

STAGE Master 2 - Data Scientist

L'objectif du stage est d'adapter et d'optimiser des réseaux de neurones convolutifs de segmentation sur une très grosse base de données d'images 3D et 4D.

Contexte : Cette offre de stage s'inscrit dans le contexte général de la compréhension du développement embryonnaire de *Phallusia Mammillatta*, espèce de mammifères marins de la famille des ascidies.

Nous sommes récemment parvenu à obtenir des images 4D (3D plus la dimension temporelle) des embryons de *Phallusia mammillatta* de haute qualité grâce à un marqueur fluorescent introduit dans les membranes des cellules embryonnaires et en utilisant la microscopie à nappe de lumière (ou SPIM pour Selective Plane Illumination Microscopy) [1]. Ces images ont ensuite été soumises à un algorithme ad hoc de segmentation nommé ASTEC[2] qui a permis d'obtenir une base de données annotées contenant les images correspondantes des régions cellulaires segmentées de manière individuelle.

Cette base de données peut désormais être utilisée pour concevoir et entraîner un réseau de neurones performant pour segmenter individuellement chacune des cellules d'embryon par apprentissage. Le Mask R-CNN [3] semble être le réseau de neurones adapté à cette tâche, pourtant, il n'existe à l'heure actuelle aucune implémentation 4D de ce réseau. Un travail de recherche visant à adapter le Mask R-CNN aux images 4D des embryons de *Phallusia mammillatta* est en cours au sein de l'équipe ICAR du Laboratoire d'Informatique, de Robotique et de Mécanique de Montpellier (LIRMM).

But du stage : Le stage consistera à proposer des modifications du Mask R-CNN en visant deux objectifs :

- (i) adapter la complexité du réseau à des données simples, ici des images de cellules assez similaires, alors que le Mask R-CNN possède la puissance de reconnaître, localiser et segmenter des objets dans une image parmi 80 classes ;
- (ii) améliorer la gestion de la mémoire utilisée par ce réseau lors des entraînements.

La pertinence de ces modifications sera évaluée en confrontant la performance de ces nouveaux réseaux avec un réseau de référence.

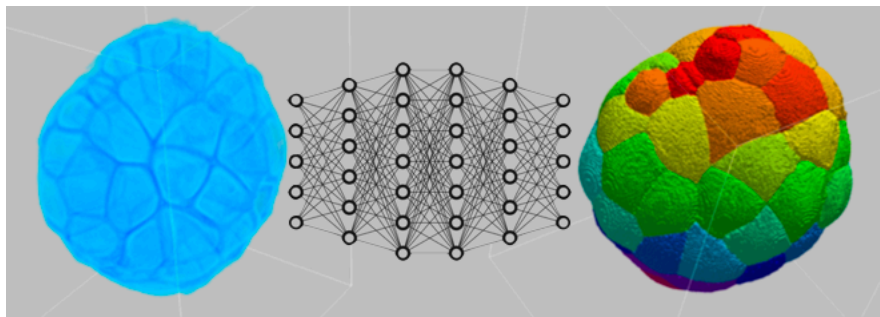
Profil recherché : Le candidat devra posséder de solides connaissances en apprentissage automatique, avec une connaissance au moins grossière des réseaux de l'état de l'art tel que le Mask R-CNN ou le YOLO. Il devra être familier avec Tensor Flow, ou à défaut Keras. Il devra se montrer créatif et force de proposition. Des connaissances de biologie du développement ne sont pas exigées mais une appétence pour les problématiques pluridisciplinaires serait avantageuse.

Contacteur : Gabriel David (gabriel.david@lirmm.fr) ou Emmanuel Faure (emmanuel.faure@lirmm.fr) pour plus de détails.

[1] Vidéo de vulgarisation : <https://vimeo.com/438580822>

[2] Guignard et al. Science 2020, <https://www.science.org/lookup/doi/10.1126/science.aar5663>

[3] Mask R-CNN, <https://arxiv.org/pdf/1703.06870.pdf>



STAGE Master 2 - Data Scientist

The objective of the internship is to adapt and optimize convolutional neural networks for segmentation on a very large database of 3D and 4D images.

Context: This internship offer is part of the general context of the embryonic development of *Phallusia Mammillatta*, a species of marine mammal of the ascidian family.

We recently succeeded in obtaining high quality 4D images (3D plus the temporal dimension) of *Phallusia mammillatta* embryos thanks to a fluorescent marker introduced in the membranes of the embryonic cells and by using SPIM (Selective Plane Illumination Microscopy) [1]. These images were then submitted to an ad hoc segmentation algorithm named ASTEC[2] which resulted in an annotated database containing the corresponding images of individually segmented cell regions.

This database can now be used to design and train a powerful neural network to segment embryo cell instances without going through ASTEC which is an expensive procedure. The Mask R-CNN [3] appears to be the right neural network for this task, yet there is currently no 4D implementation of this network. A research work aiming at adapting the Mask R-CNN to 4D images of *Phallusia mammillatta* embryos is in progress within the ICAR team at the Laboratoire d'Informatique, de Robotique et de Mécanique de Montpellier (LIRMM).

Internship goal: The internship will consist in proposing modifications to the Mask R-CNN with two objectives:

- (i) to adapt the complexity of the network to simple data, here images of similar cells, while the Mask R-CNN has the power to recognize, localize and segment objects in an image among 80 classes;
- (ii) to improve the management of the memory used by this network during training.

The relevance of these modifications will be evaluated by comparing the performance of these new networks with a reference network.

Searched candidate: The candidate should have a strong background in machine learning, with at least a rough knowledge of state-of-the-art networks such as Mask R-CNN or YOLO. He/she should be familiar with Tensor Flow or Keras. He/she will have to be creative and have the ability to make proposals. Knowledge of developmental biology is not required, but an appetite for multidisciplinary issues would be advantageous.

Please contact: Gabriel David (gabriel.david@lirmm.fr) or Emmanuel Faure (emmanuel.faure@lirmm.fr) for extensive details.

[1] Outreach video (in French): <https://vimeo.com/438580822>

[2] Guignard et al. Science 2020, <https://www.science.org/lookup/doi/10.1126/science.aar5663>

[3] Mask R-CNN 2018, <https://arxiv.org/pdf/1703.06870.pdf>