

**PROJET DE THESE :** Cartographie haute résolution des réseaux de neurones par microscopie optique en recherche préclinique  
High resolution mapping of neural networks by optical microscopy in preclinical research

**MOTS-CLES :** traitement de l'image, modélisation 3D, segmentation, machine learning

**DIRECTEUR DE THESE :** Thierry DELZESCAUX ([thierry.delzescaux@cea.fr](mailto:thierry.delzescaux@cea.fr))

**ÉQUIPE D'ACCUEIL :** CEA-MIRCent, LMN  
(Laboratoire des Maladies Neurodégénératives),  
18, route du Panorama - BP N° 6  
92265 - Fontenay aux Roses Cedex

<http://i2bm.cea.fr/drf/i2bm/Pages/mircen.aspx>

**FINANCEMENT :** Programme Paris Region PhD (PhD2), **Île-de-France**

**PARTENAIRE SOCIO-ECONOMIQUE :** société NEOXIA (co-encadrant Cédric CLOUCHOUX)

**DUREE DE LA THESE :** 36 mois (date de début 10/2019)

**UNIVERSITE / ECOLE DOCTORALE :** Université Paris Saclay, ED EOBE (ED 575)

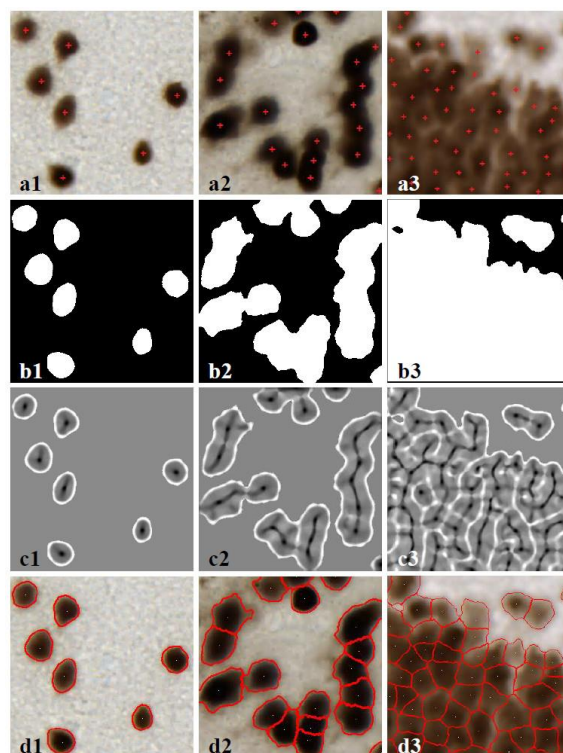
**RESUME :** L'objectif de ce travail de thèse sera de développer des outils logiciels originaux pour étudier des populations de neurones à l'échelle de cerveaux entiers en recherche préclinique. Les images seront acquises par microscopie optique à très fort grossissement pour permettre une individualisation à l'échelle des neurones. Le dénombrement ainsi qu'une modélisation de la cytoarchitecture de ces cellules fourniront des informations nouvelles d'un grand intérêt dans des cas sains (anatomie, développement) mais aussi pathologiques (maladies neurodégénératives, mort neuronale). Des techniques avancées de traitement de l'image seront utilisées pour mener ces analyses (apprentissage machine, calcul intensif). Enfin, une validation des résultats obtenus sur des données multi-espèces produites au CEA-MIRCent sera réalisée en comparant les résultats obtenus avec la technique de référence de stéréologie (comptage cellulaire) ainsi qu'avec des segmentations anatomiques de référence (expert neuroanatomiste, atlas numériques).

**SUMMARY:** The objective of this thesis work will be to develop original software tools to study whole brain neuron populations in preclinical research. The images will be acquired by optical microscopy at very high magnification to allow individualization at the scale of neurons. The counting and modelling of the cytoarchitecture of these cells will provide new information of great interest in healthy cases (anatomy, development) but also pathological cases (neurodegenerative diseases, neuronal death). Advanced image processing techniques will be used to conduct these analyses (machine learning, high performance computing). Finally, validation of the results obtained on multi-species data produced at CEA-MIRCent will be carried out by comparing the results obtained with the stereological reference technique (cell counting) and with anatomical segmentations (neuroanatomist expert, digital atlases).

### INFORMATION DETAILLÉE

- **Contexte :** Les neurones constituent une composante primordiale du cerveau. A ce titre de nombreuses recherches ont été menées en neurosciences pour dresser des cartographies de ces populations de cellules, mieux comprendre leur développement, leurs caractéristiques ainsi que leur organisation. Ces études ont été menées dans des cas sains mais également pathologiques comme les maladies neurodégénératives (Alzheimer, Parkinson, Huntington, etc.). La plupart des études se sont limitées à étudier quelques régions d'intérêt du cerveau du fait de la dimension très importante des données acquises à l'échelle cellulaire par microscopie optique (une seule coupe histologique peu atteindre plus de 150 giga-octets pour un cerveau de macaque) [Walløe *et al.*, Front. Hum. Neurosci., 2014]. De plus, il est particulièrement difficile de segmenter les neurones (intensité de marquage variable) et d'individualiser ceux qui sont accolés en 2D dans des zones denses [You *et al.*, ICIP 2016] (Figure 1). Malgré ces limitations fortes, il existe aujourd'hui une attente très forte de la communauté des neurosciences pour avoir accès à des solutions algorithmiques efficaces qui permettraient de traiter au niveau cellulaire de très grands volumes de données histologiques, idéalement à l'échelle de cerveaux entiers ce qui n'a jamais été réalisé jusqu'à présent. La technique de référence utilisée aujourd'hui par les biologistes pour effectuer des comptages cellulaires est la stéréologie [West *et al.*, Anat. Rec., 1991]. Cette technique est longue et fastidieuse à réaliser ce qui limite son champ d'application à grande échelle et elle ne fournit qu'une information globale du nombre d'objets au niveau de régions anatomiques (pas d'information locale accessible). Dans ce contexte, le développement de techniques automatiques d'analyse d'image à haut débit ouvrirait de nombreux champs d'applications en neurosciences.

**Figure 1:** Neuron segmentation results obtained in images presenting different levels of fraction of neuron surface occupation. a1-a3) Original images. Red crosses are neuron centroids marked by expert. a1) a few individual neurons, a2) several touching neurons, a3) lots of touching neurons. b1-b3) Classification results obtained by random forest. c1-c3) Two iterations min-max cartographies i.e. d1-d3) Individualization results obtained by our method (from You et al., ICIP conference, 2016).



- **Descriptif :** L'objectif de ce travail de thèse sera de fournir un ensemble logiciel de traitement de l'image intégré à BrainVISA (<http://brainvisa.info>) pour étudier les neurones par microscopie optique à grande échelle dans des modèles expérimentaux (rongeur, primate) dans des cas sains et pathologiques (modèles de neurodégénérescence, modèles de lésion excitotoxique, etc.). La création de cartographies à l'échelle cellulaire de la distribution des neurones dans des coupes entières de cerveau est impossible à réaliser aujourd'hui du fait de la très grande quantité de données numériques à gérer et du manque d'outils logiciels adaptés. L'accès à ce type d'information serait d'un grand intérêt dans le domaine des neurosciences (développement, anatomie, vieillissement). De même, la quantification de la mort neuronale est primordiale dans le cadre de l'étude des maladies neurodégénératives pour suivre l'évolution de la pathologie voire pour évaluer l'effet de thérapies (Alzheimer, Parkinson et Huntington).

Les développements méthodologiques à réaliser au sein de notre laboratoire concerneront la segmentation couleur par apprentissage machine (*machine learning / deep learning*) et l'individualisation des neurones. Des spécifications importantes devront être mises en place pour optimiser l'étape de segmentation en adaptant le choix et le nombre des caractéristiques (texture, colorimétrie) ainsi que les techniques d'apprentissage machine à utiliser. Des résultats préliminaires prometteurs ont été obtenus à l'échelle de quelques régions anatomiques d'intérêt chez le macaque mais l'extension à l'échelle de cerveaux entiers et la mise en production dans le cadre plus large d'études précliniques sont des défis qui n'ont pas été relevés à ce jour [Amunts et al., Science, 2013]. Un des verrous à lever dans ce travail sera d'étendre les fonctionnalités d'analyse à des données très massives pouvant aller de quelques centaines de giga-octets à plusieurs dizaines de téraoctets. Ceci impliquera de trouver des stratégies ad-hoc pour adapter les codes en tenant compte des contraintes matérielles (utilisation du cluster de calcul CPU du Très Grand Centre de Calcul du CEA –TGCC, Bruyères-le-Châtel). Ce passage à l'échelle nécessitera de gérer de nouveaux problèmes tels que la présence de cellules déformées ou altérées (énuclées), les variations d'intensité de marquage ou les biais d'illumination intra- et inter-coupes qui constituent des problèmes majeurs pour l'analyse de données histologiques [Roy et al., Micron, 2018]. Les méthodes qui ont été proposées dans la littérature ont souvent été spécifiées pour des types de données très particulières (application biologique, marquage histologique et région anatomique) et ne peuvent prendre en compte la grande variabilité des conformations cellulaires présentes à l'échelle de cerveaux entiers. L'étude prévue permettra de s'intéresser à ces problèmes connus mais également de dresser un inventaire plus exhaustif de conformations problématiques. Nos développements seront prioritairement réalisés sur des coupes histologiques 2D pour bénéficier des forts grossissements disponibles en microscopie optique (x20, x40 ; échelle cellulaire) et la possible extension à la 3D sera réalisée par l'acquisition de séries de coupes qui seront reconstruites en 3D a posteriori. L'équipe traitement de l'image de MIRCen justifie dans ce domaine d'une expertise reconnue au niveau international. Une fois les neurones individualisés à grande échelle, un maillage de ce réseau pourra être construit et sera le support d'une modélisation anatomique/géométrique des tissus. Des atlas numériques de cerveaux produits par IRM haut champ seront utilisés pour superviser l'exploitation des données 3D cellulaires produites afin d'identifier les caractéristiques signatures des principales régions anatomiques du cerveau (Figure 2, A1). La possibilité d'enrichir de façon rétroactive ces atlas numériques grâce aux données histologiques sera envisagée en tirant bénéfice de l'information cellulaire rendue accessible (définition de sous-structures basée sur la cytoarchitecture) (Figure 2, A2-A3). Des ressources informatiques importantes de stockage et de calcul seront allouées à cette recherche à travers un accès aux supercalculateurs du Très Grand Centre de Calcul du CEA à Bruyères-le-Châtel.



**Figure 2:** A1) MRI / digital brain atlas, A2) NeuN histological brain section and A3) expected segmentation.

La validation de ce travail sera réalisée en collaboration étroite avec les neurobiologistes de la plateforme d'histologie de MIRcen pour mener la première campagne exploratoire de cartographies de populations de neurones produites chez le macaque. Il sera ainsi possible d'étudier la cytoarchitecture du cerveau (atlas anatomique, organisation des couches corticales, etc.) mais également la mort neuronale dans des modèles primates de maladies neurodégénératives (Huntington, Parkinson). La validation des résultats obtenus sera réalisée par comparaison des résultats de référence obtenus par stéréologie. De plus, une validation multi-espèces des développements réalisés au cours de cette thèse sera menée chez le rongeur sur un modèle murin de neurodégénérescence [Brureau *et al.*, *Neurobiol Dis.*, 2017]. Il sera ainsi possible de tester le niveau de généralité des méthodes développées. Cette étude pilote dans le domaine permettra d'évaluer le niveau de performance de comptage neuronal par rapport à la stéréologie et d'établir le cadre d'utilisation de ces nouvelles approches (applications biologiques éligibles, perspectives de développement).

- **Compétences requises :** Au cours de ce travail de thèse, le candidat sera amené à interagir principalement avec l'équipe de traitement de l'image de MIRcen (informaticiens, méthodologistes en traitement de l'image, etc.) et les neurobiologistes de la plateforme d'histologie. 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**, d'avoir des connaissances dans le domaine de la parallélisation et du calcul haute performance (CPU). Des connaissances générales des techniques de traitement de l'image (segmentation, *machine learning – deep learning*) et d'un outil de **gestion de sources (subversion/github)** serait un plus. Il est également nécessaire de savoir utiliser les suites bureautiques standards (Open Office, Office). Le doctorant 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.

- **Bibliographie :**

- [1] Amunts *et al.*, "BigBrain: An Ultrahigh-Resolution 3D Human Brain Model," **Science**, 2013.
- [2] Brureau *et al.*, "NF-L in cerebrospinal fluid and serum is a biomarker of neuronal damage in an inducible mouse model of neurodegeneration", **Neurobiol Dis.**, 2017
- [3] Roy *et al.*, "A study about color normalization methods for histopathology images", **Micron**, 2018.
- [4] Walløe *et al.*, "Stereological estimation of total cell numbers in the human cerebral and cerebellar cortex," **Front. Hum. Neurosci.**, 2014.
- [5] West *et al.*, "Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator," **Anat. Rec.**, 1991.
- [6] You Z. *et al.*, "Automated cell individualization and counting in cerebral microscopic images", **conf. ICIP**, 2016.