Benchmarking of Machine Learning and Conventional Image Segmentation Techniques on 3D X-ray Microscopy Data

Matthew Andrew[1,*] and Benjamin Hornberger[1]

[1.] Carl Zeiss Microscopy, Pleasanton, CA, USA
* Corresponding author, matthew.andrew@zeiss.com

Extracting relevant information from images is a key part of any microscopy workflow. While visual interpretation by the user is still relevant, extracting quantitative information is becoming more and more important. In absorption contrast X-ray microscopy (XRM) images, pixels (or voxels, in 3D datasets) typically have a grayscale value representing the relative absorption (or transmission) of the X-ray beam at that location of the sample. An important step toward quantitative information in many applications is image segmentation, where each pixel or voxel in the raw image is assigned to one of several discrete "phases" representing the different components or materials present in the sample. A simple example is a porous material consisting of two phases: solid and pore. Once segmented, data can then be analyzed in terms of volume fraction of the various phases, pore connectivity, particle size distribution etc., or can be used as input to computational models for purposes such as flow simulation.

The most basic segmentation method is histogram thresholding, where each phase is assigned to a certain range of gray scale values in the image histogram. This works well for low noise data and few phases that are clearly separated as peaks in the histogram. It works less well for noisy data, when artifacts are present, or when the various phases differ not in gray scale but other attributes such as texture. Pre-/post-processing, as well as more advanced segmentation methods like watershed, can help, but results still depend strongly on the quality of the initial data and user expertise. These conventional methods often struggle segmenting images where the human eye and brain can still separate the different phases. However, accurate segmentation is important to obtain correct results in subsequent analysis.

The application of machine learning (ML) techniques has grown exponentially over the last few years in many fields including image processing. Several tools, both free / open source [1] and commercial, are now available for scientific image segmentation using ML. While an in-depth coverage of the underlying algorithms is beyond the scope of this paper, there is a substantial difference from the user perspective: Rather than employing a user-defined sequence of image processing functions, where the parameters are carefully tuned to achieve the desired result, the algorithm is trained by hand-labeling features representing the different phases on a small part of the dataset. This training can then be applied to the whole dataset or even a large number of images of the same type to complete segmentation.

In this paper, we benchmark ML based segmentation (using ZEISS ZEN Intellesis software [2]) against two traditional techniques: global multi-Otsu thresholding [3] (using ImageJ) and seeded watershed region growing [4] (using ORS Dragonfly Pro [5]). A low noise 3D XRM datasets of a Bentheimer sandstone sample was acquired using a ZEISS Xradia 520 Versa microscope and reconstructed using a conventional filtered back-projection algorithm (8.96 µm voxel size, 1024^3 voxels). This dataset was segmented into pore and grain phases using ZEN Intellesis, the result of which was considered the "ground truth". The dataset was then forward projected and noise of four different levels added in the projection domain. The data were then reconstructed again into five different synthetic 3D datasets (no noise plus four different noise levels) and segmented with the three methods listed above. For each

segmented dataset, the number of misclassified voxels compared to the "ground truth" was computed.

The results are shown in Fig. 1. In summary, at low noise levels all methods work similarly well and the number of misclassified voxels is in the low single percent range. However, when the noise level reaches 5%, the different methods begin to deviate. At 10% noise, the number of misclassified voxels for ML segmentation is still only 5%, while it is 16% for Otsu thresholding and 22% for watershed.

In conclusion, we demonstrate that machine-learning based segmentation produces superior results and is more tolerant to noise and artifacts compared to traditional segmentation methods such as histogram thresholding and watershed. A more detailed description of the method and the results, along with additional examples from Focused Ion Beam-Scanning Electron Microscope (FIB-SEM) tomography, including texture as a contrast mechanism, can be found in Ref. [6].
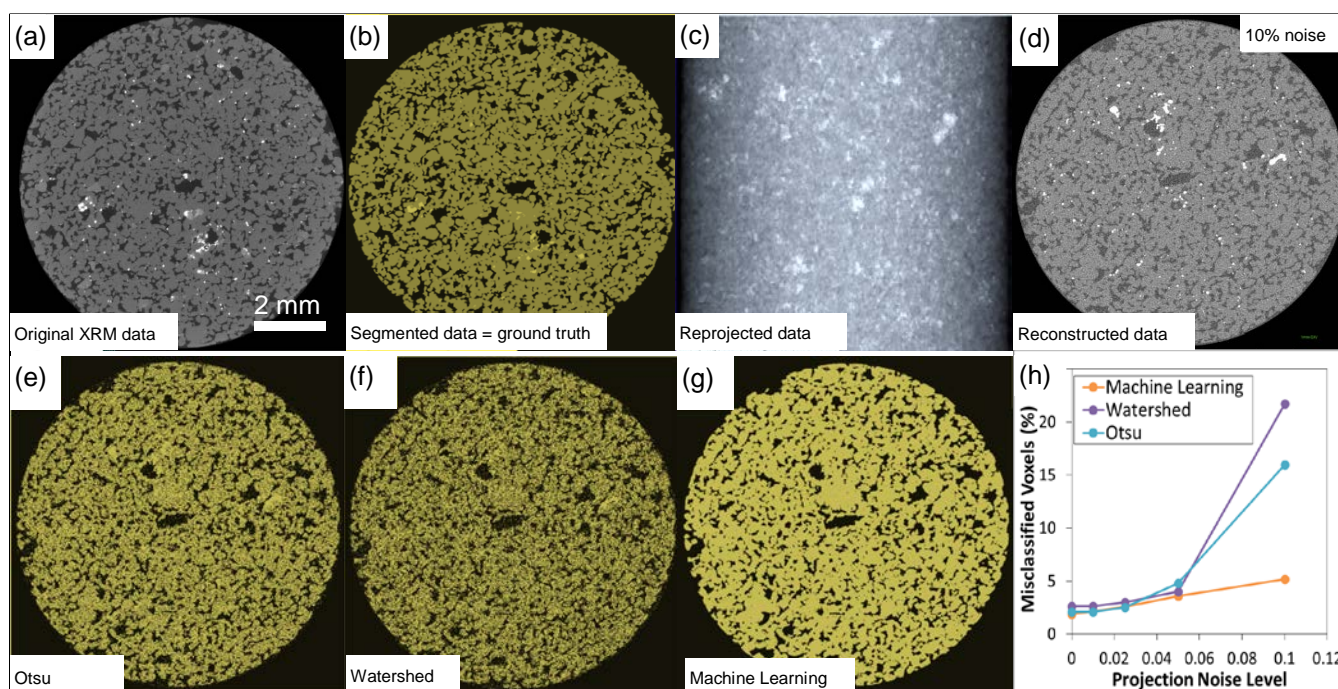

Figure 1. Example images (adapted from [6]): (a) Original XRM data (one slice from 3D volume). Black: pore space; gray: grain phase; white: inclusions. (b) Segmented data, considered "ground truth". (c) Reprojected data. (d) Reconstructed data with 10% noise. (e) Data with 10% noise, segmented using multi-Otsu technique. (f) Same data, segmented with watershed. (g) Same data, segmented with machine learning algorithm. (h) Plot of fraction of misclassified voxels as function of noise level. Scale bar 2 mm (applies to images a, b, d, e, f, g). Further images can be found in Ref. [6].

References:

[1] E.g., https://imagej.net/Trainable_Weka_Segmentation, or http://ilastik.org/
[2] https://www.zeiss.com/microscopy/int/website/landingpages/zen-intellesis.html
[3] N Otsu, IEEE Trans. Syst. Man Cybern. **9** (1979) p. 62
[4] AC Jones *et al.*, Biomaterials **28** (2007), p. 2491
[5] http:// info.dragonfly-pro.com/home.html
[6] M Andrew, *submitted*