**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 Florian Chabot
Mohamed Tamaazousti
Tél +33 (0)1 69 08 02 88
E-mail florian.chabot@cea.fr
mohamed.tamaazousti@cea.fr

**Stage 2024** Réf : LVA-24-S1

# Estimation d'incertitude dans les modèles de perception 3D

## 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

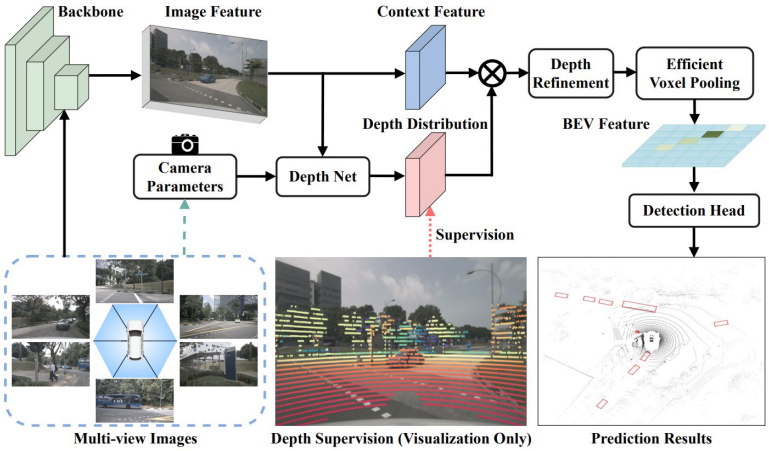
Ce stage s'inscrit dans le contexte de la perception 3D de l'environnement pour le véhicule autonome. Pour cette application, des modèles basés sur l'apprentissages profond sont développés afin d'extraire les informations nécessaires à la compréhension de la scène (détection des objets en 3D, segmentation sémantique de la route …) à partir de capteurs embarqués sur le véhicules (camera, LIDAR). Bien que ces modèles atteignent de bonnes performances, ils n'offrent pas actuellement une mesure permettant de caractériser la confiance qu'ils ont dans leurs prédictions. Cette notion de confiance est cependant nécessaire pour construire des systèmes sécurisés tels que le véhicule automatisé.

L'objectif de ce stage est de proposer une nouvelle méthode d'estimation de l'incertitude pour les modèles de perception BeV (Bird eye View) [1, 2, 3]. Pour y parvenir, le stagiaire devra :

- Étudier l'état de l'art des modèles de perception 3D et les méthodes d'estimation d'incertitude existant dans d'autres cas d'usages [4, 5, 6].
- Adapter une ou plusieurs de ces méthodes de prédiction d'incertitude dans le cadre de la perception 3D BEV. Cette première implémentation permettra d'avoir une baseline à laquelle se comparer.
- Proposer une nouvelle idée et l'implémenter.
- Mener des expériences pour valider sa méthode et se comparer à l'existant.

**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 Florian Chabot
Mohamed Tamaazousti
Tél +33 (0)1 69 08 02 88
E-mail florian.chabot@cea.fr
mohamed.tamaazousti@cea.fr




***Architecture de BEVDepth* [1] *pour la detection d'objets 3D***

## Références

[1] Li, Yinhao, et al. BEVdepth: Acquisition of reliable depth for multi-view 3d object detection. AAAI 2023

[2] Zhijian, Liu et al. BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation, ICRA 2023

[3] Zhou, Brady et al. Cross-view Transformers for real-time Map-view Semantic Segmentation, CVPR 2022

[4] Poggi, Matteo et al. On the uncertainty of self-supervised monocular depth estimation, CVPR 2020

[5] Kendall, Alex et al. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics, CVPR 2018

[6] Kendall, Alex et al. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? NIPS 2017

| | |
|---|---|
| **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 :**

- 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 Adrien Maglo
Mohamed Chaouch
Tél +33 (0)1 69 08 02 88
E-mail adrien.maglo@cea.fr
mohamed.chaouch@cea.fr

**Stage 2024** Réf : LVA-24-S2

# Génération automatique d'images de matches 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

Ces deux dernières années ont vu le développement rapide des intelligences artificielles (IA) génératives d'images. Des produits commerciaux tels que Dall-E [dalle23] ou Midjourney [midjourney23] permettent de générer des images à partir de consignes en langage naturel. Des modèles ouverts tels que Stable Diffusion [rombach22] sont ensuite apparus tandis que d'autres ont aussi étendus ces travaux à la génération de vidéos [singer2022]. Toutes ces approches produisent des résultats impressionnants. Cependant, elles font appels à des modèles de très grande taille qui nécessitent beaucoup de données pour être entraînés.

Ce sujet de stage porte sur la génération d'images réalistes de match de sport collectifs. À partir d'un nombre limité de données décrivant la scène (position des joueurs, de la balle...), le modèle avec sa connaissance acquise du sport et du contexte doit être capable de générer une image synthétique correspondant aux consignes.

Comme le contexte des images générés est réduit (match de sport collectif dans un stade), l'objectif est que les modèles développés aient un comportement frugal en quantité de données d'entraînement ainsi qu'en ressources de calcul. À partir d'un nombre limités d'images, la méthode développée devra être capable de reproduire l'environnement du stade ainsi que de générer des représentations réalistes des joueurs, arbitres ainsi que la trajectoire du ballon.

## Pistes de recherche

Nous pensons que pour mener à bien ce projet de recherche, les tâches suivantes devront être réalisées :

- Dans un premier temps, nous nous intéresserons à la génération d'images statiques de stades sans joueurs à partir d'une simple consigne textuelle.
- Nous pourrons ensuite tenter de conditionner la génération d'images avec un point de vue de caméra. Ce conditionnement pourra par exemple s'inspirer de Controlnet [zhang23] en utilisant un quadrillage uniforme de points clefs sur le terrain.
- Nous pourrons ensuite ajouter des joueurs en conditionnant sur leur position 2D sur le terrain et leur couleur de maillot.
- Un conditionnement sur les squelettes 2D des joueurs dans l'image permettra de contraindre leur attitude corporelle.

**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 Adrien Maglo
Mohamed Chaouch
Tél +33 (0)1 69 08 02 88
E-mail adrien.maglo@cea.fr
mohamed.chaouch@cea.fr

- En fin de stage, il sera possible d'aborder l'extension de la méthode à la génération de vidéos.

Durant ces dernières années, le laboratoire s'est progressivement construit une expertise sur la calibration de terrains de sports collectifs [maglo22calib], le suivi de joueurs [maglo22tracking, maglo23tracking2] ainsi que l'analyse du geste sportif. Ainsi les méthodes de suivi individuel permettent de connaître la position d'un joueur dans l'image et donc de collecter un ensemble de données d'apprentissage sur son apparence. De même, les méthodes de calibration automatiques de terrain permettent d'estimer le point de vue de la caméra et donc de potentiellement construire un jeu de données de conditionnement. Il existe aussi dans l'état de l'art des méthodes performantes capables extraire la posture d'une personne [xu22].

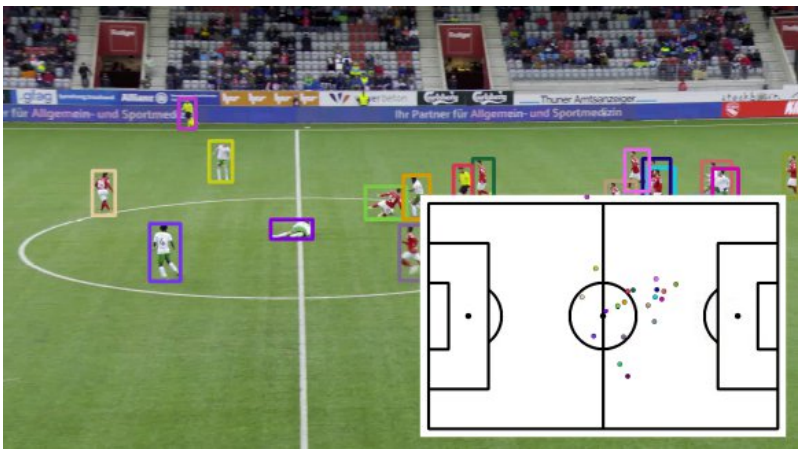


**Illustration des méthodes de suivi de joueurs et de calibration de terrain développées par l'équipe du laboratoire.**

Des jeux de données publics disponibles de séquences de football, basket ou Volleyball [cioppa22, cui23] permettront d'entraîner et de tester les modèles développés.

## Références:

**[cui23]**Cui, Y., Zeng, C., Zhao, X., Yang, Y., Wu, G., & Wang, L. (2023). SportsMOT: A Large Multi-Object Tracking Dataset in Multiple Sports Scenes. arXiv preprint arXiv:2304.05170.
**[cioppa22]**Cioppa, A., Giancola, S., Deliege, A., Kang, L., Zhou, X., Cheng, Z., ... & Van Droogenbroeck, M. (2022). Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3491-3502).
**[dalle23]**https://openai.com/dall-e-2
**[maglo22calib]**Maglo, A., Orcesi, A., & Pham, Q. C. (2022, October). KaliCalib: A Framework for Basketball Court Registration. In Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports (pp. 111-116).
**[maglo22tracking]**Maglo, A., Orcesi, A., & Pham, Q. C. (2022). Efficient tracking of team sport players with few game-specific annotations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3461-3471).
**[maglo23tracking2]**Maglo, A., Orcesi, A., Denize, J., & Pham, Q. C. (2023). Individual Locating of Soccer Players from a Single Moving View.*Sensors*,*23*(18), 7938.
**[midjourney23]**https://www.midjourney.com
**[rombach22]**Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).
**[singer22]**Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., ... & Taigman, Y. (2022). Make-a-video: Text-to-video generation without text-video data. arXiv preprint arXiv:2209.14792.
**[xu22]**Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. arXiv preprint arXiv:2204.12484.
**[zhang23]**Zhang, L., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. arXiv preprint arXiv:2302.05543.

**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 Adrien Maglo
Mohamed Chaouch
Tél +33 (0)1 69 08 02 88
E-mail adrien.maglo@cea.fr
mohamed.chaouch@cea.fr

| | |
|---|---|
| **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 :**

- 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 Guillaume Lapouge
Romain Dupont
E-mail guillaume.lapouge@cea.fr
romain.dupont@cea.fr

**Stage 2024** Réf : LVA-24-S3

# Towards a 3D real-time semantics-infused scene reconstruction

## Presentation of the host laboratory

Based in Paris-Saclay campus, CEA-LIST is one of four technological research institutes of CEA TECH, the technological research direction of CEA. Dedicated to intelligent digital systems, it contributes to the competitiveness of companies via research and knowledge transfers. The expertise and competences of the 800 research engineers and technicians at CEA-LIST help more than 200 companies in France and abroad every year on subjects categorized over 4 programs and 9 technological platforms. 21 start-ups have been created since 2003.

The Computer Vision and Machine Learning for scene understanding laboratory addresses computer vision subjects with a stronger emphasis on four axes:

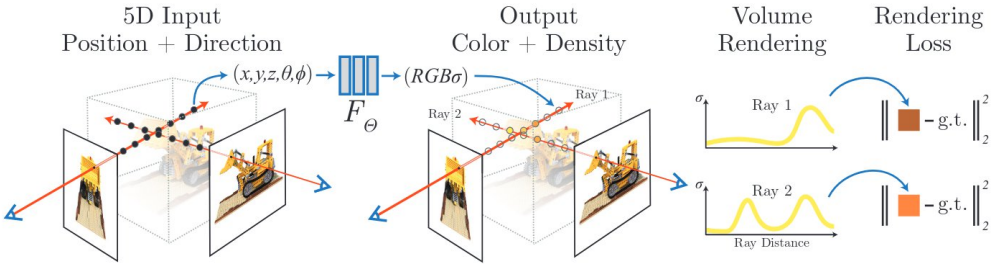
- Recognition (detection or segmentation of objects and persons)
- Behavior analysis (action and gesture recognition, anomalous behavior of individuals or crowds)
- Smart annotation (large scale annotation of 2D and 3D data using semi-supervised methods)
- Perception and decision-making (Markovian decision processes, navigation)
- The intern will join a team composed of 30 researchers (research engineers, PhD students, interns) and will be able to interact with peers working on related subjects and methods.

## Internship context

3D reconstruction has historically been made in an explicit fashion, where 3D objects are either modeled by a point cloud, a mesh or voxels. These explicit representations involve sparse data that represent complex geometry.

Recently, implicit representations have met with great success. Complex 3D scenes are implicitely represented by a function which outputs, for each position in the 3D world, a density/occupancy (=1 if we are inside of an object, =0 else).

Such functions can be learned by Neural Radiance Fields (NeRF) [1]. It is a 3D scene representation technique that uses neural networks to reconstruct detailed 3D scenes from 2D images. It models the scene's volumetric data and appearance (color + density), allowing for realistic 3D scene rendering. Since NeRF, many advances have been made and even model dynamic scenes [2].

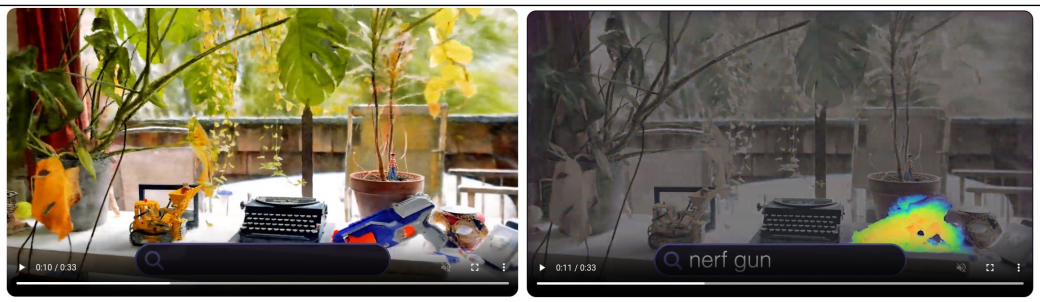



**3D representation infused with semantics (left), object retrieval with the text query « nerf gun » (right) from NeRF [1]**

However, these 3D representations do not explicitely model the scene semantics as they are purely reconstruction oriented. Recently, semantics have been infused in these 3D representations by leveraging 2D foundation models and distillating their features in the 3D scene allowing for surprisingly effective text-based object retrieval [3].

**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 Guillaume Lapouge
Romain Dupont
E-mail guillaume.lapouge@cea.fr
romain.dupont@cea.fr




**3D representation infused with semantics (left), object retrieval with the text query « nerf gun » (right) from LeRF [3]**

As performant as these methods are, their training is long and costly and the inference time slow. Recents efforts to accelerate them include Gaussian Splatting [4] which models the 3D scene by « painting it » with 3D Gaussians in space. However, such representations have not yet been used for semantics learning, as it presents its own sets of challenges.

## Internship objectives

This internship will be a part of a trans-laboratory effort to model 3D scenes and inject semantic information in that model. Such a 3D representation contains information on geometry, color and semantics and can be used for a wide variety of tasks requiring some level of scene understanding, such as object detection, segmentation or text query-based object retrieval.

In particular, in this internship, the work will be focused on the autonomous driving use case, where we want to reconstruct the dynamic scene in which the car drives in an efficient, if possible real-time manner.

In this internship, we will work on the following points :

- Efficiently learn 3D urban scenes representations that contain semantics information, leveraging fast learning methods such as gaussian splatting
- Study the object definition with respect to a chosen scale problem, that arises when an object is composed of smaller parts

Potentially, the reconstruction of dynamic scenes , with moving parts or objects could be studied.

## References

[1] MILDENHALL, Ben, SRINIVASAN, Pratul P., TANCIK, Matthew, et al. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 2021, vol. 65, no 1, p. 99-106.

[2] PUMAROLA, Albert, CORONA, Enric, PONS-MOLL, Gerard, et al. D-nerf: Neural radiance fields for dynamic scenes. In : Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. p. 10318-10327.

[3] KERR, Justin, KIM, Chung Min, GOLDBERG, Ken, et al. Lerf: Language embedded radiance fields. In : Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. p. 19729-19739.

[4] KERBL, Bernhard, KOPANAS, Georgios, LEIMKÜHLER, Thomas, et al. 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (ToG), 2023, vol. 42, no 4, p. 1-14.

| | |
|---|---|
| **Internship for:** | Engineer, Master 2 |
| This internship may eventually be followed by a PhD or an Engineering position in our laboratory | |
| **Duration:** | ~6 months |
| **Remuneration** : | between 700 € and 1300 € depending on academic background |
| **Required skills :**<br>- Computer Vision<br>- Machine Learning (especially Deep Learning)<br>- 3D geometry understanding<br>- Python, C/C++ (optional)<br>- Deep Learning Framework (PyTorch or 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 Bertrand Luvison
Mohamed Chaouch
Tél +33 (0)1 69 08 01 17
E-mail bertrand.luvison@cea.fr
mohamed.chaouch@cea.fr

**Stage 2024** Réf : LVA-24-S4

# Prédiction de posture pour l'anticipation d'intention

## 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.

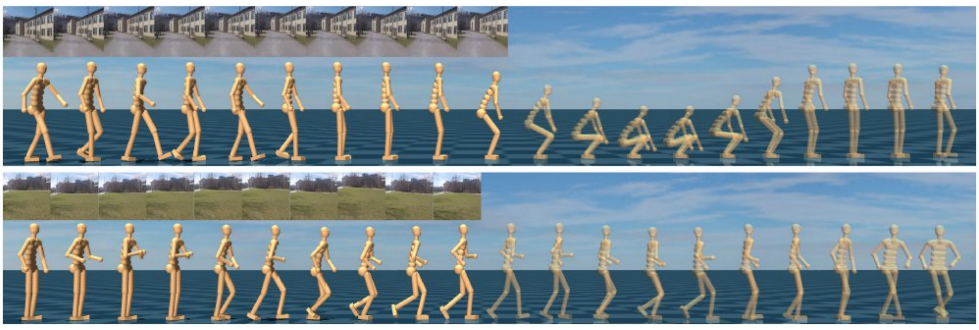
Le Laboratoire de Vision et Apprentissage pour l'analyse de scène (LVA) mène ses recherches dans le domaine de la Vision par Ordinateur (Computer Vision) selon quatre axes principaux :

- 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 estimateur et un prédicteur de posture 3D à partir de séquences vidéo. Cette recherche de l'anticipation dans la posture a de nombreux cas d'usage à commencer par le véhicule autonome, la cobotique ou de manière générale, toute application nécessitant une forte réactivité d'un système face à la présence d'une personne.

Cet estimateur devra être en mesure de réaliser son estimation même avec des informations temporelles partielles, en particulier lorsque les informations manquantes sont celles du futur proche. Après une évaluation des approches de l'état de l'art en estimation de squelette 3D à partir d'architecture spatio-temporelle [1] ainsi que l'étude des paradigmes d'analyse prédictive (« pose forecasting ») [2], l'objectif de ce stage sera de proposer une approche permettant d'anticiper au maximum la gestuelle humaine avec la plus grande précision possible. Pour cela de nombreuses bases de données existent, Human3.6M [3], AMASS [4], etc. A l'issue des développements, il sera important d'évaluer les performances de l'approche dans contextes variés tels que le véhicule autonome, l'assistance des personnes âgées ou la robotique. Les travaux menés durant le stage pourront faire l'objet d'une publication scientifique.



**Exemple de déformation 3D intra-classes (source [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 Bertrand Luvison
Mohamed Chaouch
Tél +33 (0)1 69 08 01 17
E-mail bertrand.luvison@cea.fr
mohamed.chaouch@cea.fr

## Keywords

Human 3D pose estimation, 3D pose forecasting, temporal modelling, attentional mechanism, deep learning, supervised learning.

## Références

[1] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, et Y. Wang, « MotionBERT: A Unified Perspective on Learning Human Motion Representations », ICCV, 2023.
[2] T. Sofianos, A. Sampieri, L. Franco, et F. Galasso, « Space-Time-Separable Graph Convolutional Network for Pose Forecasting », in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE, oct. 2021, p. 11189-11198. doi: 10.1109/ICCV48922.2021.01102.
[3] C. Ionescu, D. Papava, V. Olaru, et C. Sminchisescu, « Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments », IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, nº 7, p. 1325-1339, juill. 2014, doi: 10.1109/TPAMI.2013.248.
[4] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, et M. Black, « AMASS: Archive of Motion Capture As Surface Shapes », in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South): IEEE, oct. 2019, p. 5441-5450. doi: 10.1109/ICCV.2019.00554.
[5] Y. Yuan et K. Kitani, « Ego-Pose Estimation and Forecasting As Real-Time PD Control », in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South): IEEE, oct. 2019, p. 10081-10091. doi: 10.1109/ICCV.2019.01018.

| | |
|---|---|
| **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 :**<br>- Vision par ordinateur<br>- Apprentissage automatique (deep learning)<br>- Géométrie 3D et Reconnaissance de formes<br>- Python, C/C++<br>- Maîtrise d'un framework d'apprentissage profond (en particulier PyTorch ou Tensorflow) | |

**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 Hejer AMMAR
Romaric AUDIGER
Guillaume LAPOUGE
E-mail hejer.ammar@cea.fr
romaric.audigier@cea.fr
guillaume.lapouge@cea.fr

**Stage 2024** Réf : LVA-24-S5

# IA de confiance : Vers une détection d'objets robuste à un contexte nouveau, dans un monde ouvert

## 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].

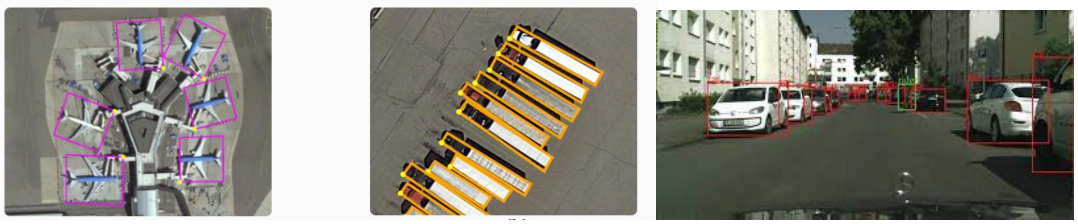

**Exemple de détection robuste selon plusieurs paradigmes**

Plusieurs méthodes ont étudié ce sujet dans le cadre de la classification d'images. 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, encore trop peu de travaux traitent l'OOD dans le cadre de la détection d'objets [2,3,4,5] et sont difficilement comparables du fait des différences de protocoles et de métriques d'évaluation. En outre, le comportement des détecteurs d'objets OOD dans des conditions plus réalistes, comme celles du monde ouvert, et d'une éventuelle confrontation à de nouveaux contextes spécifiques, tels que ceux des images aériennes [6], n'a pas été étudié. Ces contraintes opérationnelles

**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 Hejer AMMAR
Romaric AUDIGER
Guillaume LAPOUGE
E-mail hejer.ammar@cea.fr
romaric.audigier@cea.fr
guillaume.lapouge@cea.fr

posent donc de nouveaux défis.



**Exemple de détections dans des images aériennes issues de DOTA [6] et Cityscapes [7]**

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) et d'appliquer cette méthode dans des contextes nouveaux. Plusieurs axes d'étude seront explorés:

- Concevoir un détecteur d'objets conscient de l'existence de l'inconnu (*unknown-aware*). 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 ».
- Appliquer cette méthode à des contextes nouveaux (images aériennes vs images routières par exemple).
- Etendre la méthode vers un apprentissage *Open World*, où les nouvelles classes d'objets découvertes sont intégrées itérativement au modèle.

La publication d'articles scientifiques sera encouragée.

## 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] Wang, Yanghao and Yue, Zhongqi and Hua, Xian-Sheng and Zhang, Hanwang, Random Boxes Are Open-world Object Detectors, ICCV 2023
[3] Han, Jiaming and Ren, Yuqiang and Ding, Jian and Pan, Xingjia and Yan, Ke and Xia, Gui-Song, Expanding low-density latent regions for open-set object detection, CVPR 2022
[4] Maaz, Muhammad and Rasheed, Hanoona and Khan, Salman and Khan, Fahad Shahbaz and Anwer, Rao Muhammad and Yang, Ming-Hsuan, Class-agnostic object detection with multi-modal transformer, ECCV 2022
[5] Gupta, Akshita and Narayan, Sanath and Joseph, KJ and Khan, Salman and Khan, Fahad Shahbaz and Shah, Mubarak, Ow-detr: Open-world detection transformer, CVPR 2022
[6] XIA, Gui-Song, BAI, Xiang, DING, Jian, et al. DOTA: A large-scale dataset for object detection in aerial images. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 3974-3983.
[7] CORDTS, Marius, OMRAN, Mohamed, RAMOS, Sebastian, et al. The cityscapes dataset for semantic urban scene understanding. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 3213-3223.

| | |
|---|---|
| **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 :**<br>- Vision par ordinateur, reconnaissance de formes<br>- Apprentissage automatique (*deep learning*)<br>- Python, C/C++<br>- 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 Florian Chabot
Hejer Ammar
Tél +33 (0)1 69 08 02 88
E-mail florian.chabot@cea.fr
hejer.ammar@cea.Fr

**Stage 2024** Réf : LVA-24-S6

# In the pursuit of true motion: compensation of camera movement for better scene recognition

## Presentation of the host laboratory

Based in Paris-Saclay campus, CEA-LIST is one of four technological research institutes of CEA TECH, the technological research direction of CEA. Dedicated to intelligent digital systems, it contributes to the competitiveness of companies via research and knowledge transfers. The expertise and competences of the 800 research engineers and technicians at CEA-LIST help more than 200 companies in France and abroad every year on subjects categorized over 4 programs and 9 technological platforms. 21 start-ups have been created since 2003.

The Computer Vision and Machine Learning for scene understanding laboratory addresses computer vision subjects with a stronger emphasis on four axes:

- Recognition (detection or segmentation of objects and persons)
- Behavior analysis (action and gesture recognition, anomalous behavior of individuals or crowds)
- Smart annotation (large scale annotation of 2D and 3D data using semi-supervised methods)
- Perception and decision-making (Markovian decision processes, navigation)
- The intern will join a team composed of 30 researchers (research engineers, PhD students, interns) and will be able to interact with peers working on related subjects and methods.

## Context

Optical flow is the motion information that describes the displacement of each pixel between two frames of a video. In the case of a fixed camera, optical flow can be used to highlight moving objects, which are the regions of interest for various vision applications: moving objects segmentation, anomaly detection, etc. However, when analyzing dynamic scenes (i.e. observed by a moving camera), the optical flow becomes a combination of the movement of the objects in the scene along with the camera motion. This causes the background to be also mobile, making applications based on optical flow challenging.

Several state-of-the-art methods still use an uncompensated optical flow for solving vision tasks [1,2], but this comes at the cost of tedious filtering of the noise generated by background motion. Other methods have emphasized the importance of handling moving background early on [3], in motion maps, but propose normalizations that attenuate camera motion without handling it robustly. In addition, geometric image registration methods such as homography [4], which are effective in simple scenarios, are not adapted for non-planar scenes with complex camera movements, as in the case of road scenes.

The aim of this internship is to design and develop a deep learning-based method capable of separating optical flow from camera movement, also known as rigid flow or ego-motion. See illustration below.

**Illustration of the benefit of ego-motion compensation for scene recognition: optical flow map before (middle) and after compensation (right).**

## Objectives of the internship

After conducting a scientific survey on rigid flow compensation, the trainee will be able to investigate several areas of study:

- Explore supervised methods for learning the separation of optical and rigid flow, on synthetic datasets (ground truth available from previous work).
- Develop a rigid flow compensation method capable of exploiting camera poses and self-

**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 Florian Chabot
Hejer Ammar
Tél +33 (0)1 69 08 02 88
E-mail florian.chabot@cea.fr
hejer.ammar@cea.Fr

supervised depth maps. This step aims to alleviate the constraint of having a perfect depth map (ground truth).

- To take this a step further, we will look at the case of camera pose unavailability, by investigating methods for estimating the camera's trajectory (ego-motion).

## References

[1] A. Dave, P. Tokmakov, and D. Ramanan. Towards segmenting anything that moves. In IEEE/CVF InternationalConference on Computer Vision Workshop (ICCVW), 2019.
[2] Zhipeng Bao, Pavel Tokmakov, Allan Jabri, Yu-Xiong Wang,Adrien Gaidon, and Martial Hebert. Discorying object that can move. In CVPR, 2022
[3] Xinyu Zhang, Abdeslam Boularias. Optical Flow boosts Unsupervised Localization and Segmentation. In IROS, 2023.
[4] Yaqing Ding, Daniel Barath, and Zuzana Kukelova. Homography-Based Egomotion Estimation Using Gravity and SIFT Features. In Computer Vision – ACCV 2020

## Keywords

Optical flow, moving objects, dynamic scene, deep learning, ego-motion compensation

| | |
|---|---|
| **Internship for:** | Engineer, Master 2 |
| This internship may eventually be followed by a PhD or an Engineering position in our laboratory | |
| **Duration:** | ~6 months |
| **Remuneration** : | between 700 € and 1300 € depending on academic background |
| **Required skills :**<br>- Computer Vision<br>- Machine Learning (especially Deep Learning)<br>- 3D geometry understanding<br>- Python, C/C++ (optional)<br>- Deep Learning Framework (PyTorch or 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 Fabrice Mayran De Chamisso
Aleksandr Setkov
Tél +33 (0)1 69 08 02 88
E-mail fabrice.mayran-de-chamisso@cea.fr
aleksandr.setkov@cea.fr

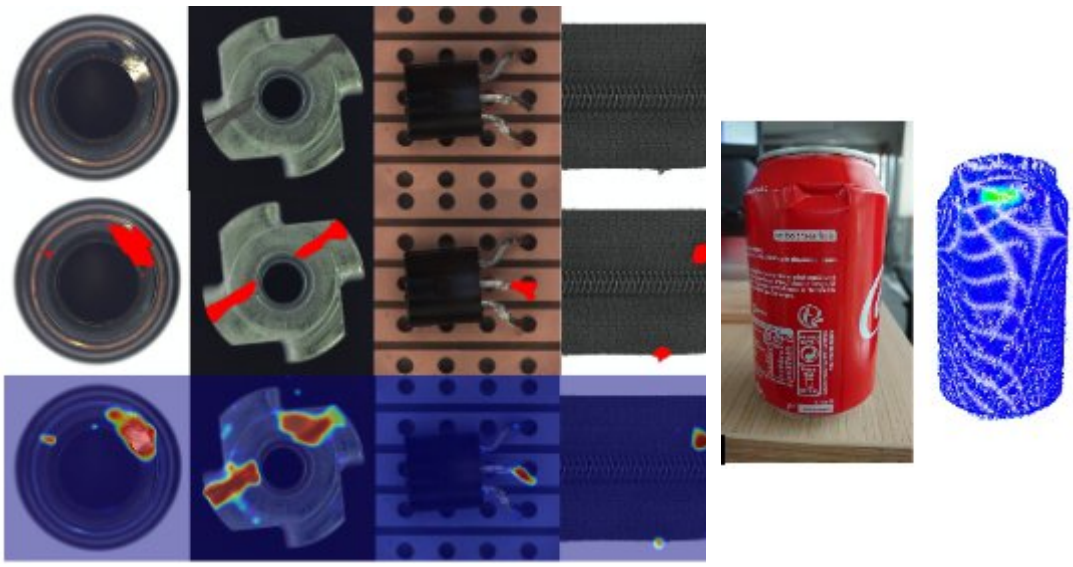
**Stage 2024** Réf : LVA-24-S7

# Détection d'anomalies images dans des objets par apprentissage profond sur peu de données

## Contexte du stage

La détection d'anomalies (défauts notamment) dans des objets est une thématique de recherche importante ainsi qu'un enjeu industriel. Détecter une anomalie tôt dans un processus de production représente un enjeu écologique (arrêt et correction avant de produire des éléments inutilisables, gains de matière première) et financier (coût de la matière première et temps machine). Cela vaut aussi pour le domaine agroalimentaire, où par exemple détecter de la moisissure sur un fruit et l'éliminer peut permettre d'éviter la contamination du cageot complet.

L'une des problématiques rencontrées est que la collecte de données présentant des anomalies est complexe (notamment en environnement industriel) et fastidieuse. Et ce sans même compter la nécessité d'annoter lesdites données. La tendance générale de l'état de l'art est donc soit de réduire le nombre de données en entrée nécessaire pour entraîner de l'IA, soit de pouvoir générer de la donnée de synthèse réaliste (par exemple, rendu photoréaliste), soit de pouvoir transférer un apprentissage plus facilement à une nouvelle donnée. Notre service s'intéresse naturellement à ces trois problématiques : modélisation (pour le rendu) et approches utilisant peu de données tout en permettant de segmenter et localiser des défauts géométriques et colorimétriques par géométrie et par apprentissage (outils statistiques, deep learning).



**A gauche: Détection d'anomalies par apprentissage (images sources, vérité terrain, anomalies détectées). A droite : détection d'anomalies géométriques**

## Objectifs du stage

Le premier objectif de ce stage sera d'étudier et évaluer des approches de l'état de l'art récentes ([1,2,3,4]) de détection de défauts industriels par apprentissage (non-supervisée et/ou One-Class) ainsi que d'améliorer une ou plusieurs méthodes développées au CEA [5] :

- Trouver les jeux de données pertinents et les approches utilisées pour les traiter (par exemple, le dataset MVTEC Anomaly detection et ses descendants)
- Mettre en place et comparer les approches (évaluation qualitative et quantitative)
- Prendre en main les approches développées par le CEA et les comparer à l'état de l'art
- Le second objectif de ce stage consistera à améliorer les approches CEA, par exemple :
- Proposer des modifications sur la donnée (augmentations, …) et sur l'approche (structure

**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 Fabrice Mayran De Chamisso
Aleksandr Setkov
Tél +33 (0)1 69 08 02 88
E-mail fabrice.mayran-de-chamisso@cea.fr
aleksandr.setkov@cea.fr

de réseau, procédure d'entraînement, …)

- Constituer un dataset pour la détection de défauts permettant d'entraîner les approches (le service dispose d'un robot UR10 permettant d'automatiser certaines tâches)
- Proposer une nouvelle méthode innovante couplant géométrie et deep learning afin d'atteindre et dépasser l'état de l'art
- Évaluer les performances obtenues.

## Références

[1] Božič, J.; Tabernik, D. & Skočaj, D., Mixed supervision for surface-defect detection: From weakly to fully supervised learning , Computers in Industry, 2021, 129, 103459

[2] C.-L. Li, K. Sohn, J. Yoon, T. Pfister, CutPaste: Self-Supervised Learning for Anomaly Detection and Localization, CVPR, 2021

[3] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, Beyond Dents and Scratches: Logical Constraints in Unsupervised Anomaly Detection and Localization, IJCV, 2022

[4] D. A. Gudovskiy, S. Ishizaka, K. Kozuka, CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows, WACV 2022

[5] T. Defard , A. Setkov , A. Loesch , R. Audigier, PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization, ICPR Workshops, 2021

## Compétences

Le candidat devra disposer des connaissances en vision par ordinateur et en machine learning, d'une bonne maîtrise du langage Python et du framework d'apprentissage Pytorch. La maîtrise d'autres frameworks (Tensorflow 2, etc.) est un plus. La connaissance du C++ et des bonnes pratiques du développement logiciel est souhaitable. Le candidat disposera d'une grande autonomie, un esprit d'initiative est attendu.

| | |
|---|---|
| **Formation / Niveau d'étude** | Ingénieur, Master 2 / Bac+5 |
| **Possibilité poursuite** | Oui, en thèse ou CDD selon profil. |
| **Durée** | 6 mois |
| **Lieu** | Palaiseau (91) – Centre d'intégration de Nano-INNOV |
| **Indemnités de stage** | Entre 700 € et 1400 € suivant formation.<br>Aide au logement / transport / restauration. |

## Candidatures

- Joindre CV + lettre de motivation + tous les bulletins de notes post-bac à **fabrice.mayran-de-chamisso@cea.fr et aleksandr.setkov@cea.fr** avec le nom du stage auquel vous postulez
- Ne pas hésiter à détailler les projets ou cours auxquels vous avez participé
- Indiquer les dates de début/fin de stage envisagées.
- Ce stage pourra prendre une orientation recherche ou industrie en fonction du profil du candidat
- En raison du grand nombre de candidatures reçues, merci de considérer une absence de réponse sous trois semaines comme une réponse négative.

**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

**Stage 2024** Réf : LVA-24-S8

# UTILISATION DES MODÈLES MULTI-MODAUX POUR LE CONTRÔLE ROBOTIQUE AVANCÉ

## 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 travaux de recherche dans le domaine des modèles de perception pour la prise de décision se concentrent sur le développement d'algorithmes de contrôle robustes basés sur l'apprentissage par renforcement et l'observation visuelle. Nos objectifs incluent la création de solutions de pointe pour la manipulation robotique et la navigation autonome, en partenariat avec des acteurs industriels. Dans ce contexte, nous sommes confrontés à plusieurs défis scientifiques, notamment l'extraction d'informations pertinentes à partir de données visuelles et la modélisation dynamique du monde pour une planification robuste.

Nous invitons un chercheur stagiaire à rejoindre notre équipe de chercheurs travaillant sur ces sujets stimulants. L'accent sera mis sur l'exploration de l'application des modèles multi-modaux, des Large Language Models (LLMs) et des Visual Language Models (VLMs) pour le contrôle robotique avancé.

## Rôle

En tant que stagiaire, vous collaborerez étroitement avec notre équipe de chercheurs pour :

- Étudier l'état de l'art des LLMs et des VLMs, et identifier les méthodes les plus pertinentes pour le contrôle robotique.
- Implémenter ces méthodes et mener des expériences d'apprentissage pour évaluer leur efficacité dans des scénarios de manipulation robotique.
- Proposer des améliorations méthodologiques et algorithmiques pour optimiser les performances et la robustesse des systèmes robotiques.
- Présenter les résultats de vos recherches à l'équipe et contribuer à la rédaction de publications scientifiques.

## Profil Requis

Pour postuler à ce stage, vous devez :

- Être inscrit dans un programme de Master de recherche avec une spécialisation en intelligence artificielle, apprentissage statistique ou apprentissage par renforcement.
- Avoir une expérience prouvée ou un intérêt marqué pour la recherche en intelligence artificielle et en apprentissage par renforcement.
- Être familier avec les outils et frameworks de développement de modèles d'apprentissage (Jax).

**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 Romaric AUDIGIER
Angélique LOESCH
E-mail romaric.audigier@cea.fr
angelique.loesch@cea.fr

**Stage 2024** Réf : LVA-24-S9

# IA de confiance : Défense des modèles d'analyse de scène contre les attaques antagonistes

## 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

Dans de nombreuses applications de vision par ordinateur, des briques d'analyse de scène comme la segmentation sémantique, la détection et la reconnaissance d'objets, ou la reconnaissance de pose sont nécessaires. Les réseaux de neurones profonds sont aujourd'hui parmi les modèles les plus efficaces pour effectuer un grand nombre de tâches de vision. Cependant, ceux-ci peuvent être vulnérables face à des attaques antagonistes (*adversarial attacks*) : En effet, il est possible d'ajouter aux données d'entrée certaines perturbations imperceptibles par l'œil humain qui mettent à mal les résultats lors de l'inférence faite par le réseau de neurones. Or, une garantie de résultats fiables est capitale pour les systèmes de décision où les failles de sécurité sont critiques (ex : applications comme le véhicule autonome, la reconnaissance d'objets en surveillance aérienne, ou la recherche de personnes/véhicules en vidéosurveillance).

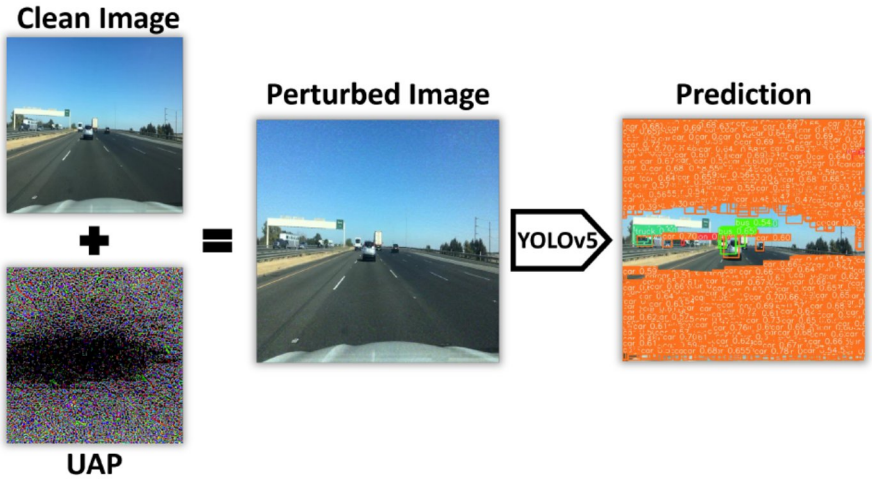



***Fig.1 : Illustration de l'effet de l'UAP (Universal Adversarial Perturbation) [7] sur un détecteur YOLOv5***

**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 Romaric AUDIGIER
Angélique LOESCH
E-mail romaric.audigier@cea.fr
angelique.loesch@cea.fr

Différents types d'attaques adversaires et de défense ont été proposés, le plus souvent pour le problème de classification (d'images [1] notamment). Quelques travaux ont abordé l'attaque des plongements optimisés par apprentissage de métrique, notamment utilisés pour les tâches de type ensemble-ouvert comme la réidentification d'objets [2], la reconnaissance faciale ou la recherche d'images par le contenu. Les types d'attaques se sont multipliés, qu'il s'agisse d'attaques universelles ou optimisées sur une instance particulière. Les défenses proposées doivent faire face à de nouvelles menaces sans trop sacrifier les performances initiales du modèle. Récemment, des attaques de modèles ont été proposées pour la tâche composée de détection d'objets [3-11]. Les attaques ou défenses de modèles multitâches comme ceux de recherches d'objets (personnes, véhicules...) dans les images sont encore rares.

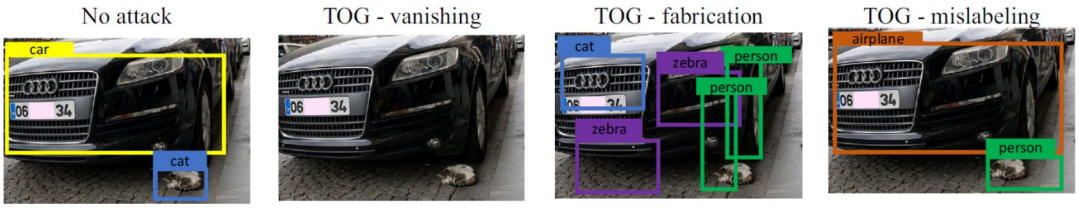



***Fig.2 : Exemple d'attaques TOG (Targeted adversarial Objectness Gradient) [9]***


***Fig.3 : Exemple de « capes d'invisibilité » pour un détecteur YOLO en vues frontales [8] ou aériennes [11]***

## Objectif

L'objectif de ce stage est d'étudier et de proposer différentes attaques et défenses applicables aux briques d'analyse de scène, notamment celles de détection d'objets et de recherche d'instance d'objet dans les images.

- Divers types d'attaques seront étudiés, qu'il s'agisse d'attaques au niveau de l'image ou dans le monde physique [8,10,11].
- Différentes approches de défense seront étudiées afin d'évaluer quels sont les meilleurs compromis entre performances et robustesse aux attaques.
- Les méthodes pourront être testées et évaluées dans des contextes variés (ex : véhicule autonome terrestre [9], vues aériennes [10-11]).
- La publication d'articles scientifiques sera encouragée.

**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 Romaric AUDIGIER
Angélique LOESCH
E-mail romaric.audigier@cea.fr
angelique.loesch@cea.fr

## Références

[1] Bouniot et al., "Optimal transport as a defense against adversarial attacks", ICPR 2021.
[2] Bouniot et al., "Vulnerability of person reidentification models to metric adversarial attacks", IEEE/CVF CVPR-Workshops 2020.
[3] Mi et al., "Adversarial examples based on object detection tasks: A survey", Neurocomputing,vol. 519, 2023.
[4] Byun et al., "Improving the Transferability of Targeted Adversarial Examples through Object-Based Diverse Input", IEEE/CVF CVPR 2022.
[5] Cai et al., "Context-Aware Transfer Attacks for Object Detection", AAAI 2022.
[6] Yin et al., "ADC: Adversarial attacks against object Detection that evade Context consistency checks", IEEE/CVF WACV 2022.
[7] Shapira et al., "Phantom Sponges: Exploiting Non-Maximum Suppression to Attack Deep Object Detectors," IEEE/CVF WACV 2023.
[8] Wu et al., "Making an invisibility cloak: Real world adversarial attacks on object detectors", ECCV 2020.
[9] Amirkhani et al., "A survey on adversarial attacks and defenses for object detection and their applications in autonomous vehicles", The Visual Computer, 2022.
[10] Du et al., "Physical Adversarial Attacks on an Aerial Imagery Object Detector", IEEE/CVF WACV 2022.
[11] Zhang et al., "Adversarial Patch Attack on Multi-Scale Object Detection for UAV Remote Sensing Images", Remote Sensing, vol.14, 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 :**

- Vision par ordinateur, reconnaissance de formes
- Apprentissage automatique (*deep learning*)
- Python
- Maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch)