



list
cea tech

Offres de stages 2021

AI - Deep Learning

Computer Vision - Scene Understanding

- 1 Apprentissage de représentation et apprentissage par renforcement pour la navigation autonome
- 2 Few-Shot Learning : Comment apprendre avec peu de données annotées ?
- 3 Suivi d'objets par ré-identification
- 4 Prédiction de cartes de profondeur à partir d'images monoculaires
- 5 Segmentation panoptique 3D
- 6 Suivi d'instances dans les données 3D LIDAR par apprentissage profond
- 7 Détection d'événements anormaux dans les vidéos
- 8 Adaptation de domaine pour la détection et la segmentation: du capteur pinhole au capteur fisheye
- 9 Suivi visuel par segmentation d'instances et apprentissage profond



Laboratoire de Vision et d'Apprentissage
pour l'analyse de scène
www.kalisteo.eu
e-mail: lva-stages@cea.fr



STAGE 2021

Réf : LVA-2021-S1

Apprentissage de représentation et apprentissage par renforcement pour la navigation autonome

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

L'apprentissage par renforcement a permis de développer des algorithmes capables de battre les humains dans différentes tâches de contrôle et prise de décision comme les jeux Atari ou les jeux de plateau (Go, Echecs). Son utilisation pour la conduite autonome est prometteur mais souffre de plusieurs problèmes :

- Il est difficile de résoudre le problème d'optimisation associé à ces algorithmes quand les espaces d'état et d'action sont de très grande dimension.
- La phase d'apprentissage nécessite beaucoup d'interaction entre l'agent et l'environnement. C'est un processus difficile à mettre en œuvre dans le cas de la conduite autonome.
- Les algorithmes d'apprentissage par renforcement sont souvent perçus comme une boîte noire. Or les notions d'explicabilité et d'interprétabilité sont primordiales pour faire accepter ces solutions dans le monde de l'automobile et par les usagers.

Ce stage a pour objectif d'apporter des solutions à ces problèmes. Notamment, en étudiant les méthodes d'apprentissage de représentation pour réduire la dimensionnalité des espaces d'état et d'action. Pour limiter l'interaction avec l'environnement il est possible de s'appuyer sur les techniques d'apprentissage hors-ligne (offline reinforcement learning). Finalement, la notion d'interprétabilité peut être introduite dans ces algorithmes en ajoutant des tâches à niveau sémantique élevé dans ces algorithmes (prédiction de trajectoire, segmentation de la scène, ...). Il sera demandé au stagiaire de faire une étude bibliographique, de proposer et d'implémenter une solution aux différents problèmes soulevés et finalement procéder à une validation expérimentale.

Keywords : *computer vision, deep learning, reinforcement learning, off-line reinforcement learning, self-supervised learning, interpretability.*



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

Contact: Jaonary Rabarisoa
Jaonary.rabarisoa@cea.fr
+33 (0)1 69 08 01 29

Références

- [1] Learning by cheating. Dian Chen et al. <https://arxiv.org/abs/1912.12294>.
- [2] Interpretable End-to-end Urban Autonomous Driving with Latent Deep Reinforcement Learning. Jianyu Chen et al. <https://arxiv.org/abs/2001.08726>.
- [3] An Optimistic Perspective on Offline Reinforcement Learning. Rishabh Agarwal et al. <https://arxiv.org/abs/1907.04543>.

Profil du candidat/de la candidate

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



Few-Shot Learning : Comment apprendre avec peu de données annotées ?

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

De nombreuses tâches en vision par ordinateur sont aujourd'hui réalisées avec succès par les méthodes d'apprentissage profond (*deep learning*). C'est le cas par exemple de la détection et la reconnaissance d'objets dans les images, la classification d'images, la reconnaissance de personnes ou d'activités... utiles dans divers domaines d'application (vidéosurveillance, conduite autonome, imagerie médicale, etc.). L'inconvénient de ces approches basées réseaux de neurones profonds est que leur entraînement supervisé requiert de grandes quantités de données annotées. D'une part, l'annotation manuelle de données est une tâche longue et coûteuse. D'autre part, ces données peuvent être rares ou difficiles à collecter. Il est alors nécessaire d'avoir des méthodes d'apprentissage qui se basent sur très peu d'exemples.

Deux stratégies principales se distinguent. L'une consiste à apprendre sur une seule tâche globale des caractéristiques suffisamment génériques pour résoudre la tâche élémentaire cible (*feature reuse*). Il s'agit alors de régler finement les dernières couches (*fine-tuning*) d'un réseau de neurones existant sur les données rares ciblées [Wang20]. L'utilisation de données augmentées ou non annotées (lorsque celles-ci ne sont pas rares) rend alors l'apprentissage plus générique. L'autre stratégie est le paradigme de méta-apprentissage (*meta-learning*). Il consiste à apprendre à apprendre de multiples tâches élémentaires à partir de peu de données, pour ensuite s'adapter facilement à la tâche élémentaire ciblée [Finn17]. Dans le cas où les données sont nombreuses mais les annotations rares, certaines méthodes exploitent l'information des données non annotées [Ren18, Gidaris19].

L'objectif de ce stage est d'explorer des stratégies basées sur les paradigmes d'apprentissage semi-supervisé et/ou auto-supervisé qui permettent de répondre à la problématique du *few-shot learning*. On s'intéressera, en particulier, aux tâches de classification d'images et de détection d'objets [Karlinsky19, Yan19]. Dans un premier temps, l'appropriation des méthodes de l'état de l'art permettra d'analyser leurs avantages et leurs limitations. Dans un second temps, de nouvelles méthodes seront développées. Elles seront ensuite évaluées et comparées à l'état de l'art sur divers jeux de données afin de quantifier l'amélioration des performances, la dépendance au nombre d'échantillons annotés/non-annotés, ainsi que la capacité de généralisation face aux types de données. Les résultats du stage pourront faire l'objet de publication.

Keywords : *computer vision, deep learning, few-shot learning, meta-learning, semi-supervised learning, self-supervised learning.*

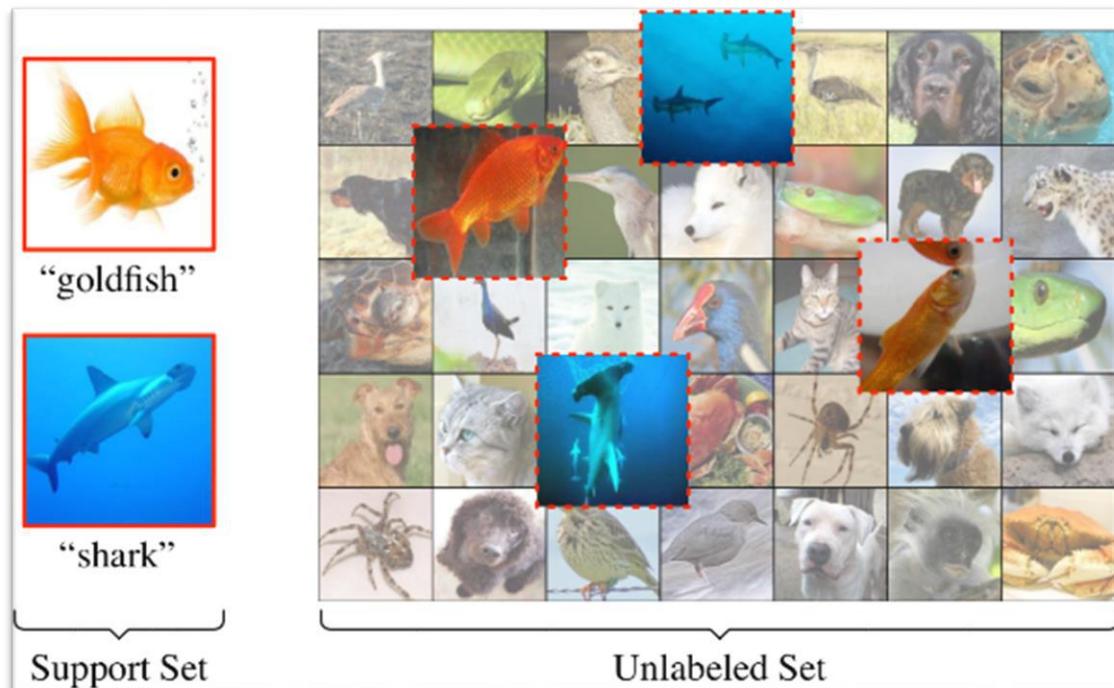


Fig. 1 La méthode [Ren18] propose une approche semi-supervisée : Apprendre à reconnaître de nouvelles classes d'objets à partir de peu d'exemples annotés et des données non annotées, dont de multiples distracteurs.

Références

- [Wang20] Wang X. et al. (2020) Frustratingly Simple Few-Shot Object Detection. *ICML*.
- [Finn17] Finn C. et al. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *ICML*.
- [Ren18] Ren M et al. (2018) Meta-learning for semi-supervised few-shot classification. *ICLR*.
- [Gidaris19] Gidaris S. et al. (2019) Boosting Few-Shot Visual Learning with Self-Supervision. *ICCV*.
- [Karlinsky19] Karlinsky L. et al. (2019) Repmet : Representative-based metric learning for classification and few-shot object detection. *CVPR*.
- [Yan19] Yan X. et al. (2019) Meta R-CNN : Towards General Solver for Instance-Level Low-Shot Learning. *ICCV*.

Profil du candidat/de la candidate

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



CEA List

Laboratoire de Vision et d'Ingénierie des Contenus

Centre de Saclay 91191 Gif-sur-Yvette France

<http://www.kalisteo.eu>

Contact Bertrand LUVISON

Tél +33 (0)1 69 08 01 37

E-mail bertrand.luvison@cea.fr

STAGE 2021

Réf : AS-21-S3

Suivi d'objets par ré-identification

Présentation du laboratoire d'accueil

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

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

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

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

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

Description du stage

Le suivi de personne est une problématique d'étude de la vision par ordinateur depuis plusieurs décennies. Historiquement très tourné vers le suivi visuel avec des méthodes de filtrage diverses, les méthodes plus récentes [1] ont fait la place belle au suivi par détection grâce aux très grands progrès réalisés en détection d'objet via les méthodes de *deep learning*. Cependant malgré l'amélioration des performances de détection, la problématique de suivi ajoute des difficultés qui lui sont propres : gérer les apparitions, les disparitions, correctement ré-identifier les personnes lors de croisement ou de disparition ponctuelle, etc. Pour cela, différents mécanismes sont associés les uns aux autres pour aboutir au système de suivi complet. On peut citer les heuristiques de création/destruction bien souvent ajustées au contexte d'utilisation ou les méthodes de ré-identification qui à partir de deux images de personne issues de la même caméra à des instants différents ou de caméras différentes cherchent à dire s'il s'agit des mêmes personnes.

L'objectif de ce stage est de chercher à unifier détection et ré-identification voire création/destruction de piste dans une seule et même architecture de réseau de neurones profond. Il faudra pour cela, concevoir et proposer une architecture capable de répondre au problème de suivi dans son ensemble en s'appuyant sur les récentes approches [2] ainsi que sur les nombreux datasets publics de suivi et de ré-identification disponibles. Le stagiaire pourra s'appuyer sur l'existant du laboratoire, aussi bien en terme de technologies de détection d'objets [2], de ré-identification [3] que de suivi.

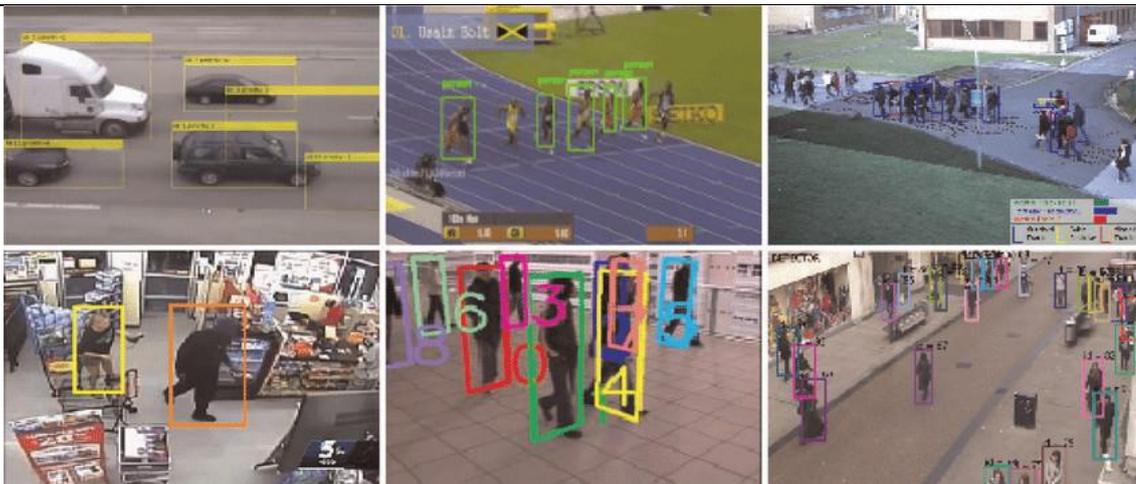


Figure 1. Exemple d'application du suivi multi-personnes.

[1] W. Nicolai, B. Alex and P. Dietrich, "Simple Online and Realtime Tracking with a Deep Association Metric", IEEE International Conference on Image Processing ICIP, 2017
 [2] Y. Zhang, C. Wang, X. Wang, W. Zeng, W. Liu, "FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking", <https://arxiv.org/abs/2004.01888>, 2020
 [3] F. Chabot, Q.C. Pham and M. Chaouch, "LapNet : Automatic Balanced Loss and Optimal Assignment for Real-Time Dense Object Detection", <https://arxiv.org/abs/1911.01149>, 2020
 [4] A. Loesch, J. Rabarisoa and R. Audigier, "End-To-End Person Search Sequentially Trained On Aggregated Dataset", IEEE International Conference on Image Processing (ICIP), 2019

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



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
Tél +33 (0)1 69 08 02 88
E-mail florian.chabot@cea.fr

STAGE 2021

Réf : LVA-21-S4

Prédiction de cartes de profondeur à partir d'images monoculaires

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 :

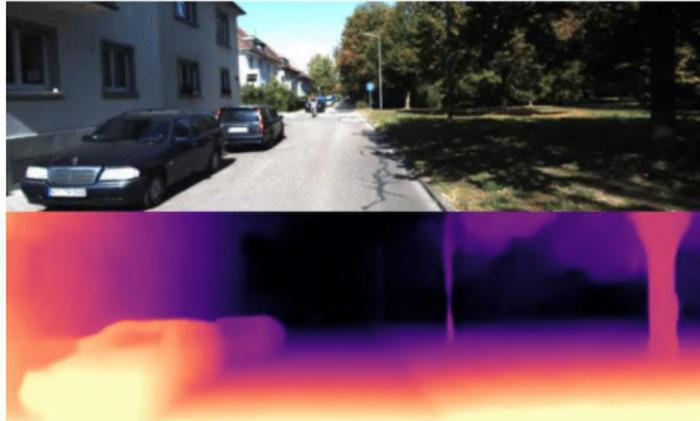
- La reconnaissance visuelle (détection et/ou segmentation d'objets, de personnes, de patterns ; détection d'anomalies ; caractérisation)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, de comportements anormaux ou spécifiques pour des individus, un groupe, une foule)
- 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

L'objectif de ce stage est de proposer une méthode hybride (basée sur l'apprentissage profond supervisé et auto-supervisé) pour la prédiction de cartes de profondeur à partir d'une image pour des applications de type véhicules autonomes. Les approches supervisées récentes pour résoudre cette tâche utilisent généralement un nuage de points 3D (LIDAR) comme vérité terrain pour superviser l'apprentissage [1, 2]. Cependant, l'information provenant du LIDAR est incomplète, sparse et parfois bruitée ce qui ne permet pas de prédire avec précision la profondeur sur certaines zones de l'image. Une autre famille de méthodes s'intéresse à l'analyse du mouvement entre deux images pour inférer la profondeur par des mécanismes d'apprentissage auto-supervisé [3]. Ces approches ont l'avantage de n'utiliser que des données non-annotées pour entraîner le modèle. Cependant, elles présentent deux défauts majeurs :

- l'incapacité d'apprendre correctement la profondeur quand les objets de la scène sont en mouvement
- la profondeur est estimée à un facteur d'échelle prêt

Pour pallier ces problèmes, le stagiaire devra dans un premier temps investiguer l'ajout de l'auto-supervision dans les approches supervisées afin de d'améliorer l'estimation de la profondeur. Dans un second temps, il devra analyser les cas d'erreurs et proposer des solutions pour les résoudre. Enfin, l'algorithme développé devra être comparé aux méthodes existantes sur des benchmarks académiques.



Estimation de profondeur monoculaire par apprentissage auto-supervisé [3]

Références

- [1] Fu et al. Deep Ordinal Regression Network for Monocular Depth Estimation, CVPR 2018
- [2] Diaz et al. Soft Labels for Ordinal Regression, CVPR 2019
- [3] Godard et al. Digging into Self-Supervised Monocular Depth Prediction, ICCV 2019

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



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 Nicolas Granger

Tél +33 (0)1 69 08 11 32

E-mail nicolas.granger@cea.fr

STAGE 2021

Réf : LVA-21-S5

Segmentation panoptique 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.

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

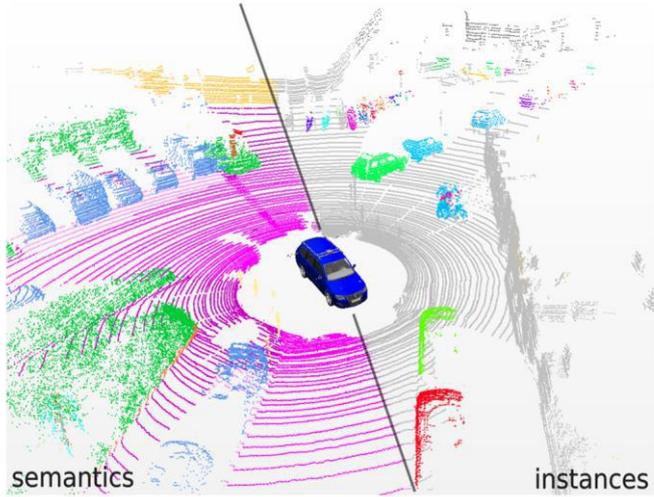
Le Laboratoire de Vision et Apprentissage pour l'analyse de scène (LVA) mène ses recherches dans le domaine de la Vision par Ordinateur (Computer Vision) selon quatre axes principaux :

- 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

L'objectif de ce stage est de proposer une nouvelle méthode de segmentation panoptique d'un nuage de point 3D. La segmentation panoptique 3D [2, 3, 4] prédit une carte de segmentation regroupant deux types d'information: la segmentation sémantique et la segmentation d'instances. La segmentation sémantique consiste à attribuer à chaque élément du nuage de point la classe auquel il appartient (arbre, route, voiture, personne...) tandis que la segmentation d'instances produit un masque pour chaque objet dans la scène. Le regroupement de ces deux types de segmentation fournit une description 3D très riche ce qui s'avère très utile dans un certain nombre d'application (notamment la conduite autonome, cartographie aérienne).

Le stagiaire devra s'inspirer des algorithmes de reconnaissance 3D existants et des approches de segmentation panoptique appliqués aux images pour proposer une approche de segmentation panoptique 3D.



Segmentation panoptique 3D [1]

Références

[1] Behley et al. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences, ICCV 2019

[2] Wu et al. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud, ICRA 2018

[3] Alonso et al. 3D-MiniNet: Learning a 2D Representation from Point Clouds for Fast and Efficient 3D LIDAR Semantic Segmentation , IROS 2020

[4] Biasutti et al. LU-Net: An Efficient Network for 3D LiDAR Point Cloud Semantic Segmentation Based on End-to-End-Learned 3D Features and U-Net, ICCVW, 2019

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



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 Camille Dupont

Tél +33 (0)1 69 08 27 03

E-mail camille.dupont@cea.fr

STAGE 2021

Réf : LVA-21-S6

Suivi d'instances dans les données 3D LiDAR par apprentissage profond

Présentation du laboratoire d'accueil

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

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

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

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

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

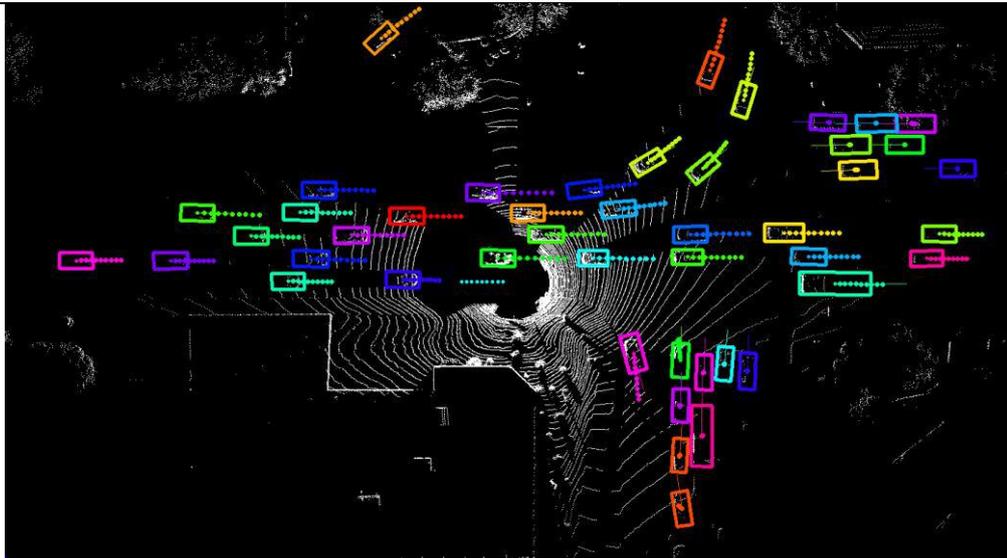
Description du stage

Les données LiDAR sous forme de nuages de points 3D constituent une source d'information robuste pour la classification et le positionnement d'objets dans une scène.

Ces dernières années, des progrès considérables ont été réalisés sur les méthodes de détection à base de réseaux de neurones, à la fois en termes de coût de calcul et en métrique d'évaluation [4,6,7]. Ces améliorations s'appuient aussi sur de nouveaux jeux de données plus grands et plus diversifiés [1,2,5] qui permettent de mieux entraîner et évaluer les modèles. Ces nouveaux datasets permettent aussi d'étudier de nouvelles problématiques comme le suivi (*tracking*) temporel des objets et l'anticipation de trajectoires, qui pourra notamment trouver son application dans le contexte automobile pour l'aide à la conduite ou la conduite autonome.

À ce jour, l'état de l'art en suivi multi-objet dans les nuages de points repose largement sur la performance d'un détecteur, dont les prédictions successives sont combinées avec un algorithme de suivi simple [7].

Ce stage a donc pour objectif d'étudier l'utilisation de méthodes d'apprentissage pour améliorer les performances du suivi temporel d'objets dans des nuages de points 3D [7], en particulier sur les cas difficiles (occultation, objet peu détaillé, etc.).



Exemple de prédictions de pistes d'instances 3D [3]

Références

- [1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," CVPR, 2012.
- [2] P. Sun et al., Scalability in Perception for Autonomous Driving: Waymo Open Dataset. 2019.
- [3] W. Luo, B. Yang, and R. Urtasun, "Fast and Furious: Real Time End-to-End 3D Detection, Tracking and Motion Forecasting with a Single Convolutional Net," CVPR, 2018.
- [4] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection," CVPR, 2018.
- [5] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," arXiv:1903.11027, 2019.
- [6] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection from Point Clouds," arXiv:1812.05784, 2018.
- [7] S. Shi et al., "PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection," CVPR, 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 - C/C++, Python - La maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch) est un plus. 	



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

STAGE 2021

Réf : LVA-21-S7

Détection d'évènements anormaux dans les vidéos

Présentation du laboratoire d'accueil

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

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

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

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

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

Description du stage

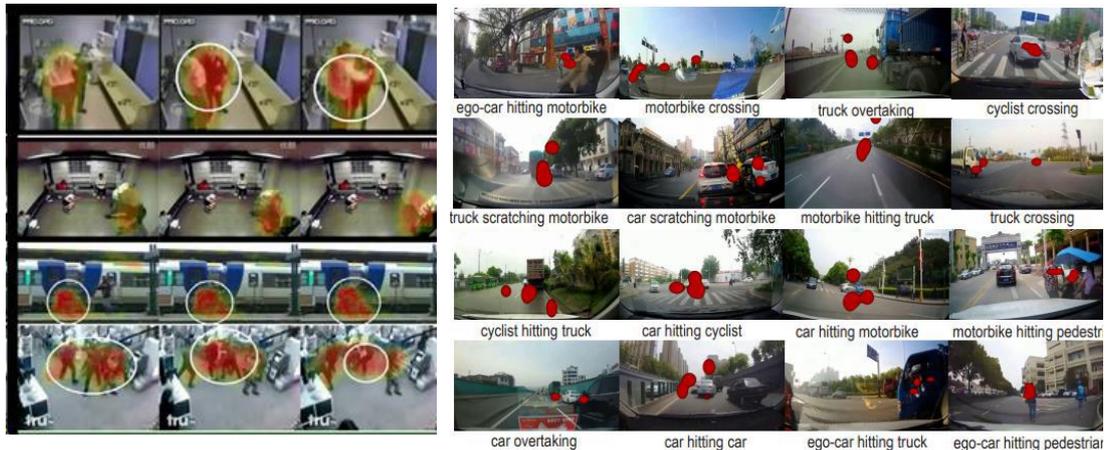
La reconnaissance d'évènements anormaux dans des vidéos par apprentissage profond est une fonctionnalité cruciale pour les applications de vidéo surveillance, de sécurité routière ou encore de conduite autonome. Il existe cependant encore beaucoup de verrous techniques et scientifiques. Les principaux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

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

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

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

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



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

Références

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

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

STAGE 2021

Réf : AS-21-S8

Adaptation de domaine pour la détection et la segmentation : du capteur pinhole au capteur fisheye

Présentation du laboratoire d'accueil

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

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

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

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

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

Description du stage

La détection et la segmentation sont des tâches de vision par ordinateur de mieux en mieux maîtrisées sur des images issues de capteur *pinhole*, c'est-à-dire des images classiques, sans beaucoup de distorsions sur les bords. En effet, pour apprendre des modèles de détection ou de segmentation, il est nécessaire d'utiliser les grandes bases de données publiques d'images annotées, ces datasets sont constitués d'images de la vie courante, issues de capteur *pinhole*.

Cependant, les caméras *fish-eye* sont de plus en plus utilisées, que ce soit pour la vidéo surveillance, le véhicule autonome ou l'industrie. En effet, ce type de caméra possède l'avantage de couvrir un large champ de la scène à elle toute seule. L'inconvénient est que les images issues de capteurs *fish-eye* présentent d'importantes distorsions.

Les modèles de détections et de segmentations d'objets appris sur des images *pinhole* ne sont plus aussi performants sur ce type d'images à moins qu'elles soient préalablement redressées. Or le mécanisme de redressement des images est coûteux en temps de calcul et ne peut s'appliquer sur toute l'image sans perdre une partie du champ de vision.



Exemple de démonstration du laboratoire utilisant des caméras *pinhole*. Nous souhaitons appliquer cette démo sur la vue *fish-eye* de droite.

L'état de l'art sur l'adaptation de domaine des modèles convolutifs est de plus en plus riche. Cependant il concerne essentiellement l'adaptation entre deux datasets dont les contextes sont différents mais dont les images sont issues de capteurs *pinhole* pour les deux. Par exemple, [1] présente une étude sur l'adaptation de domaine pour la détection d'objet, [2] et [3] proposent des méthodes concernant la segmentation d'instances.

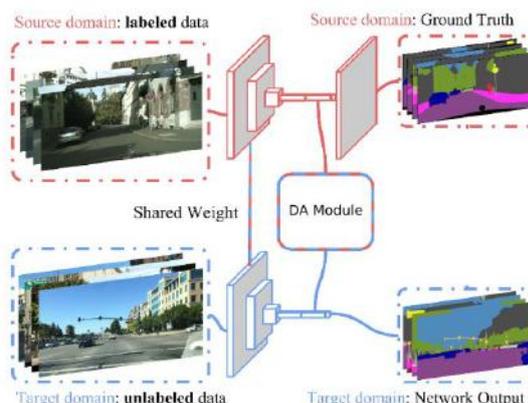


Illustration d'un réseau de neurone pour l'adaptation de domaine en segmentation sémantique (figure modifiée de [2])

L'objectif du stage est donc de mettre en place une méthode d'adaptation de domaine qui permet d'appliquer les modèles appris sur des images *pinhole* sur des images *fisheye*. La première mission du stagiaire sera de prendre connaissance de l'état de l'art en adaptation de domaine de manière à proposer une première méthode pour la détection d'objet dans les images *fisheye*. Dans un second temps, la méthode pourra être étendue à la segmentation. Les résultats obtenus pourront faire l'objet de publications scientifiques. Ce stage ouvre la possibilité de poursuite en thèse sur cette problématique.

- [1] Li, W., Li, F., Luo, Y., & Wang, P. (2020). Deep Domain Adaptive Object Detection: a Survey. arXiv preprint arXiv:2002.06797.
- [2] J. Hoffman, D. Wang, F. Yu, and T. Darrell. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649, 2016
- [3] M. Bassetton, U. Michieli, G. Agresti, and P. Zanuttigh. Unsupervised domain adaptation for semantic segmentation of urban scenes. In CVPRW , 2019.

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

STAGE 2021

Réf : AS-21-S9

Suivi visuel par segmentation d'instances par apprentissage profond

Présentation du laboratoire d'accueil

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

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

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

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

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

Description du stage

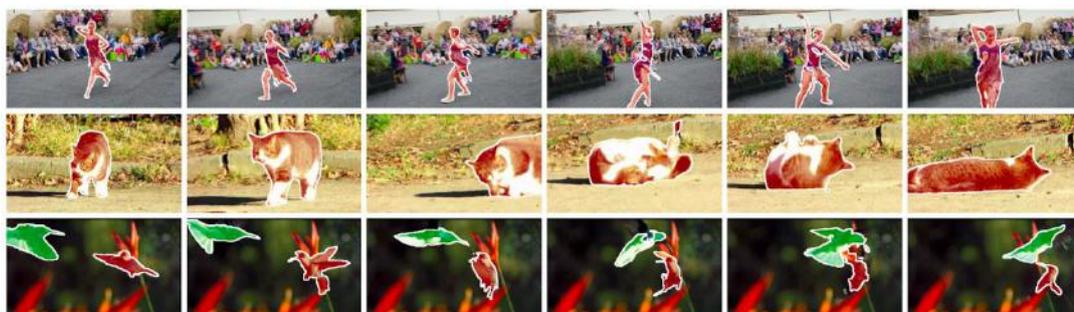


Figure 1. Exemples de résultats de suivi de segmentation d'instance avec la méthode CINM [1] sur les datasets DAVIS 2016 [2], Youtube-Objects [3] et SegTrack v2 [4].

Le suivi d'objets dans une image est une problématique très étudiée dans le domaine de la vision par ordinateur. Cette technique est utilisée dans de nombreux domaines d'application, tels que la vidéosurveillance et la conduite autonome. Nous pouvons distinguer deux différentes familles d'algorithmes de suivi d'objets: le suivi « par détection » qui repose sur une association a posteriori d'instances détectées image par image et le suivi « visuel » qui propage directement la détection d'un objet d'une image à l'autre. Le laboratoire travaillant sur l'analyse de vidéos de sports collectifs, nous avons besoin de suivre des objets particulièrement difficiles à détecter image par image tels des balles ou des joueurs.



Figure 2. Illustration du suivi de balles de volleyball¹

L'objectif de ce stage est donc de développer un algorithme de suivi visuel d'instances robuste aux occultations partielles voire intégrales souvent retrouvées dans les scènes de sport collectif. Une première approche pourra reposer sur un suivi visuel de boîtes englobantes autour d'un objet d'intérêt cible. Le stagiaire aura pour première mission d'étudier les approches de l'état de l'art, et de concevoir une nouvelle méthode dans le cadre de vidéos de sport collectif. Une deuxième phase du stage consistera à étendre l'algorithme conçu au suivi visuel pixellique de la cible (« segmentation d'instance »). Il pourra aussi effectuer une étude comparative avec les méthodes existantes sur des bases publiques [1].

[1] L. Bao, B. Wu and W. Liu. CNN in MRF: Video Object Segmentation via Inference in A CNN-Based Higher-Order Spatio-Temporal MRF (2018)

[2] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In CVPR (2016)

[3] S. D. Jain and K. Grauman. Supervoxel-consistent foreground propagation in video. In ECCV, 2014

[4] F. Li, T. Kim, A. Humayun, D. Tsai, and J. M. Rehg. Video segmentation by tracking many figure-ground segments. In ICCV (2013)

¹ <https://towardsdatascience.com/ball-tracking-in-volleyball-with-opencv-and-tensorflow-3d6e857bd2e7>

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