Quantitative Big Imaging: From Images to Statistics

Kevin Mader[1, 2]

1 ETH Zurich **2** 4Quant Ltd.

Summary

Imaging is a well established field and is rapidly growing as technological improvements push the limits of resolution in space, time, material and functional sensitivity. These improvements have meant bigger, more diverse datasets being acquired at an ever increasing rate. With methods varying from focused ion beams to X-rays to magnetic resonance, the sources for these images are exceptionally heterogeneous; however, the tools and techniques for processing these images and transforming them into quantitative, biologically or materially meaningful information are similar (Zitnik et al. 2018). The module consists of equal parts theory and practical analysis of first synthetic and then real imaging datasets. Basic aspects of image processing are covered such as filtering, thresholding, and morphology (Soille 2002). From these concepts a series of tools will be developed for analyzing arbitrary images in a very generic manner. Specifically a series of methods will be covered, e.g. characterizing shape, thickness, tortuosity, alignment, and spatial distribution of material features like pores (Claude 2008). From these metrics the statistics aspect of the module will be developed where reproducibility, robustness, and sensitivity will be investigated in order to accurately determine the precision and accuracy of these quantitative measurements (Smith-Spangler 2012) (Alden and Read 2013). A major emphasis of the module will be scalability and the tools of the 'Big Data' trend will be discussed and how cluster, cloud, and new high-performance large dataset techniques can be applied to analyze imaging datasets (Mader and Stampanoni 2016). In addition, given the importance of multi-scale systems, a data-management and analysis approach based on modern databases will be presented for storing complex hierarchical information in a flexible manner (Altintas 2013)(Ollion et al. 2013). Finally as a concluding project the students will apply the learned methods on real experimental data from the latest 3D experiments taken from either their own work / research or partnered with an experimental imaging group. The module provides the necessary background to perform the quantitative evaluation of complicated 3D imaging data in a minimally subjective or arbitrary manner to answer questions coming from the fields of physics, biology, medicine, material science, and paleontology.

Statement of Need

The module and associated techniques serve to offer a basis for understanding the process and mindset of image analysis. The focus is to clarify which steps are taken, in which order, for what purpose. The goal is to provide the consumer of the module with the ability to address a wide variety of different problems and have a process for deciding which methods to pursue and how to best evaluate those methods in a qualitative manner. These abstract concepts are brought into tangible problems and solutions using popular competitions on Kaggle and the PyData toolset (numpy, scipy, scikit-learn, scikit-image).

In addition certain lectures will include portions of Tensorflow, PyTorch, and ITK code which is better suited for some problems.

This module explicitly does not go into the details of implementation or execution of the various tasks and leaves that exercise for other much better suited modules and courses.

References

Alden, Kieran, and Mark Read. 2013. "Computing: Scientific software needs quality control." *Nature* 502 (7472). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.:448. https://doi.org/10.1038/502448d.

Altintas, Ilkay. 2013. "Workflow-driven programming paradigms for distributed analysis of biological big data." In *2013 Ieee 3rd International Conference on Computational Advances in Bio and Medical Sciences (Iccabs)*, 1–1. IEEE. https://doi.org/10.1109/ICCABS.2013.6629243.

Claude, Julien. 2008. *Morphometrics with R.* http://books.google.ch/books/about/Morphometrics

Mader, Kevin, and Marco Stampanoni. 2016. "Moving image analysis to the cloud: A case study with a genome-scale tomographic study." In, 1696:020045. AIP Publishing. https://doi.org/10.1063/1.4937539.

Ollion, Jean, Julien Cochennec, François Loll, Christophe Escudé, and Thomas Boudier. 2013. "TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization." *Bioinformatics (Oxford, England)* 29 (14):1840–1. https://doi.org/10.1093/bioinformatics/btt276.

Smith-Spangler, Crystal M. 2012. "Transparency and reproducible research in modeling: why we need it and how to get there." *Medical Decision Making : An International Journal of the Society for Medical Decision Making* 32 (5):663–6. https://doi.org/10.1177/0272989X12458977.

Soille, P. 2002. *Morphological image analysis: principles and applications.* http://dl.acm.org/citation.cfm?id=773286.

Zitnik, Marinka, Francis Nguyen, Bo Wang, Jure Leskovec, Anna Goldenberg, and Michael M. Hoffman. 2018. "Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities," June. http://arxiv.org/abs/1807.00123.