Abstract:

Improvements in medical image acquisition technologies make ever larger image volumes available for diagnostic analysis. Detection of relevant visual patterns in medical images calls for training supervised predictors on these large image sets. However, data annotation capabilities do not keep pace with data generation, especially in areas such as cancer diagnosis where only experienced domain experts can provide the requisite information. My research addresses this annotation scarcity problem of the present big data age from diverse perspectives, such as multiple instance learning, transfer learning, and active learning. In this talk, I will present the new Bayesian models I have developed to facilitate annotation of various medical image modalities. I will demonstrate:

i) how Gaussian Processes can accurately detect live cell events from few annotations, requiring less than 3% of a training set to be annotated,

ii) how they can effectively model big data (> 1M data points) based only on weak labels,

iii) how they can be assembled into deep network architectures that achieve unprecedented accuracy on certain hard tasks, such as survival time prediction of Barrett’s cancer patients from histopathological tissue images.

The sharp increase in available data volumes has recently led to a paradigm shift in machine learning research. We see thrilling outcomes of this shift in computer vision, finance, and recommender systems. I will conclude my talk by sketching my ideas on how big data modeling can shape the future of science and how the Bayesian approach can play a key role towards this target.

Host: Yongluan Zhou