Review: Detecting micro fractures: A comprehensive comparison of conventional and machine-learning based segmentation methods

Description of the work and recommendation

The aim of the paper is to provide a comprehensive comparison of the chosen methods for image segmentation which can be used to guide the reader in developing their own workflow. The novelty of the paper is in this comparison, rather than the methods themselves.

I found the manuscript to be well-written and well-structured. The methods used are described in detail and the results are clearly presented. I feel some additions are required to provide the "comprehensive comparison" and "guide towards an individualized segmentation workflow" as promised in the abstract.

I think this manuscript merits publication following some extension of the evaluation of the different methods and discussion of their use for different materials and modelling use cases.

Specific comments

Introduction

The notion of 'segmentation' could have more introduction. The corresponding paragraph in the introduction goes straight into implementation, without a description of what it is and why it is needed.

It would be useful to point to some studies in which images of fractured rock have successfully been analysed for different purposes, and mention the methods they used. Some examples:

- Fredrich, J. T., & Wong, T. (1986). Micromechanics of thermally induced cracking in three crustal rocks. *Journal of Geophysical Research*, *91*(B12), 12743–12743. https://doi.org/10.1029/JB091iB12p12743
- Arena, A., Delle Piane, C., & Sarout, J. (2014). A new computational approach to cracks quantification from 2D image analysis: Application to micro-cracks description in rocks. *Computers & Geosciences*, *66*, 106–120. https://doi.org/10.1016/j.cageo.2014.01.007
- Griffiths, L., Heap, M. J., Baud, P., & Schmittbuhl, J. (2017). Quantification of microcrack characteristics and implications for stiffness and strength of granite.

International Journal of Rock Mechanics and Mining Sciences, *100*, 138– 150. https://doi.org/10.1016/j.ijrmms.2017.10.013

- Healy, D., Rizzo, R. E., Cornwell, D. G., Farrell, N. J. C., Watkins, H., Timms, N. E., et al. (2017). FracPaQ: A MATLAB™ toolbox for the quantification of fracture patterns. *Journal of Structural Geology*, *95*, 1–16. https://doi.org/10.1016/j.jsg.2016.12.003 And following papers using this tool.

There has been a lot of work done in civil engineering on detecting fractures in images of concrete, building materials, pavements etc. that could deserve mention. Some examples:

- Yamaguchi, T., & Hashimoto, S. (2010). Fast crack detection method for large-size concrete surface images using percolation-based image processing. *Machine Vision and Applications*, *21*(5), 797–809. https://doi.org/10.1007/s00138-009-0189-8
- Xing, C., Huang, J., Xu, Y., Shu, J., & Zhao, C. (2018). Research on crack extraction based on the improved tensor voting algorithm. *Arabian Journal of Geosciences*, *11*(13), 342. https://doi.org/10.1007/s12517-018-3676-2
- Nguyen, T. S., Avila, M., & Begot, S. (2009). Automatic detection and classification of defect on road pavement using anisotropy measure. In *2009 17th European Signal Processing Conference* (pp. 617–621). IEEE.

It would also be useful to hear of new ML developments since U-net that have been applied to this problem. E.g. a lot has been done in edge detection.

Methods

What pre-processing did you apply to the CT image? Beam hardening correction? Ring artefact removal? Could total variation denoising have been used to correct for changes in brightness? If not, why not?

If the results of thresholding are fine for training the U-net, why not also use these data as input to the watershed segmentation? I'd be very interested in seeing how the watershed compares, as from my perspective this is the most commonly used method and feels missing from this study.

There are implementation details in the appendix that should appear in the main text, especially regarding limitations and extra steps required for the U-net.

It would help to have a high-resolution zoomed image of the "ground truth" and the original image – how good is the ground truth? What porosity does the ground truth give? How much could it be improved if annotated manually?

The code doesn't seem to be available yet at the linked repository, I see just the – did I understand correctly

Results and discussion

Evaluation the different methods

More explanation could be provided regarding the assessment of the quality of the segmentation results. In the introduction, emphasis is put on the importance of correctly mapping a fracture network and determining connectivity of fractures. But in these results, the scalar porosity is used as an evaluating metric. Which are more relevant for understanding stress, strain, fluid flow etc.? Might the local threshold + morphological erosion of the segmented image result in the best model for flow simulation, for example?

It would help to see the distribution of segmented object sizes, to get an idea of how connected the fractures are for each method. How connected are the fractures expected to be?

Measurements of aperture are discussed (p. 15), but it should be highlighted that these are qualitative assessments (if I understood correctly). Why not measure some of the apertures for a more quantitative comparison?

What are the next steps required to calculate parameters such as crack density, geometries etc. that may be used in modelling?

Regarding porosity: It would be a helpful comparison to determine the porosity from the greyscale values of the filtered image, as in Pini, R., & Madonna, C. (2016). Moving across scales: a quantitative assessment of X-ray CT to measure the porosity of rocks. *Journal of Porous Materials*, *23*(2), 325–338. https://doi.org/10.1007/s10934-015-0085-8.

Suggestions for which models to use and when

How will these methods work for materials that contain different minerals? What about for fractures which have very different sizes? And 'discontinuous' fractures at the resolution of the image?

It would be helpful to see more discussion on how best to produce the training data for the machine learning models: eroding / dilating to ensure connectivity and correct line thickness, removing small objects that have been misclassified, and even manual corrections. In this paper, a key aspect for the ML-based methods is the enhancement of the training data resulting from the active contouring, but data augmentation is described only in the appendix.