

STAGE de MASTER 2 « DATA SCIENTIST »

SUJET: *Apprentissage Profond* pour le suivi systématique de cellules d'embryons vivants imagés par microscopie.

L'apprentissage automatique et plus particulièrement l'apprentissage profond (*Deep Learning*) donnent des résultats très prometteurs dans divers domaines de la recherche en biologie. Le but du stage est d'appliquer l'apprentissage profond au domaine de l'imagerie biologique, et plus particulièrement à la segmentation cellulaire 4D (3D plus temps), la reconnaissance et le suivi de chaque cellule dans des embryons vivants en développement. L'apprentissage se fera sur une base de données préexistante de plus de 10 embryons entièrement segmentés (> 10000 cellules suivies). L'objectif est de créer un suivi des cellules basé uniquement sur l'apprentissage en profondeur. Durant ce stage, en fonction du profil et des intérêts du stagiaire, nous pourrons :

- Implémenter et adapter à notre base de données les méthodes d'apprentissage en profondeur basées sur l'identification d'objets individuels (*instance segmentation*).
- Explorer les aspects théoriques du concept de mémoire dans les réseaux neuronaux récurrents (LSTM) et les adapter au suivi cellulaire dans des images 3D + temps.
- Utiliser l'apprentissage en profondeur pour prédire et définir les règles de la division cellulaire au cours de l'embryogenèse.

Ce stage de 5 à 6 mois peut déboucher sur un projet de thèse, combinant potentiellement des approches informatiques avec la validation expérimentale des prédictions.

PROFIL ET COMPÉTENCES DE RECHERCHE :

Le stage est principalement destiné aux étudiants ayant une formation initiale en mathématiques et en informatique et un intérêt pour la biologie, mais aussi à des biologistes ayant de très bonnes compétences en programmation informatique et en analyse statistique. Esprit d'équipe, autonomie, dynamisme et créativité seraient appréciés. La maîtrise de l'anglais ou du français technique doit être suffisante pour rédiger la documentation technique et interagir verbalement quotidiennement.

Langage de programmation requis : *python*

L'étudiant sélectionné aura des connaissances de base sur au moins l'un des sujets suivants : la théorie de l'apprentissage automatique, le traitement d'image, le calcul haute performance ou l'utilisation de bibliothèques d'apprentissage en profondeur (par exemple Keras).

LE CONTEXTE :

Le stage sera basé à Montpellier et l'encadrement sera assuré par un informaticien Emmanuel Faure (CNRS) et un biologiste Patrick Lemaire (CNRS).
Email: emmanuel.faure@lirmm.fr et patrick.lemaire@crbm.cnrs.fr

Emmanuel Faure, chercheur CNRS, (LIRMM, Montpellier <https://www.lirmm.fr/icar/>), est un *Data Scientist* utilisant les techniques d'apprentissage automatique récentes pour comprendre la biologie quantitative et la médecine.

L'équipe de Patrick Lemaire, chercheur CNRS, (CRBM, Montpellier, France, <http://www.crbm.cnrs.fr/en/recherche/equipes/>) étudie le développement embryonnaire animal, en utilisant les ascidies comme système modèle.

REFERENCES:

- Contact-dependent cell communications drive morphological invariance during ascidian embryogenesis. Leo Guignard, Ulla-Maj Fiuza, et al.
bioRxiv 238741; doi: <https://doi.org/10.1101/238741>
- Convolutional Neural Network-Based Instance Segmentation Algorithm to Acquire Quantitative Criteria of Early Mouse Development. Yuta Tokuoka, Takahiro G Yamada, Noriko Hiroi, Tetsuya, J Kobayashi, Kazuo Yamagata, Akira Funahashi
bioRxiv 324186; doi: <https://doi.org/10.1101/324186>

DATA SCIENTIST MASTER 2 INTERNSHIP

SUBJECT: Deep Learning for Cell Tracking on 3D + time live embryos.

Machine Learning and more specifically Deep Learning give very promising results in various domains of biological research. The aim of the internship is to apply Deep learning to the field of biological imaging, and more specifically to 4D (3D plus time) cell segmentation, the recognition and tracking of each cell and its progeny in live developing embryos. The internship will start from a preexisting database of more than 10 fully-segmented embryos (>10000 tracked cells). We would like to create an end-to-end method for complex 3D cell tracking using deep learning. In this internship, depending on the profile and interests of the intern, we can:

- Implement and adapt to our dataset existing end-to-end deep learning methods for instance segmentation (the identification of individual objects).
- Explore theoretical aspects of the adaptation of the memory concept in Recurrent Neural Networks (LSTM) to cell tracking in 3D+time images.
- Use deep learning to predict and define the rules of cell division during embryogenesis.

This 5-6 months internship can lead to a PhD project, potentially combining computer approaches with the experimental validation of predictions.

PROFILE AND RESEARCH SKILLS:

The internship is mainly intended for students with an initial training in maths and computer science, though biologists with very good skills in computer programming and statistical analysis could be considered. An interest in biology is welcome. Team spirit, autonomy, dynamism, creativity would be appreciated. Proficiency in technical English or French must be sufficient to write technical documentation, and to interact verbally daily. Required programming language: python
The selected student should have basic knowledge of at least one of the following topics: machine learning theory, image processing, high performance computing or the use of deep learning libraries (e.g. Keras).

CONTEXT:

The intern will be based in Montpellier, and the supervision will be provided by a computer scientist Emmanuel Faure (CNRS) and a biologist Patrick Lemaire (CNRS)
Mails : emmanuel.faure@lirmm.fr and patrick.lemaire@crbm.cnrs.fr

Emmanuel Faure, (LIRMM Montpellier <https://www.lirmm.fr/icar/>) is a Data Scientist using recent machine learning techniques to understand quantitative biology and medicine.

Patrick Lemaire's team (CRBM, Montpellier, France; <http://www.crbm.cnrs.fr/en/recherche/equipes/>) focuses on animal embryonic development, using ascidians as a model system.

REFERENCES:

- Contact-dependent cell communications drive morphological invariance during ascidian embryogenesis. Leo Guignard, Ulla-Maj Fiuza, et al.
bioRxiv 238741; doi: <https://doi.org/10.1101/238741>
- Convolutional Neural Network-Based Instance Segmentation Algorithm to Acquire Quantitative Criteria of Early Mouse Development. Yuta Tokuoka, Takahiro G Yamada, Noriko Hiroi, Tetsuya, J Kobayashi, Kazuo Yamagata, Akira Funahashi
bioRxiv 324186; doi: <https://doi.org/10.1101/324186>