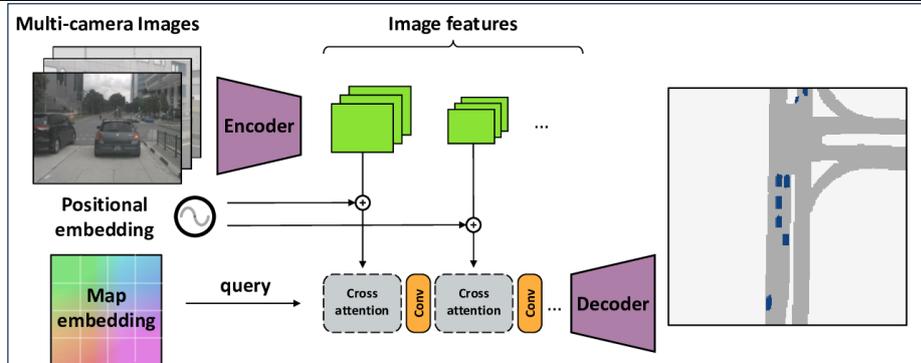
Le passage des représentations apprises au niveau image vers le repère BEV est alors un point complexe. En effet, connaissant les paramètres intrinsèques et extrinsèques de chaque caméra, les représentations associées à un pixel donné ne sont pas associées à une unique position dans le repère global, mais à un rayon optique (vecteur dans l'espace).

Un certain nombre de modèles se sont alors concentrés sur l'apprentissage conjoint de représentations image et de profondeur estimée, pour pouvoir alors transposer les représentations images dans le repère BEV sans ambiguïté. Ces méthodes souffrent cependant de plusieurs limites, notamment un temps de calcul et/ou une utilisation de mémoire très élevés.

Plus récemment, des modèles comme [1] ont proposé d'utiliser une architecture *Transformer* pour apprendre cette transposition de manière directe. Pour cela, des *positional encodings* représentés par les rayons optiques dans le repère global sont associés à chaque pixel de chaque image, à la manière d'autres modèles réalisant des tâches directement au niveau image [2, 3].



CEA List Laboratoire de Vision et Apprentissage pour l'analyse de scène Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu	Contact Valentin Belissen, Florian Chabot Tél +33 (0)1 69 08 33 63 E-mail valentin.belissen@cea.fr florian.chabot@cea.fr
Réf : LVA-23-S1	



Segmentation de la route en vue aérienne à partir d'images multiples, via une architecture Transformer [1]

Nous pouvons alors nous demander si un tel modèle peut apprendre à généraliser à des dispositions et des types de caméras qui différencieraient au sein des données d'apprentissage, ou encore entre les données d'apprentissage et les données d'inférence.

Objectif du stage

Le stage consistera à tenter de répondre à la question précédente, et sera centré sur l'apprentissage généralisable, robuste, léger et rapide de représentations BEV.

Le ou la stagiaire pourra commencer par établir différents ensembles de données hétérogènes et mettre en place un processus d'évaluation. Par la suite, l'implémentation de modèles inspirés de l'état de l'art pourra permettre de tester la robustesse de la méthode face à des dispositions ou des types de caméras variables. L'amélioration de ces modèles voire la mise en place de modèles palliant les problèmes rencontrés seront attendues.

Une extension à ce travail pourra également consister à évaluer la possibilité que le réseau apprenne de lui-même les *positional encodings* associés à chaque pixel de chaque image. En effet, les calibrations intrinsèques et extrinsèques ne sont fiables qu'à une certaine précision et peuvent évoluer dans le temps.

Références

- [1] B. Zhou, P. Krähenbühl, *Cross-view Transformers for real-time Map-view Semantic Segmentation*, CVPR 2022
- [2] J. M. Facil, B. Ummenhofer, H. Zhou, L. Montesano, T. Brox, and J. Civera. *CAM-Convs : Camera-Aware Multi-Scale Convolutions for Single-View Depth*, 2019.
- [3] V. R. Kumar, S. Yogamani, H. Rashed, G. Sitsu, C. Witt, I. Leang, S. Milz, and P. Mäder. *OmniDet : Surround View Cameras based Multi-task Visual Perception Network for Autonomous Driving*. RAL 2021.

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- Python, C/C++- Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch)	



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 Bertrand Luvison
Mohamed Chaouch
Tél +33 (0)1 69 08 01 17
E-mail bertrand.luvison@cea.fr
mohamed.chaouch@cea.fr

Réf : LVA-23-S2

STAGE 2023

Recherche d'une représentation robuste 3D d'une classe sémantique à partir d'une collection de modèles CAO

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

Ce stage s'inscrit dans un projet de recherche dont l'objectif est de mettre au point un détecteur à classes paramétrables par des modèles CAO en se basant sur les mécanismes d'attention et les réseaux de neurones profonds. Ce détecteur qui prend en entrée une image et un ensemble de modèle CAO doit retrouver dans l'image la présence ou non de ces objets. Les modèles CAO donnés en entrée ne correspondant pas nécessairement exactement à la forme de l'objet présent dans l'image, le réseau doit être en mesure d'accepter une tolérance à la forme dans sa reconnaissance. La pertinence, la robustesse et la compacité de la représentation des modèles CAO dans l'approche sont des points clés pour le bon fonctionnement de ce détecteur.

L'objectif de ce stage est de concevoir et développer une méthode de recherche d'une représentation robuste d'une classe sémantique à partir d'une collection de modèles CAO. Concrètement, les deux questions suivantes seront investiguées : (1) Quelle est la représentation commune aux modèles CAO d'une même classe sémantique, qui soit invariante aux déformations géométriques ? (2) Faut-il avoir une ou plusieurs représentations 3D pour une classe sémantique d'objets ?

Un premier axe de recherche consistera à déterminer les sous-catégories d'une classe d'objets (ex. chaise, voiture, avion, etc...) au sens de la forme, à partir d'une collection de modèles CAO. Une réflexion sur les critères de clustering définissant la notion de bon représentant devra être définie ainsi que la caractérisation de l'éloignement à ce représentant et son implication sur la forme. Un deuxième axe consistera à rechercher une représentation 3D d'une classe d'objets par une modélisation robuste aux déformations sur l'espace des formes tout en conservant la structuration sémantique de la classe. Cette modélisation devra permettre de trouver le ou les meilleurs représentant(s) géométrique(s) des différentes classes sémantiques. Pour avoir cette modélisation, un point-clé sera de saisir les déformations entre différents modèles d'une même classe. Certaines méthodes par apprentissage profond encodent l'information de façon globale dans un espace de représentation latent où les régularisations [2], les interpolations [6] et les paramétrisations [4] permettent de générer des formes



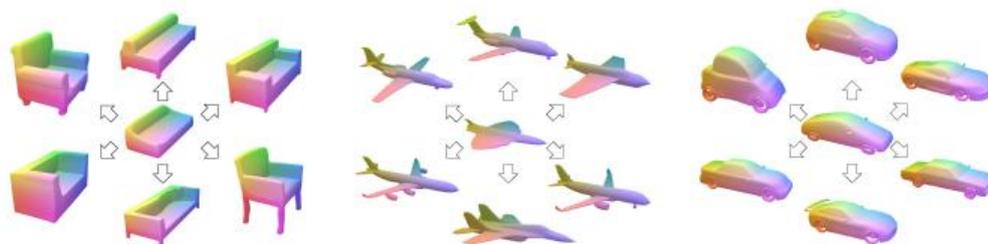
CEA List
Service d'Intelligence Artificielle pour le Langage et la Vision
Centre de Saclay 91191 Gif-sur-Yvette France
<http://www.kalisteo.eu>

Contact Bertrand Luvison
Mohamed Chaouch
Tél +33 (0)1 69 08 01 17
E-mail bertrand.luvison@cea.fr
mohamed.chaouch@cea.fr

Réf : LVA-23-S2

cohérentes appartenant à la bonne classe. D'autres méthodes proposent des représentations structurées : le modèle 3D est représenté par un ensemble de formes locales déformables [1], de structure globale avec des déformations locales qui portent un sens sémantique [3,5], ou des représentations hiérarchiques [1,7].

Dans un premier temps, le candidat étudiera les méthodes de l'état de l'art de recherche de représentation 3D. Ces approches de l'état de l'art devront être évaluées sur la base de données ShapeNet en portant un intérêt particulier autour de la modélisation de déformation 3D, de sa quantification et de son impact dans l'espace 2D après rendu. Cela permettra de caractériser l'adéquation de ces modélisations pour une intégration dans un détecteur d'objets à classes paramétrables. Le candidat sera également invité à proposer une nouvelle méthode de recherche de représentation à partir d'une collection de modèles CAO. Les travaux menés durant le stage pourront faire l'objet d'une publication scientifique.



Exemple de déformation 3D intra-classes (source [8])

Keywords

3D shape representation, 3D shape retrieval, 3D shape clustering, attentional mechanism, deep learning, supervised learning, zero-shot learning.

Références

- [1] *Local Deep Implicit Functions for 3D Shape*, K. Genova, F. Cole, A. Sud, A. Sarna, T. Funkhouser CVPR 2020
- [2] *Learning Implicit Fields for Generative Shape Modeling*, Z.Chen, H.Zhang, CVPR 2019
- [3] *Deformed Implicit Field: Modeling 3D Shapes with Learned Dense Correspondence*, Yu Deng, Jiaolong Yang, Xin Tong, CVPR 2021
- [4] *ShapeFlow: Learnable Deformations Among 3D Shapes*, C. Jiang, J. Huang, A. Tagliasacchi, L. Guibas, NeurIPS 2020
- [5] *KeypointDeformer: Unsupervised 3D Keypoint Discovery for Shape Control*, Tomas Jakab, Richard Tucker, Ameesh Makadia, Jiajun Wu, Noah Snively, Angjoo Kanazawa, CVPR 2021
- [6] *Representing Shape Collections with Alignment-Aware Linear Models*, Romain Loiseau, Tom Monnier, Mathieu Aubry, Loïc Landrieu, 3DV 2021
- [7] *RIM-Net: Recursive Implicit Fields for Unsupervised Learning of Hierarchical Shape Structures*, C. Niu, M. Li, K. Xu, H. Zhang, CVPR, 2022
- [8] *Deep Implicit Templates for 3D Shape Representation*, Z. Zheng, T. Yu, Q. Dai, Y. Liu, CVPR 2021

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)- Géométrie 3D et Reconnaissance de formes- Python, C/C++- Maîtrise d'un framework d'apprentissage profond (en particulier PyTorch ou Tensorflow)	



CEA List Service d'Intelligence Artificielle pour le Langage et la Vision Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu Réf : LVA-23-S3	Contact Florian Chabot Tél +33 (0)1 69 08 02 88 E-mail florian.chabot@cea.fr
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------

STAGE 2023

Estimation de flux optique à partir de séquences vidéo de plus de trois images et d'architecture transformers

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.

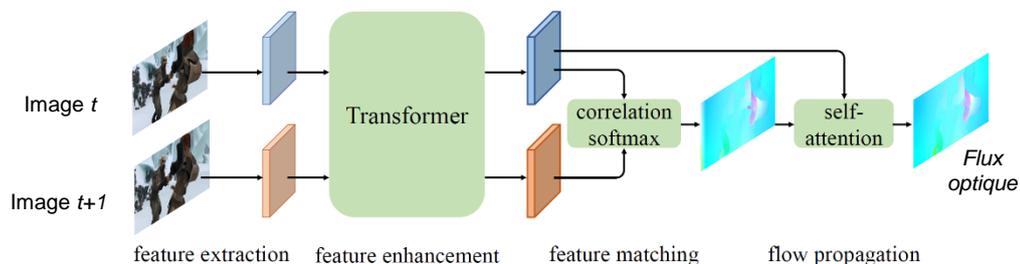
L'expertise et les compétences développées par les 800 ingénieurs-chercheurs et techniciens du CEA List permettent à l'Institut d'accompagner chaque année plus de 200 entreprises françaises et étrangères sur des projets de recherche appliquée s'appuyant sur 4 programmes et 9 plateformes technologiques.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

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)
- Les modèles de perception pour l'aide à la décision.

Description du stage

Le flux optique (déplacement pixellique entre deux images consécutives) est une modalité très utilisée en vision par ordinateur (tracking, détection d'anomalies, représentation de vidéos...) qui a largement bénéficié de l'essor du deep learning. Cependant, la difficulté d'obtenir des labels dans le cas de datasets de vidéos photo-réalistes, a poussé au développement de méthodes non-supervisées [1, 2] moins fiables que les méthodes supervisées notamment pour les pixels occultés. Une solution pour les robustifier consiste à augmenter le nombre d'images en entrée du réseau le passant de deux à trois [3, 4]. Le but de ce stage serait de généraliser cette approche en développant une méthode qui prend en entrée une séquence vidéo de longueur variable (au moins quatre images) et estime en sortie le flux optique entre tous les couples d'images consécutives. Plus concrètement, il s'agit d'adapter l'architecture à base de transformers (GMFlow [5]) qui se prête bien au problème.



Architecture de GMFlow [5]

Dans un premier temps, le stagiaire devra développer cette nouvelle architecture et la faire fonctionner dans le cadre d'un apprentissage supervisé. Un des principaux défis sera de parvenir à contenir le coût en calcul. Dans un second temps, la méthode devra être adaptée pour pouvoir être



CEA List Service d'Intelligence Artificielle pour le Langage et la Vision Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu Réf : LVA-23-S3	Contact Florian Chabot Tél +33 (0)1 69 08 02 88 E-mail florian.chabot@cea.fr
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------

entraînée de façon non-supervisée. Enfin, un mécanisme de knowledge distillation pourra être utilisé afin de transférer les connaissances de ce réseau à un autre réseau d'estimation de flux optique plus simple (type RAFT [6]).

Références

- [1] Z. Bian et al., *Learning Pixel Trajectories with Multiscale Contrastive Random Walks*, CVPR 2022
- [2] A. Stone et al., *SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping*. CVPR 2021
- [3] P. Liu et al., *SelfFlow: Self-Supervised Learning of Optical Flow*, CVPR 2019
- [4] J. Janai et al., *Unsupervised Learning of Multi-Frame Optical Flow with Occlusions*, EVVC 2018
- [5] H. Xu et al., *GMFlow: Learning Optical Flow via Global Matching*. CVPR 2022
- [6] Z. Teed et al., *RAFT: Recurrent All Pairs Field Transforms for Optical Flow*. ECCV 2020

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- Python, C/C++- Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch)	

**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

Réf : LVA-23-S4

STAGE 2023**Détection d'évènements anormaux dans les scènes dynamiques****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 en conduite autonome, en sécurité routière ou encore en vidéoprotection / vidéosurveillance. Il existe cependant encore beaucoup de verrous techniques et scientifiques. Les principaux sont les suivants:

- La dépendance au contexte : Un évènement normal dans un contexte peut ne pas l'être dans un autre.
- L'hétérogénéité des évènements anormaux : comment décrire et caractériser un évènement anormal ?
- La rareté des évènements anormaux : souvent les données ne sont pas disponibles ou disponibles en petites quantités.
- La complexité de la scène : l'arrière-plan peut être dynamique et complexe à analyser.
- 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.
- Le temps de traitement et la latence : une réponse temps réel est requise dans certaines applications comme la conduite autonome.

Les méthodes par apprentissage profond 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 modéliser la classe normale pour en déduire les évènements anormaux qui sont trop différents du modèle de normalité [1,2]. Différents paradigmes (*Auto-encoders*, *Generative Adversarial Networks*, tâches prétextes auto-supervisées...) permettent cette modélisation de façon non supervisée.

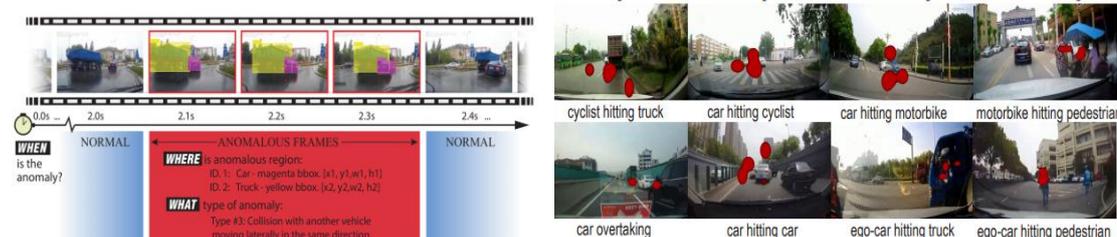


CEA List
Service d'Intelligence Artificielle pour le Langage et la Vision
Centre de Saclay 91191 Gif-sur-Yvette France
<http://www.kalisteo.eu>

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

Réf : LVA-23-S4

L'objectif du stage est de proposer une méthode pour modéliser la normalité dans le cadre de caméra dynamique (*dashcam* embarquée dans une voiture autonome, un robot ou un drone, par exemple) afin de détecter spatialement et temporellement les événements anormaux. Le candidat étudiera pour cela les caractéristiques de la scène : analyse des apparences, des mouvements (égocentriques et externes), trajectoires et positionnements en fonction du contexte (*background*), interactions entre les différents objets de la scène. Les caractéristiques les plus pertinentes seront ensuite utilisées et des mécanismes de fusion seront investigués pour modéliser la normalité afin de pouvoir détecter un vaste champ d'anomalies. La publication des résultats sera encouragée.



Exemples des scénarios d'accidents dans les vidéos routières égocentriques [3,4]

Références

- [1] "Object-centric and memory-guided normality reconstruction for video anomaly detection", Khalil Bergaoui, Yassine Naji, Aleksandr Setkov, Angélique Loesch, Michèle Gouiffès, Romaric Audigier. ICIP 2022.
- [2] "Future frame prediction for anomaly detection – a new baseline.", Liu, W., Luo, W., Lian, D., Gao, S., CVPR 2018.
- [3] "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, arXiv preprint 2020, <https://arxiv.org/abs/2004.03044>.
- [4] "Dada: A largescale benchmark and model for driver attention prediction in accidental scenarios", Jianwu Fang, Dingxin Yan, Jiahuan Qiao, and Jianru Xue, arXiv preprint 2019, <https://arxiv.org/abs/1912.12148>.

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- Deep learning- La maîtrise de Python et d'un framework d'apprentissage (en particulier, Pytorch)	

**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 Jaonary Rababarisoa

Tél +33 (0)1 69 08 01 29

E-mail jaonary.rabarisoa@cea.fr

Réf : LVA-23-S5

STAGE 2023**Stage de recherche scientifique : apprentissage par renforcement****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.

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)
- Les modèles de perception pour l'aide à la décision.

Description du stage

Nos actions de recherche sur les modèles de perception pour l'aide à la décision se focalisent sur le développement d'algorithmes de contrôle robuste à partir d'une observation visuelle basés l'apprentissage par renforcement. L'objectif est de proposer des solutions de manipulation robotique et de navigation autonome performantes pour nos partenaires industriels. Plusieurs défis scientifiques sont à relever, comme l'extraction d'information pertinente à partir d'une image ou d'une vidéo, ou encore l'apprentissage du modèle de dynamique du monde « world model » pour une planification robuste. Dans ce cadre, nous proposons à un chercheur stagiaire d'intégrer notre équipe de chercheurs travaillant sur le sujet et de contribuer à relever ces défis.

Rôle

En tant que stagiaire vous collaborez avec nos chercheurs pour :

- Étudier l'état de l'art et identifier les méthodes pertinentes pour le problème en question
- Implémenter les méthodes et conduire des expériences d'apprentissage
- Proposer des améliorations méthodologiques et algorithmiques
- Présenter les résultats de vos recherches.

Profil

Pour candidater vous devez :

- Être en master recherche avec une dominante : intelligence artificielle, apprentissage statistique, apprentissage par renforcement.
- Démontrer une capacité à conduire des recherches dans le domaine de l'intelligence artificielle et un fort intérêt pour le sujet de l'apprentissage par renforcement.
- Être familier avec les outils de développement de modèle d'apprentissage (Jax, Tensorflow, Pytorch)



CEA List Service d'Intelligence Artificielle pour le Langage et la Vision Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu Réf : LVA-23-S6	Contact Mohamed Chaouch Guillaume Lapouge Tél +33 (0)1 69 08 01 17 E-mail mohamed.chaouch@cea.fr guillaume.lapouge@cea.fr
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

STAGE 2023

Reconstruction d'un objet 3D à partir d'une seule image

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

Reconstruire une forme 3D à partir d'une seule image est une tâche de vision difficile, qui nécessite de modéliser la géométrie et la structure complète de la forme en 3D, en se basant sur les attributs sémantiques de l'image 2D. La difficulté principale est « comment extraire la forme, malgré la présence d'autres composantes comme la texture, la lumière et la pose ». Pour ce faire, la majorité des approches de l'état de l'art utilisent l'apprentissage profond à base de réseaux de neurones convolutifs [1]. Récemment, les architectures avec des mécanismes d'attention (transformer) ont émergé comme une alternative aux CNNs pour résoudre les tâches de vision [2], et ont été adaptées avec succès à la tâche de reconstruction 3D [3]. Une autre difficulté de la reconstruction d'un objet 3D à partir d'une seule image, réside dans l'ambiguïté à estimer les surfaces opposées, auto-occultées ou partiellement observées [3]. Une approche pour pallier ces problèmes serait d'exploiter des mécanismes d'attention, tels ceux trouvés dans les transformers, afin de repérer des caractéristiques géométriques et structurelles, comme des symétries planaires et axiales dans [4,5].

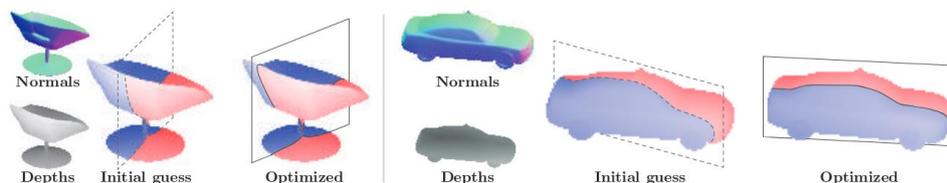


Figure 1 : Détection de plan de symétrie [4]

Par manque de données réelles, la plupart des méthodes proposées adoptent une approche supervisée sur des datasets synthétiques. Afin de pallier ce problème, des méthodes de reconstruction non supervisées ont vu le jour avec notamment une reconstruction 3D puis une mise en correspondance de la forme 3D reconstruite avec sa projection en 2D [5].



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 Mohamed Chaouch Guillaume Lapouge Tél +33 (0)1 69 08 01 17 E-mail mohamed.chaouch@cea.fr guillaume.lapouge@cea.fr
Réf : LVA-23-S6	

L'objectif de ce stage est de concevoir et développer une méthode de reconstruction d'une forme 3D à partir d'une seule image 2D. Concrètement, les problématiques suivantes seront investiguées :

- l'attention géométrique et structurelles (invariances et contraintes géométriques, cohérence géométrique, redondances dans la structure, etc) ;
- les tâches annexes à prédire ;
- l'estimation de la bounding box 3D ;
- les approches non supervisées ;
- la pertinence du point de vue versus la fiabilité de la reconstruction.

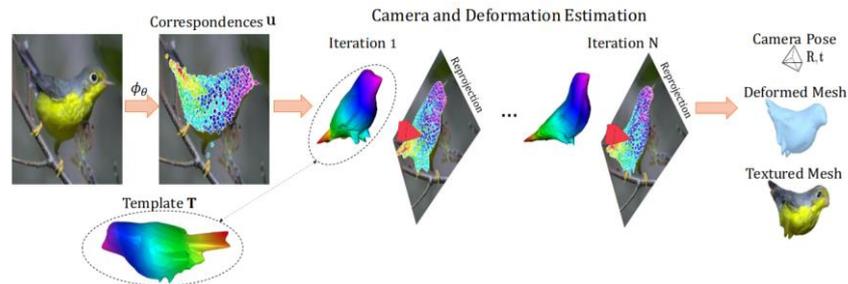


Figure 2 : Reconstruction 3D par correspondance 2D/3D [5]

Dans un premier temps, le candidat étudiera les méthodes de l'état de l'art de la reconstruction d'un objet 3D à partir d'une seule image. Ensuite, il évaluera leurs performances sur les deux datasets, ShapeNet et Pix3D, et identifiera les différents problèmes. Le candidat sera également invité à proposer une nouvelle méthode de reconstruction supervisée. Éventuellement, les travaux pourront porter sur le non supervisé. Les travaux menés durant le stage pourront faire l'objet d'une publication scientifique.

Keywords

3D shape reconstruction, 3d shape characterization, Shape analysis, attentional mechanism, deep learning, supervised learning, zero-shot learning.

Références

[1] WANG, Nanyang, ZHANG, Yinda, LI, Zhuwen, et al. *Pixel2mesh: Generating 3d mesh models from single rgb images*. In : *Proceedings of the European conference on computer vision (ECCV)*. 2018. p. 52-67.

[2] DOSOVITSKIY, Alexey, BEYER, Lucas, KOLESNIKOV, Alexander, et al. *An image is worth 16x16 words: Transformers for image recognition at scale*. *arXiv preprint arXiv:2010.11929*, 2020.

[3] LI, Xi et KUANG, Ping. *3D-VRVT: 3D Voxel Reconstruction from A Single Image with Vision Transformer*. In : *2021 International Conference on Culture-oriented Science & Technology (ICCST)*. IEEE, 2021. p. 343-348.

[4] YAO, Yuan, SCHERTLER, Nico, ROSALES, Enrique, et al. *Front2back: Single view 3d shape reconstruction via front to back prediction*. In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020. p. 531-540.

[5] KOKKINOS, Filippos et KOKKINOS, Iasonas. *To The Point: Correspondence-driven monocular 3D category reconstruction*. *Advances in Neural Information Processing Systems*, 2021, vol. 34, p. 7760-7772.

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) - Géométrie 3D et reconnaissance de formes - Python, des connaissances en C++ seraient un plus - Maîtrise d'un framework d'apprentissage profond (en particulier PyTorch ou Tensorflow) 	

**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 Astrid Sabourin

Tél +33 (0)1 69 08 33 63

E-mail Astrid.sabourin@cea.fr

Réf : LVA-23-S7

STAGE 2023**Détection et suivi du porteur de ballon
dans des vidéos de matchs de sports collectifs****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.

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)
- Les modèles de perception pour l'aide à la décision.

Description du stage

Déterminer le porteur du ballon est une tâche essentielle à l'analyse d'un match de sport collectif. En effet, l'analyse des décisions individuelles mais également la validation de la réalisation de la stratégie du groupe reposent en grande partie sur la détection du porteur de la balle. Dans ce stage nous nous intéressons au Basketball, au Handball ainsi qu'au Rugby.

Différentes approches basées sur des modèles d'apprentissage profond d'analyse d'images ou de vidéos sont envisageables :

- Afin de déterminer quel joueur possède la balle, l'estimation de la position du ballon est une information importante. Cependant cette détection est loin d'être évidente car la balle est souvent en partie occultée et le flou de bougé lié aux déplacements rapides des joueurs et de la balle elle-même lors d'une passe, déforme l'apparence visuelle du ballon. La méthode développée pourra s'inspirer des travaux de *Zandycke et al.* [4, 5]. Le laboratoire possède une base de données annotées d'image avec la position du ballon pour apprendre un tel modèle.
- La détection des joueurs est une tâche aujourd'hui maîtrisée mais déterminer qui possède la balle à partir de la position de celle-ci n'est pas évident pour des sports où les joueurs sont proches les uns des autres. L'hypothèse de [1] reposant sur le fait que le joueur le plus proche du ballon est le porteur de la balle n'est plus valable lorsque les joueurs s'occultent suivant l'axe de la caméra comme c'est le cas pour l'image suivante :



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 Astrid Sabourin Tél +33 (0)1 69 08 33 63 E-mail Astrid.sabourin@cea.fr
Réf : LVA-23-S7	



- Pour remédier à ces ambiguïtés, la calibration du terrain est nécessaire pour appliquer un raisonnement en 3D et prendre en compte la profondeur dans l'image [2].
- Enfin, une analyse temporelle basée sur le suivi des joueurs [3] est indispensable pour consolider les trous de détection lorsque le ballon n'est pas visible.

Le stage a pour objectif d'étudier les différentes briques technologiques nécessaires à la détection du porteur du ballon puis de les intégrer ensemble intelligemment pour obtenir une méthode finale capable de déterminer le porteur du ballon à chaque instant d'une vidéo. La méthode sera développée dans un premier temps pour le Basketball de la manière la plus générique possible puis sera ensuite testée et adaptée si besoin au Handball et au Rugby.

Références

[1] Martin, Z., Hendricks, S., & Patel, A. (2021). Automated tackle injury risk assessment in contact-based sports-a rugby union example. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4594-4603).

[2] Maglo, A., Orcesi, A., & Pham, Q. C. (2022). KaliCalib: A Framework for Basketball Court Registration. *arXiv preprint arXiv:2209.07795*.

[3] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Yuan, Z., Luo, P., ... & Wang, X. (2021). Bytetrack: Multi-object tracking by associating every detection box. *arXiv preprint arXiv:2110.06864*.

[4] Van Zandycke, G., & De Vleeschouwer, C. (2019, October). Real-time CNN-based segmentation architecture for ball detection in a single view setup. In *Proceedings Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports* (pp. 51-58).

[5] Van Zandycke, G., & De Vleeschouwer, C. (2022). 3D Ball Localization From a Single Calibrated Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3472-3480).

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- Python, C/C++- Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch)	



CEA List
Laboratoire de Vision et Apprentissage
pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
<http://www.kalisteo.eu>

Contact Nicolas Granger
Tél +33 (0)1 69 08 11 32
E-mail nicolas.granger@cea.fr

Réf : LVA-23-S8

Stage 2023

Méthodes Non-Supervisées Appliquées aux Vision Transformers pour la localisation d'objets

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).

Contexte

Les méthodes d'apprentissage non-supervisées pour les réseaux de neurones profonds ont connu un essor important ces dernières années, notamment dans le contexte de la vision. Elles permettent d'une part d'améliorer les performances des méthodes existantes en tirant parti de vastes corpus de données non-annotées, mais répondent aussi à un besoin croissant d'adapter des modèles à de nouveaux cas d'application rapidement et à moindres frais.

L'état de l'art se distribue essentiellement en deux familles : les approches par reconstruction [4] qui forcent le modèle à synthétiser une représentation implicite de ses entrées, les méthodes contrastives ou par clustering [1,2] qui poussent le modèle à structurer les exemples dans l'espace des représentations. Dans les deux cas, l'utilisation d'augmentations de données s'avère crucial pour imprimer dans le modèle la connaissance d'invariances et assurer une bonne généralisation.

L'essentiel des méthodes de l'état de l'art se concentre sur la tâche de classification, ce qui se justifie en partie par la plus grande facilité à mettre en évidence les contributions sur de larges jeux de données. Ces méthodes et jeux de données se montrent aussi plus efficaces que des méthodes ad-hoc [1,3] pour pré-entraîner des modèles de détection ou de segmentation sémantique sur le jeu de données cible, car il est en général plus petit.

On entrevoit cependant plusieurs limitations potentielles dans le cadre de tâches comme la détection, le suivi ou la segmentation d'instance :

- L'information de localisation contenue dans les données n'est pas optimisée
- Les augmentations (déformation, zoom, etc.) ne préservent pas l'information de localisation ou doivent être exclues
- Les méthodes optimisent principalement la prédiction du modèle et non les couches intermédiaires.



CEA List
Laboratoire de Vision et Apprentissage
pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
<http://www.kalisteo.eu>

Contact Nicolas Granger
Tél +33 (0)1 69 08 11 32
E-mail nicolas.granger@cea.fr

Réf : LVA-23-S8

Objectif du stage

L'objectif de ce stage est d'améliorer l'utilisation de méthodes contrastives pour des modèles de localisation. On envisage une approche complémentaire aux méthodes de l'état de l'art, visant les couches intermédiaires du modèle et plus particulièrement le module d'attention des Transformers.

Après un bref état de l'art, le/la stagiaire manipulera les bases de code publique servant de référence pour des contributions en se concentrant sur les modèles à base de Transformers. Les améliorations seront ensuite implémentées de manière incrémentale, en adaptant dans un premier temps des idées issues de l'état de l'art lié au sujet, puis en proposant de nouvelles approches.

Références

[1] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised Learning of Visual Features by Contrasting Cluster Assignments," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020*.

[2] Z. Dai, B. Cai, Y. Lin, and J. Chen, "UP-DETR: Unsupervised Pre-Training for Object Detection With Transformers," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021, 2021*, pp. 1601–1610.

[3] Z. Xie, Y. Lin, Z. Zhang, Y. Cao, S. Lin, and H. Hu, "Propagate Yourself: Exploring Pixel-Level Consistency for Unsupervised Visual Representation Learning," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021, 2021*, pp. 16684–16693.

[4] Z. Li et al., "MST: Masked Self-Supervised Transformer for Visual Representation," in *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual, 2021*, pp. 13165–13176.

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- Programmation et algorithmique (Python)- Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch)	



CEA List Laboratoire de Vision et Apprentissage pour l'analyse de scène Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu Réf : LVA-23-S9	Contact Hejer AMMAR Romaric AUDIGER Guillaume LAPOUGE E-mail hejer.ammar@cea.fr romaric.audigier@cea.fr guillaume.lapouge@cea.fr
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Stage 2023

IA de confiance : Vers une détection d'objets robuste face à l'inconnu

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).

Contexte

Le comportement de réseaux de neurones supervisés pour reconnaître certaines classes précises est incertain lorsqu'il est soumis à des classes jamais vues auparavant. Pouvoir détecter que la classe d'un objet est différente de celles connues lors de l'apprentissage, est un challenge appelé *Out Of Distribution detection* (OOD) qui est important pour assurer un déploiement sûr des modèles d'IA. Une IA de confiance est cruciale dans des applications critiques comme la conduite autonome. En effet, l'une des problématiques les plus importantes pour ces applications est le manque de signaux de supervision sur des données non vues en apprentissage qui peut générer des prédictions erronées mais confiantes sur des données hors-distribution (OOD) [1].



Prédiction fautive mais confiante d'un réseau qui n'a pas été supervisé sur la classe « élan » [4]

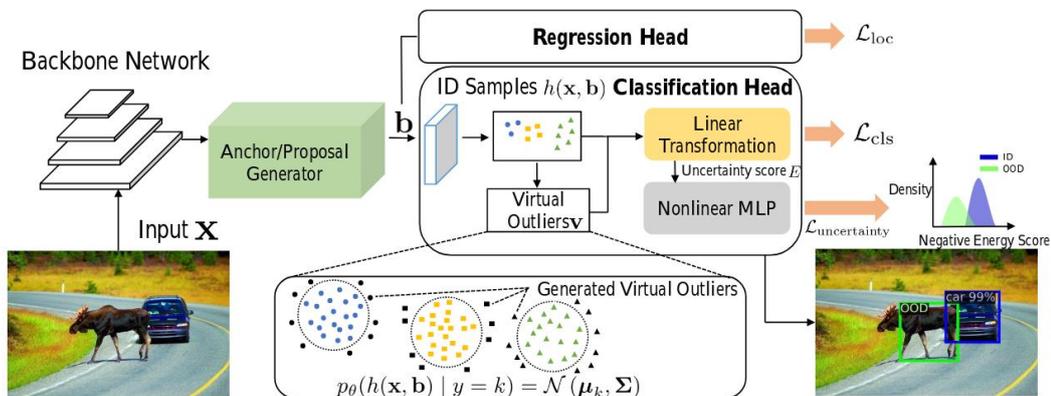
Plusieurs méthodes ont étudié ce sujet dans le cadre de la classification d'images [2, 3]. Ces approches peuvent être globalement classées en approches *post-hoc* et approches de régularisation de modèle en intégrant différentes formes d'*outliers* comme des données OOD (via des Generative Adversarial Networks, l'addition de bruit, *mixup*, etc). Néanmoins très peu de travaux traitent l'OOD dans le cadre de la détection d'objets [4,5].



CEA List Laboratoire de Vision et Apprentissage pour l'analyse de scène Centre de Saclay 91191 Gif-sur-Yvette France http://www.kalisteo.eu Réf : LVA-23-S9	Contact Hejer AMMAR Romaric AUDIGER Guillaume LAPOUGE E-mail hejer.ammar@cea.fr romaric.audigier@cea.fr guillaume.lapouge@cea.fr
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Le but de ce stage est de concevoir et de développer une méthode de détection d'objets capable de différencier avec confiance les objets connus « *In distribution* » (ID) des inconnus « *Out Of Distribution* » (OOD). Plusieurs axes d'étude seront explorés:

- Concevoir un détecteur d'objets conscient de l'existence de l'inconnu (*unknown-aware*) qui soit capable d'améliorer les résultats à la fois sur les bases de données sources et inconnues ;
- Étudier les différentes méthodes de proposition d'exemples « négatifs » ajoutés à l'apprentissage. Les méthodes de détection d'objet *Open Set* pourront ainsi être investiguées. Celles-ci permettent la détection non seulement des objets de classes connues mais aussi des autres objets présents de classes « inconnues ».
- Étudier la classification ID / OOD à différents niveaux de granularité : entre classes, sous-classes, voire instances d'objets (exemple : voitures vs personnes, voitures de différentes marques, personnes d'identités différentes).



Exemple de détection d'objets avec détection d'OOD (VOS [4])

Références

[1] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE CVPR*, 2015.

[2] SUN, Yiyou, GUO, Chuan, et LI, Yixuan. React: Out-of-distribution detection with rectified activations. In *Advances in Neural Information Processing Systems*, vol. 34, pp. 144-157, 2021.

[3] Peyman Morteza and Yixuan Li. Provable guarantees for understanding out-of-distribution detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[4] Xuefeng Du, Zhaoning Wang, Mu Cai, and Yixuan Li. Vos: Learning what you don't know by virtual outlier synthesis. In *International Conference on Learning Representations (ICLR)*, 2022.

[5] DU, Xuefeng, WANG, Xin, GOZUM, Gabriel, et al. Unknown-Aware Object Detection: Learning What You Don't Know from Videos in the Wild. In *IEEE/CVF CVPR*, 2022.

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, reconnaissance de formes - Apprentissage automatique (<i>deep learning</i>) - Python, C/C++ - Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch) 	