



Apprentissage profond joint en vision et langage

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 d'Analyse Sémantique des Textes et des Images (LASTI) est une équipe de 25 personnes (chercheurs, ingénieurs, doctorants) menant des travaux de recherche sur les technologies de description et de compréhension du contenu multimédia (image, texte, parole) et des documents multilingues, en particulier à grande échelle. Les enjeux scientifiques sont :

- développer des algorithmes efficaces et robustes pour l'analyse et l'extraction de contenu multimédia, leur classification et analyse sémantique.
- reconstitution ou fusion de données hétérogènes pour interpréter des scènes ou documents.
- développer des méthodes et des outils pour la construction, la formalisation et l'organisation des ressources et connaissances nécessaires au fonctionnement de ces algorithmes.
- intégrer plusieurs de ces briques technologiques afin d'accéder à l'information et répondre à un besoin utilisateur (moteurs de recherche, chatbot, rapports synthétiques de veille)

Description du stage

Un problème central de l'apprentissage profond est que les méthodes développées nécessitent généralement un grand nombre de données annotées. En pratique, obtenir de telles données et les annoter est très coûteux et constitue donc un verrou majeur pour les industriels souhaitant bénéficier de ces technologies. Il existe plusieurs approches pour pallier ce problème, que le CEA LIST a déjà exploré, notamment le transfert d'apprentissage [1], l'apprentissage à partir de données bruitées collectées automatiquement [2], l'apprentissage incrémental [3] ou encore l'apprentissage sans exemple d'apprentissage [4].

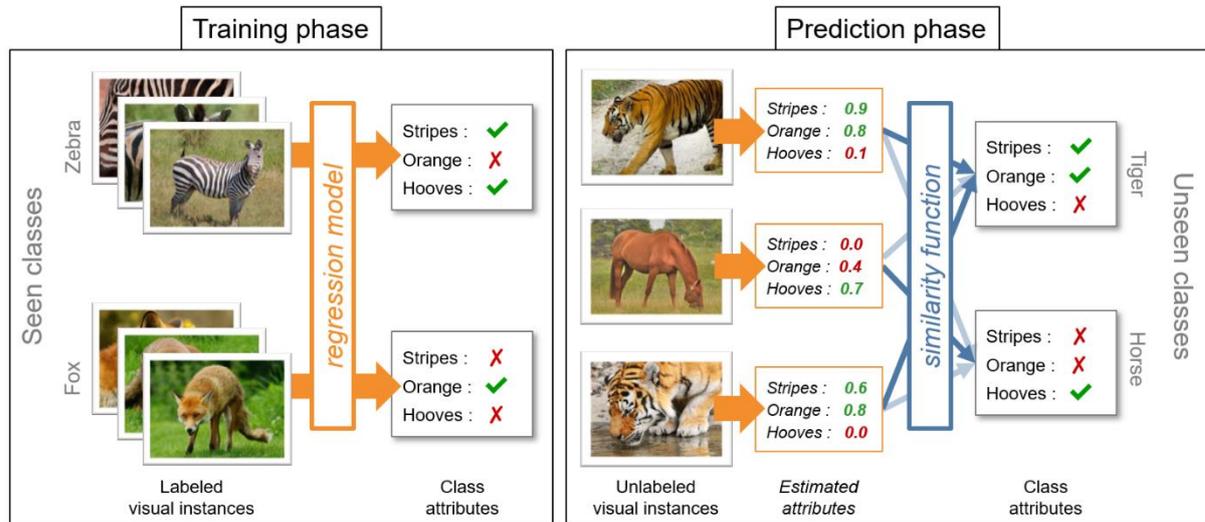
Dans le cadre de ce stage, nous proposons d'explorer l'apprentissage joint de plusieurs tâches [5]. Des travaux ont déjà été menés au CEA LIST dans cette direction, mais il reste de nombreuses pistes à explorer. En particulier, il serait intéressant de diversifier radicalement les domaines d'application, en apprenant conjointement des tâches de vision par ordinateur et de traitement automatique des langues. Fondamentalement, ces travaux permettent de mieux comprendre les interactions entre le codage selon plusieurs modalités (texte, image) d'une même information. Au-delà de l'intérêt scientifique, le potentiel applicatif de ces approches, en termes d'économie de données et d'amélioration des performances est très prometteur.

Concernant le calendrier, la personne recrutée commencera par se familiariser avec le domaine et les outils disponibles. Cela permettra de mettre en place une chaîne de traitement et d'évaluation basique puis d'identifier les points à améliorer. La suite du stage sera constituée de cycles idée/implémentation/évaluation qui seront, le cas échéant, partiellement rédigés de manière à s'insérer aisément dans le rapport final de stage ou une publication scientifique.

Idéalement, à l'issue de ce stage, une publication sera soumise dans une conférence prestigieuse de vision par ordinateur, de traitement automatique des langues ou d'apprentissage (CVPR, ICCV, ACL, EMNLP, ICLR, NeurIPS...).

Mots-clés :

vision par ordinateur, traitement automatique des langues, apprentissage profond, apprentissage multi-tâches.



Le zero-shot learning est une tâche d'apprentissage mêlant représentation visuelle et sémantique (textuelle). Voir [4] en particulier.

Références

- [1] Y. Tamaazousti, H. Le Borgne, C. Hudelot, M.E.A Seddik, M. Tamaazousti, Learning More Universal Representations for Transfer-Learning, IEEE T. Pattern Analysis and Machine Intelligence (PAMI), 2019 (pre-print)
- [2] P. Vo, A. Ginsca, H. Le Borgne and A. Popescu, Harnessing Noisy Web Images for Deep Representation, Computer Vision and Image Understanding, Volume 164, November 2017, Pages 68-81
- [3] E. Belouadah, A. Popescu, DeeSIL: Deep-Shallow Incremental Learning, 5th TASK-CV: Transferring and Adapting Source Knowledge in Computer Vision and 2nd VisDA Challenge (Workshop of the European Conference on computer Vision), 14 sept. 2018, Munich, Germany, 2018
- [4] Y. Le Cacheux, H. Le Borgne and M. Crucianu. Modeling Inter and Intra-Class Relations in the Triplet Loss for Zero-Shot Learning. In Proceedings of the IEEE International Conference on Computer Vision, ICCV, Seoul, Korea, Oct. 27 - Nov. 2, 2019
- [5] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In European Conference on Computer Vision. Springer, 2018.

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, Reconnaissance de formes ; éventuellement Traitement automatique des langues. - Apprentissage automatique (deep learning) : architectures neuronales, optimisation, régularisation,... - Python et environnement Linux - La maîtrise d'un framework d'apprentissage profond (PyTorch ou Tensorflow) est un plus. 	