Climate
of the Past

Open Access

Discussions

EGU

CPD

Interactive
comment

Interactive comment on "A new automated radiolarian image acquisition, stacking, processing, segmentation, and identification workflow" *by* Martin Tetard et al.

Martin Tetard et al.

tetard@cerege.fr

Received and published: 26 August 2020

Interactive comment on "A new automated radiolarian image acquisition, stacking, processing, segmentation, and identification workflow" by Martin Tetard et al.

Anonymous referee.

Received and published: 18 August 2020

Comment: Structure/composition of the Paper: While the paper looks overwhelming due to the number of pages, the content is actually concise. The information written there is not too long nor short. The structure of the paper follows the usual format

(introduction, methodology, results and discussion).

Response: Thank you very much for your review. Indeed, lots of things are discussed in the manuscript, and it is difficult to reduce it too much, but we try to keep things concise.

Comment: Methodology/approach to the Problem: Their workflow is a whole and complete system, which starts at image acquisition and ends at classification. The workflow hopes to ease the tedious task of identifying specimen from samples, which requires extensive and consistent taxonomic knowledge of the observer to correctly identify radiolarians. Research was done well, as it can be seen that they have explained the steps in great detail (including the measurements).

Response: We aimed at developing a complete system for radiolarian research by using the expertise developed at CEREGE for several decades now. Everything was detailed as much as possible to enable scientists from other laboratories to use this method for their research.

Comment: As for the AI specific topic, they have used a usual Deep Learning approach. They have used ResNet and finetune it on their dataset. They have also included non-Radiolarian classes, which I believe provided an edge especially since they will be classifying things straight from the image acquisition (w/o humans to remove the non-Radiolarian particles). They have acquired a lot of samples, so they did not struggle that much on this part. Overall what I can see here is that the acquisition and segmentation of images are more tedious than the actual training and classification of Radiolarians.

Response: Indeed, non radiolarian classes were used to enable the system to identify the sediment material straight from the vial. Acquisition and segmentation parts were indeed the more tedious parts due to the complex morphology and composition of radiolarians shells.

Comment: For the purpose of training the Convolutional Neural Network (CNN) for identification of Radiolarians, they developed and released AutoRadio (Automated Radiolarian). To encourage participation and contributions on adding more images to AutoRadio, they provided a very detailed protocol to standardize the way of obtaining images. Even the file for 3D printing Decanter, used to prepare the slides, is provided for everyone to use. The repository for Decanter also includes a video for the modified random settling protocol.

Response: We tried to make this whole new protocol as accessible as possible for future radiolarian studies.

Comment: It is suggested that a section briefly discussing the convolutional neural network model should have been included. The approach fundamentally relies on the model and hence it is necessary to detail how it is applied so as to properly justify the solution for automated identification. As such, the section shall essentially include the following: CNN overview, model architecture, and training approach (transfer learning, loss, etc.).

Response: A small discussion was added as suggested by reviewer 1 and anonymous reviewer. More details about the used CNN are provided.

Comment: A minor concern is that I noticed that the Random Settling Protocol, as discussed starting in line 95 and the video (https://www.youtube.com/watch?v=veRmKI4rGTo) differ in the series of steps taken for the preparation of the radiolarian slides. I recognize that some steps are possibly not filmed for brevity, and the difference in steps might suggest that what is written on paper may not be strictly followed. But the motivated reader who wishes to contribute and follow the protocol may feel confused at first. I also noticed that in the video, the sample taken only amounted to 0.1 mg, but in the protocol the recommended amount is 0.6 mg (line 130, step #15), as it corresponds to the best compromise to ensure that a sufficient number of radiolarian specimens are covered

and at the same time the specimens are not crowded and not touching one another, as discussed in subsequent sections that overlaps might affect the ability of the workflow to identify radiolarians. What I thought is that in cases where the amount of samples is limited, taking at most 0.6 mg would be enough. All things taken, the inclusion of the video is very helpful.

Response: The video shows how to use the decanter. The part were samples are chemically prepared is thus not included in the video. The preparation protocol was very slightly emended since the original publication and now matches the protocol that was actually used to prepare 400 samples for the next study and that is visible in the video. Regarding the amount of sample taken in the video (0.1 mg), it was an annotation mistake that was corrected. It is indeed recommended to use between 0.6 and 1.0 (not 0.1) mg of material. This was corrected in the text and in the video.

Comment: Another concern is about imbalance in classes, which is actually common among Radiolarian studies. Reading on the documentation of ParticleTrieur, the recommended number of images per class is 50 at minimum and preferably at least 200 images per class, which can be very difficult to achieve especially on rare radiolarian species. Commonly, data augmentation is performed to address the issue of class imbalance. But augmenting the data has to ensure that variations applied to the image still preserve the class/label after applying transformations. Hence, careful application of augmenting data must be ensured. ParticleTrieur also makes use of weighted loss functions, which is another good way of handling class imbalance.

Response: Indeed, some radiolarian species can be very rare to tricky to found. 200 specimens are recommended per class, 50 is a minimum for accurate results, and here we decided to use classes with at least 10 specimens to train as many classes as possible where more images will be progressively added. This even if these classes are not very accurate, the system can start to recognize them and already help with the identification. The presence of classes with few images do not decrease the accuracy of the overall network and do not affect the other classes.

C4

Comment: I agree that ideally, adding more data on rare species would improve the trained model so paper emphasizing the possibility of collaboration through adding more images to AutoRadio and detailing on how one can contribute is really a good step.

Response: More images will be progressively added to the database as we process samples, to cover the rare and under-represented species. We also encourage people to send us their own pictures to cover the morphological variability that is not due to actual change in the shape of specimens, but to different acquisition settings, material, and so on. With active collaboration, the database should be expanding quickly.

Comment: Discussion of the Results: They have used the usual metrics (Accuracy, Precision, Recall, Confusion Matrix). The results were good since image acquisition and segmentation methodology is already profound, their data is quite large (17k samples total), and they have reported an overall accuracy of 90%.

Response: Again, we thank the reviewer for its comments about the MS.

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2020-76, 2020.