Clim. Past Discuss.,
https://doi.org/10.5194/cp-2020-76-RC2, 2020

Climate of the Past Discussions

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

Thore Friesenhagen (Referee)

thore.friesenhagen@unibas.ch

Received and published: 14 August 2020

1. General Comments

The development of an automated system for the collection of radiolarian census-data using a neuronal network is a consequent step to give over time-consuming workflows to machines. Posting the codes as well as the organisation of a discretionary image-based radiolarian (training) dataset are a good practice, but it also means that maintaining the dataset will be one of the most important tasks for the future. The manuscript is well done and requires only minor revisions. The following annotations and questions should be considered and/or answered in the final publication.

Printer-friendly version

Discussion paper

2. Scientific Questions

2.1. Convolutional Neuronal Network

I miss a short introduction about neuronal networks and the “k-nearest neighbours” algorithm for readers who are not familiar with these terms.

2.2. CNN Database

As mentioned in the script’s introduction, one and the same specimen may be referred to different species/classes depending on the experience, subjective interpretation and/or taxonomic “education” of the researcher (e.g. Fenton et al. (2018) for planktonic foraminifera). Thus, to reduce the number of possibly mistakenly identified specimens in the training dataset, having at least one more taxonomic expert checking the correctness of the species determination of specimens within the dataset could increase the reliability of the dataset.

2.3. Image Stacking

Does the transparency of the radiolarian shells produce any problems for the image stacking, especially in case of smaller and more delicate specimens? Figure 3j) shows a specimen of the species Collosphaera tuberosa. Its contours are diffuse. Is this a common “problem” for this species? Does this affect the identification accuracy for this species and may be one reason for the relatively high value of confusion with Solenosphaera zanguebarica?

2.4. Early Ontogenetic Stages/Juveniles

The collection of census data for planktonic foraminifera avoid juvenile specimens (e.g. Davis et al. (2019) only investigated the >125μm fraction), because their identification is often very difficult (Fenton et al., 2018). Is there a lower size limitation for radiolarian specimens to be detected and identified by the new system? Is the system able to distinguish between early ontogenetic stages and broken specimens? Does the size of specimens affect the accuracy of the automated species determination?

2.5. Image Acquisition

What is the procedure for (intact) specimens which extend over the borders of the 324 FOV and are parted/bisected? Is the program able to identify these specimens as being intact? In this case, are these specimens prevented from being “double-counted” by the system?

2.6. Influence of Orientation

Closely related species tend to show a similar morphology and are often only distinguishable by details. Since the sample preparation bases on random settling, the orientation of a single specimens may not be ideal to enable the program to recognise these morphological details. What is the procedure for specimens which do not show an ideal orientation for determination?

2.7. Morphological Measurements

I give the authors credit for implementing morphometric measurements. In combination with census data they may provide additional and valuable information for palaeoenvironmental reconstructions and evolutionary studies. Although this paper clearly focuses on the collection of census data, the accuracy of the morphometric measurements should be given as well. To what extend do differences in specimen orientation affect the accuracy and intraspecific comparability?

3. Technical Corrections

3.1. Text

-l. 62-63: The sentence contains two times the phrase “promising results”.

-l. 86: A comma is missing after “6.3ka”. “[. . .] 3-4cm, (6.3ka[,] de Garidel-Thoron et al., 2005) [. . .]”

-l. 273: there is a closing bracket at the end of the sentence, but I could not figure out the corresponding, opening counterpart. “[. . .]and that 150 images represent the

original dataset for this class; Fig. 5 green square). [. . .]"

-l. 359: The semicolon may be replaced by a closing bracket. "[. . .] palaeoenvironmental proxies such as SSTs (e.g., radiolarian-based palaeotemperatures for [. . .], Kamikuri, 2017;[)] and paleoproductivity [. . .]

3.2. Figures

-Fig. 2: A space is missing in the text for step 7. "7.[ ]Identification of every single particle using a trained CNN."

-Fig. 4: The printed version is difficult to read, because the font size of the species names is relatively small. The digital figure requires a lot of scrolling.

-Fig. 5: Several names of species overlap and make it impossible to read them.

-Fig. 6 e,f: The percentage numbers are difficult to read, because they overlap with black bars within the figure.

4. References

Davis, Catherine V., et al. "Seasonal and interannual changes in planktic foraminiferal fluxes and species composition in Guaymas Basin, Gulf of California." Marine Micropaleontology 149 (2019): 75-88.

Fenton, Isabel S., et al. "Factors affecting consistency and accuracy in identifying modern macroperforate planktonic foraminifera." Journal of Micropalaeontology 37.2 (2018): 431-443.

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