Group meeting

Camelyon 2016

Naylor Peter

08 March 2016

Outline

Camelyon16

ISBI-2016

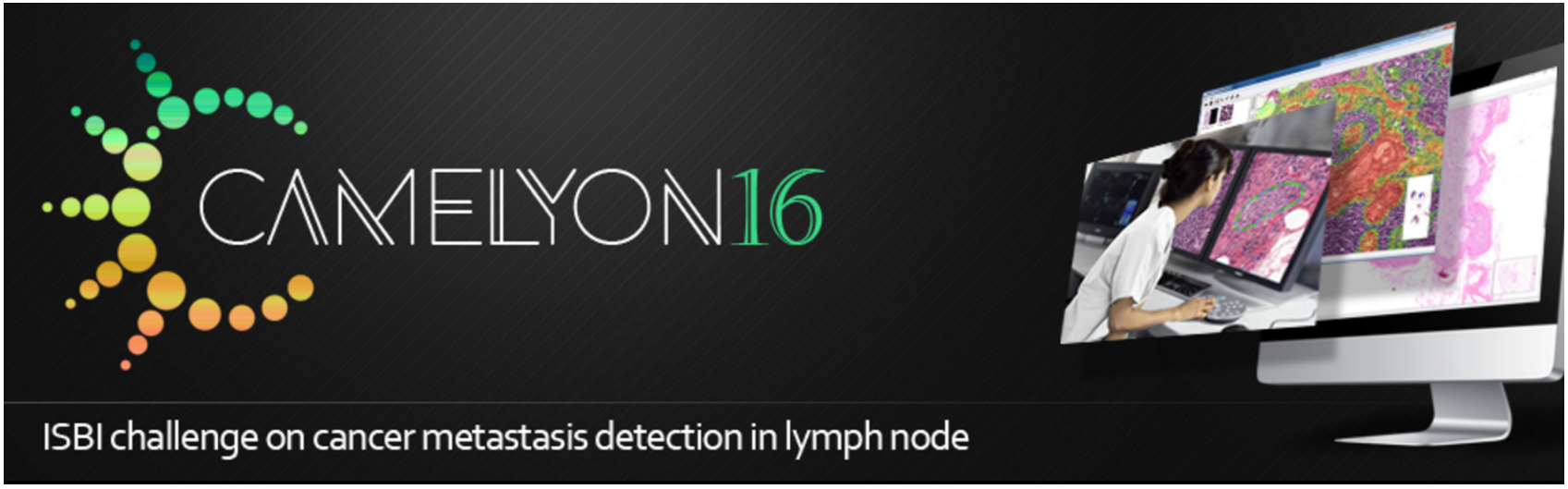


Figure : Official logo

Challenge apart of ISBI 2016.

Important date:

- 1st April 2016, submission deadline.
- 13th to 16th April: ISBI 2016.

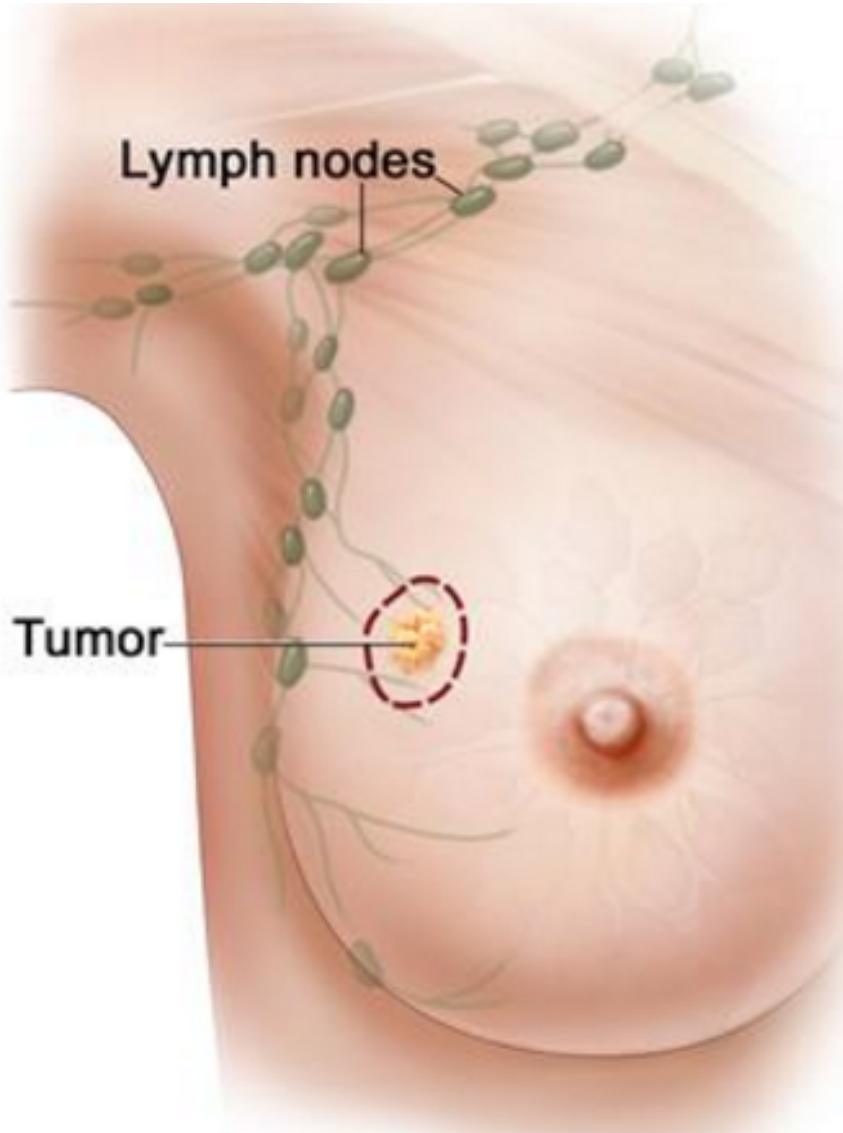
Camelyon16

Goal:

- Detection of micro- and macro-metastases in lymph node digitized images.
- Automated detection of metatases in hematoxylin and eosin (H&E) stained whole-slide images of lymph node sections.

Motivation:

- Lymph node metastases occur in most cancer types.
- Lymph nodes in the underarm are the first place cancer is likely to spread.
- The prognosis is poorer when cancer has spread there.



Camelyon16

Goal:

- Detection of micro- and macro-metastases in lymph node digitized images.
- Automated detection of metastases in hematoxylin and eosin (H&E) stained whole-slide images of lymph node sections.

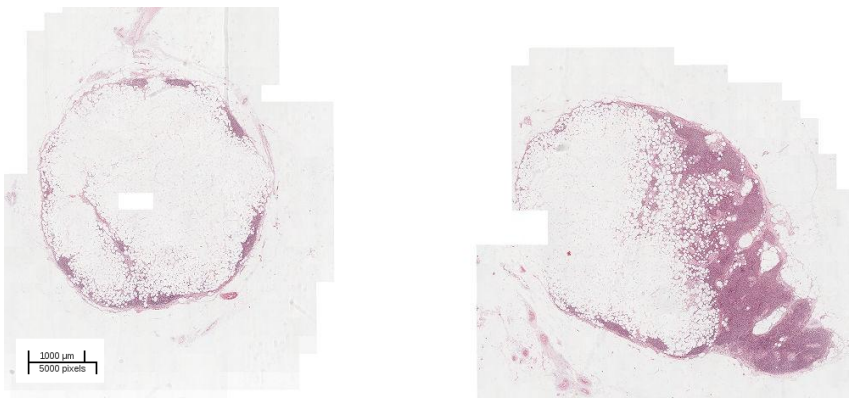


Figure : Normal 38

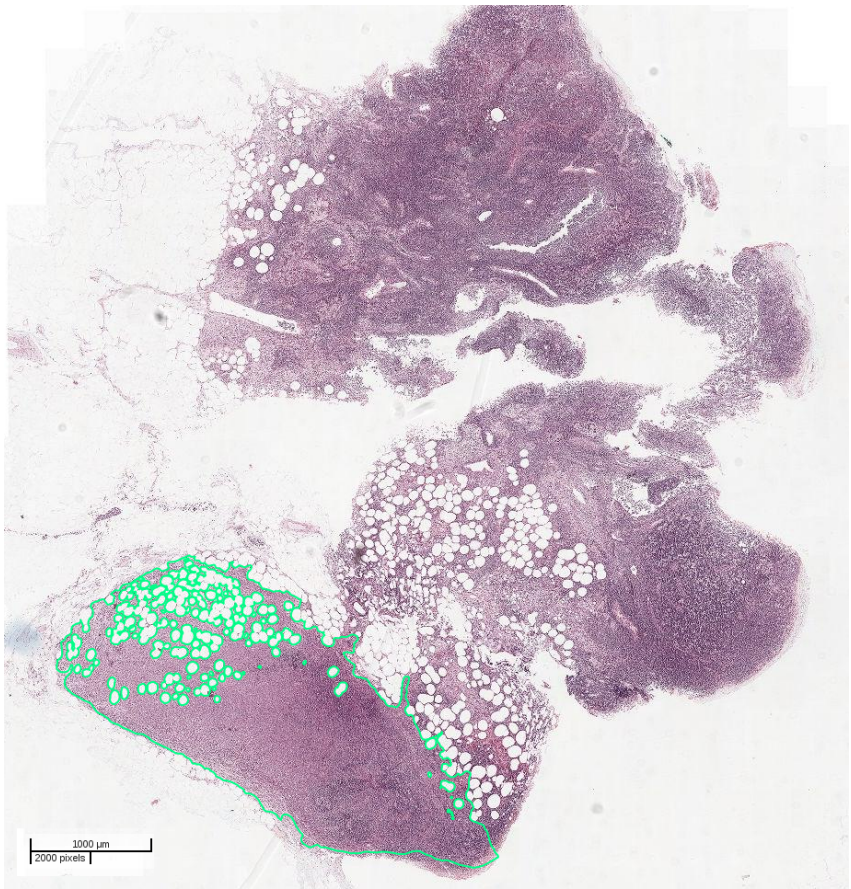


Figure : Tumor 34

Metastases detected in green

Evaluation

Two evaluation metrics and two leader boards.

Slides based:

- Binary classification of whether or not a slide contains metastases.
- Evaluating with the area under the ROC curve.

Region based:

- Correctly detecting metastases within slides.
- Evaluation with the FROC curve (free-response receiver operating characteristic)

Outline

The data set

Data set provided:

- 160 Normal slides.
- 110 Tumor slides.
  - Huge images with very high precisions, using c++ library openslide.
  - One image compressed with JPEG2000: $\sim$ 2/3Gb.
  - Uncompressed at approx. 10Gb.
  - Highest resolution : $\sim 96256 \times 218624$.
  - Lowest resolution : $\sim 188 \times 427$.

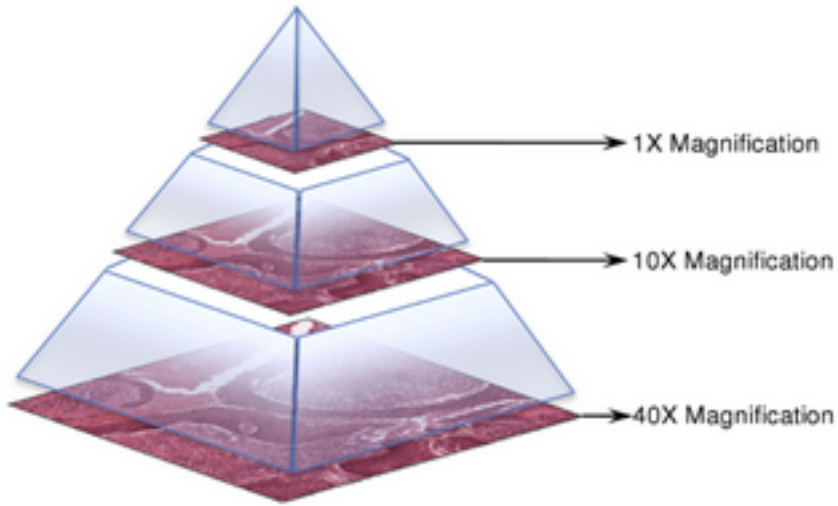


Figure : Pyramid data structure

Between 8 and 10 different resolutions

Example

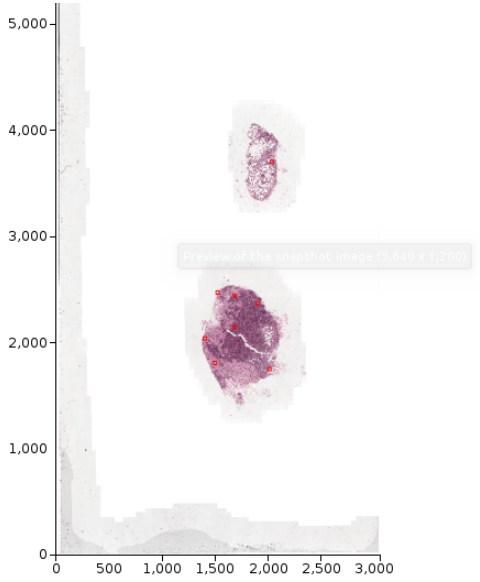


Figure : Tumor 31

7000 x 3500, resolution 6

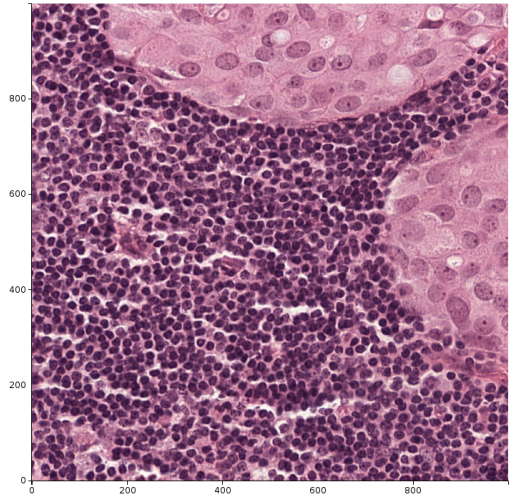


Figure : Sub-image of Tumor 31, highest resolution

1000 x 1000, resolution 0

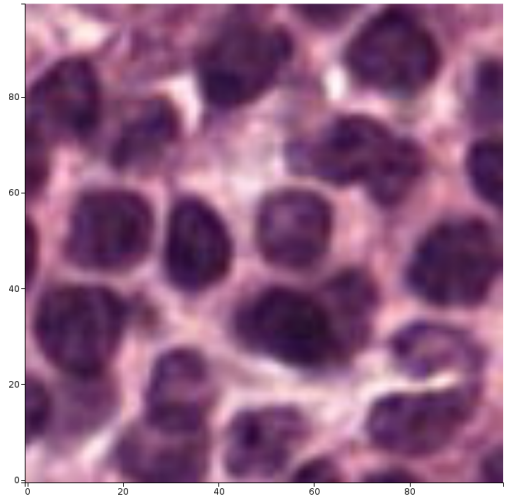


Figure : Sub-image of Tumor 31, highest resolution (zoom)

100 x 100, resolution 0

Outline

Images

Very large number of samples:

- In pixel classification, each pixel is an instance, $n \gg p$.
- Appropriate subsampling methods. In particular here, first subsampling for the "sub-images". Second subsampling given a particular "sub-image".

Outline

Summary

workflow.png

Outline

Subsampling

Subsampling 1 (region of interest detection over slides) Trying to find interesting part of the image.
We can divide sub-images in 4 groups.

- Only metastasis tissue.
- Only normal tissue.
- Centered on the boarder metastasis tissue/normal tissue.
- Centered on the boarder tissue/background.

Subsampling

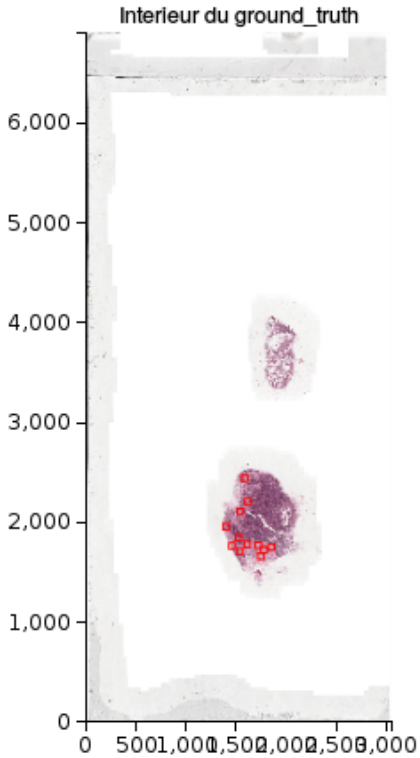


Figure : Only metastasis regions

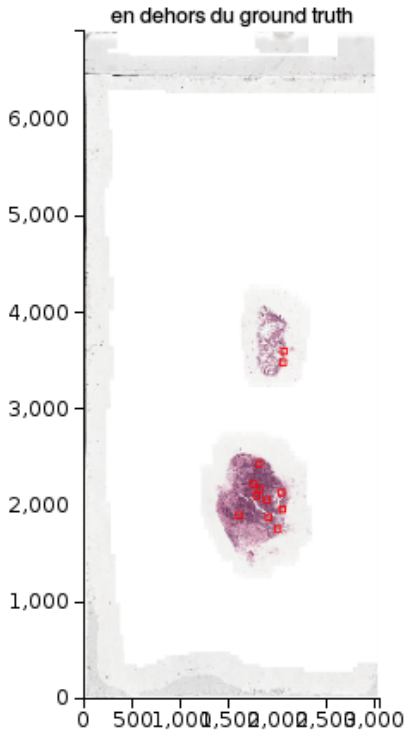


Figure : Only metastasis-free regions

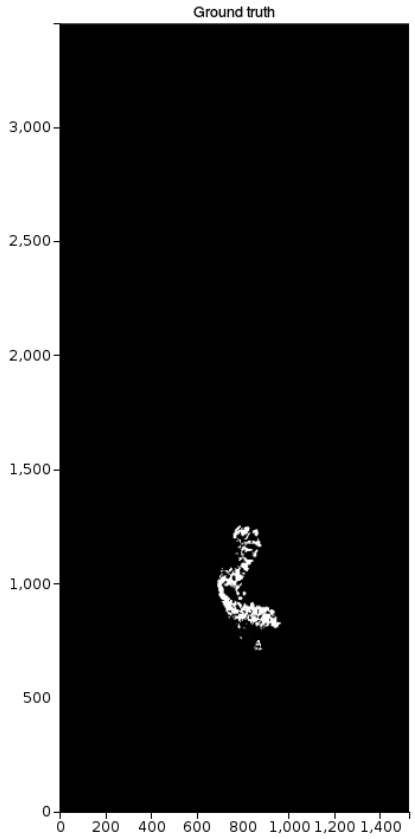


Figure : Ground truth

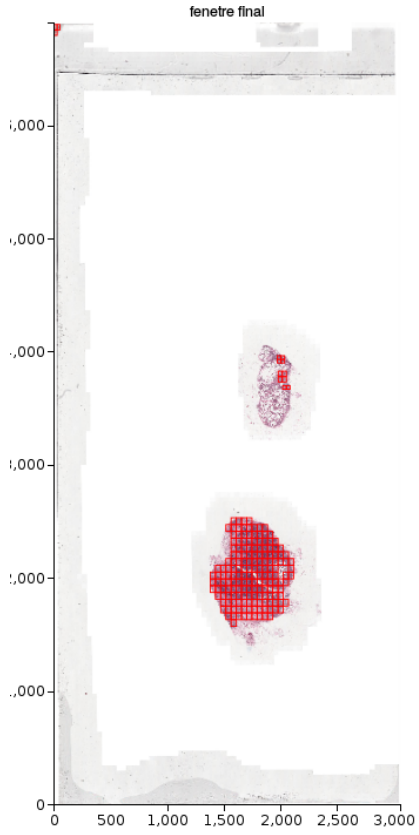


Figure : Grid partition

Outline

Ilastik

Ilastik : software for interactive image classification, segmentation and analysis.
Using features from ilastik (implemented in vigra, c++ library)

- Color/Intensity:
  - Gaussian Smoothing
- Edge:
  - Laplacian of Gaussian
  - Difference of Gaussians
  - Gaussian Gradiant Magnitude
- Texture:
  - Structure Tensor Eigenvalues
  - Hessian of Gaussian Eigenvalues

Figure : Ilastik

Ilastik - examples

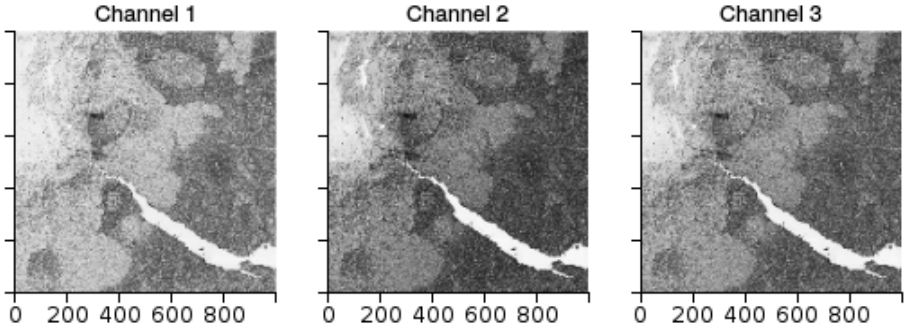


Figure : Gaussian smoothing, $\sigma = 0.7$

Figure : Ilastik

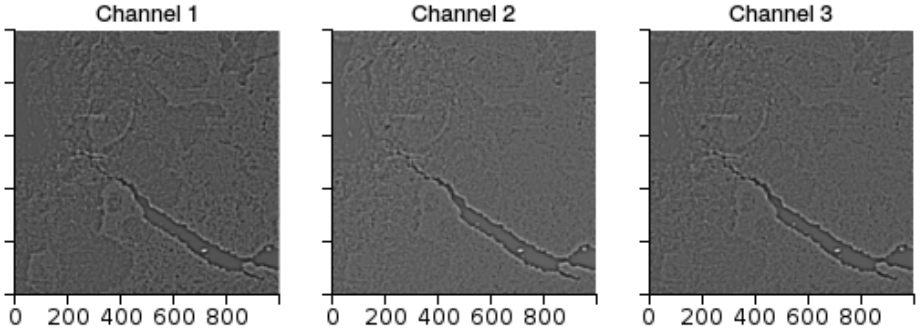


Figure : Laplacian of Gaussian, $\sigma = 5$

Outline

Superpixels/Waterpixels

Superpixels: Regions resulting from a low-level segmentation.
They have these proprieties: homogeneity, connected partitions, adherence to object boundaries, regularity.

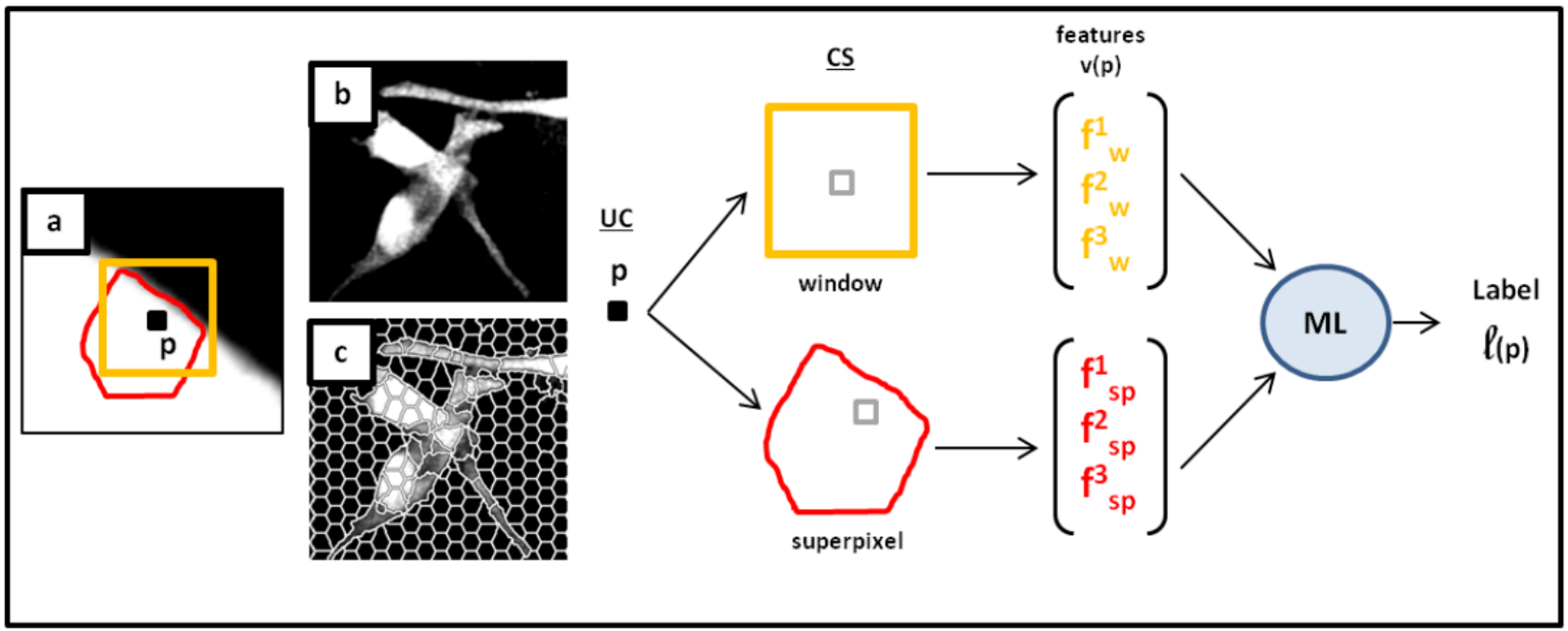


Figure : Illustration of how superpixels are used

Paper: *Waterpixels*, V. Machairas, M. Faessel, D. Cardenas-Pena, T. Chabardes, T. Walter and E. Decencire.

Waterpixels - examples

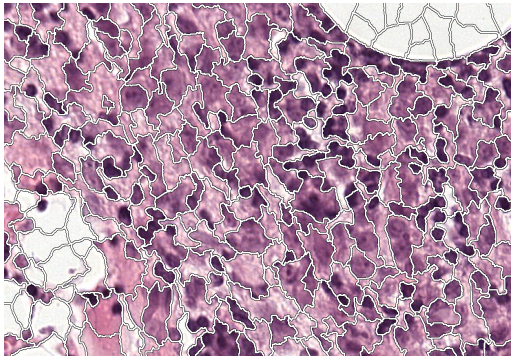


Figure : Waterpixels at resolution 0

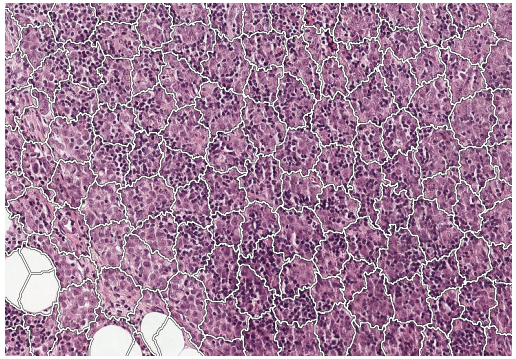


Figure : Waterpixels at resolution 2

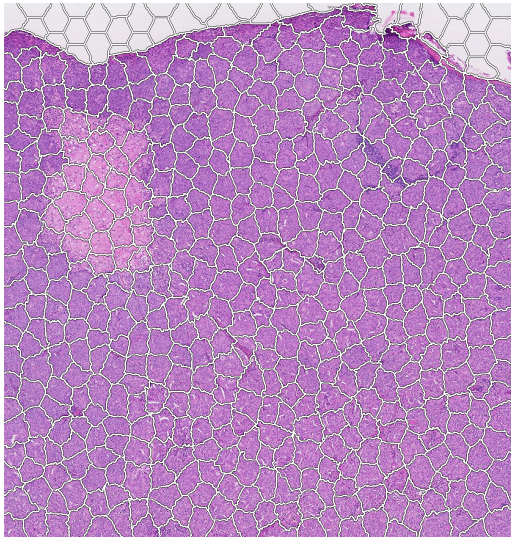


Figure : Waterpixels at resolution 4

Outline

Subsampling 2

Pixel classification.

Still very large data-set.

260 slides $\rightarrow$ 100 sub-images per slides $\rightarrow$ 1 000 000 pixels per sub-images.

Need for a second resampling.

Resampling randomly? smartly, via unsupervised methods?

Outline

Pixel based classifier

Finding metastasis regions:
Each pixel, will be considered metastasis if it belongs to a metastasis region.
Model: random forest.
Detecting metastasis patients: Each pixel is or not a metastasis but it also belongs to bigger group. This larger group, the slide, is annotated.
Model: Multiple instance learning random forest.
Cross-validation/evaluation:
One slide out scheme.

Multiple instance learning

Notations:

- Pixels are a pair $(x_i, y_i) \in \mathbb{R}^d \times \{-1, +1\}$.
- Slides are bags of pixels: $B_I = \{x_i, i \in I\}$, and we have $Y_I = 1$ if there is at least one x_i in B_I that is positive, otherwise $Y_I = -1$.
- Constraint to add to an optimization problems:

$$\sum \frac{y_i + 1}{2} \geqslant 1, \forall I \ s.t \ Y_I = 1 \text{ and } y_i = -1, \forall I \ s.t \ Y_I = -1$$

Implementation : MISVM, Multiple-Instance Support Vector Machines by Gary Doran. Python package.
Paper: *Support Vector Machines for Multiple Instance learning*, S. Andrews, I. Tsochantaridis and T. Hofmann.