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

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

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

Abstract

Emerging serial section light microscopy platforms, like Knife-Edge Scanning Microscopy (KESM), generate high-resolution data sets at a rate exceeding 1 TB / cm^3 of tissue, and can generate 3D voxel data of an entire mouse organ like a brain or kidney. Not only is this technique much faster than imaging slides manually on a traditional microtome it generates much larger and better statistically sampled data sets. These large datasets require new and innovative infrastructure to support the development and deployment of automated segmentation algorithms. In this paper we briefly describe the KESM, our analysis infrastructure, and our validation procedures for automated tissue segmentation routines. We demonstrate multiparametric quantification of vasculature across large sample volumes and the validation of segmentation of these volumes using both trained pathologists and un-trained workers. This type of validated vascular analysis is useful for understanding tumor angiogenesis, arteriosclerosis, vasculopathies, and neurodegenerative diseases. To that end, recent studies have shown an increase in blood vessel density & reduction in blood vessel diameter in the striatum of mice with Huntington's Disease [1,2]. We show that our validated mouse model recapitulate and expand on these findings across larger volumes.

Index Terms- Knife Edge Scanning Microscopy(KESM), vasculature, quantification, validation

INTRODUCTION

Light microscopy is a useful tool for investigating microscopic structures and pathological alterations in both human and animal models of disease. However, due to tedious manual interventions, quantification of histopathologic markers is classically performed on only a few 2D tissue sections, thus restricting measurements and observations to limited portions of the sample volume. Serial section microscopy platforms, like KESM [3] offer the potential for the generation of high-resolution terabyte scaled data sets. These datasets accompanied with automated identification and quantification of cytological and morphological features in spatial context across entire tissue blocks can have many advantages. It can be useful for transforming the way doctors and researchers examine tissue by taking the costly and slow technique of manual tissue sectioning and interpretation and replacing it with precise robotics and computer vision.

KESM Imaging

The KESM is an automated, serial section microscopy system that is capable of simultaneous tissue sectioning and imaging. An objective is pointed at the bevel of a diamond knife. The sample is clamped to a precision XYZ stage, which moves against the knife blade as shown in

figure 1. A line-scanning camera images through the objective; images are tiled into 256x256 pixel images, and stored on AWS S3 with access via AWS DynamoDB records. Samples are imaged with voxel size 0.7 x 0.7 x 5 µm per voxel. Full sample volume on average is 0.81 cm^3.

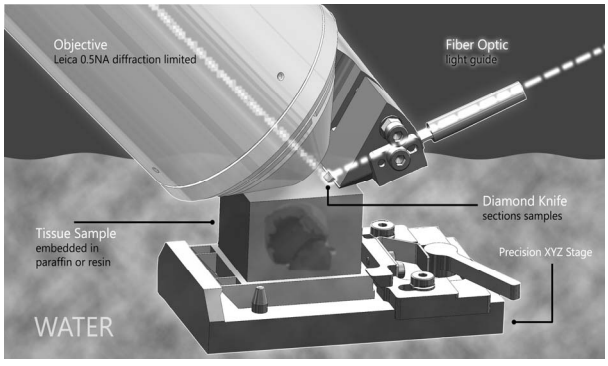

Figure 1: Knife Edge Scanning Microscope

The quantification and mapping of vasculature has wide applications: imaging of tumor xenografts for screening of targeted therapies, reconstruction of ischemic stroke models, understanding of vascular diseases such as hypertension, arteriolosclerosis [4], periventricular venous collagenosis [5], and tumor induced angiogenesis [6]. and in mapping neural innervation patterns for surgical planning [7]. Additionally, quantification, mapping of vasculature can be used to investigate the plasticity of cerebral micro-vessels influenced by learning, emotions, injuries or

environment.

In this paper, we demonstrate quantification of murine brain vasculature using computer vision across large sample volumes obtained using KESM imaging and processed using a spatially-aware distributed computing infrastructure. Implementation of a noise-reduction algorithm followed by connected component analysis and feature extraction enable enhanced localization of murine vasculature.

The object detection is validated by collecting ground truth annotations from a board-certified pathologist and crowdsourced annotations from untrained workers which are subjected to a rigorous quality assurance process by the pathologist. The combinatory effect of accessing a smaller sample size of expert annotations, and force-multiplying the scale of that set through crowdsourced labeling platforms creates the opportunity for terascale algorithm validation.

In a controlled experiment investigating the role of vascular remodeling in neurodegeneration, we compare Wild Type (WT) and Huntington's Disease (HD) R6/2 mouse striatum vasculature across a 250 µm z-stack ROI. We found greater density, higher vessel count, and smaller average diameter in the HD mice against the WT, which is consistent with previous literature. We have open-sourced this data which can be accessed at http://www.3scan.com/technology/opendata/

METHODS

Vascular perfusion

Six validation, four control and four R6/J mice were anesthetized using isoflorane gas. The heart was exposed and the apex snipped; a blunted 18-gauge cannula was inserted into the left ventricle. Mice were perfused at 300 mmHg, first with PBS until perfusate ran clear, then with 200mL 10% neutral buffered formalin. Finally, they were perfused with 10 mL India ink. Brains were dissected, processed in a Sakura Tissue-tek E300 tissue processor, and then embedded in paraffin.

Image Acquisition and Storage

Each knife pass of KESM produces a single image called a slice which is hundreds of megabytes in size. Each layer in z of a tissue sample is much bigger than the knife, so a z section is composed of slices to cover the entire face of the z stack. Samples are sliced into many thousands of faces of thickness ranging from 5-10 um. Each tissue sample after imaging is 10s of terabytes in size.

To enable terabyte-scale medical image processing we implemented a software infrastructure which works as as volumetric spatial database and distributed computing pipeline. The infrastructure is designed to store large collections of spatially organized data and distribute the processing of these data to a high throughput, cloud-based computing environment. A single position in the spatially-aware storage system defines a space in which many millions of separate but spatially related pieces of data can be stored for random access real time retrieval and processing. This brings us to the steps in the image analysis pipeline that quantify vasculature in this paper.

Image Analysis

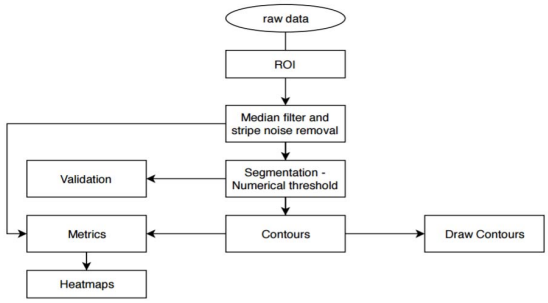

Figure 2: Steps in Analysis Pipeline

Image tiles and volumes are analyzed with a map-reduce infrastructure described in the above section that maintains a global coordinate system and allows large-scale distributed analysis of vasculature in tiles. The raw sample images are passed through an image processing pipeline that includes steps for image analysis and vessel quantification. The steps in this pipeline are as shown in figure 2. The first step in the pipeline is finding a Region of Interest (ROI). This currently is a manual step done by histologists to segregate a region like striatum affected due to HD. The striatum ROI selected for the 8 mice in this paper(4 HD, 4 WT) is 2048 x 2048 x 50(0.21 GigaVoxels = 0.514 mm^3) on average. The second step involves removing noisy outliers in grayscale images using a median filter and hough line based stripe noise detection algorithm. The grayscale data after removing artifacts is sent to a binarization filter which sets vessels to 255's and background/tissue to 0's based on a threshold. Results of the segmentation i.e binary images are used to detect contours around elliptical cross section of vessels across a 2D tile. These contours are used to define and obtain metrics that are presented in Results section. Definitions of these metrics are as follows

1) Vessel density - Ratio of total number of pixels that are recognized as vessels or foreground to the total number of pixels in a region.
2) Vessel diameter (um) - Vessel diameters are estimated as the diameter of the circle whose area is same as the contour area.
3) Vessel contour count - Number of vessel contours counted per tile

Ground Truth Data Collection

In order to perform an initial performance validation of our segmentation algorithm, six control mice were sacrificed and perfused using our standard methods and then imaged

on a KESM. A random sampling of 90, 3 x 3 tiles sections (768 x 768) from the whole brain were taken for labeling by a board-certified pathologist. The pathologist used a software provided by the vendor *named* FigureEight (subsequently called F8 for clarity) to generate raster pixel mask outlining different aspects of a semantic segmentation ontology. The pathologist engaged in multiple rounds of pilot training sessions with the crowd labor force on F8 where example data to the crowd labor pool on F8. The results of the pilot jobs were visually and subjectively evaluated by the pathologist and poorly performing workers were removed from the labor pool. The remaining workers were given additional feedback on their performance so they could improve. This process continued until the overall results seemed satisfactory. While the annotation of background paraffin enabled singular validation of the noise reduction techniques, the definite and probable (definite and likely) vasculature labels provided us with a strict and permissive approach to validating the segmentation algorithm. For the purposes of our data collection a "strict" interpretation included any area where there was visible dark india ink present while the permissive definition include potential lumen and areas that appeared to be poorly perfused vessel. A parameter sweep of different thresholding values and coefficients associated with binarization and regularization/false positive filtering was used to compare the isolation of vasculature against the permissive and strict data provided by the pathologist. We used a combination of Accuracy, Matthews Correlation Coefficient, and F1-score to determine the optimal parameter set for correspondence with professionally labeled ground truth, as demonstrated in figures 3 and 4.

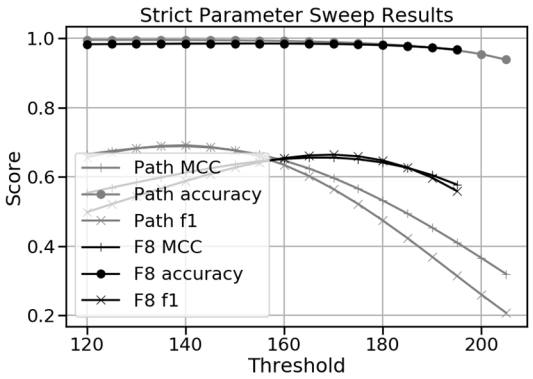

Figure 3: Performance of segmentation against pathologist and crowd labor data using the "strict" interpretation of vasculature.

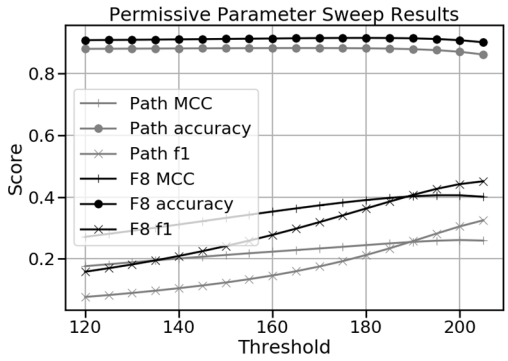

Figure 4: Performance of segmentation against pathologist and crowd labor data using the "permissive" interpretation of vasculature.

Validation Data Collection

In order to further validate the deployment of the automated segmentation system on a large sample, we needed to force multiply the amount of annotations provided by the pathologist with a sub-linear labor cost. We used the professional labels as examples to the vendor's crowdsourced annotation platform. The untrained, crowdsourced human workers used the professionally labeled examples to increase the sample size of ground truth data by an additional 16200 tiles (sections 1800 3 x 3 tile sections, or 16200 tiles). We employed a manual quality assurance process to ensure label fidelity in a manner similar to the ground truth data collection

The major discrepancy between the two data sets in the strict category seems to be due to a bias of the untrained labelers to be less strict than our pathologist; this is to say they were often less precise in their labeling and labels would often bleed into the vessel lumen and beyond as observed qualitatively. The lower overall scores in the permissive category appear to be due to very lightly colored small vessels, vessel periphery, and partial lumen in the serial sections. These areas are often similarly colored to the embedding media, and would explain the monotonic growth as threshold increases. Moreover, the increased number of lighter pixels in the permissive set seems to skew the results.

This validation data is available at http://www.3scan.com/technology/opendata/. Our intent with open sourcing these datasets in threefold. We believe the large amount of semantic segmentation data (both pathologist and crowd labor labeled) can be used to train deep learning models to detect vasculature and also further This validation data is available at evaluate the efficacy of this approach for microscopy data. Moreover, both the raw and processed brain vasculature data represent a significant baseline data upon which novel segmentation, 3D reconstruction techniques can be vetted.

RESULTS

To further investigate the soundness of our methods we chose to perform validation by replication, that is to say if our automated analysis was correct it should replicate some of the results of [1,2]. To this end we chose a a conservative threshold between our two validation sets of 155 and then applied our validation pipeline to four wild type and four HD model mice in rectangular prism section of the brain that approximated the mouse striatum. After automated processing, probability density functions of vessel density and vessel count were calculated across the cohort and are as shown in figures 5 and 6. These findings would recapitulate and further support the work in [1,2]. The difference and advantage of our automated analysis is that it can be rapidly applied over larger sections of region of interest. Hence we were able to analyze HD mice over 50 sections in z in the striatum as compared to 3 sections in z at random sampling points in striatum in [1, 2]. This data is publicly visible and explorable at the following URL http://www.3scan.com/technology/opendata/

CONCLUSION

Whole organ imaging and analysis via advanced microscopic methods like KESM facilitates large sample volumetric analysis of features within tissue, providing an ability to characterize vasculature as well as identify pathological states within those networks. Comparison of vascular features between control and test mouse brain reveals significant differences between the anatomic areas analyzed within target tissue.

The work presented here enables observations within large-volume samples, provides opportunities for computer vision-based quantification, and high-resolution analysis of vasculature. Identifying the characteristics of healthy vascular tissue will help in digitally-assisted diagnosis of pathological states, such as cardiovascular disease, diabetic complications, and vasculitides.

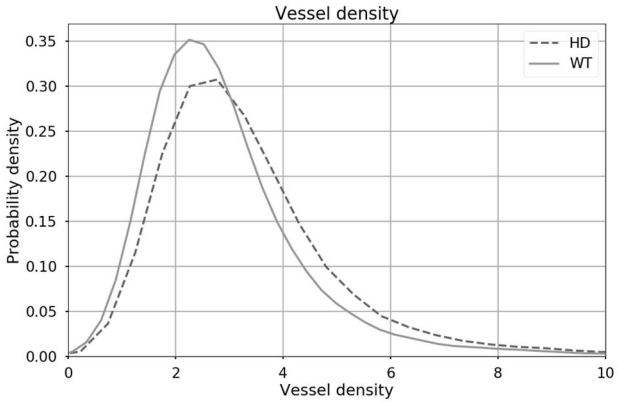

Figure 5: Graph showing the vessel density against the probability density in HD and WT mice. Taken as a population, HD mice show greater vessel density and higher density than WT mice.

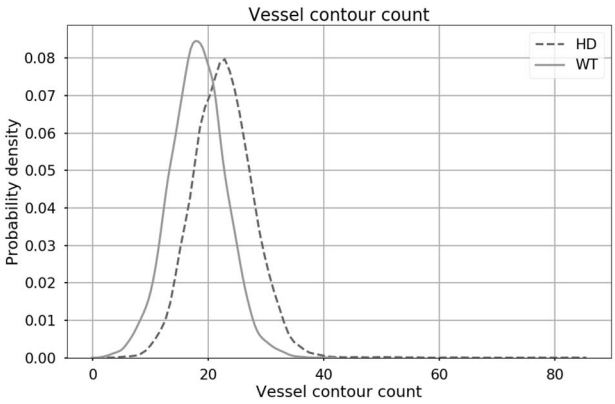

Figure 6: Graph showing the vessel contours count against the probability density in HD and WT mice. Taken as a population, HD mice show greater vessel contour count and higher density than WT mice.

References

1. Neurovascular abnormalities in humans and mice with Huntington'sdisease.Chien-Yuan Lin, Yi-Hua Hsu, Ming-Huang Lin, Tsai-Hui Yang, Hui-Mei Chen, Yu-Chen Chen, Han-Yun Hsiao, Chiao-Chi Chen, Yijuang Chern, Chen Chang Exp Neurol. 2013 Dec; 250: 20–30. Published online 2013 Sep 10. doi: 10.1016/j.expneurol.2013.08.019
2. Cerebrovascular and blood-brain barrier impairments in Huntington's disease: Potential implications for its pathophysiology. Janelle Drouin-Ouellet, Stephen J. Sawiak, Giulia Cisbani, Marie Lagacé, Wei-Li Kuan, Martine Saint-Pierre, Richard J. Dury, Wael Alata, Isabelle ReferencesSt-Amour, Sarah L. Mason, et al. Ann Neurol. 2015 Aug; 78(2): 160–177. Published online 2015 Apr 9. doi: 10.1002/ana.24406
3. Knife-Edge Scanning Microscopy for Bright-Field Multi-cubic Centimeter Microscopic Analysis of Microvasculature. Michael J. Pesavento, Caroline Miller, Katy Pelton, Molly Maloof, Corey E Monteith, Venkata Vemuri , and Megan Klimen, July 2017, doi:10.1017/S1551929517000645
4. Ali Can, Hong Shen, Turner, Tanenbaum, & Roysam. (1999). Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. IEEE Transactions on Information Technology in Biomedicine, 3(2), 125-138. doi:10.1109/4233.767088
5. Kanbay, M., Sanchez-Lozada, L., Franco, M., Madero, M., Solak, Y., Rodriguez-Iturbe, B., Johnson,R.J.(2010). Microvascular disease and its role in the brain and cardiovascular system: a potential role for uric acid as a cardiorenal toxin. Nephrology Dialysis Transplantation, 26(2), 430-437. doi:10.1093/ndt/gfq635
6. Brown, W. R., & Thore, C.R. (2011). Review: Cerebral microvascular pathology in aging and neurodegeneration. Neuropathology and Applied Neurobiology, 37(1), 56-74. doi:10.1111/j.1365-2990.2010.01139.x
7. Nishida, N., Yano, H., Nishida, T., Kamura, T., & Kojiro, M. (2006). Angiogenesis in cancer. Vascular Health and Risk Management, 2(3), 213-219. doi:10.2147/vhrm.2006.2.3.213