EGU2020-16034
https://doi.org/10.5194/egusphere-egu2020-16034
EGU General Assembly 2020

A new automated radiolarian image acquisition, processing and identification workflow

Martin Tetard, Ross Marchant, Yves Gally, Thibault de Garidel-Thoron, and Luc Beaufort
Aix Marseille Univ, CNRS, IRD, Coll France, INRAE, CEREGE, Aix-en-Provence, France (tetard@cerege.fr)

Identification of microfossils is usually done by taxonomist experts and requires significant systematic knowledge and time, as about 300 specimens per sample are commonly identified for statistically reliable studies. Radiolaria are no exception, and their utility has been demonstrated for a long time in biostratigraphy using the presence / absence of some species, as well in palaeoceanographic reconstructions (past productivity, temperature, and water masses variability). Traditionally, these studies have required the manual identification of numerous species in a lot of samples under a transmitted light microscope, which is very time consuming. Furthermore, identification may differ between operators, biasing reproducibility. Recent technological advances in image acquisition, processing, and recognition now enable automated procedures for this process, from microscopic slide field-of-view acquisition to taxonomic identification.

A new workflow was developed for radiolarian acquisition, processing and identification. Firstly, a new protocol was developed as a proposed standard methodology for preparing radiolarian microscopic slides. We mount 8 samples per slide (using 12x12 mm cover slides) on which radiolarians were randomly and uniformly decanted using a new 3D-printed decanter that minimizes the loss of material. The slides are then automatically imaged using an automated transmitted light microscope. About 500 individual radiolarian specimens (excluding the broken and overlaying specimens) are recovered (about 4000 specimens per slide) from 3375 original fields of view (15 images z-stacked per FOV x 225 FOVs) per sample, after which automated image processing and segmentation is performed using a custom plugin developed for the ImageJ software. Each image is then classified using a convolutional neural network (CNN) trained on a database of radiolarian images.

To create the CNN classification stage, a dedicated software program, ParticleTrieur, was used to annotate a large dataset of radiolarian taxa (currently more than 27488 images, corresponding to 101 classes, from Neogene to recent). This software enables the visualisation and assignation of radiolarian pictures to defined taxa by progressively learning and suggesting taxa labels based on previous labelling. This database was then used to train a CNN (convolutional neural network) for

the automated taxonomical identification stage. After fusing classes containing less than 10 images into a single “other” class, 69 classes were trained to be recognised with an overall accuracy of 93 %. This new workflow will now be used on a Miocene to Recent sedimentary record from the IODP expedition 363 (Core U1488A), recovered in the West Pacific Warm Pool.