

Sujet de thèse Unité Evo-Eco-Paléo – campagne 2020

Université : Université de Lille

Ecole doctorale: ED104 Sciences de la matière, du rayonnement et de l'environnement (SMRE)

Filière doctorale : Géosciences Ecologie Paléontologie Océanographie

Titre de la thèse : Intelligence ARTIFICIELLE appliquée à la détermination des radiolaires fossiles; Implications taxonomiques, biostratigraphiques et évolutives.

Direction de thèse : Taniel DANELIAN & Pierre BOULET

Email de contact: taniel.danelian@univ-lille.fr

Laboratoire(s) de Rattachement : Unité Evolution, Ecologie et Paléontologie, UMR CNRS 8198

Financement acquis ou demandé : Financement isite ULNE PEARL ou Financement 50% Région et 50% financement Président (Univ. Lille)

SUJET DE THESE

Objectifs : Les données paléontologiques sont essentielles pour obtenir des informations sur les changements de paléobiodiversité à travers le temps géologique et la réponse biotique aux événements passés de changement global. Les enquêtes sont généralement menées à travers l'analyse de larges bases de données, dans lesquelles l'occurrence des espèces est de prime importance. Cependant, la collecte de nombreuses données au niveau des espèces demande beaucoup de travail; de plus, ces données peuvent être partiellement influencées par l'opinion des taxonomistes. L'utilisation de réseaux de neurones peut être la solution pour étudier de grands ensembles de données plus rapidement et objectivement. Ce projet de thèse appliquera la modélisation neuromorphique à la détermination des radiolaires. Le projet sera dirigé conjointement par les Professeurs Taniel Danelian (micropaléontologue, spécialiste des radiolaires, laboratoire Evo-Eco-Paléo) et Pierre Boulet (spécialiste de l'informatique neuromorphique, travaillant au laboratoire CRISTAL).

Le projet se focalisera sur deux principales questions scientifiques:

- (1) Comment l'apprentissage des réseaux neuronaux peut-il atteindre une précision de détermination des espèces de radiolaires éocènes égale à celle d'un spécialiste sur la taxonomie des radiolaires éocènes?
- (2) Comment la détermination des réseaux neuronaux se compare-t-elle aux résultats des études morphométriques classiques?

Nous comparerons l'apprentissage profond et le renforcement des réseaux de neurones pour ces tâches.

L'apprentissage profond (deep learning) est la technologie dominante et les réseaux de neurones dopés ont le potentiel à nous amener à un apprentissage non supervisé et une implémentation matérielle à très faible puissance, permettant ainsi la construction de microscopes de classification autonomes.

La possibilité existe pour une visite d'un mois au Musée d'Histoire Naturelle de Berlin. Le doctorant sera également impliqué dans l'alliance d'intelligence artificielle humAIn des Hauts-de-France.

Compétences recherchées : Nous nous attendons à ce que le candidat possède une expérience solide en paléontologie ou biologie et un intérêt pour l'informatique. La maîtrise d'au moins un langage de programmation informatique et une compréhension des réseaux de neurones ou de la classification numérique d'images seraient un atout.

Mots clefs : Micropaléontologie, apprentissage automatique, réseaux de neurones, taxonomie

Publication du laboratoire sur le sujet :

Danelian T., Saint-Martin S. & Blanc-Valleron M.-M. (2007) - Middle Eocene radiolarian and diatom accumulation in the equatorial Atlantic (Demerara Rise, ODP Leg 207). Possible links with climatic and palaeoceanographic changes – *Comptes Rendus Palevol.*, 6: 103-114.

Danelian T. & MacLeod N. (2019) – Morphometric analysis of two Eocene related radiolarian species of the *Podocyrtes* (*Lampterium*) lineage. – *Paleontological Research* 23, 314-330.

Renaudie J., Danelian T., Saint Martin S., Le Callonnec L. & Tribovillard N. (2010) – Siliceous phytoplankton response to a Middle Eocene warming event recorded in the tropical Atlantic (Demerara Rise, ODP Site 1260A). *Palaeogeog., Palaeoclim., Palaeoecology* 286: 121-134.

PhD SUBJECT

ARTIFICIAL intelligence in radiolarian fossil identification; taxonomic, biostratigraphic and Evolutionary implications

Paleontological data are essential to obtain insights on changes in paleobiodiversity through the geological time and the biotic response to past events of global change. Investigations are usually carried out through the analysis of large databases, in which the record of species level occurrences is of prime importance. However, the collection of numerous data at species level is labour intensive; also, taxonomic occurrence data may be influenced to some extent by the opinion of taxonomists. The use of neural networks may be the solution to study big datasets more rapidly and objectively. This PhD project will apply neuromorphic modelling in radiolarian identification. The project will be jointly supervised by Prof. Taniel Danelian, a radiolarian micropaleontologist (UMR Evo-Eco-Paléo) and Prof. Pierre Boulet, a computer scientist specialist of neuromorphic computing, working at CRISTAL (UMR 9189, ULille/CNRS/Centrale Lille).

The project will focus in exploring two main scientific questions:

(1) How can neural network learning achieve equal accuracy in the identification of Middle Eocene radiolarian species as an Eocene specialist in radiolarian taxonomy?

(2) How does neural network identification compare with the results of classical morphometric studies?

We will compare deep learning and spiking neural networks for these tasks. Deep learning is the mainstream technology and spiking neural networks have the potential for unsupervised learning and ultralow power hardware implementation allowing the building of autonomous classifying microscopes.

The possibility exists for a month's visit at the Natural History Museum in Berlin. The PhD student will also be involved in the humAI artificial intelligence alliance of the Hauts-de-France region.

Requested skills: We expect the candidate to have a strong background in paleontology/biology and an interest in computer science. Proficiency with at least one computer programming language and an understanding of neural networks or image and numerical classification would be an asset.

Key words : Micropaleontology, Machine learning, Neuron networks, Taxonomy.