review of ImageGrains 2.0: Improved precision and generalization for grain segmentation, by Mair et al., submitted to Earth Surface Dynamics

In the paper, Mair et al. present a new method of image segmentation dedicated to the detection of sediments in 2D or 3D images (including CT scans) based on deep-learning algorithms. They build on previous work by their team and take advantages of up-to-date image segmentation methods (SAM, Cellpose-SAM) to provide ImageGrains 2.0. On average, this new algorithm outperforms previous methods and provides grain distributions estimated in very good agreement with the ground truth (manually annotated images). Overall, this is a very good, well-written manuscript supported by appropriate and clear figures.

However, I think the manuscript is a bit too focus on the method and the increase in precision it brings with respect to previous studies, at the cost of a general contextualisation. In the current form, I feel that it is difficult from someone that is not really into to grain-size measurements and deep-learning to really get the novelty and interest of this work. In fact, why we need such a new method is mostly addressed in a few lines at the very beginning and a bit in the discussion. Why we need to improve the precision on grain-size distribution is not really addressed (and I think this is a real question). Therefore, I think that the manuscript could benefit from some additions in particular in the introduction and in the Discussion, related to the use of grain sizes distributions, in order to reach the broad audience of Earth Surface Dynamics and to highlight the interest of this new approach (which I find really cool and promising).

Introduction

As mentioned above, the manuscript could better present the need for refined grain segmentation method, beyond the fact that better segmentation leads to higher accuracy in grain diameters and shape descriptors (this point is very well explained latter in the text, this is great). Why does it matter ? What can we do with an increase resolution / what do we miss with current methods ?
In a provocative way, my question is, if I want to get a grain-size distribution, why should I get into deep-learning rather than taking a ruler and a pen ? I think your method is really efficient and nice and the manuscript should state this point more clearly.

L. 42 and following: You mentioned manual measurements and imagery but I think 3D methods should be mentioned as well here (e.g., based on roughness, Tario-Vasquez et al, 2017, on point clouds, Steer et al, 2022)

In addition, the method presented here is based on image segmentation, ie on areal measurements that are not equivalent to a volumetric or surfacic sample (see for example Bunte and Abt on this question). Yet, all the grains are not detected on the image so that it is not a real areal distribution, nor a grid one. This is ok if you want to do relative sizes and to compare the evolution of a given site, without interested for the physical value, but this is far less correct if someone wants to do some physical calculations or numerical simulations. This is a classic limitation of such approach that must be addressed somewhere in the paper.

l. 83 new types of imagery: I think this is a real addition of your method, this is great and clever. You could develop this part a bit more in the results and discussion.

Methods

I understand that the database has been presented in another paper and that you do not want to do it again. However, I felt it a bit confusing to have so many reference to previous papers by your team. Many readers will not have read these papers or will not remember the specifics of them. Therefore, I think that section 2.1 could be reframed in order to be more accessible (less references, less acronymes and a more general explanation of the type of rocks that are in the images - this is partly done in the discussion, it could be done here).

2.2.2 I like this section and the method. Yet, in line with my comments on the Introduction, it could be a great addition to the manuscript to state more clearly why you use these metrics (beyond the description of the shape, why do they matter for the study of natural systems).

2.3 and 2.5 these sections are well written but might be a bit hard to follow for someone not familiar with deep-learning. I don't think this is your job to explain DL but you could explain a bit the choice of the parameters and how it could impact the segmentation, and what is a precision.

l. 214 you could explicite what king of context are in these subsets as it is not clear from the name only (and it's not ideal to move to Supplement to get a sense of it).

Results

l.284 are the 18 500 masks 63% of the ROIs or do you use 6% of 18 500 ? please rephrase to avoid confusion.

3.2 the accuracy of your new method is higher than previous ones, yet, it is still somehow limited (cf table 1, figure 4). I would appreciate that the authors acknowledge it more clearly in this section.

3.3 overall, you are able to derive grain-size distributions that match ground truth values in most cases. However, how to deal with the other situation ? is there a way to anticipate that the results are shifted ? I understand that if you work on a large collection of data, knowing that about 90% of the distributions are correct can be satisfying. But if you have only one or two images, this is more tricky. This should be addressed here (if you have materials to do so) or in the Discussion.

Discussion

As mentioned previously, I miss a section dedicated to the 2D approach in terms of grain-size description (areal measurements, comparison with manual measurements, etc). In addition, I would like the authors to comment on the fact that some grains are not segmented (either because they are too small or because it is too difficult to segment them from the images), therefore, they do not retrieve a real areal distribution (Bunte and Abt 2001 would be a very classic reference to start this discussion).

l 486 missing space between SAM and (Chan et al) + extra dot at the end of the line.

l. 503 this is a great result of your work. Do you have any idea on why this value of 68% ? Could you suggest some recommendations for image acquisition to maximize the likelihood to reach these 68% ?

l 537 fro -> for

l 557 could you elaborate a bit on what you mean by image types that are substantially different" ?

4.4.2 I would encourage the authors, if possible, to mention the environmental costs of DL methods

4.4.3 in line with my comment on the Introduction, I think that the manuscript could do a better job at presenting why grain size distributions and metrics are of interest, and what can be learn from such data set.

I am confident the authors can address all of these comments and suggestions, so I'm looking forward to read the revised version.

Yours,

Laure Guerit
CNRS Géosciences Rennes