Localization of Diagnostically Relevant Regions of Interest in Whole Slide Images

Ezgi Mercan$^1$, Selim Aksoy$^2$, Linda G. Shapiro$^1$, Donald L. Weaver$^3$, Tad Brunye$^4$, Joann G. Elmore$^5$

$^1$Dept. of Computer Science and Engineering, University of Washington, Seattle, WA
$^2$Dept. of Computer Engineering, Bilkent University, Ankara, Turkey
$^3$Dept. of Pathology, University of Vermont, Burlington, VT
$^4$Dept. of Psychology, Tufts University, Medford, MA
$^5$Dept. of Medicine, University of Washington, Seattle, WA
Outline

1. Motivation

2. Methodology
   1. Learning a Visual Dictionary
   2. Training using expert ROIs
   3. Testing using tracking data
      • Interpretation of tracking data and computation of evaluation data set

3. Experiments and Results
Bigger Picture

• Our Digipath grant from the U.S. National Cancer Institute is to study the scanning behaviors of pathologists while they diagnose breast biopsy slides.

• We use a mouse tracking system to record their panning and zooming behaviors.

• They also draw a rectangle on the one area of each slide they think was most important to diagnosis.

• The long-term goal of our work is understand what pathologists are doing when they diagnose both correctly and incorrectly.
Motivation

• Pathologists study whole slide biopsy images in order to make a diagnosis.

• They stop at and/or zoom on regions of interest they think are important for diagnosis.

• We would like to understand and partially automate this process using the rectangles they draw and the mouse tracking system.

![our predicted probability map](image1)

![ground truth from viewport logs](image2)
Localization of Regions of Interest (ROIs) in Whole Slide Images

• Goal: To predict regions that attract pathologists’ attention in whole slide images.

• Method:
  • Extract diagnostically relevant ROIs/positive samples
  • Model ROIs using image features and bag-of-words
  • Train a classifier to detect ROIs in a new image
  • Test on images already viewed by pathologists
1. Learning Visual Dictionary

- Each 120x120 pixel patch is a *word*.
- Words are obtained from ROIs marked by experts.
- We calculated **color** (L*a*b*) and **texture** (LBP of H&E channels) **histograms** from RGB images for each word.
- Using k-means clustering we obtained a visual dictionary.
2. Training: Sliding Window

- We used a sliding window approach to cut 3600x3600 pixel overlapping windows (*bags*).
- Each bag contains 30x30=900 words and is represented as a histogram of k words from the visual dictionary.
- **Bags in ROIs marked by experts are positive examples** for diagnostically relevant regions. **Bags outside the expert ROIs are negative examples.**
3. Testing

whole slide image

3600x3600 pixel sliding windows

probability of being a ROI

visual dictionary

bag of words for each sliding window

Binary Classifier (LogReg and SVM)

trained on expert ROIs
Evaluation Data Set

• Each expert pathologist was allowed to draw one rectangle on the digital whole slide image after interpreting and diagnosing the case.

• However, each whole slide contains more than one diagnostically relevant area — either confirming or refuting the final diagnosis.

• We hypothesize that these areas attract the attention of pathologist during interpretation and can be calculated from viewing logs.
Viewport Log Analysis

• *Where did pathologists really look?*

• **viewport**: a rectangular part of the image the pathologist sees on the screen

• **viewport logs**: a new entry when viewport changes
  
  • **panning**: changes the location of viewport rectangle
  
  • **zooming in/out**: changes the size of viewport rectangle
Zoom Level

The value of the magnification per each log entry during an image interpretation session. It ranges from 1 to 60. Higher the zoom, the smaller the rectangle.
Displacement

The distance between the locations of viewport rectangles from two consecutive log entries. It’s measured in pixels over the actual whole slide image.
Duration

The amount of time a pathologist spent looking at each rectangle that corresponds to a log entry. A log entry is created whenever the viewport changes.

8/26/2014
Zoom Peaks:
Local maxima in zoom levels.

\[ \text{zoom}(i) > \text{zoom}(i + 1) \]
\[ \text{zoom}(i) > \text{zoom}(i - 1) \]
Slow Pannings:
Constant zoom, small changes in displacement.
Fixations:
Large amount of time spent on a rectangle.

Fixation:
\[ \text{duration}(i) > 2 \text{ sec} \]
Union of three defines diagnostically important areas
Experiments and Results

• 20 high resolution breast pathology images from 5 diagnostic categories: Non-proliferative changes, proliferative changes, (ADH) atypical ductal hyperplasia, (DCIS) ductal carcinoma in-situ, and invasive cancer.

• Each of 3 experts provided 1 hand marked ROI and 1 viewport log per image.

• 10-fold cross-validation experiments with L1- regularized logistic regression and SVM. The differences between Logistic Regression and SVM are statistically significant.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Dictionary Size</th>
<th>L<em>a</em>b* Color Features</th>
<th>LBP Texture Features</th>
<th>LBP + L<em>a</em>b* Color + Texture Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogReg</td>
<td>K=50</td>
<td>73.70</td>
<td>72.80</td>
<td>73.13</td>
</tr>
<tr>
<td></td>
<td>K=100</td>
<td>73.38</td>
<td>75.52</td>
<td>75.01</td>
</tr>
<tr>
<td>SVM</td>
<td>K=50</td>
<td>78.73</td>
<td>79.60</td>
<td>79.16</td>
</tr>
<tr>
<td></td>
<td>K=100</td>
<td>78.57</td>
<td>78.88</td>
<td>78.63</td>
</tr>
</tbody>
</table>
Evaluation

- Probabilities predicted by SVM using **color**
- Probabilities predicted by SVM using **texture**
- Probabilities predicted by SVM using **color + texture**
- **Ground truth** from viewport analysis
Evaluation

ground truth from viewport analysis

probabilities predicted by SVM using texture
Discussion

• A method to predict regions of interest in whole slide images:
  • learns a visual dictionary of different patches
  • models the regions as the frequencies of visual words

• 80% overlap between predictions and the actual regions pathologists look at
Discussion

• **In training**, hand-marked ROIs are used because the actual *relevant regions* retrieved from viewport logs include a lot of background.

• **In testing**, ground truth consists of everything pathologists see on the screen, including background around edges, decreasing accuracy.
Future Work

• Additional features based on superpixels.
• Add eye-tracking to the mouse tracking data.
• Apply multi-instance learning using multiple ROIs with different labels from the same image.
• Construct graphs from the tracking data that summarize how each pathologist scans each slide.
• Study data from 100 community pathologists in addition to the current 3 expert pathologists.

This work was supported by the National Cancer Institute under Grant No. R01-CA172343 and Grant No. K05-CA104699. S. Aksoy was supported by a Fulbright Visiting Scholar grant.