Keynote: Automatic Segmentation and Quantification of TB Scale Volumetric Murine Brain Vasculature Data

Katherine Scott[1], Venkata. N.P. Vemuri[2], Hunter Jackson[3],

1. Analysis Lead, 2. Research Scientist 3. Data Scientist, 3Scan, San Francisco, CA, USA

Abstract 3Scan is a computational pathology company which uses robotics, chemistry, and computational platforms transforming our ability to understand and analyze diseased tissue. The robotic microscopes allows 3Scan to image a tissue volume of up 5 cm^3 with micron resolution using a standard RGB line scan imager, creating data sets of up to 5 TB per sample and over 5 PB of data collected to date. In addition to technical innovation internally 3Scan is committed to open science and is using legal and technical processes to make the data available to the wider research community where they can pose scientific questions.

Recently, we deployed a prototype big-data infrastructure and our first product that uses this infrastructure. This product is called the 3Scan Vascular Fingerprint™ (VFP) and it allows us to precisely segment and quantify murine vasculature and present in an easily accessible way to scientists for further exploration. In this keynote we'll discuss how our software infrastructure is designed to allow us to exploit aspects of volumetric raster data in a novel way. Moreover we'll show how this infrastructure allows our customers (that is to say researchers and scientists) to quickly distill a large volume sample into smaller regions and quickly calculate statistical values across multiple brain regions and across the whole brain as necessary.

As part of the product development cycle we put the system through a rigorous validation cycle. This validation process required us to create a large dataset of 3D murine vasculature along with hand annotated sections of vasculature for validating our approach using both crowd labor and licenced pathologists. Using this data we were able to benchmark how well our segmentation algorithms perform against naive human labelers and board certified pathologists. To validate the scientific accuracy and usefulness of our approach we then compared our metrics between murine brain vasculature models of Huntington's Disease and control mice. We were able to show expected changes in vasculature within HD mice as established in existing literature.

Inherent in this validation process, and in the spirit of Open Science, 3Scan is proud to announce that we are releasing this vasculature data from about 30 mice which sums close to 30-40 TB as "Murine Brain Vasculature Baseline" under a Open Database share-alike license. This vasculature baseline can serve as an initial set of controls for brain vasculature research; and as a benchmark for comparison in murine disease models.

This keynote will also discuss the architectural decisions that allowed us to make advances in processing our large spatial image datasets and how we exploited open data formats to make progress. To this end we will look at the pros and cons of some of the existing open data