

Stage de M2 (ou dernière année d'école d'ingénieurs)
Apprentissage géométrique profond pour la physique des hautes énergies

Au cours de la dernière décennie, le domaine de la science des données et de l'apprentissage automatique a connu une révolution en particulier à travers l'essor des méthodologies d'apprentissage profond. À l'expérience ATLAS [1] du Grand Collisionneur de Hadrons (LHC) au CERN [2], des approches innovantes basées sur l'apprentissage profond géométrique (GDL) [3-4] sont explorées pour faire face à l'afflux de données attendu dans les prochaines années, où le taux de collisions au LHC sera augmenté bien au-delà du niveau initial [5]. Le Laboratoire des 2 Infinis - Toulouse (L2IT) est un jeune laboratoire créé en 2020 à Toulouse pour mener des recherches en physique fondamentale avec de nouvelles approches numériques et théoriques pour l'analyse des données. Au sein de la collaboration ATLAS, l'équipe L2IT joue un rôle majeur dans la recherche et le développement pour construire un algorithme de nouvelle génération basé sur les réseaux neuronaux graphiques (GNN) pour la reconstruction de trajectoire des particules chargées. Ce travail s'effectue en collaboration avec l'équipe du projet Exa.Trkx du Lawrence Berkeley National Laboratory (LBNL) [6] et a donné lieu à des publications dans des conférences internationales majeures [7-9].

Nous proposons une opportunité de stage pour les étudiants poursuivant un master, ou en dernière année d'école d'ingénieurs, pour rejoindre notre équipe de recherche et développement en apprentissage automatique (ML R&D). Ce stage se concentrera sur l'exploration et l'évaluation de nouvelles architectures GNN opérant dans un contexte de graphe multi-échelle et multi-topologique, dans le but d'atteindre une reconstruction de traces de particules hautement efficace.

Résumé des tâches du stage :

- Revue de la littérature sur les réseaux neuronaux graphiques (GNN)
- Implémentation, entraînement et évaluation de modèles GNN avancés opérant dans un contexte de graphe multi-échelle et multi-topologique
- Présentation des résultats lors des réunions de collaboration hebdomadaires
- Intégration du code dans le framework open-source de la collaboration [10]
- Écriture d'un rapport ou d'un article scientifique détaillant les résultats obtenus

Pourquoi nous rejoindre :

Si vous êtes étudiant en informatique, en science des données ou en physique avec une forte spécialisation en apprentissage automatique et que vous cherchez à effectuer votre stage dans un environnement de recherche académique intellectuellement stimulant, cette opportunité est pour vous. Le stage se déroulera dans l'équipe "Calcul, Algorithme et Données" du L2IT, sous la supervision directe d'un ingénieur de recherche en informatique du CNRS expert en calcul scientifique. Vous aurez la chance d'appliquer et d'améliorer considérablement vos compétences les méthodes d'apprentissage profond et les architectures avancées de réseaux neuronaux dans un environnement de R&D dynamique et stimulant. Votre travail s'inscrira dans un effort de recherche crucial pour la recherche fondamentale en physique au sein d'une collaboration internationale majeure. Le L2IT a régulièrement des opportunités pour des thèses de doctorat.

Compétences requises :

- Niveau avancé en Python et ses packages scientifiques/science des données (numpy, pandas, etc.)
- Solide connaissance de l'apprentissage profond, avec au moins une expérience de projet avec PyTorch
- Compétences secondaires : maîtrise d'environnement sous Linux, gestion de version Git
- Autonomie et créativité

Contact: Sylvain Caillou, sylvain.caillou@l2it.in2p3.fr

Site web du L2IT: <https://www.l2it.in2p3.fr>

- [1] ATLAS experiment, url: <https://atlas.cern>
- [2] CERN, url: <https://home.cern/>
- [3] G. DeZoort, P.W. Battaglia, C. Biscarat, J.R. Vlimant, Nature Rev. Phys. 5, 281 (2023)
- [4] J. Shlomi, P. Battaglia, J.R. Vlimant, Machine Learning: Science and Technology 2, (2020)
- [5] Apollinari et al., High-Luminosity Large Hadron Collider (HL-LHC), url: <https://cds.cern.ch/record/2284929>
- [6] HEP advanced tracking algorithms at the exascale (Project Exa.TrkX), url: <https://exatrnx.github.io/>
- [7] C. Rougier et al., ATLAS ITk Track Reconstruction with a GNN-based Pipeline, CTD2022, Princeton (USA), url: <https://doi.org/10.5281/zenodo.8187248>
- [8] X. Ju et al., Physics Performance of the ATLAS GNN4ITk Track Reconstruction Chain, CHEP2023, Norfolk (USA), url: <https://indico.jlab.org/event/459/contributions/11414/>
- [9] S. Caillou et al., Novel fully-heterogeneous GNN designs for track reconstruction at the HL-LHC, CHEP2023, Norfolk (USA), url: <https://indico.jlab.org/event/459/contributions/11713/>
- [10] GNN4ITk Team, A Charged-particle geOmetric Reconstruction Network (ACORN), 2023, url: <https://github.com/GNN4ITkTeam/CommonFramework>