



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

Contact Steve Bourgeois
Vincent Gay-Bellile
Florian Chabot
Tél +33 (0)1 69 08 02 88
E-mail Steve.bourgeois@cea.fr
Vincent.gay.bellile@cea.fr
Florian.chabot@cea.fr

THESE 2022

Réf : LVML-LVA-22-T5

Reconstruction 3D panoptique d'images 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

La thèse se déroulera au sein du Service Intelligence Artificielle pour le Langage et la Vision, à cheval sur les laboratoires LVA et LVML.

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)

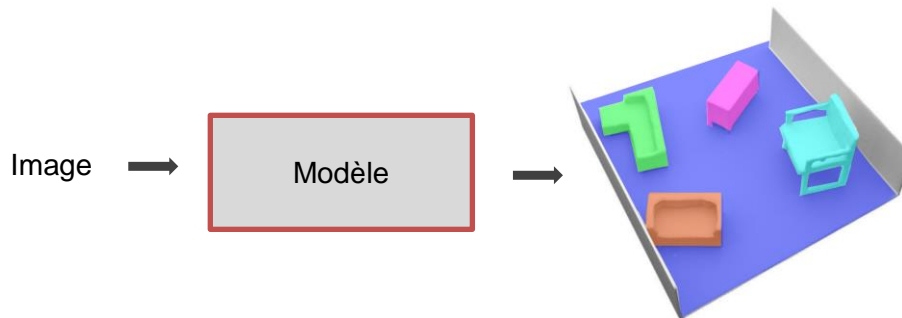
Le Laboratoire de Vision pour la Modélisation et la Localisation (LVML) mène ses recherches dans le domaine de la Vision par Ordinateur (Computer Vision) selon quatre axes principaux :

- La géolocalisation (localisation en environnement inconnu, localisation environnement urbain, SLAM, fusion de capteurs,...)
- La localisation 3D et la caractérisation d'objets (détection et localisation dans un nuage de points 3D, localisation et suivi 3D dans une séquence vidéo, contrôle de défauts, imagerie multi-spectrale)
- La reconstruction 3D (reconstruction 3D dense à partir d'une collection d'images,...)
- Méthodes d'apprentissage et apprentissage profond (méthodes génératives, compression de réseaux de neurones,...)

Objectifs

Les récents progrès de l'analyse sémantique d'images (détection d'objets, segmentation sémantique ou panoptique,...) ont été obtenus grâce aux méthodes d'apprentissage automatique profond, et en particulier grâce à la capacité des réseaux de neurones à extraire d'une image une représentation compacte et facile à interpréter. Ces progrès ont été à l'origine de nombreuses avancées dans des domaines applicatifs tels que le véhicule autonome, la robotique autonome,... Plus récemment, dans le domaine de la reconstruction 3D de scène et de la synthèse d'image, les représentations implicites 3D neurales [1,2,3] ont gagné en popularité en raison de leurs performances et de leur compacité mémoire. Il est donc légitime de se demander si les représentations implicites 3D permettront des avancées notables dans l'interprétation du contenu d'une scène 3D, et c'est à cette question que cette thèse tentera de répondre. Plus précisément, l'objectif de la thèse est de proposer une nouvelle approche, basée sur les modèles 3D implicites, permettant à la fois de reconstruire en 3D la scène observée mais

aussi d'en analyser le contenu.



Étant donné une ou plusieurs images, la méthode développée devra être capable de prédire un maillage 3D précis ainsi que le segmentation panoptique 3D des structures et objets de la scène.

Pistes de recherche

Un premier axe de recherche consistera à étudier l'interprétation a posteriori d'une représentation 3D implicite. Il s'agira ici d'extraire une information sémantique 3D (détection 3D d'objet, segmentation sémantique 3D, etc.) à partir d'une représentation implicite 3D n'ayant pas été entraînée spécifiquement pour cette tâche. Il s'agira ici d'étudier si, en plus d'être compact en mémoire, une représentation implicite 3D peut être efficacement interprétée par un algorithme d'apprentissage sans avoir à repasser par une représentation explicite.

Un second axe de recherche consistera à proposer une approche de reconstruction 3D sémantisée. Contrairement aux méthodes de l'état de l'art [4] consistant à d'abord extraire l'information sémantique en 2D (segmentation panoptique 2D, détection d'objet 2D) puis à fusionner celles-ci dans une reconstruction 3D, cette thèse aura pour objectif de reconstruire simultanément la géométrie de la scène et sa sémantique 3D, permettant au processus de reconstruction de tirer bénéfice de l'information de géométrie 3D pour améliorer l'analyse sémantique, et inversement de tirer bénéfice de l'information sémantique pour améliorer la reconstruction de la géométrie 3D. Idéalement, la solution proposée fournira une reconstruction 3D implicite intégrant une information de segmentation panoptique 3D des structures et objets composant la scène, et cette reconstruction pourra être obtenue de manière incrémentale (ie. Intégrer une nouvelle image dans le processus de reconstruction ne doit pas impliquer de recommencer le processus de reconstruction depuis le début).

Qu'il s'agisse du premier ou second axe de recherche, la problématique de la constitution du jeu de données d'apprentissage sera une question essentielle. Le doctorant devra proposer des solutions pour pallier cette difficulté via, par exemple, l'utilisation de moteurs de simulation de données et/ou d'autres formes d'apprentissage telle que l'auto-supervision.

Références:

[1] Convolutional Occupancy Networks



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

Contact Steve Bourgeois
Vincent Gay-Bellile
Florian Chabot
Tél +33 (0)1 69 08 02 88
E-mail Steve.bourgeois@cea.fr
Vincent.gay.bellile@cea.fr
Florian.chabot@cea.fr

[2] Efficient Geometry-aware 3D Generative Adversarial Networks

[3] Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

[4] Panoptic Neural Field :A Semantic Object-Aware Neural Scene Representation

Niveau demandé :	Ingénieur, Master 2
Durée :	3 ans
Rémunération :	entre 1800 € et 2000 €.
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.	