



STAGE 2020

Réf : LVA-2020-S6

## Ré-identification de personnes et adaptabilité inter-domaines via modèles génératifs

### 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 :

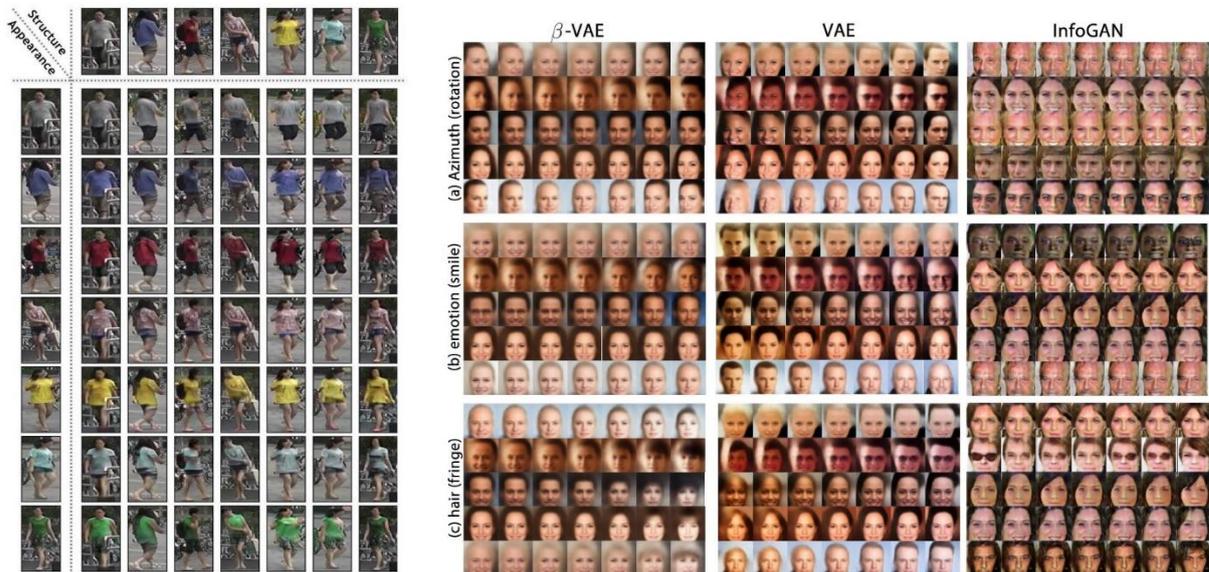
- La reconnaissance visuelle (détection et/ou segmentation d'objets, de personnes, de patterns ; détection d'anomalies ; caractérisation)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, de comportements anormaux ou spécifiques pour des individus, un groupe, une foule)
- 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 ré-identification automatique de personnes vues par des caméras est une fonctionnalité clé pour les applications de vidéo-protection. Elle consiste à retrouver les occurrences d'une personne dans un ensemble d'images. Malgré les nombreux travaux sur la problématique, la modélisation de l'apparence des personnes reste un défi. En effet, elle doit pouvoir discriminer des personnes distinctes malgré leurs éventuelles similitudes, tout en étant robuste face à la forte variabilité de l'apparence visuelle d'une même personne. 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.

Dans ce stage, on s'intéressera au problème d'adaptabilité des modèles d'apparence à un domaine cible dont on ne possède que des données non annotées. Pour appréhender ce problème d'adaptation de domaine, on explorera les modèles génératifs tels que le *Conditional Cycle GAN (Generative Adversarial Network)* [Lu18], l'*InfoGAN* [Chen16] ou le  $\beta$ -VAE (*Variational Auto-Encoder*) [Higgins17]. Ces méthodes de *deep learning* permettent la génération d'images de façon contrôlée par contrainte sur les variables de l'espace latent, afin d'imposer, par exemple, un style sur les images générées ou des attributs sémantiques [Zheng19]. En effet, ces méthodes cherchent à apprendre de manière non supervisée des représentations des données par des concepts indépendants et interprétables.

L'objectif de ce stage sera d'étudier les contraintes sur l'espace latent des modèles génératifs pour des jeux de données d'images de personnes. Tout d'abord, on utilisera les approches de l'état de l'art et on explorera leurs limites dans notre cadre applicatif. Ensuite, on proposera de nouvelles méthodes ou des améliorations qui permettent de transférer de manière non supervisée ou semi-supervisée des modèles de reconnaissance de personnes vers un domaine cible donné. La publication de brevets et d'articles sur les travaux menés sera encouragée.



A gauche : illustration issue de [Zheng19] : Exemples d'images générées par la méthode DG-Net qui sépare les concepts d'apparence et de structure.

A droite : illustration issue de [Higgins17] : Génération d'images par manipulation des variables latentes apprises sur le jeu de données de visages *celebA*.

### Mots-clés :

vision par ordinateur, apprentissage profond, modèles génératifs, adaptation de domaine.

### Références

[Chen16] Chen et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. NIPS 2016.

[Higgins17] Higgins et al.  $\beta$ -VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. ICLR 2017.

[Lu18] Lu et al. Attribute-Guided Face Generation Using Conditional CycleGAN. ECCV 2018.

[Zheng19] Zheng et al. Joint Discriminative and Generative Learning for Person Re-identification. CVPR 2019.

### Profil du candidat/de la candidate

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