list
cea tech

Offres de stages 2022

AI - Deep Learning

Computer Vision - Scene Understanding

1. IA frugale pour la ré-identification d'objets par l'adaptation de domaine non supervisée
2. Apprentissage incrémental pour l'analyse de scène
3. Apport de l'attention spatio-temporelle pour la reconnaissance d'action deans une séquence vidéo
4. Reconnaissance visuelle dans les nuages de points 3D avec des modèles Tranformers
5. Adaptation des méthodes de reconnaissance visuelle pour divers points de vue
6. Segmentation interactive auto-supervisée

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

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 Angélique Loesch
Tél +33 (0)1 69 08 89 07
E-mail angelique.loesch@cea.fr

STAGE 2022 Réf : LVA-22-S1

IA frugale pour la ré-identification d'objets par l'adaptation de domaine non supervisée

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

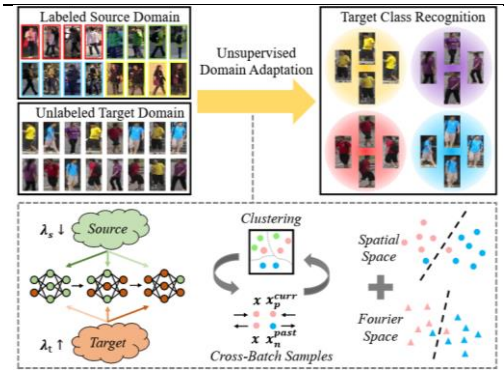
La ré-identification automatique d'objets (personnes ou voitures par exemple) vus par des caméras est une fonctionnalité clé pour les applications de vidéo protection ou encore de surveillance routière. Ce problème en ensemble ouvert consiste à retrouver les occurrences d'un objet dans un ensemble d'images, problème que l'on retrouve également dans des applications de recherche d'images ou d'instances. Malgré les nombreux travaux sur la problématique ces dernières années, la modélisation de l'apparence d'un objet reste un défi. En effet, elle doit pouvoir discriminer des objets distincts (malgré leurs éventuelles similitudes) tout en étant robuste face à la forte variabilité de leur apparence visuelle (due aux postures si on s'intéresse à des personnes, aux points de vue, aux conditions d'illumination, à la sensibilité de la caméra, à sa résolution, …). Si les méthodes d'apprentissage profond supervisé ont fortement amélioré les performances de ré-identification sur certains jeux de données académiques, leur mise en œuvre dans un contexte opérationnel demeure difficile. En effet, un modèle appris sur un jeu de données est très souvent peu performant s'il est appliqué tel quel sur d'autres jeux de données. Par ailleurs, annoter manuellement les données du domaine cible est une tâche fastidieuse donc coûteuse.

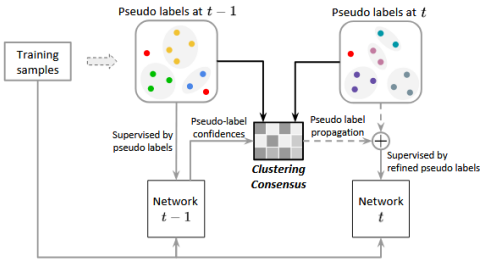
Nous nous intéresserons dans ce stage au problème d'adaptabilité des modèles d'apparence visuelle à un domaine cible dont on ne possède que des données non annotées. Cette adaptation de domaine non supervisée vise à transférer les connaissances acquises dans un domaine source annoté afin de mesurer correctement les affinités entre les instances dans le domaine cible [1,2,3,4]. L'influence de l'ensemble de données source peut varier selon les méthodes proposées, allant même jusqu'à être écartée dans les dernières approches complètement non supervisées [5,6,7].

L'objectif du stage est, dans un premier temps, d'étudier plusieurs méthodes de l'état de l'art. Le candidat devra les évaluer afin d'apprécier leurs avantages et leurs limitations. Dans un deuxième temps, le candidat devra étudier dans quelles mesures les méthodes auto-supervisées viennent supplanter les approches par adaptation de domaine non supervisée. Le candidat sera également invité à proposer des améliorations aux méthodes de l'état de l'art pour pallier un ou plusieurs problèmes identifiés. Les travaux menés durant le stage pourront faire l'objet de publications scientifiques.

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 Angélique Loesch
Tél +33 (0)1 69 08 89 07
E-mail angelique.loesch@cea.fr





Exemples d'une méthode de réidentification d'objets par adaptation de domaine non supervisée à gauche [1] et d'une méthode de réidentification d'objets complètement non supervisée à droite [5].

Keywords

Object re-identification, image retrieval, deep learning, unsupervised domain adaptation, self-supervised learning, frugal AI.

Références

[1] Isobe, T., Li, D., Tian, L., Chen, W., Shan, Y., & Wang, S. (2021). Towards discriminative representation learning for unsupervised person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8526-8536).

[2] Dubourvieux, F., Audigier, R., Loesch, A., Ainouz, S., & Canu, S. (2021, January). Unsupervised domain adaptation for person re-identification through source-guided pseudo-labeling. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 4957-4964). IEEE.

[3] Ge, Y., Chen, D., & Li, H. (2019, September). Mutual Mean-Teaching: Pseudo Label Refinery for Unsupervised Domain Adaptation on Person Re-identification. In *International Conference on Learning Representations*.

[4] Song, L., Wang, C., Zhang, L., Du, B., Zhang, Q., Huang, C., & Wang, X. (2020). Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, *102*, 107173.

[5] Zhang, X., Ge, Y., Qiao, Y., & Li, H. (2021). Refining Pseudo Labels with Clustering Consensus over Generations for Unsupervised Object Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3436-3445).

[6] Ge, Y., Zhu, F., Chen, D., Zhao, R., & Li, H. (2020). Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In Adv. Neural Inform.Process. Syst., 2020

[7] Chen, H., Wang, Y., Lagadec, B., Dantcheva, A., & Bremond, F. (2021). Joint Generative and Contrastive Learning for Unsupervised Person Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2004-2013).

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

Contacts Romaric Audigier
Hejer Ammar

E-mail romaric.audigier@cea.fr
hejer.ammar@cea.fr

STAGE 2022 Ref : LVA-22-S2

Incremental Learning for Scene Analysis

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

Incrementally adapting an existing object detection model to detect new unseen classes with severe memory and computational constraints is a critical capacity in real-word applications such as robotics, self-driving vehicles or video surveillance. However, while human beings can easily recognize new objects continuously without forgetting the old knowledge, deep learning models can suffer from 'catastrophic forgetting'. In fact, adding new classes without using the old training dataset can cause a big degradation of performance on the original set of classes.

To overcome this issue, several methods use a memory buffer to save a set of the old dataset and re-use it to retrain the model with the new classes [1] or extend the model architecture by adding other detection heads. Others focus essentially on regularizing the training to minimize the discrepancy between responses for the old and the updated model [2]. The results of these methods are still limited compared to the models trained jointly with all the dataset. Recent methods identify instances of unknown objects as unknown and subsequently learn to recognize them when training data progressively arrive without retraining from scratch [3].

While various studies are conducted on image classification and object detection, only few methods [4,5] focus on incremental learning for other scene analysis tasks like semantic segmentation. However, semantic segmentation is a key task that computer vision systems must face frequently in various applications.

Objectives of the internship

- Analyze existing incremental learning for object detection and semantic segmentation methods and point their limitations.
- Propose and develop an incremental learning method with severe memory and computational constraints.
- Evaluate the developed method on public datasets (e.g. PASCAL VOC, MsCOCO).
- Publication of results will be encouraged.

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

Contacts Romaric Audigier
Hejer Ammar

E-mail romaric.audigier@cea.fr
hejer.ammar@cea.fr

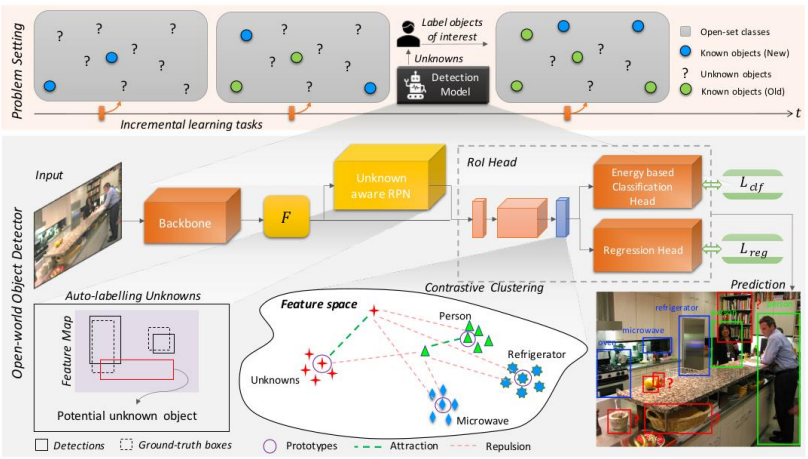


Figure 2: *Approach Overview: Top row:* At each incremental learning step, the model identifies unknown objects (denoted by '?'), which are progressively labelled (as blue circles) and added to the existing knowledge base (green circles). *Bottom row:* Our open world object detection model identifies potential unknown objects using an energy-based classification head and the unknown-aware RPN. Further, we perform contrastive learning in the feature space to learn discriminative clusters and can flexibly add new classes in a continual manner without forgetting the previous classes.

Figure (extracted from [3]): An example of incremental learning approach for object detection

References

[1] Konstantin Shmelkov, Cordelia Schmid, Karteek Alahari (2017) Incremental Learning of Object Detectors without Catastrophic Forgetting, *2017 IEEE International Conference on Computer Vision (ICCV)*.

[2] Shieh, J.-L.; Haq, Q.M.u.; Haq, M.A.; Karam, S.; Chondro, P.; Gao, D.-Q.; Ruan, S.-J (2020) Continual Learning Strategy in One-Stage Object Detection Framework Based on Experience Replay for Autonomous Driving Vehicle, *Sensors* 20, no. 23: 6777.

[3] K. J. Joseph, Salman H. Khan, Fahad Shahbaz Khan, Vineeth N. Balasubramanian (2021) Towards Open World Object Detection, *CVPR.*

[4] Umberto Michieli and Pietro Zanuttigh (2019) Incremental Learning Techniques for Semantic Segmentation, *ICCV.*

[5] Umberto Michieli, Pietro Zanuttigh (2021) Knowledge Distillation for Incremental Learning in Semantic Segmentation, *Computer Vision and Image Understanding (CVIU)*, Vol. 205.

Keywords

Object detection, semantic segmentation, deep learning, incremental learning, knowledge distillation.

Required level:	Engineer, Master 2
This internship opens the possibility of pursuing a thesis and R&D engineer in our laboratory.	
Duration :	6 months
Remuneration:	between 700 € and 1300 € depending on the training.
Required Skills : - Computer vision - Machine learning (deep learning) - Shape recognition - Python, C / C ++ - Mastery of a deep learning framework (in particular Tensorflow or PyTorch)	

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

Contact Jaonary Rabarisoa
Tél +33 (0)1 69 08 01 29
E-mail jaonary.rabarisoa@cea.fr

STAGE 2022 Réf : LVA-22-S3

Apport de l'attention spatio-temporelle pour la reconnaissance d'action deans une séquence vidéo

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

La reconnaissance d'action dans une vidéo est une tâche de vision par ordinateur à la base de plusieurs applications (analyse sportif, vidéo projection, système de recommandation ...). La majorité des approches utilisent l'apprentissage profond à base d'architectures convolutives (*Convolutional Neural Networks*). Ces derniers temps les architectures à base d'attention (*transformers*) ont émergé comme une alternative performante aux CNN pour résoudre les tâches de vision. Pour l'analyse vidéo en particulier, les *transformers* offrent un moyen naturel de gérer le lien spatio-temporel entre les objets présents dans une séquence. Cependant, pour être efficace, ils nécessitent beaucoup plus de données annotées que les CNN pendant la phase d'apprentissage. L'étape de pré-apprentissage (initialisation des poids) devient primordiale pour réussir cette phase.

L'objectif de ce stage est d'étudier l'utilisation des réseaux de neurones à base de *transformer* pour la classification de séquence vidéo. Il s'agira principalement de répondre aux questions suivantes :

- Comment découper une séquence vidéo pour que les blocs d'attention soit les plus pertinents possible?
- Comment pré-apprendre le réseau afin de palier au manque de données annotées?

La méthode proposée sera évaluée sur des données d'analyse de geste sportif, comme le tennis par exemple. Des dataset public tel que Thetis [7] ou d'autres données interne au laboratoire pour mettre en place l'architecture basé transformer puis d'autres dataset seront utilisés pour mettre en évidence la stratégie de pré-apprentissage du réseau dans une situation d'adaptation de domaine.

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 Rabarisoa
Tél +33 (0)1 69 08 01 29
E-mail jaonary.rabarisoa@cea.fr

Exemples du dataset video Thetis.

Exemples de reconnaissance de coup pendant un match de tennis.

Keywords

Action recognition, visual transformers, domain adaptation, sport analysis.

Références

[1] Temporal Contrastive Pretraining for Video Action Recognition. G. Lorre et al. WACV 2020.
[2] Spatiotemporal Contrastive Video Representation Learning. R. Qian et al. https://arxiv.org/abs/2008.03800
[3] An Image is Worth 16x16 Words. Transformers for Image Recognition at Scale. Alexey Dosovitskiy et al. https://arxiv.org/abs/2010.11929
[4] ViViT: A Video Vision Transformer. A. Arnab et al. https://arxiv.org/abs/2103.15691
[5] A Large-Scale Study on Unsupervised Spatiotemporal Representation Learning. C. Feichtenhofer et al. CVPR 2021.
[6] Video Action Transformer Network. R. Girdhar et al. CVPR 2019
[7] http://thetis.image.ece.ntua.gr/

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

Contacts Nicolas Granger
Mohamed Chaouch

E-mail nicolas.granger@cea.fr
mohamed.chaouch@cea.fr

STAGE 2022 Réf : LVA-22-S4

3D Point Cloud Perception with Transformer Models

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 "Laboratoire Vision et Apprentissage pour l'analyse de scenes" 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.

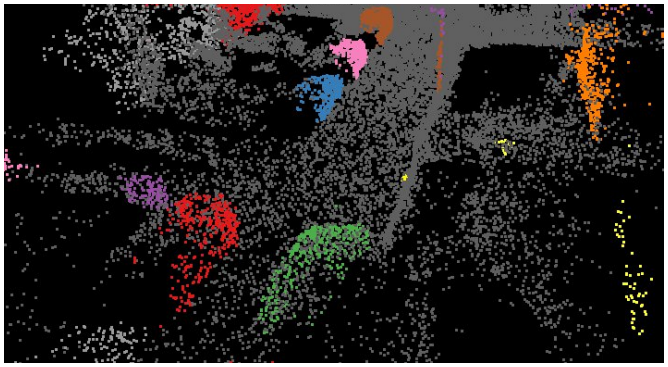


Figure 1 - Sample semantic segmentation from KITTI360 dataset

Context

LiDAR or ToF sensors evolve rapidly and become more ubiquitous, similarly to color cameras a couple of decades ago. Point clouds differ largely from images in their structure as they are largely unstructured in the form of a set of points, have a lower resolution compared to cameras for now, but provide an accurate point coordinates in 3D which helps eliminate many ambiguities related to relative and absolute positioning.

So far, the state of the art on modeling 3D point clouds has mirrored advances made on image models using one of the following strategies: up-lift a 2D model to 3D using for instance 3D convolutions [2], down-cast the 3D space to 2D via depth map projections [4] or bird-eye-views for flat scenes [3]. A third category of models addresses 3D point clouds directly as a set prediction problem [1,6].

In the mean-time Transformer models have successfully applied set-based logic onto several data modalities such as text, images, speech, and more recently point clouds. On the latter category however, we observe that the proposed solutions are still in a preliminary stage and have not yet addressed some of the challenges raised by 3D point cloud modeling. Indeed, 3D point clouds provide a sparse information within large scenes leading to computational challenges. Moreover, 3D point clouds benefit from strong localization properties that have not

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

Contacts Nicolas Granger
Mohamed Chaouch

E-mail nicolas.granger@cea.fr
mohamed.chaouch@cea.fr

yet been transposed into the models.

Objectives of this internship

Starting from the current state of the art models in 3D detection, including Transformer models, the intern will research the integration of recent contributions from the field (ex: [7]) to improve the processing and training efficiency of Transformer models on 3D point clouds. In addition, this internship will seek new contributions specifically focused on the point cloud modality with the following suggested axes of research:

- Integration of occlusion hypotheses
- Stronger utilization of invariance hypotheses in model design
- Spatial coherence in time

References

[1] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017. doi: 10.1109/cvpr.2017.16.
[2] C. Choy, J. Gwak, and S. Savarese, "4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks," presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019. doi: 10.1109/cvpr.2019.00319.
[3] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 12689–12697. doi: 10.1109/CVPR.2019.01298.
[4] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2019, Macau, SAR, China, November 3-8, 2019*, 2019, pp. 4213–4220. doi: 10.1109/IROS40897.2019.8967762.
[5] I. Misra, R. Girdhar, and A. Joulin, "An End-to-End Transformer Model for 3D Object Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 2906–2917.
[6] J. Xu, R. Zhang, J. Dou, Y. Zhu, J. Sun, and S. Pu, "RPVNet: A Deep and Efficient Range-Point-Voxel Fusion Network for LiDAR Point Cloud Segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 16024–16033.
[7] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection," 2021. [Online]. Available: https://openreview.net/forum?id=gZ9hCDWe6ke

Keywords

object detection, semantic segmentation, deep learning, Lidar, 3D, tranformer

Required level:	Engineer, Master 2
This internship opens the possibility of pursuing a thesis and R&D engineer in our laboratory.	
Duration :	6 months
Remuneration:	between 700 € and 1300 € depending on the training.
Required Skills : - Computer vision - Machine learning (deep learning) - Shape recognition - Proficiency in programming (Python) - Mastery of a deep learning framework (in particular PyTorch)	

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

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

STAGE 2022 Réf : LVA-22-S5

Adaptation des méthodes de reconnaissance visuelle pour divers points de vue

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

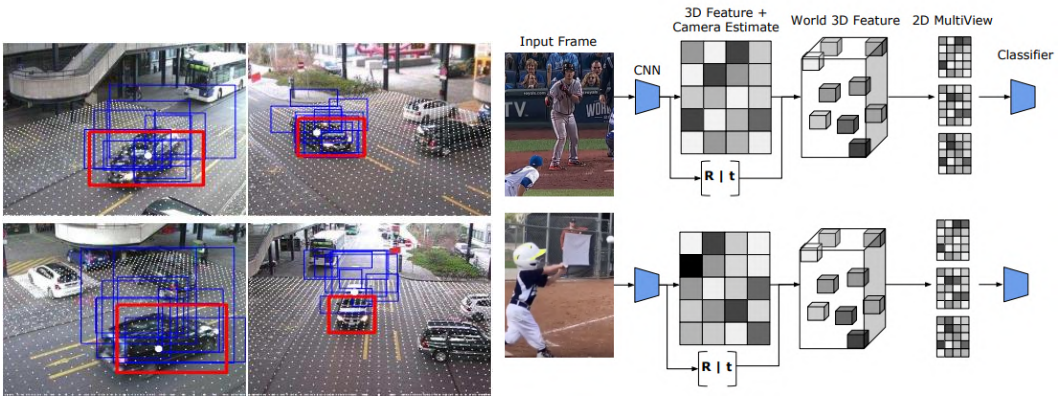
Les méthodes de reconnaissance visuelle (détection d'objets, segmentation sémantique et d'instances, etc.) ont connu un grand essor ces dernières années atteignant de très bonnes performances sur la plupart des datasets publics. Cependant, leurs performances baissent drastiquement quand elles sont testées sur les données différentes de celles d'apprentissage. Les méthodes d'adaptation de domaine ont pour but de réduire la différence entre les domaines, mais leur application est souvent limitée quand la différence entre les domaines est dû aux changements de points de vue. En effet, les transformations géométriques qui apparaissent sur les images quand la caméra change de position représentent un verrou pour les méthodes d'apprentissage automatique actuelles. L'objectif du stage est d'explorer les méthodes de reconnaissance visuelle qui prennent en compte les transformations géométriques provoquées par le changement de point de vue de la caméra afin de rendre les modèles invariants ou robustes à de tels changements.

Nous nous intéressons dans ce stage aux modèles d'auto-calibration [1] capables de prédire, à partir d'une image, les paramètres extrinsèques de la caméra. Ces paramètres peuvent être utilisés dans les apprentissages afin d'obtenir des *features* 3D qui ne dépendent pas du point de vue [2]. Ces *features* serviront dans les tâches finales de reconnaissance visuelle, telles que la détection d'objet, reconnaissance d'action et d'interaction [2,5].

L'objectif du stage est, dans un premier temps, d'étudier plusieurs méthodes de l'état de l'art de l'auto-calibration et adaptation de domaine en utilisant l'information de calibration estimée. Ensuite, le candidat devra évaluer leurs performances sur des datasets multi-vues publics [3,4]. Le candidat sera également invité à proposer des améliorations aux méthodes de l'état de l'art pour pallier un ou plusieurs problèmes identifiés ou adapter les méthodes à d'autres tâches de reconnaissance visuelle [5]. Les travaux menés durant le stage pourront faire l'objet de publications scientifiques.

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



L'image de gauche : exemples d'images du dataset multiview multi-classe pour la détection d'objets [3]. L'image de droite : la représentation des features 3D utilisée pour la reconnaissance d'actions invariante aux points de vues [2].

Keywords

Viewpoint estimation, domain adaptation, camera calibration, camera self-calibration, object detection, action recognition, interaction detection.

Références

[1] G. Iyer, R. K. Ram., J. K. Murthy, K. M. Krishna. CalibNet: Geometrically Supervised Extrinsic Calibration using 3D Spatial Transformer Networks. In *Proccedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[2] AJ Piergiovanni, M. S. Ryoo. Recognizing Actions in Videos from Unseen Viewpoints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[3] G. R. Noguera; X. B. Bosch; H. B. Shitrit; P. Fua, Conditional Random Fields for Multi-Camera Object Detection, In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2011.

[4] G. Vaquette, A. Orcesi, L. Lucat, C. Achard. The DAily Home LIfe Activity Dataset: A High Semantic Activity Dataset for Online Recognition. . In *Proceedings of the IEEE 12th International Conference on Automatic Face & Gesture Recognition*, 2017.

[5] S. Chafik, A. Orcesi, R. Audigier, B. Luvison. Classifying All Interacting Pairs in a Single Shot**.** In *Proceedings of the IEEE Winter Conference on Application of Computer Vision*, 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 : - 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

Contacts Camille Dupont
Jaonary Rabarisoa

E-mail camille.dupont@cea.fr
Jaonary.rabarisoa@cea.fr

STAGE 2022 Réf : LVA-22-S6

Self-Supervised Interactive Segmentation

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

As deep learning gains popularity, the need for large amounts of annotated images has never been greater. Annotating images is a tedious and time-consuming task, especially in the field of image segmentation where human annotators have to draw complex polygons around all sorts of objects. Interactive segmentation can considerably reduce the amount of time needed to annotate a dataset by making the process of annotating images much easier for annotators. In another area, video or image editing software also use interactive segmentation to help artists precisely select objects within images. However, such interactive segmentation algorithms heavily rely on large annotated datasets to train and are therefore highly dependent on the type and quality of those given annotations. To overcome this issue, several methods choose to exploit synthetically created datasets from « png » images in order to obtain perfectly delineated objects [1]. While they already obtain a significant gain in accuracy, such images come in limited numbers. Meanwhile, self-supervised learning techniques have recently proved very successful in the extraction of meaningful feature maps for downstream tasks such as segmentation [2].

Objectives of this internship

Based on these observations, the objectives of the internship are :

– Analyze existing self-supervised « pre-tasks » and choose the most relevant to generate features which discriminate well objects in a scene

– Conceive and develop an interactive segmentation algorithm, which combines such features with sparse human interactions in order to generate fine-grained segmentation masks. To achieve such a goal, the intern may look into classical image segmentation tools and adapt their use of either texture or edge information (e.g. Magic Wand, Intelligent Scissors, Graph-Cut) to « deep » embeddings.

– Evaluate the developed method on standard interactive segmentation benchmarks (SBD, MsCOCO, DAVIS).

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

Contacts Camille Dupont
Jaonary Rabarisoa

E-mail camille.dupont@cea.fr
Jaonary.rabarisoa@cea.fr

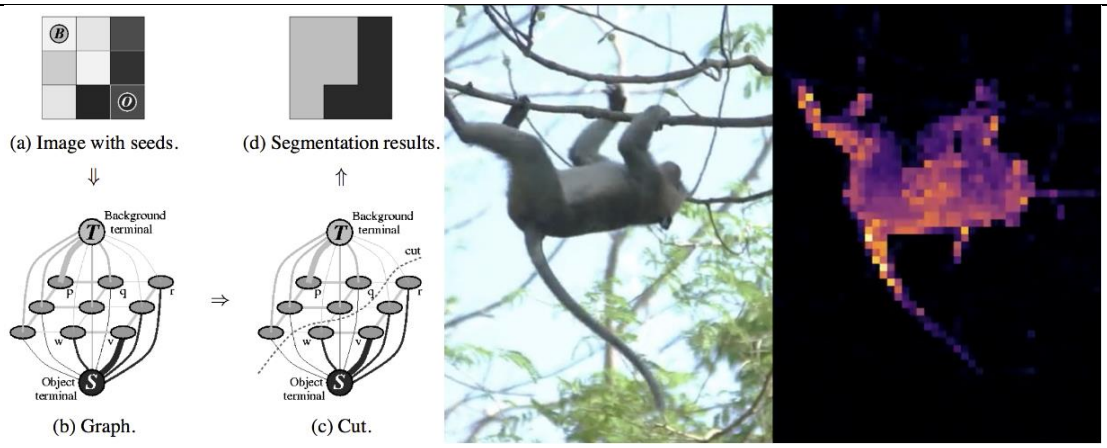


Figure 1 - LEFT Image segmentation as a graph cut [3] : the user provides some prior knowledge about the object and background through the O (Object) and B (Background) and the labelling of the pixels is then solved through a cut on the graph. RIGHT Attention heatmap generated with s Self-supervised Vision Transformer (DINO) [4]

Keywords

instance segmentation, deep learning, self-supervised learning.

References

[1] Marco Forte, Brian Price, Scott Cohen, Ning Xu, François Pitié. Getting to 99% Accuracy in Interactive Segmentation (2020), https://arxiv.org/abs/2003.07932
[2] Xie, Zhenda and Lin, Yutong and Zhang, Zheng and Cao, Yue and Lin, Stephen and Hu, Han. Propagate Yourself: Exploring Pixel-Level Consistency for Unsupervised Visual Representation Learning (CVPR 2021)
[3] Boykov and Jolly, "Interactive Graph Cuts" (ICCV 2001)
[4] Emerging Properties in Self-Supervised Vision Transformers (ICCV 2021)
[5] Duchenne et al. Segmentation by transduction (CVPR 2008).

Required level:	Engineer, Master 2
This internship opens the possibility of pursuing a thesis and R&D engineer in our laboratory.	
Duration :	6 months
Remuneration:	between 700 € and 1300 € depending on the training.
Required Skills : - Computer vision - Machine learning (deep learning) - Shape recognition - Proficiency in programming (Python) - Mastery of a deep learning framework (in particular PyTorch or Tensorflow)	