

SUJET DE STAGE : Création d'une infrastructure logicielle de test pour la segmentation cellulaire par IA de données cérébrales acquises par microscopie optique

Responsables : Thierry Delzescaux –PhD, Directeur de recherche CEA, HDR- (thierry.delzescaux@cea.fr), Nicolas Souedet, responsable logiciel (nicolas.souedet@cea.fr)

Equipe d'accueil :

CEA-MIRCent, LMN
Laboratoire des Maladies Neurodégénératives
Equipe traitement de l'image
18, route du Panorama - BP N° 6
92265 - Fontenay aux Roses Cedex

<https://jacob.cea.fr/drf/ifrancoisjacob/Pages/Departements/MIRCent/themes/alzheimer-vieillessement-cerebral-modelisation.aspx>

Durée de stage : 6 mois

Description du sujet de stage : L'objectif de ce stage ingénieur / Master 2 sera de spécifier et de mettre en œuvre une infrastructure logicielle générique permettant de recenser et de tester les méthodes de référence de **segmentation** cellulaire par **deep learning** disponibles dans la communauté. Les méthodes identifiées devront permettre de segmenter les différents **types cellulaires** d'intérêt en neurosciences (neurones, astrocytes, cellules gliales ...) à partir de données de **tissus cérébraux 2D acquises par microscopie optique** en fond clair ou en fluorescence. L'étude automatisée et à grande échelle des populations de cellules et de leurs propriétés géométriques (taille, forme, distribution) fournit des biomarqueurs d'un grand intérêt pour caractériser avec précision les modèles de **maladies neurodégénératives** et également pour mieux évaluer et valider de nouvelles approches thérapeutiques.

Une étude approfondie des méthodes de l'**état de l'art** de segmentation sémantique et par instance des cellules sera réalisée dans un premier temps en se concentrant sur les dernières techniques de *Deep Learning* (*Cellpose* [1], architectures *U-Net* [2], Ma [3], etc.).

Par la suite, l'organisation d'une **base de données** 2D de référence sera spécifiée (ontologie des données). Il en découlera la création d'une base de référence (images de microscopie, segmentations de référence) constituée de données issues des projets de l'équipe et de données disponibles identifiées lors de la phase de recherche bibliographique (challenges internationaux : Data Bowl Science, Monuseg, Cellpose, NeurIPS, etc.).

En parallèle, une **organisation logicielle** sera décrite afin de pérenniser les logiciels identifiés et de les partager au sein de l'équipe et du laboratoire. Les solutions envisagées pourraient se baser sur un gestionnaire de paquets de type [mamba](#), [pixi](#), etc. et/ou des logiciels de containerisation ([apptainer](#)). Les logiciels seront déployés selon l'infrastructure décrite et utilisés pour évaluer de façon rigoureuse et quantitative leur capacité à réaliser les tâches : 1) de segmentation sémantique (segmentation des pixels en un nombre limité de classes –neurones, astrocytes, fond, etc.-) et 2) de segmentation d'instances (individualisation des composantes cellulaires des tissus étudiés).

Le.a candidat.e aura accès à des ressources computationnelles (station de travail, moyens de calcul intensif CPU/GPU voire cluster de calcul) ainsi qu'à un encadrement expert adaptés au projet.

Le stage aboutira à la mise en place et à l'évaluation quantitative des principales méthodes identifiées au sein de l'infrastructure créée. Cette infrastructure logicielle aura pour vocation à être pérennisée au sein du laboratoire et à être étendue à d'autres technique d'IA ce qui permettra d'avoir une plateforme opérationnelle d'évaluation des méthodes de références ainsi que les toutes dernières de l'état de l'art qui seront intégrées au fur et à mesure. L'adjonction de bases de données ouvrira des perspectives pour faire des comparaisons de méthodes de façon rapide et efficace.

Compétences requises : Au cours de ce stage, le.a candidat.e sera amené.e à interagir principalement avec l'équipe de traitement de l'image de MIRCent (informaticiens, méthodologistes en traitement de l'image, etc.) et les neurobiologistes des plateformes d'histologie et de microscopie. Il est demandé d'avoir une très bonne connaissance des environnements **Linux** et **Windows**, de maîtriser la programmation dans les **langages C, C++** et **Python**. Des connaissances générales des techniques de traitement de l'image (segmentation, *machine learning* – *deep learning*) et d'un outil de **gestion de sources** ([github](#)) serait un plus. Il est également nécessaire de savoir utiliser les suites bureautiques standards (Open Office, Office). Le.a stagiaire bénéficiera pour réaliser ces développements des savoirs faire, de l'encadrement des équipes du CEA et de l'infrastructure informatique existante (serveurs de calculs internes, plateforme logicielle). De bonnes capacités d'**adaptation** à des environnements multidisciplinaires, de **coordination** et de **travail en équipe** sont également requises pour ce projet.

Formation requise : Ecole ingénieur, Master de recherche dans les sciences de l'information, traitement de l'image et du signal (salaire calculé en fonction de la formation du/de la candidat.e).

Possibilité de poursuivre en CDD dans la continuité du sujet de stage ou sur des thématiques d'intelligence artificielle / de calcul intensif voire en thèse en fonction des résultats du stage.

Bibliographie : [1] Stringer C. *et al.*, Nat Methods, 2021. [2] Falk T. *et al.*, Nat Methods. 2019.[3] Ma *et al.*, Nat Methods, 2024