Spectral Clustering for automated segmentation

Silvia Miramontes[1,2] and Daniela Ushizima[2,3]

[1]Department of Mathematics, University of California, Berkeley, CA

[2]Berkeley Institute of Data Science, University of California, Berkeley, CA

[3]Lawrence Berkeley National Laboratory, Berkeley, CA

Abstract

While object detection has impacted several industries, much remains unanswered in cancer cell detection using Pap smear microscopy. This project focuses on exploring **machine learning algorithms to detect blobs** and crop ROI before scrutinizing cells for abnormal traits. Using digital images of cervical cells, our goal is to determine the location of nucleus, which exhibits a range of purple staining intensities. After non-local means preprocessing of 2D color micrographs, we ran the Felsenszwalb's [1]. efficient **graph-based image segmentation**, which produced an over-segmentation of a RGB image using a fast, minimum spanning tree based clustering on the image grid. We also inspected the blob morphology.

Why Cell Detection?

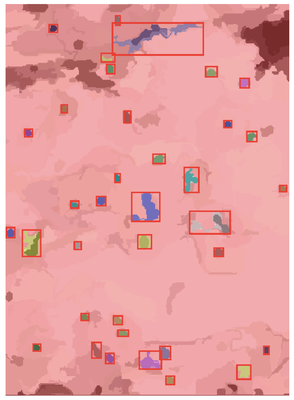


Fig 1. Cell Detection

From the available 15,000 fields of Pap smear slides coverage area, less than a hundred is often inspected for normal and abnormal cells. Our **multiscale region proposal** aims to reduce data and detect blobs to support **cytopathology analytics**.

Cervical Cancer

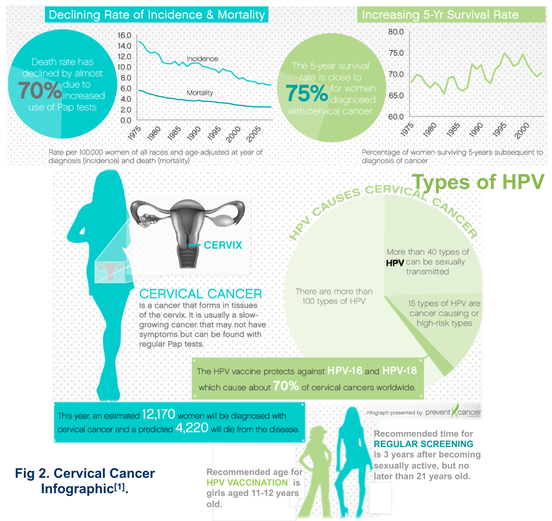


Fig 2. Cervical Cancer Infographic[1].

Cell Detection Results

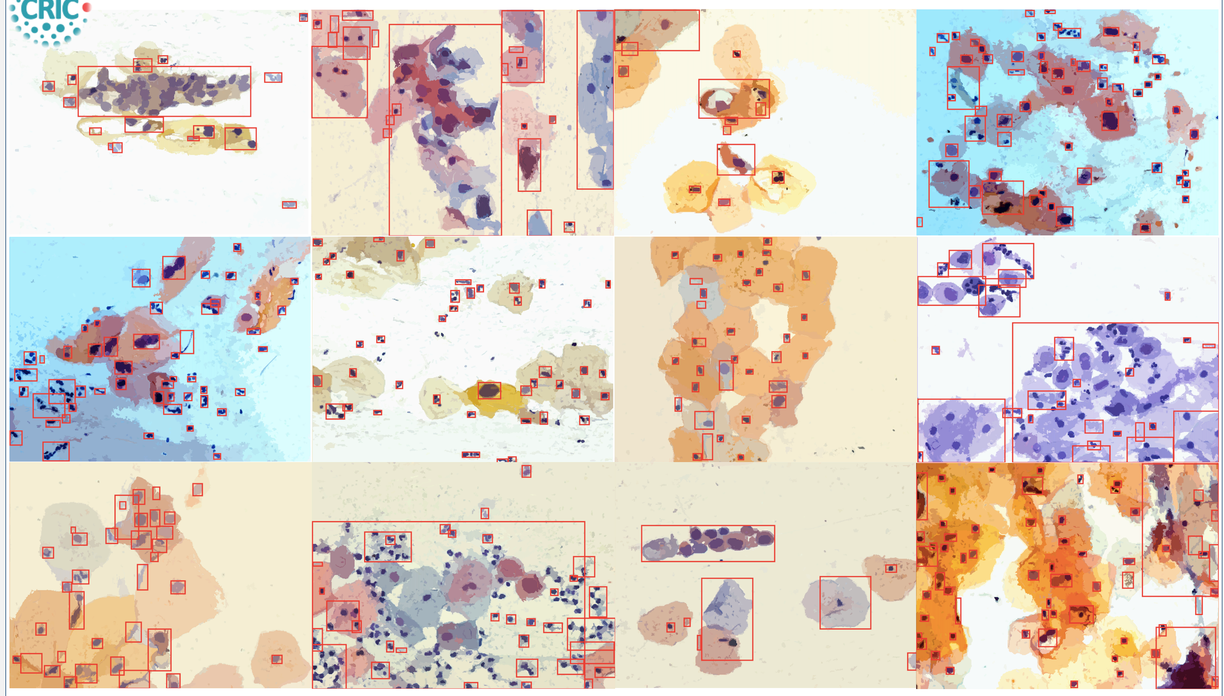
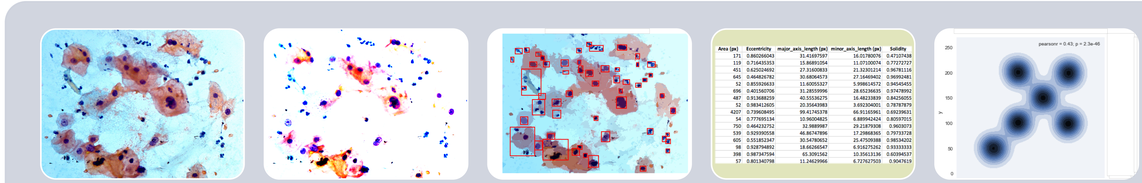


Fig 3. Multiscale region proposal (MRP) from CRIC digital Pap Smear images.

Methodology



Digital Pap Smear
- Raw image: 40X
- Ground truth
- Priors on morphology and staining

Preprocessing
- Colormaps inspection
- Non-linear filters
- Non-local means

Region Proposal
- Graph-based segmentation
- Superpixel
- Isodata
- Labeling

Metrology
- Refinement by priors
- Nuclei detection
- Shape analytics: area, solidity, eccentricity, major axis, etc.

Data Mining
- Comparative evaluation: experimental data vs. expert manual curation
- Clustering
- Visualization

Fig.4. Our method begins with the preprocessing of the CRIC images to later apply unsupervised machine learning such as Felzenszawlb's graph-based algorithm to obtain ROIs where a cancer cell may be located. We finalize this process by storing morphological metrics about detected nucleus candidates. Preliminary ground-truth evaluation took place through the mining of manually curated nucleus, and comparison with MRP.

Results

According to ground-truth, morphological analysis of blobs from digital Pap smear using MRP showed encouraging results, but will need to be combined to textural descriptors to eliminate regions of overstaining (Fig.3 right bottom)

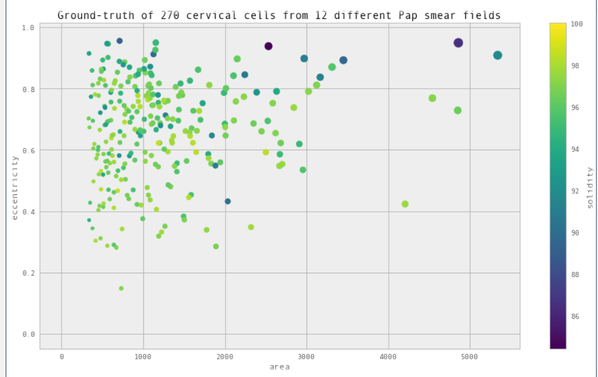


Fig 5. Morphological Analysis of nuclei ground-truth

Conclusions and Future Work

- We introduced a multiscale region proposal algorithm for blob detection after work this summer of 2018;
- Next, we will compare experimental results with ground-truth provided by 3 pathologists;
- We will investigate mechanisms to construct emulators of **self-driving microscopes** by designing methods that can deliver the most **relevant** information to domain scientists:
  - Regions of interest associated to science domain;
  - Dashboard with key information about ROIs;
  - Metrics and relative comparisons.

Acknowledgements

Special thanks to the Brazilian Center for Recognition and Inspection of Cells (**CRIC**) for donating a small subset of digital Pap smear images, so that this project was possible. Also to Diana Lizarraga, the UC Leads program, and BIDS CerviCal project.

Citations

[1] Efficient graph-based image segmentation, Felzenszwalb, P.F. and Huttenlocher, D.P. International Journal of Computer Vision, 2004.

[2] Figure 2: Infograph on Cervical Cancer: "Foundation Debuts Cervical Cancer Prevention Infographic." Infographic." Prevent Cancer Foundation Community Grants Comments.