University of Stuttgart
Germany

Institute of Applied Mechanics (CE)

Prof. Dr.-Ing. Holger Steeb

Pfaffenwaldring 7
70569 Stuttgart
T +49 711 685-66346
F +49 711 685-66347

steeb@mechbau.uni-stuttgart.de
www.mechbau.uni-stuttgart.de/ls2

University of Stuttgart • Pfaffenwaldring 7 • 70569 Stuttgart • Germany
Prof. Dr.-Ing. Holger Steeb / Institute of Applied Mechanics (CE)

08.08.2022

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

Response Letter

Dear Reviewers,

First and foremost, we would like to thank again both reviewers for their very thorough assessment of our submission. In the following, we have summarized the remarks of Reviewer #1 with a corresponding statement (blue).

Bank
Baden-Württembergische Bank
Stuttgart – BW-Bank

IBAN
DE51 6005 0101 7871 5216 87

SWIFT/BIC
SOLADEST600

VAT identification number
DE147794196

Zertifikat seit 2012
audit familiengerechte hochschule

Reviewer #1

Specific comments

Introduction

1. 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

   -> We appreciate the reviewer's comment. We have now added the relevant information in our revised manuscript. More specifically, in the 4th paragraph of the introduction we have added the following:

   <page 3, line 61>
   "The extraction of features of interest in images ought to be carried out for further evaluations. This can be achieved by performing image segmentation, which is a clear pixel-wise classification of image objects. For instance, the classification of the pixels belonging to the pore space or the solid matrix, results in the effective estimation of porosity."

2. 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 MATLABTM 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.

   ->We would like to thank the reviewer for providing us with the relevant information. We have now included the aforementioned works in the revised manuscript, and also provided a brief description of these works, as follows:

   <page 5, line 144>
   "Eventually, the introduced various segmentation techniques were used in order to characterize fractures, i.e., crack density, length, orientation and aperture. In the works of Fredrich and Wong (1986),Healy et al. (2017), the authors binarized their high resolution images by manually annotating the fracture traces. Griffiths et al. (2017) classified fractures by applying the watershed segmentation method. Arena et al. (2014) segmented fractures by using the Trainable Weka platform Arganda-Carreras et al. (2017) Given that the estimation of such parameters in the studies relied on binarized images, by following differently adopted segmentation methods, it is important to compare the effectiveness and efficiency of the relevant segmentation methods."

3. 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.

-> We have now included the works mentioned by the reviewer in the revised manuscript as follows:

<page 1, line 21>
"In civil engineering, detecting fractures on building materials Yamaguchi et al (2010). Xing et al. (2018) or road pavement Nguyen et al. (2009) is important for maintenance and inspection. "

4. 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.

-> We have now included the recently developed models in our revised manuscript as follows:

<page 5, line 128>
"With the help of those CNN architectures, exploratory models were proposed and studied for the detection and segmentation of cracks from image data. In the work of Rezaie et al. (2020), the U-net and VGG-16 combined architecture, called TernausNet, were employed. The VGG-16 Simonyan and Zisserman (2014) is also a CNN-type architecture consisting of sequential convolutional layers with multiple pooling layers, which can extract features hierarchically. Fei et al. (2020), introduced CrackNet-V, which utilizes a VGG-16 like architecture. The CrackNet-V contains only sequential convolutional layers without the pooling layers.The authors showed that the model was beneficial to detect fine and shallow cracks. Recently, Li et al. (2022), proposed the SoUNet (Side-output U-net) model. The model consists of two parts; one is a conventional U-net part which performs a segmentation task. The other is the side-output part which detects edges of segmented results. By utilizing both, they improved their segmentation result."

Methods

1. 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?

-> We would like to thank the reviewer for this comment. During the XRCT data acquisition, an aluminum plate with a 0.5 mm thickness plate was used as a hardware-based filter. In order to compensate for the remaining beam-hardening effect and ring artefacts in XRCT data, simple beam-hardening correction and ring artefacts removal methods were used in the reconstruction procedure. Both methods are supported by the commercial software Octopus. The used parameters can be found in a separate dataset https://doi.org/10.18419/darus-682. This has now been addressed in the revised manuscript in appendix.A :Technical specification, as follows:

<page 18, line 412>
"In order to reduce the beam hardening effect, the 0.5 mm thickness of the aluminum plate was used as a filter during the acquisition."

<page 18, line 414>
"With the same software, the simple beam hardening correction and ring artefact removal were conducted after employing FBP."

In order to deal with the inherent noise in image data after reconstruction, we used the Adaptive Manifold Non-Local Mean (AMNLM) filtering method, which is available in the commercial software AVIZO 2019.02. This denoising method was adopted in this study after some comparisons with the other filtering methods such as anisotropic diffusion, median, Gaussian blurring and standard non-local mean filter. The adopted AMNLM method outperformed the others in terms of noise reduction and edge conservation.

We also tried the total variation method on our dataset as well. It was certainly possible to suppress the noise effectively by applying the method. A typical example is shown right after:

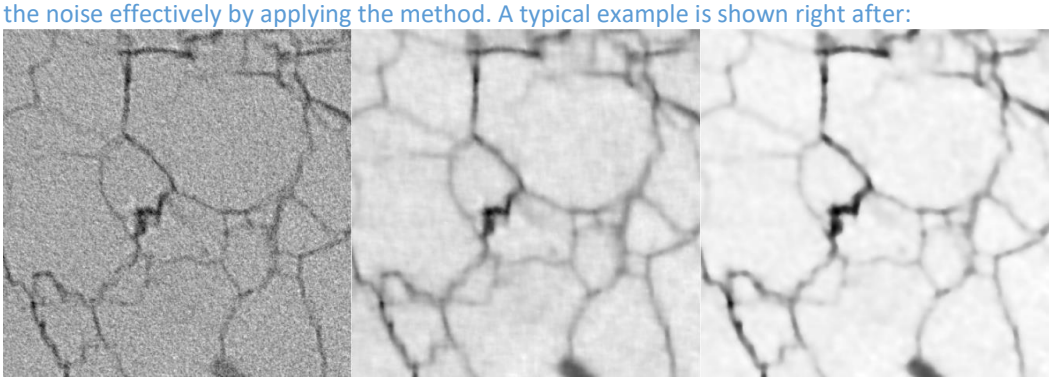


Figure 1.The comparison between original (left), AMNLM (middle) and total variation denoised image (right). After applying both filtering methods, the noise in the matrix is reduced while the boundaries of the fractures are conserved.

The total variation denoising was applied by using the 'denoise_tv_chambolle' function in skimage library in python (https://scikit-image.org/docs/stable/api/skimage.restoration.html#skimage.restoration.denoise_tv_chambolle). The result showed that we can also expect a "noise-clean" image by applying the total variation denoising method. Given that we could not quantify a qualitative difference between the two methods, as they seemed to be equivalent, we chose to proceed with the AMNLM filtering method.

It was also possible to obtain the segmented fracture network by applying the local threshold method on the denoised image. At first glance, the result from the total variation of the denoised image seems to have failed to identify relatively faint fractures. However, we expect that this can be potentially improved by tuning the parameters used in the denoising and segmentation methods.

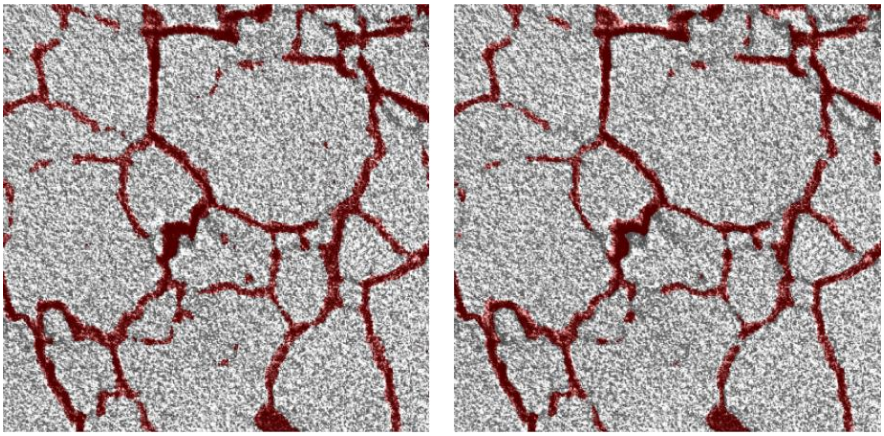


Figure 2. The local threshold results applied to amnlm (left) and tv denoised (right) images. The faint fractures are better visible with the AMNLM method.

2. 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.

-> It is indeed true that the watershed method is a popular and powerful method to identify the cracks, or more specifically the boundaries of grains. As we show in the figure below, the method provided us with clear grain boundaries and was also able to capture most of the cracks within the image data.

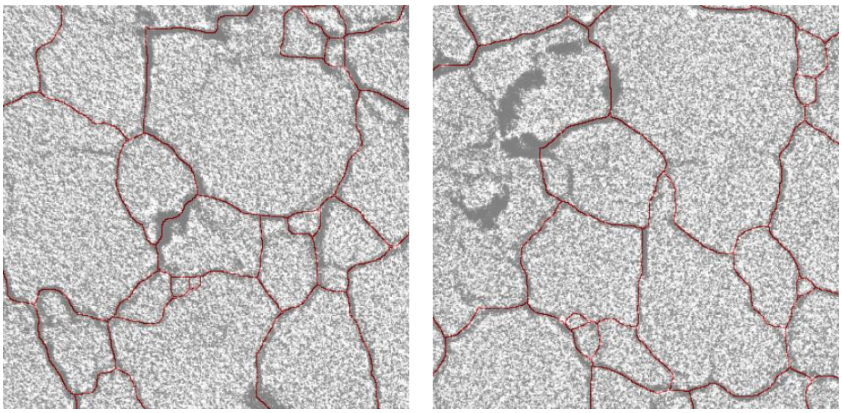


Figure 3. The results of the conventional watershed segmentation to capture the grain boundaries (marked with red color) where it successfully captures most fractures (left) and fails to detect some cracks (right).

With some extensions, such as combining additional steps, or using an advanced watershed approach, e.g. the marker-based watershed segmentation method, not only the boundaries but also the thickness of the fractures could be identified.

For instance, many studies (Lisa et al.(2021), Reinhardt et al.(2022), Lv et al.(2022))
used the marker-based watershed method which is supported by the commercial software AVIZO. We also applied this to our data set by following the workflow as follows:
(1). We obtained the intensity gradients of denoised (AMNLM filter) images in order to define a “gray area” where we are uncertain about the phase, either a fracture or the matrix.
(2). We marked each phase, fracture and matrix. In fracture phase marking, the result of the local thresholding method was used after applying an erosion operation in order to mark only certain fracture areas in the image. In matrix phase marking, we used the images, applied dilation and invert operation on the result of the local thresholding method.
(3). Finally, the watershed method was applied thus, the contour of each phase could be defined by assigning the gray areas to each corresponding phase according to computed gradients.
We also show an example of the results in the figure 4, which follows right after.

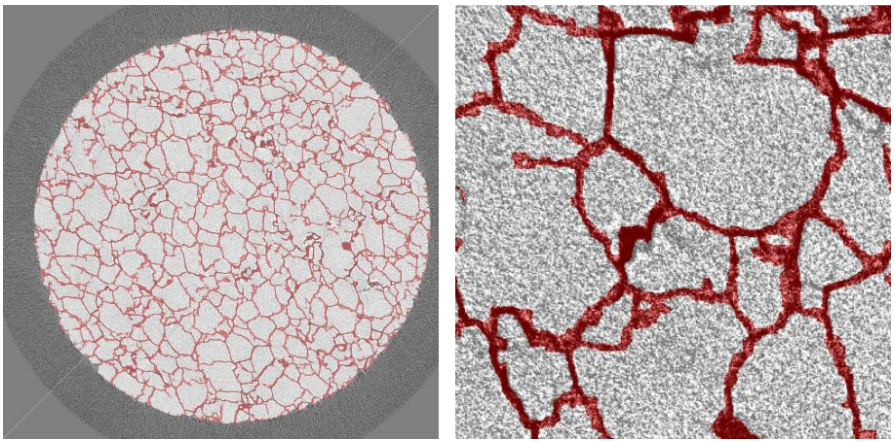


Figure 4. An example of marker-based watershed results by using the result of the local threshold as input. The identified fractures are marked in red color in overview (left) and in a magnified view (right).

With the help of this method we were able to identify the fractures in the whole region of our image data. The connectivity of segmented fractures was enhanced with the method as shown in the right figure. However, we also found a major drawback of the method in the estimation of the aperture of the segmented fracture. In the output, we were able to observe many overstated fractures and merged joints which led to its resulting porosity estimation of 20.05 [%]. By considering the quality of the segmentation output and its estimated porosity, the segmentation effectiveness of the method should be placed in between the local threshold and the Chan-vese method.

As we mentioned in the introduction, given that the marker-based watershed method relies on the gradients which were used in order to define an uncertain region (either fracture or matrix in our case), we predicted the limitation of the watershed segmentation method on our dataset due to its extremely low contrast between fracture and matrix. In addition, we experienced that the output was affected very much by the initial markers which were often mis-defined due to the sophisticated shape of fractures. This was actually one of our strong motivates to use the Chan-vese method, which is a histogram based approach instead of the marker-based watershed method.

We, of course, could have demonstrated the watershed result in our comparison study. However, with regards to the length of our manuscript, we considered it is better to show superior (Chan-vese) and inferior (Local threshold) methods instead. We also would like to add that the watershed method is rather beneficial when the fracture structure is simple and when we investigate a few cracks in a smaller domain.

3. 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.

-> We agree with the reviewer's comment. In order to facilitate the reader in understanding the U-net model, the paragraph below has been moved to the section "The U-net model" from "Appendix E: Splitting, Training, Merging for the U-net model " in our revised article.

<page 10, line 234>
"The model makes use of repeating down-scaling of the input image with the help of max-pooling layers, and up-scaling with a de-convolutional layer. Additionally, before and after each of the up-/down-scaling layers, the convolutional layers which extract the feature maps were used with an activation function which introduced a non-linearity into the model. Each of the extracted and down-scaled features were concatenated to the same size of the up-scaled features, in order to force the output pixels to be located at reasonable locations (see Figure 4)."

4. 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?

-> We prepared zoomed images of the ground truth (Chan-vese method) and the original image for comparison. The mean porosity of the adopted ground truth data was 14.88 [%].

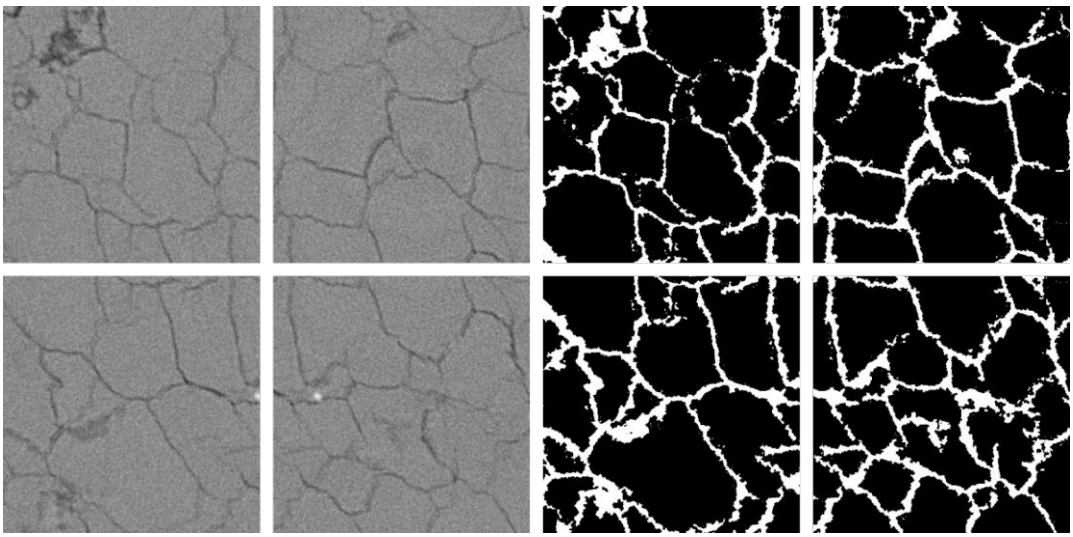


Figure 5. The examples of original and its corresponding ground truth data which were used in training of the U-net.

We have not explicitly studied the effect of manual annotation in this work, since it was very tricky to

create enough training data (approx.3000 to 4000) manually. However, we experienced that the resulting images of the ML models were affected very much with regards to the provided training data. Therefore, we expect that the resulting porosity would be closer to the estimated experimental value for porosity, if we provided the manipulated (thinning, erosioning, etc) training data to the models.

Additionally, we would like to highlight that there was always a trade-off between identifying more cracks (and pores) and reducing the aperture sizes during the model training. Given that the resolution of the original image was not fine enough to resolve the aperture of the fractures accurately, our segmentation approaches were focused on the effective identification of the cracks.

For these reasons, we concluded in using the results of the Chan-vese method, which gave us the most promising results, to the best of our knowledge, as our ground truth instead of using other possible means.

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

-> We would like to thank the reviewer for pointing this out. The code files are accessible with the full address https://doi.org/10.18419/darus-1847. We have now updated the doi accordingly in the revised article <page 18, line 393>.

Results and discussion

Evaluation the different methods

1. 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?

	LT	Sato	CV	RF	Unet
previous (rounded values from 1100 to end)	25	12	15	16	14
Vp/Vt	25.00	11.42	15.41	14.67	14.16
Vp	1.9692e+09	899892808	1.2139e+09	1.1560e+09	1.1158e+09
Vpc	1.9683e+09	875058426	1.2084e+09	1.1432e+09	1.1085e+09
Vpc/Vp	0.9995	0.9724	0.9955	0.9889	0.9935
number of objects > 1000 voxels	204	5292	226	508	299

-> We updated the estimated porosity values in table 1 of our manuscript (page 16). The red marked parameters correspond to the values previously used, and the blue marked parameters are the updated ones. The updated values were estimated for the whole volume while the previous values were evaluated in a sub-volume.

We adopted the porosity value as a standard in this comparison study, since it is a global parameter that we know the exact value of. In addition, assuming that the fractures in our dataset were well-detected by the methods adopted, the porosity values provided us with an insight into aperture information which plays a critical role in the estimation of permeability, and the mechanical response

such as its stiffness.

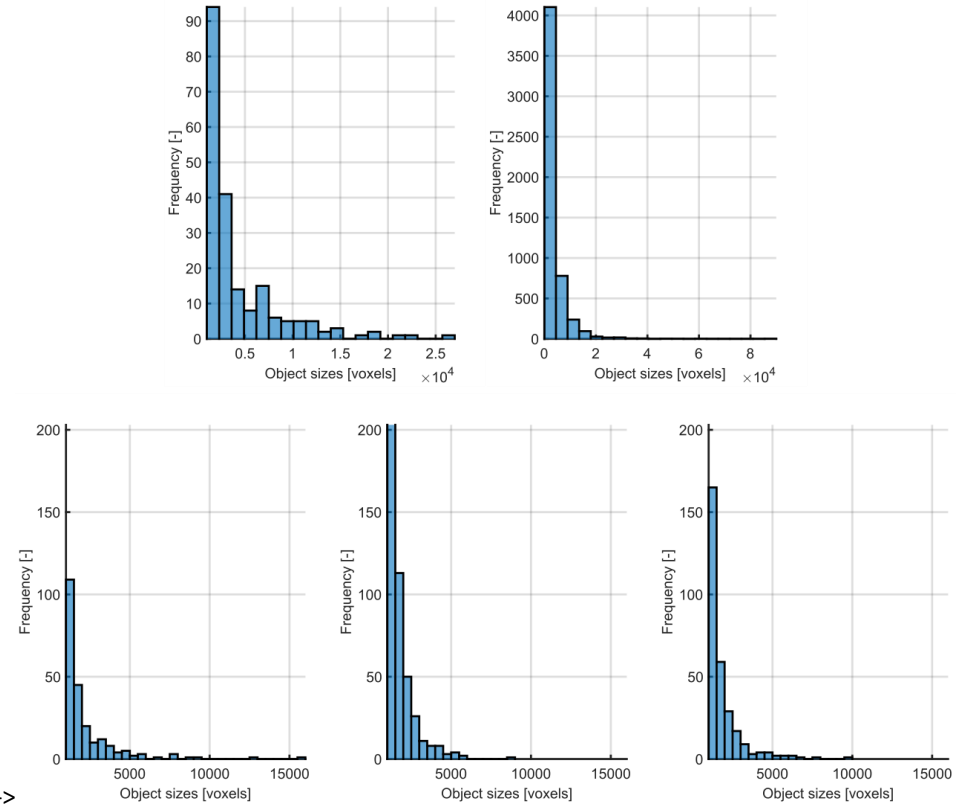
As we discussed in the introduction of our manuscript, we also agree that the connectivity of features plays a significant role to estimate permeability in a flow simulation. In order to approximate this based on the segmented geometries, we performed connectivity evaluation with the segmented fractures via the relation Vpc/Vp. The ratio Vpc/Vp provides us with the connectivity of components. Vp is the number of pore voxels and Vpc is the number of voxels of the biggest pore component, i.e., if the value is close to 1, all detected pores are connected. Falsely detected voxels were eliminated in this estimation by excluding small objects less than 1000 voxels (in 3D).

Although there were noticeable differences in the number of connected objects depending on the employed method, this was not reflected in the ratio Vpc/Vp; instead the differences in the porosity were much more profound.

By taking very detailed inspection in the ratio Vpc/Vp, the results of local threshold, chan-vese and U-net methods provided a slightly better ratio for Vpc/Vp, which was above 99%. This characteristic might be beneficial to utilize such geometry as an input of flow simulation to estimate permeability. In addition, the exaggerated apertures of the binarization results could be improved subsequently by a further imaging technique such as erosion. However, given the limitation of µXRCT to resolve such fine structure, the exact connectivity of the sample and its effect on permeability with a morphological modification to the geometry remained uncertain.

Since we believe a further numerical investigation is required to answer those scientific questions such as which method is better for a flow simulation , and our scope is on the comparison of segmentation methods in this study, we did not adopt the connectivity properties and did not conclude which output of methods is beneficial in such simulations.

2. 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?



->

Figure 6. The object size histogram of the Local threshold (top left), Sato method (top right), Chan-vese

(bottom left), Random forest (bottom middle) and U-net (bottom right). The object size smaller than 1000 voxels were excluded in this evaluation.

In the result of the local threshold method, relatively bigger sizes of objects were counted compared to other methods. In the result of the sato method, the biggest number of disconnected objects was found. The chan-vese, random forest and U-net results showed similar object size distributions.

These disconnected components were caused due to misclassification in image segmentation. In addition, some fracture connections, especially at the rim, could be lost during drilling out the sample core. We believe a numerical investigation based on obtained geometry compared with an experimental result, e.g, permeability, would be required in order to address the real connectivity in our sample.

3. 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?

-> In our study an actual measurement of aperture was not performed. The aperture size of a single fracture was estimated from the volume change (3 %) of the sample before and after quenching. Assuming that the cracks mostly appear at the boundaries of grains, the estimated aperture size was approximately 2 μm with the reported mean grain size 200 μm of Carrara marble (Saadatfar et al.(2020), Xu & Evans (2010)).

We also tried to extract quantitative aperture information by using a simple approach such as a thinning algorithm. Based on segmentation results, we adopted the “bwskel” function in matlab which skeletonized the binarized image. The results of the function are shown in figure 7, which follows right after.

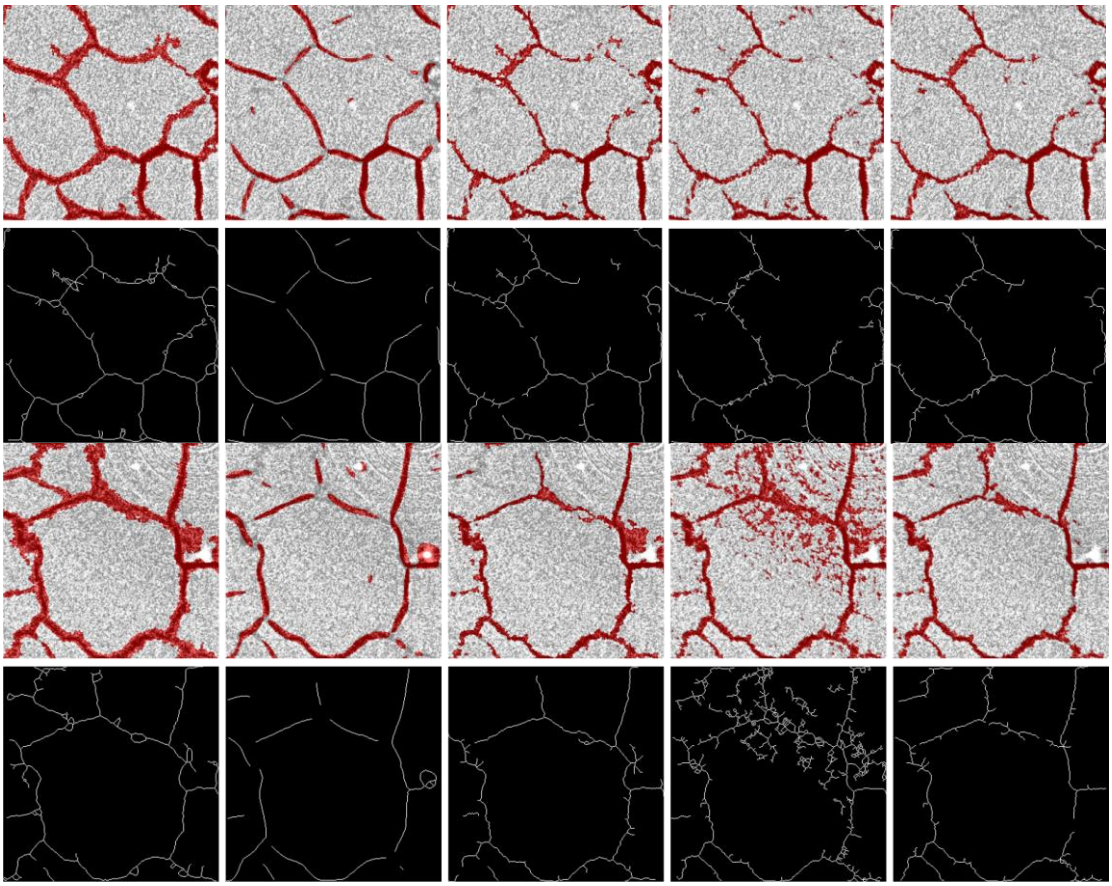


Figure 7. Examples of skeletonized subregions (top : 1000th, bottom: 2100th slice). The results are placed in order of local threshold, sato, chan-vese, random forest and U-net from left.

From the obtained skeleton, the mean aperture could be estimated by dividing the area of fractures (2D) by the length (number of skeletonized voxels). We applied the approach to the selected 11 subregions with a size of 300 by 300, where we confirmed that the approach worked fairly well. The estimated porosity, mean aperture and obtained total length of fractures are shown in the table below.

	LT	Sato	CV	RF	U-net
porosity [%]	26.21	12.35	15.93	16.21	14.93
mean aperture [μm]	18.855	17.167	14.754	12.225	13.488
The total length of fractures [mm]	5.1	2.6	3.9	4.7	4.0

Considering the estimation of the crack aperture was 2 μm, the mean apertures from all of the adopted methods were overestimating. In this comparison, the mean aperture of the random forest method showed the best result.

We would like to emphasize that this result is a rough estimation that contains defects. The problem is that the skeleton of binarized results does not provide a clear length profile in the whole region of images. As we have shown in the above figure, the skeletons contained many bifurcating branches when the shape of the fracture was sophisticated. In the worst case, it does not provide any useful information at all as the skeleton of random forest shown at the bottom of figure 7. These mis-identified branches lead the approach to overestimate the length of fractures. Consequently, the overestimation in length of fractures causes underestimation of its mean aperture. This overestimation problem was often observed in the skeleton of the local threshold and random forest results. In order to obtain reliable aperture size based on the binarized images, we believe more dedicated work is required to extract the length of fractures.

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

   -> A separation, or identification, of individual cracks is further required in order to obtain parameters such as crack density, orientation, apertures, etc. The published model for analyzing 2D fracture images, FracPaQ (Healy et al.), could be of course utilized. However, given that our segmented fractures are in practice 3D objects, a 3D approach should be employed for the calculation of the relevant parameters.

5. 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.

   -> As suggested in the work of Pini et al., we performed a sensitivity analysis associated with the value of the threshold intensity. We performed this analysis for 8 threshold values within the range from 3.15e+4 to 3.3e+4, where the step size was 0.025e+4. Since the fracture voxels in our data mostly contain the intensities within this range, these values were chosen in this analysis. By estimating porosity from the resulting binarized image corresponding to the chosen threshold value, we obtained the threshold and porosity relation as shown in the figure below with the histogram. The porosities were estimated from the whole dataset.

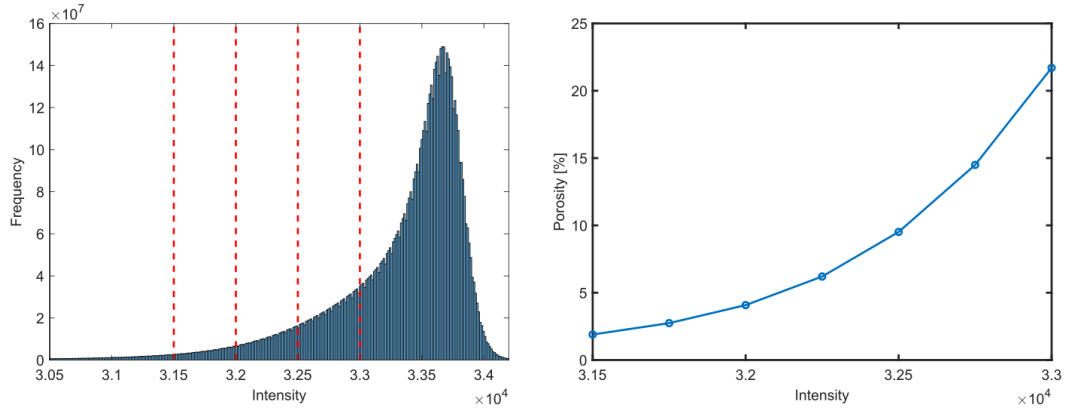


Figure 8. The histogram of the AMNLM denoised data and the porosity-threshold relation. The adopted threshold values for visualization are marked with red dashed lines in the histogram.

In addition, we adopted 4 threshold values for visualization which were 3.15e+4, 3.2e+4, 3.25e+4 and 3.3e+4. The resulting mean porosities of each selected threshold value were 1.90, 4.08, 9.52 and 21.7 [%]. Example images are shown at the below figure,

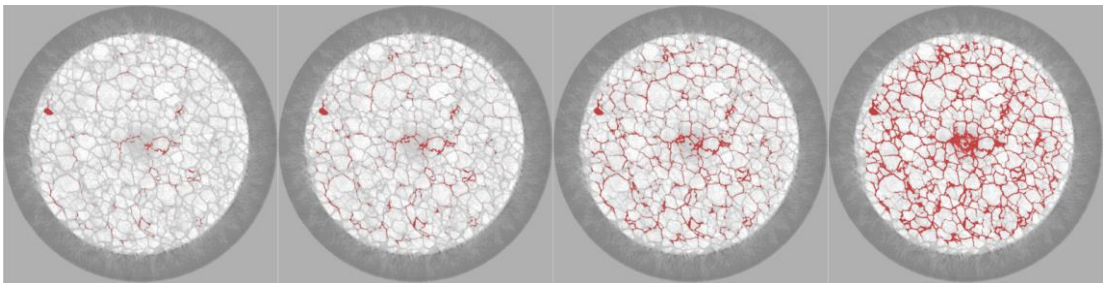


Figure 9. The example images correspond to the threshold values of 3.15e+4, 3.2e+4, 3.25e+4 and 3.3e+4 from the left. The 1000th slice of the images were chosen here.

Although the resulting porosity of the threshold 3.2e+4 provided us the closest value to the actual measurement (3 %), as we showed in the figure, the binarized image with the chosen threshold did not represent most of the fractures within the image.

On the other hand, increasing the threshold value in order to capture more fractures did not end up improving the output. The resulting image included roughly identified fractures containing significant ring artefacts. The micro cracks which had higher intensity values close to the matrix were not detected. It ended up to the resulting porosity corresponding to the threshold values to increase exponentially while the quality of the resulting image remained poor.

As we showed above, we believe the sensitivity analysis of threshold intensity compared to independent measures in this case is a bit odd since we do not have clear local minima in our histogram.

Suggestions for which models to use and when

1. 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?

-> In case the data contains a different mineral which causes a difference in brightness (intensity), we expect this would not significantly affect our comparison result since the same strategies which were used in our data should also be used. For instance, in the case of intensity based approaches like, local threshold, Chan-vese methods, we simply need to redefine the threshold value. In the case of Sato, the same structural information of fracture such as thin-long shape holds regardless of the intensity value.

In case the data includes various types of minerals, however, the fracture segmentation would be troublesome since this would require a number of additional steps to identify crack, for instance,

applying individual thresholds for different minerals. The local threshold method may not work so well in this case since the method relies on the local intensity difference. We still expect that the adopted machine learning based methods should work well. This is because these types of methods only require a ground truth. The trainable weka platform provides flexibility for such cases (assigning multi-classes). In order to use the U-net, the loss function should change to the softmax or categorical cross function for multi-class problems.

Regarding the cracks of different sizes, we anticipate the results will vary depending on the scenario. As an example, in the case of a big crack with many micro branches, the local threshold would be powerful to capture the big crack; however, the method would fail to identify most of the micro cracks nearby. This is because the result of the local threshold is affected by adjacent intensities. Due to this fact, when there is a dominant intensity value within the region of interest, a low contrast of pixels which is the case for micro cracks would not be well detected. Due to the same reason, the Chan-vese method which is an intensity based approach would need a "well treated " initial marker which highlights nearby micro cracks. Thus, for this kind of cracks, we expect that a separation in segmentation workflows for each type of crack, i.e., one for a big crack and the other for micro cracks, would be beneficial with such intensity based approaches.

In the case of the Sato method, the differences in aperture sizes are an important issue. This is because the method has to be employed based on a contrast (between fracture and matrix) amplification with Gaussian blurring. In this Gaussian blurring, the standard deviation is used as an input parameter and this has to be tuned corresponding to the width of the aperture within the image (please refer to the section Appendix:C in our manuscript). Therefore, in order to effectively detect the cracks of different sizes, a bigger range of standard deviations ought to be applied in the Gaussian blurring. In addition, the same grounds hold for the random forest method since the method also utilizes such Gaussian blurring in order to extract its feature map. However, in the case of the U-net model, the features will be extracted by convolutional layers without user intervention. Therefore, we expect the method would work without modifications.

With respect to discontinued cracks, we believe there will be no significant drawbacks with any of the adopted methods. On the other hand, if the crack is continuous, the Sato method would not be able to effectively detect branching or intersecting points as we have already shown.

2. 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.

-> In order to enhance our understanding of the adopted data augmentation for the training of the U-net, the paragraph below has moved to the section The U-net model from Appendix E: Splitting, Training, Merging for the U-net model in our revised manuscript.

<page 10, line 252>
"For the sake of accuracy, the data augmentation technique was applied on the training data-set. This allowed us to enrich the training data-set by employing a modification to the data; thus, the model could be trained with sufficient data of different variations. Consequently, the model would be trained with more trainable data. This contributed to the prevention of overfitting, which made the model capable of dealing only with a specific case. In our application, we varied the brightness of the training data. Thus, the model was able to be trained by data with variation. This was necessary to get good predictions from all cropped tiles."

Sincerely Yours,

Dongwon Lee, Nikolaos Karadimitriou, Matthias Ruf & Holger Steeb

References

- Lisa, S., Ruf, M., Steeb, H., & Quintal, B. (2021). Digital rock physics applied to squirt flow. *Geophysics*, *86*(4). https://doi.org/10.1190/geo2020-0731.1
- Lv, A., Bahaaddini, M., Masoumi, H., & Roshan, H. (2022). The combined effect of fractures and mineral content on coal hydromechanical response. *Bulletin of Engineering Geology and the Environment*, *81*(5). https://doi.org/10.1007/s10064-022-02669-0
- Reinhardt, M., Jacob, A., Sadeghnejad, S., Cappuccio, F., Arnold, P., Frank, S., Enzmann, F., & Kersten, M. (2022). Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks. *Environmental Earth Sciences*, *81*(3). https://doi.org/10.1007/s12665-021-10133-7
- Saadatfar, M., Brink, F., Latham, S., King, P., Middleton, J., Troitzsch, U., Turner, M., & Henley, R. W. (2020). High resolution 3D mapping of grain kinematics during high temperature sequestration of SO2 from flue gas by carbonate aggregates. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-58216-y
- Xu, L., & Evans, B. (2010). Strain heterogeneity in deformed Carrara marble using a microscale strain mapping technique. *Solid Earth (AGU)*, *115*(B4). https://doi.org/10.1029/2009JB006458

University of Stuttgart
Germany

Institute of Applied Mechanics (CE)

Prof. Dr.-Ing. Holger Steeb

Pfaffenwaldring 7
70569 Stuttgart
T +49 711 685-66346
F +49 711 685-66347

steeb@mechbau.uni-stuttgart.de
www.mechbau.uni-stuttgart.de/ls2

University of Stuttgart • Pfaffenwaldring 7 • 70569 Stuttgart • Germany
Prof. Dr.-Ing. Holger Steeb / Institute of Applied Mechanics (CE)

08.08.2022

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

Response Letter

Dear Reviewers,

First and foremost, we would like to thank again both reviewers for their very thorough assessment of our submission. In the following, we have summarized the remarks of Reviewer #2 with a corresponding statement (blue).

Bank
Baden-Württembergische Bank
Stuttgart – BW-Bank

IBAN
DE51 6005 0101 7871 5216 87

SWIFT/BIC
SOLADEST600

VAT identification number
DE147794196

Zertifikat seit 2012
audit familiengerechte hochschule

Reviewer #2

1. An introductory paragraph in section 2 might help to separate the 3 subsections here: sample preparation, noise reduction and segmentation methods.

-> We appreciate the reviewer's comment and we have now explicitly separated the three sections. Section 2.1 refers to sample preparation <page 6, line 164>, section 2.2 to noise reduction <page 6, line 182> and section 2.3 to segmentation methods <page 7, line 189> in the revised article.

2. Section 2.2: Could the authors please provide information regarding the parameters employed for applying the Non-local mean methods? That is, shape and diameter of the searching windows, similarity value, etc.

-> We have now added the relevant information in the revised manuscript as follows:

<page 6, line 186>
"The adopted input parameters of the filter in this study were spatial standard deviation: 5, intensity standard deviation: 0.2, size of search window: 10 and size of local neighborhood: 3."

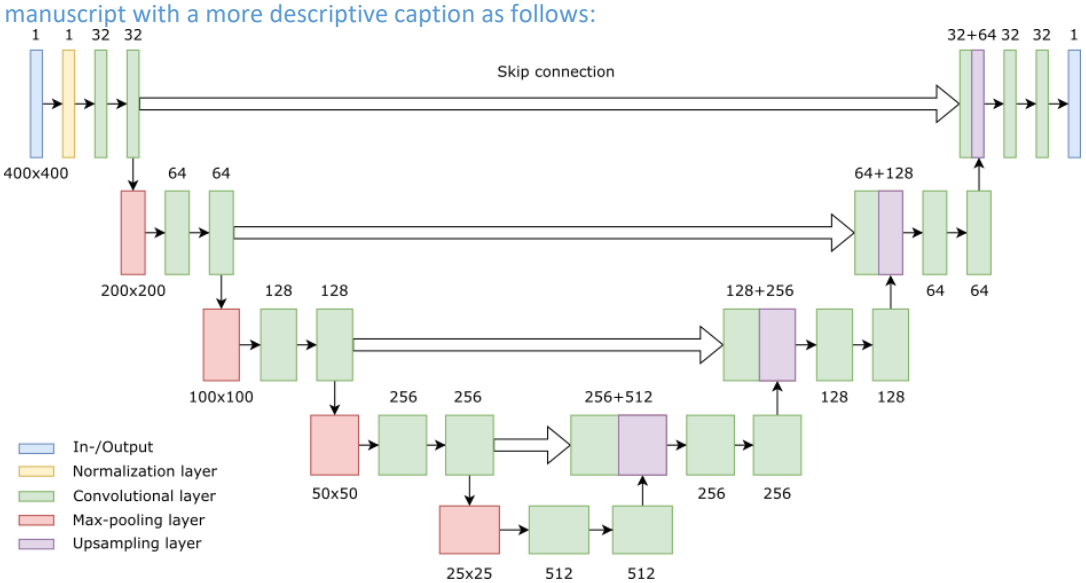
3. Sections 2.3.1/2/3. Given that the µ-CT images of the thermally treated Carrara marble sample are open to be access by the community, could the authors provide with the parameters employed to segment the data? Like this, any reader would be able to reproduce the results and work from there on improving the methods.

-> The code used for the segmentation, along with the relevant parameters, has been made available via the link in the "code availability" section. It is addressed in our revised manuscript as follows:

<page 7, line 192>
"The implemented code with detailed parameters is available via the link in the "code availability" section."

4. Figure 4: The legend item "Upsampling layer" (purple) is not in the figure. Also, the caption could be a bit more descriptive of the sketched workflow.

-> We thank the reviewer for their comment. The figure has now been modified in the revised manuscript with a more descriptive caption as follows:



Caption : *"The schematic of the used 2D u-net model. The input image was down-/upsampled with the help of Max-pooling and Upsampling layers. The features were extracted by convolutional layers. Those extracted features during down-sampling were concatenated to corresponding up-sampled features (skip connection)." <page 11>*

5. Sections 2.3.3: The authors might consider improving the explanation of the U-net model. It is the most efficient of the analyzed methods and I found this section a bit difficult to follow.

-> We appreciate the reviewer's comments. For the sake of not making the main article too extensive, we decided to keep the section short, which was obviously not the best option. In the revised article we have now moved the paragraphs below to the section "The U-net model" from "Appendix E: Splitting, Training, Merging for the U-net model". We trust that we now provide adequate information for a better understanding of the model.

<page 10, line 234>
"The model makes use of repeating down-scaling of the input image with the help of max-pooling layers, and up-scaling with a de-convolutional layer. Additionally, before and after each of the up-/down-scaling layers, the convolutional layers which extract the feature maps were used with an activation function which introduced a non-linearity into the model. Each of the extracted and down-scaled features were concatenated to the same size of the up-scaled features, in order to force the output pixels to be located at reasonable locations (see Figure 4)."

<page 10, line 252>
"For the sake of accuracy, the data augmentation technique was applied on the training data-set. This allowed us to enrich the training data-set by employing a modification to the data; thus, the model could be trained with sufficient data of different variations. Consequently, the model would be trained with more trainable data. This contributed to the prevention of overfitting, which made the model capable of dealing only with a specific case. In our application, we varied the brightness of the training data. Thus, the model was able to be trained by data with variation. This was necessary to get good predictions from all cropped tiles."

6. Have the authors tried to resample the raw images by reducing the voxel size? This might help to segment fractures thinner than 2 microns.

-> We appreciate the reviewer's comment. Given the limitation of our hardware, we were not able to physically resolve features smaller than 2um/voxel. A possible resampling option could potentially increase the image resolution artificially by using interpolation. With the help of this approach, we anticipate some fractures will be captured as thinner than 2 microns. However, we believe that a further investigation would be necessary in order to trust such captured thinner fractures since those are based on artificial intensities. Besides, the size of data would become significantly larger as the resolution increases. This will directly affect the computation time and required amount of memory for computation.

7. I'm not sure if this makes sense, but have the authors tried to re-apply the U-net model using the U-net model output as the new GT for training a new model?

-> We appreciate the reviewer's comments, and this had indeed been one of our thoughts too. However, we strongly believe that this approach would cause the mode-collapse problem which oversimplifies the training. If we repeatedly feed the model with its own products, the model rather converges to learning the simplest features instead of learning various and generalized features in the ground truth. This problem was reported and known in the GAN (Generative Adversarial Network) type of model (Li et al.(2019))

Sincerely Yours,

Dongwon Lee, Nikolaos Karadimitriou, Matthias Ruf & Holger Steeb

References

- Li, Y., Xiao, N., & Ouyang, W. (2019). Improved generative adversarial networks with reconstruction loss.

Neurocomputing, *323*(5), 363-372. https://doi.org/10.1016/j.neucom.2018.10.014