Active Learning for Pathology Image Analysis

Lee A.D. Cooper, Christina Appin, Rami Yacoub, David A. Gutman, Hyun Ju Choi, Jun Kong, Fusheng Wang, Carlos S. Moreno, Robin Bostick, Daniel J. Brat, Joel H. Saltz
Department of Biomedical Informatics, Emory University School of Medicine
Center for Comprehensive Informatics, Emory University
Introduction

• Classification in pathology imaging
  – IHC positivity, localization, morphology (mitoses)

• Traditional approaches to develop classifiers

• Active Learning - a more natural approach to classification

• Engage a broader audience in image analysis tasks

• Eventual goal to approach more difficult applications
Classification: Identify, Measure and Learn

Identify - Segmentation

Learn - Classification

Measure - Features

Size, Shape, Texture, Expression, Coexpression
Classification: Identify, Measure and Learn

Identify - Segmentation

Learn - Classification

Measure - Features

Size, Shape, Texture, Expression, Coexpression
Traditional Approaches: “The Mechanic”

- Used in many commercial packages
- Users study classifier mechanism and role and effect of parameters
- Tune parameters until desired results are achieved
- Steep learning curve
Traditional Approaches: User-Driven Sampling

- Create a set of training examples a priori

- Classifier learns training data - hope for generality

- Sampling biases
  - Users select obvious and avoid ambiguous examples
Active Learning: Iterative Feedback

• Interactively query user to label examples
  – Classifier: “I think this is a _____, am I correct?”

• Examples are selected intelligently
  – Force user to make hard decisions
  – Select examples based on a confidence score
  – Ambiguous examples help carefully define boundary

• Common in general applications like spam filtering

• Limited use in pathology imaging so far (GENIE)
The Active Learning Cycle

Processed Whole Slide → Select Examples

User Review

Yes → No

Update Classifier

Classify Whole Slide, Repeat
Our Software

- Prototype to investigate active learning for cell classification applications

- Binary classification problems

- Provides interactive review with functions for training, refresh, validation and export

- Initial examples selected randomly*
Prototype

Nuclear Positive (+)  Nuclear Negative (-)
Prototype
Prototype

<table>
<thead>
<tr>
<th>Nuclear Positive (+)</th>
<th>Nuclear Negative (-)</th>
</tr>
</thead>
</table>

Accuracy: 100 %
Experiments & Results

• Glioblastoma sections stained for mTOR, pRB, Ki67
  – Segment and extract 76 protein expression features per cell
  – Logitboost classifier, validated on 500 random samples

• Experiment 1: single positive stain
  – 99% accuracy
  – 300 cells reviewed in 4 iterations
  – 76 total corrections, 73 in first iteration

• Experiment 2: multiple positive stains
  – No convergence after 600 reviews
  – Rare phenotypes problematic for random initialization
Conclusions & Future Work

• Active learning obscures classifier mechanism, avoids bias of user-driven sampling

• Need to investigate more complex applications
  – Multiple classes
  – Classifier generality across multiple images

• Integration with Emory whole slide viewer
  – cancer.digitalslidearchive.net
  – Supports “kickstart” initialization
  – Provide context needed for judgments
Thank You!