Hands-on Workshop on Machine Learning Applied to Medical Imaging

Machine learning for histopathological images analysis

Germain Forestier[†], Jonathan Weber[†], Cédric Wemmert[⋆]

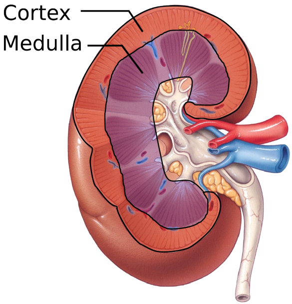
[†] IRIMAS Mulhouse, [⋆] ICube Strasbourg

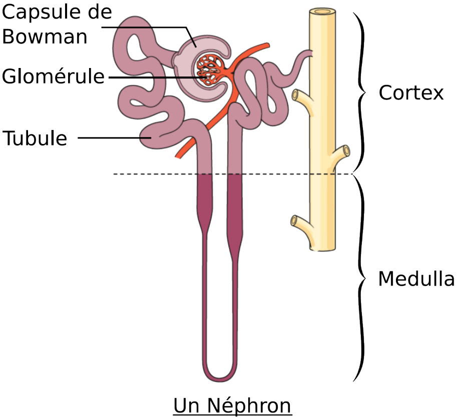
Outline

1. Histopathological images
2. Machine learning for image analysis
3. Hands-on

Histopathological images

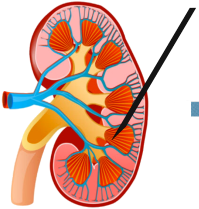
An example: the kidney



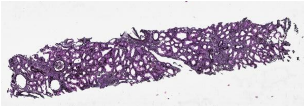


Histological slide

Biopsy



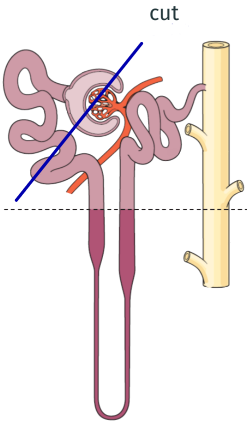
Stained biological slide

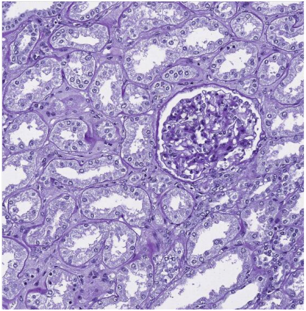


Scanned Whole Slide Image

- Many slides are consecutively cut from a piece of tissue
- A different staining is applied on each to highlight different types of cells or structures
- Very high resolution image (1 pixel = 0.25μm)

Histological slide

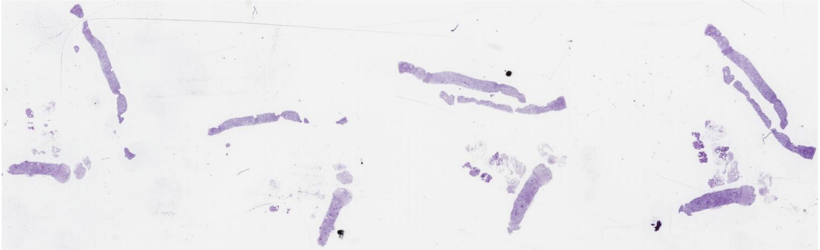




Microscope observation after staining

Histological slide

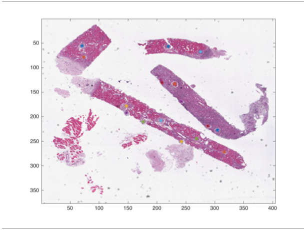
Example of consecutive slides

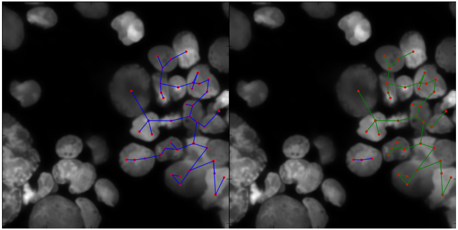


Machine learning and histopathology

Challenges

- Huge need of labeled data
- Size of the images
- High heterogeneity due to tissue preparation
- Clusters of objects of interest
- Cost of digitisation (financial and time)
- Explicability

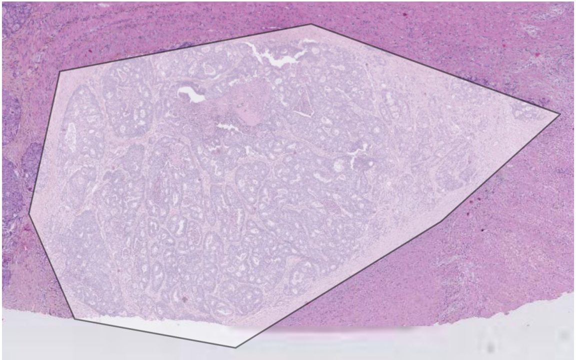




Challenges

Supervised approaches need huge amount of labeled data

- very time consuming to have high quality annotations



Challenges

Supervised approaches need huge amount of labeled data

- very time consuming to have high quality annotations
- privacy protection: medical data are sensitive and cannot be exchanged easily
- scanners are very expansive
- unbalanced classes: for many tasks (mitosis detection for example) negative labels (*no mitosis*) are much more frequent than positive labels (*mitosis*)

Challenges

Data volume

- very large and high resolution images (*e.g.* $100,000 \times 100,000$ pixels) that cannot be used entirely in memory
- resolution reduction (loss of details)
- work on extract or patches (loss of context)
- tasks executed in parallel (GPU)

Heterogenity and noise due tissue preparation

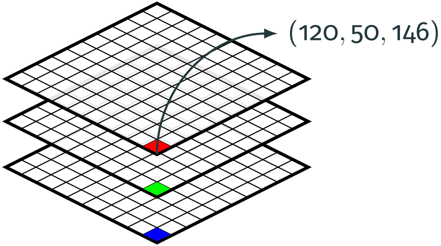
- lots of noise in the images
- a model trained on images acquired from one center is not efficient on images from another center

Machine learning for image analysis

Images as data

Images are spatial multivalued data

- each pixel is a set of spectral values




```
>>> img=skimage.io.imread('test.png')
>>> print(img.shape)
(460, 700, 3)
>>> print(img[0, 0])
[120, 50, 146]
```

Images as data

Machine learning algorithms can work on

- each pixel as an element described by its values

but...

- spatial information is lost
- cannot be used to classify images (need to have a description for each image)
- spectral information is only a piece of information contained in an image

and...

- contrary to other data such as temperature or speed, raw pixel values have no semantic meaning

Machine learning for image analysis

Main applications

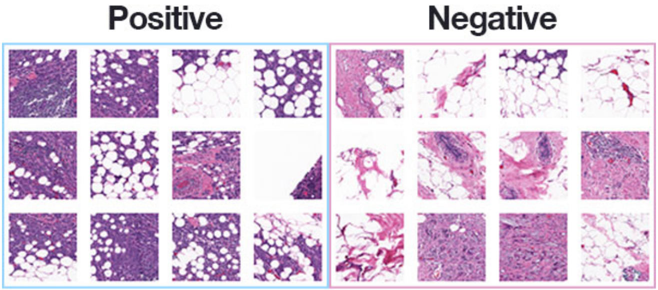
- Classification of images (or part of images)
- Objects detection
- Segmentation

Image classification

Classification

Given a set of images and a set of labels, give the corresponding label to each image

Example: Breast cancer positive or negative images



source: *The Kaggle Breast Histopathology Images dataset by Janowczyk et al.*

Image classification

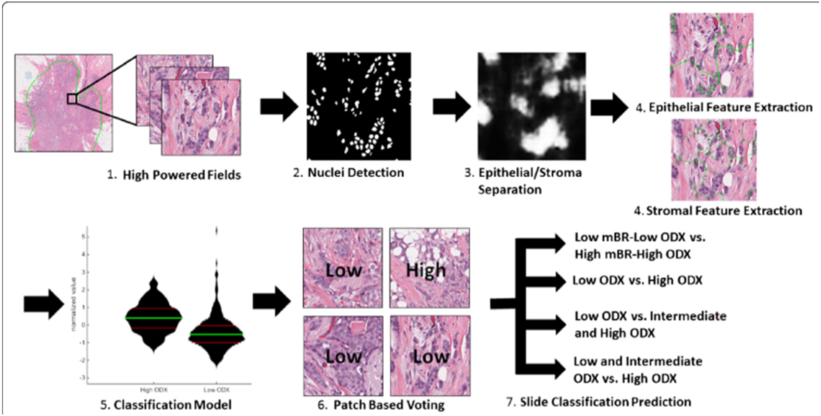
How to apply ML to images for classification?

- calculate features describing each image
- classify the images according to these new features

but...

- It's not trivial to
  - choose the features adapted to a specific problem
  - create new handcrafted features
- and features often need parameters
  - how to fix them?
  - parameters can have a great impact on efficiency

Example - Patches classification using standard descriptors



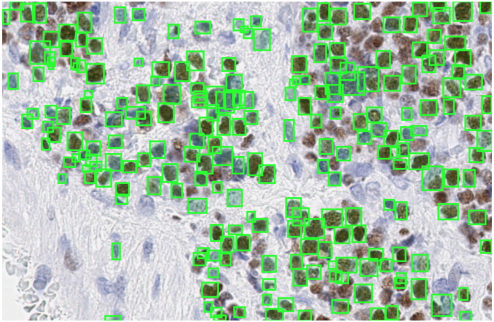
source: *Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer, Whitney et al.*

Objects detection

Detection

Given an image and examples of objects of interest, locate all instances of similar objects

Example: cell nuclei detection

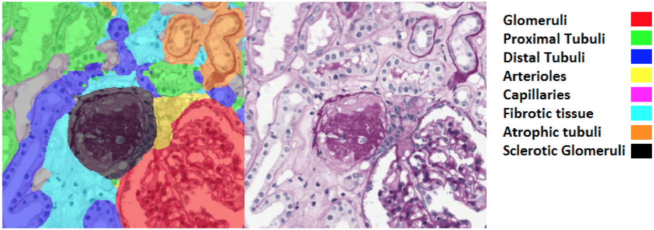


source: *Cell Nuclei Detection on Whole-Slide Histopathology Images Using HistomicsTK and Faster R-CNN Deep Learning Models, Chandradevan et al.*

Segmentation and classification

Segmentation

Given an image, examples of objects of interest and class terminology, locate **and delineate** all instances of objects and associate them to their class



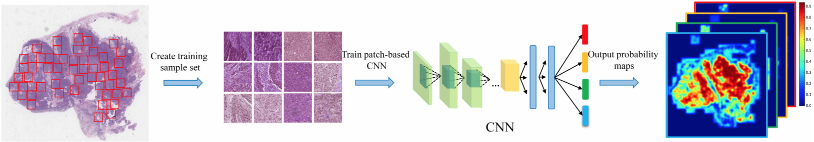
source: *Automatic segmentation of histopathological slides of renal tissue using deep learning, de Bel et al.*

Conclusion

All classical image analysis method can be used on digitized histopathological images to perform many useful tasks for the pathologists. **But** a lot of problems have to be solved as: volume, noise, inter-center heterogenity, cluster of objects of interest, availability of annotations and data, explainability, etc.

Digital histopathology

- emerging domain but still a lot of work to achieve
- huge predominance of deep learning



source: *Weakly Supervised Learning for Whole Slide Lung Cancer Image Classification, Xi Wang*

Hands-on

Hands-on

Patches classification extracted from breast tumor dataset

Binary classification

- Data loading and preparation
- Data visualisation and preprocessing
- Image features extraction
- SVM binary classification (and other methods)
- Model evaluation (accuracy, cross-validation)
- Hyperparameters estimation (grid search)

Multi-class classification

- Multi-layer Perceptron classification

This is it!

Let’s get your hands dirty ☺