**« Semi-supervised Deep Learning for image sequences: Applications to autonomous driving and remote sensing »**

**Laboratory:** IMS (Laboratoire de l'Intégration du Matériau au Système), UMR CNRS 5218
351 cours de la Libération, 33400 Talence, France

**Supervision:** Yannick Berthoumieu (supervisor) - Bordeaux INP / IMS
Guillaume Bourmaud (co-supervisor) - Bordeaux INP / IMS
Rémi Giraud (co-supervisor) - Bordeaux INP / IMS

**Funding:** 3-year PhD studentship funded by the french "Ministère de l'Enseignement supérieur, de la recherche et de l'Innovation"

**Context:** The field of computer vision has been transformed, in just a few years, by the advent of techniques known as Deep Learning (DL), especially through convolutional neural networks. This thesis will seek to develop DL techniques in the specific context of image sequences (videos, time series, etc.) where the challenges and fields of application are multiple.

In fact, the sequential nature of the data itself poses problem. On the one hand, the memory footprint of a neural network dedicated to a sequence of images can quickly become disproportionate and generalizing the convolution operation to this type of data is not trivial. On the other hand, the architectures of recurrent neural networks, such as the Long-Short-Term-Memory networks, which were historically dedicated to time series, have proven to be outdated. The question of the neural network architecture for problems involving image sequences is therefore far from being solved, and will be at the heart of this thesis.

From an application point of view, driver assistance and remote sensing systems directly depend on the study of image sequences. In the context of autonomous driving, a lot of video data has been made available in order to learn how to semantically analyze urban scenes. However, the semantic annotations on these sequences sometimes remain inaccurate. This first field of application is therefore particularly suitable for developing semi-supervised or even self-supervised DL techniques. The field of remote sensing also offers large amounts of data whose specificity still makes it difficult to use DL techniques. In fact, unlike the context of autonomous driving, temporal acquisition processes, such as SENTINEL, are periodic and spatially irregular. In addition, few semantically annotated data are available and several sensors (e.g. radar and optical) can be used.

**Qualifications:** Graduated from a Masters or engineering school, specialized in image computing and / or artificial intelligence. Solid programming skills are required (Python, C, C++), and some knowledge in image processing and deep learning frameworks (PyTorch or TensorFlow) are a significant plus. Fluency in scientific English and writing skills are also very important.

**Application:** To apply, send a file with CV, motivation letter, transcripts, or any document likely to strengthen the application (letter of recommendation, etc.) at yannick.berthoumieu@ims-bordeaux.fr

**« Apprentissage profond semi-supervisé pour les séquences d'images : Applications à la conduite autonome et à la télédétection »**

**Laboratoire :** IMS (Laboratoire de l'Intégration du Matériau au Système), UMR CNRS 5218
351 cours de la Libération, 33400 Talence, France

**Supervision :** Yannick Berthoumieu (directeur) - Bordeaux INP / IMS
Guillaume Bourmaud (co-encadrant) - Bordeaux INP / IMS
Rémi Giraud (co-encadrant) - Bordeaux INP / IMS

**Financement :** Bourse ministérielle (MESR)

**Description du sujet :**

Le domaine de la vision par ordinateur a été métamorphosé, en quelques années seulement, par l'avènement des techniques dites d'apprentissage profond ou Deep Learning (DL), notamment à travers les réseaux de neurones à convolution [1]. Cette thèse cherchera à développer et valider des techniques de DL dans le contexte spécifique des séquences d'images (vidéos, séries temporelles, prises de vues multiples, etc) où les défis et domaines d'applications sont multiples.

La prise en compte explicite de la nature séquentielle des données pose en effet problème à plusieurs titres. D'une part, l'empreinte mémoire d'un réseau de neurones dédié à une séquence d'images peut vite devenir démesurée. A cela s'ajoute la difficulté de généraliser l'opération de convolution à ce type de données. D'autre part, les architectures de réseaux de neurones récurrents, comme les réseaux Long-Short-Term-Memory (LSTM) [2], qui étaient historiquement dédiées aux séries temporelles, s'avèrent dépassées [3]. Ces architectures sont en effet progressivement remplacées, en traitement de la parole et du texte, par des réseaux non-récurrents utilisant des couches d'« attention » [4]. La question de l'architecture du réseau de neurones pour des problèmes faisant intervenir des séquences d'images est donc loin d'être résolue, et sera au coeur de cette thèse.

D'un point de vue applicatif, les systèmes d'aide à la conduite et les systèmes de télédétection dépendent directement de l'étude de séquences d'images. En effet, dans le contexte de la conduite autonome, de nombreuses données vidéos ont été rendues disponibles dans le but d'apprendre à analyser sémantiquement les scènes urbaines [5], rendant cette application privilégiée pour concevoir de nouvelles architectures de réseaux de neurones. Néanmoins, les annotations sémantiques sur ces séquences restent parfois imprécises voire incohérentes au fil de la vidéo. Ce premier domaine d'application est donc particulièrement adapté pour développer des techniques de DL semi-supervisé, voire auto-supervisé. Très attrayantes sur le papier, la mise en oeuvre de ces approches donne lieu à divers problèmes (minima locaux, pertinence du critère auto-supervisé vis-à-vis du problème considéré, etc.) pour lesquels des réponses restent à apporter dans le cadre des séquences d'images.

Le domaine de la télédétection offre également des quantités de données importantes dont la spécificité rend encore difficile l'utilisation de techniques de DL. En effet, contrairement au contexte de la conduite autonome, les processus d'acquisition temporels, tels que SENTINEL [6], sont périodiques et spatialement irréguliers. De plus, la quantité de données sémantiquement annotées reste faible. Enfin, plusieurs capteurs, radar et optique (données multi-modales), peuvent être utilisés.

L'objectif de cette thèse est donc le développement et la validation de nouveaux outils méthodologiques qui permettront de répondre aux problématiques posées par la modélisation et l'application de réseaux de neurones profonds à des séquences d'images. Les premiers travaux à mener se focaliseront sur la généralisation des modèles récents de couche d'attention et d'apprentissage auto-supervisé aux séquences d'images.

**Profil recherché :**

Diplômé de Master ou d'école d'ingénieurs, spécialisé en informatique image et/ou intelligence artificielle. Des bases techniques solides en programmation sont demandées (Python, C, C++), et quelques connaissances en traitement d'images et apprentissage profond (TensorFlow, PyTorch) sont un plus non négligeable. La maîtrise de l'anglais scientifique et des qualités rédactionnelles sont également très importantes.

**Candidature :**

Pour candidater, envoyer un dossier avec CV, lettre de motivation, relevés de notes, ainsi que toute pièce susceptible de renforcer la candidature (lettre de recommandation, etc.). Pour l'envoi des pièces demandées, ou pour toute question sur le sujet :
yannick.berthoumieu@ims-bordeaux.fr

**Références :**

[1] A. Krizhevsky, I. Sutskever, G. Hinton *: ImageNet Classification with Deep Convolutional Neural Networks.* Advances in neural information processing systems (NIPS). 2012. https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[2] S. Hochreiter, J. Schmidhuber. *Long short-term memory*. Neural computation. 1997. https://www.bioinf.jku.at/publications/older/2604.pdf

[3] E. Culurciello : *The fall of RNN / LST*M. 2018. https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, J. Jones, L. Gomez, I. Polosukhin. *Attention is all you need*. Advances in neural information processing systems (NIPS). 2017. https://arxiv.org/abs/1706.03762

[5] A. Geiger, P. Lenz, C. Stiller, R. Urtasun. *Vision meets robotics: The KITTI dataset*. The International Journal of Robotics Research. 2013. https://journals.sagepub.com/doi/full/10.1177/0278364913491297

[6] R. Torres, P. Snoeij, et al. *GMES Sentinel-1 mission*. Remote Sensing of Environment. 2012. https://www.sciencedirect.com/science/article/abs/pii/S0034425712000600