

Title: Active and Transfer Learning Methods for Computational Histology

Tissue micro-environments of critical interest like tumors, stem-cell niches, and brain tissue surrounding implanted neuroprosthetic devices are complex in structure and harbor complex processes. Understanding events and perturbations that occur in these micro-environments entails selective molecular imaging of the tissues, delineating cellular structures, and accurate cell classification. The algorithm presented in this thesis advances the state of the art in cell classification in large scale histological studies.

The core contribution of this thesis is a novel active machine learning algorithm that leverages the advances made in the fields of optimal experimental design and submodular functions. In large and diverse datasets, manually annotating examples to create a training set is effort intensive and suboptimal due to subjectivity and selection bias introduced by human experts. The proposed algorithm reduces human effort and eliminates subjectivity by actively participating in the learning process to select informative examples for the user to label. The algorithm selects multiple informative examples in a learning iteration reducing the burden of retraining the classifier multiple times. The algorithm relies on the submodularity property of the D-optimal criterion to provide performance guarantees for the examples selected for labeling. The algorithm also obviates the necessity for performing offline analysis for feature selection by using the popular LASSO technique to perform feature selection during training. Our experiments on multiple real world data from clinical studies show that the proposed active learning algorithm outperforms standard learning and other active learning frameworks.

Since histological studies involve analysis of similar cells under different conditions, the labeling effort to classify similar or related cells in different tissues or conditions can further be reduced by leveraging knowledge learned from one classification task and using it for a related task. The proposed algorithm is also extended to a *transfer learning* setting to take advantage of existing labeled data sets even when they are mismatched. When applied in transfer learning mode to endothelial cell classification problems, the algorithm consistently achieves classification accuracies greater than 90% with minimal effort. Both the algorithms have been embedded into the open source FARSIGHT toolkit with an intuitive graphical user interface that provides constant feedback about the classification process to the user.

Committee Chair: Dr. Badrinath Roysam

Place: ECE Conference Room

Committee Members: Dr. Thomas Hebert

Date : September 27th, 2012

Dr. Saurabh Prasad

Time: 10AM - 11AM

Dr. Demetrio Labate

Dr. William Lee