

DISSERTATION  
DEFENSE

*Scalable Computational pathology: From Interactive to Deep Learning*

Michael Nalisnik  
Emory University

**Abstract:** Advances in microscopy imaging and genomics have created an explosion of patient data in the pathology domain. Whole-slide images of histologic sections contain rich information describing the diverse cellular elements of tissue microenvironments. These images capture, in high resolution, the visual cues that have been the basis of pathologic diagnosis for over a century. Each whole-slide image contains billions of pixels and up to a million or more microanatomic objects whose appearances hold important prognostic information. Combining this information with genomic and clinical data provides insight into disease biology and patient outcomes. Yet, due to the size and complexity of the data, the software tools needed to allow scientists and clinicians to extract insight from these resources are non-existent or limited. Additionally, current methods utilizing humans is highly subjective and not repeatable. This work aims to address these shortcomings with a set of open-source computational pathology tools which aim to provide scalable, objective and repeatable classification of histologic entities such as cell nuclei.

We first present a comprehensive interactive machine learning framework for assembling training sets for the classification of histologic objects within whole-slide images. The system provides a complete infrastructure capable of managing the terabytes worth of images, object features, annotations and metadata in real-time. Active learning algorithms are employed to allow the user and system to work together in an intuitive manner, allowing the efficient selection of samples from unlabeled pools of objects numbering in the hundreds of millions. We demonstrate how the system can be used to phenotype microvascular structures in gliomas to predict survival, and to explore the molecular pathways associated with these phenotypes. Quantitative metrics are developed to describe these structures.

We also present a scalable, high-throughput, deep convolutional learning framework for the classification of histologic objects is presented. Due to its use of representation learning, the framework does not require the images to be segmented, instead learning optimal task-specific features in an unbiased manner. Addressing scalability, the graph-based, parallel architecture of the framework allows for the processing of large whole-slide image archives consisting of hundreds of slides and hundreds of millions of histologic objects. We explore the efficacy of various deep convolutional network architectures and demonstrate the system's capabilities classifying cell nuclei in lower grade gliomas.

Thursday, March 30, 2017, 10:00 am  
Mathematics and Science Center: W306

Advisor: Lee Cooper

MATHEMATICS AND COMPUTER SCIENCE  
EMORY UNIVERSITY