

Classification of ISUP grades from Whole Slide Images

Presentation of the Kaggle challenge

Deep Learning for Medical Imaging

Master Mathématiques Vision et Apprentissage (MVA)

03/03/2022



- 1 General information
- 2 Overview of the Challenge
- 3 Some challenges to face
- 4 Some tips

- 1 General information
- 2 Overview of the Challenge
- 3 Some challenges to face
- 4 Some tips

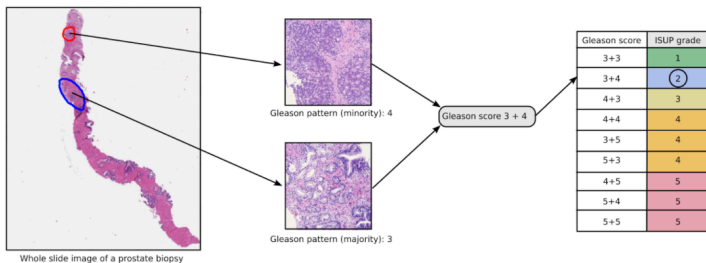
Kaggle Competition

- You should split on teams of 2 people.
- On piazza you have more information about the deliverables, dates, grading process
- Deadline for kaggle competition: 02 April
- Max 4 submission per day for test!

Kaggle Competition

- The competition is private on the following link:
<https://www.kaggle.com/t/da439cea5d9f4ca4b6054ac748f6b409>
- You will need to create a kaggle account !
- **Topic** : Classification of ISUP Grades from whole slide images.

- 1 General information
- 2 Overview of the Challenge
- 3 Some challenges to face
- 4 Some tips

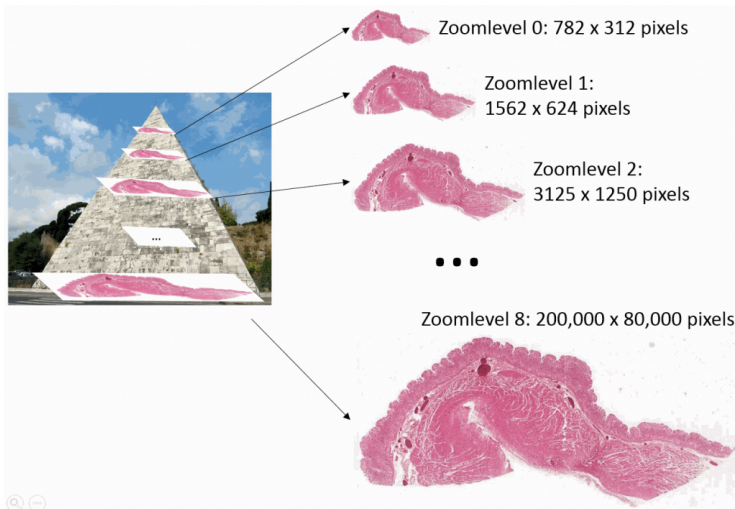


Objective

- Predict the ISUP Grade based on Histopathology images.
- Training set : 340 samples and 338 masks
- Final prediction : 86 samples

- 1 General information
- 2 Overview of the Challenge
- 3 Some challenges to face**
- 4 Some tips

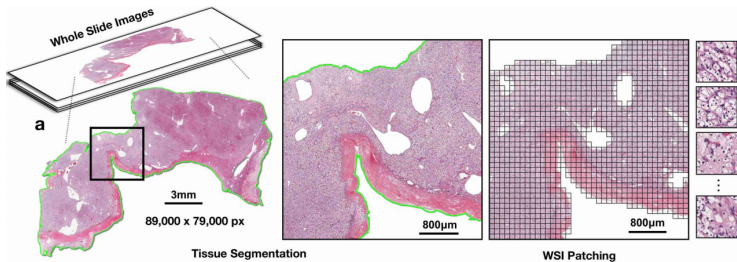
Whole Slide Images



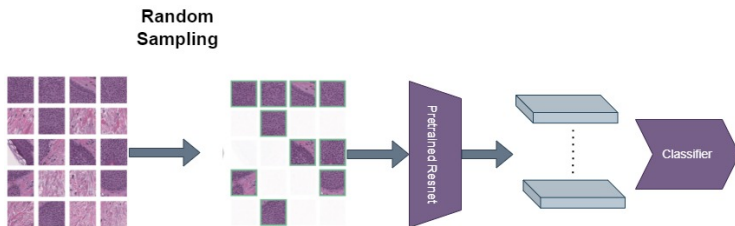
- You will not be able to feed all the image into a network at once.
- The processing time can be slow if you don't deal well with the size issue.
- Images are very large (around 50Mb in a compressed format and can go to 1Gb for uncompressed ones).
- The challenge is only a subset of the dataset, you will need to handle carefully your models for them not to overfit.

- 1 General information
- 2 Overview of the Challenge
- 3 Some challenges to face
- 4 Some tips

Baseline model (AUC = 0.67)



Baseline model (AUC = 0.67)



What is Multiple Instance Learning

- It is a form of **weakly** supervised learning
- Training instances are arranged in **sets** called bags
- A label is provided for entire bags **but not for instances**

Remark : It is **neither** supervised nor unsupervised learning.

Intuition behind Multiple Instance Learning

Serge's key-chain



Serge **cannot** enter
the *Secret Room*

Sanjoy's key-chain



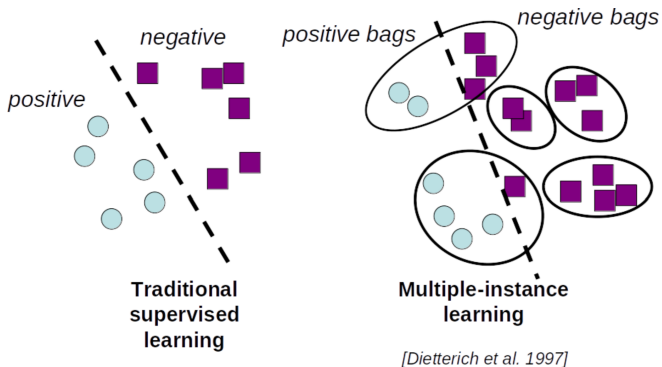
Sanjoy **can** enter
the *Secret Room*

Lawrence's key-chain



Lawrence **can** enter
the *Secret Room*

Multiple Instance Learning



Pros

- It deals with **weakly annotated** data.
- It reduces the annotation cost.
- Algorithms can learn from a greater quantity of training data

Definition of the standard MIL assumption

- Training instances are arranged in sets generally *called bags*.
- A label is given to bags but not to individual instances.
- Negative bags do not contain positive instances.
- Positive bags may contain negative and positive instances.
- Positive bags contain at least one positive instance.

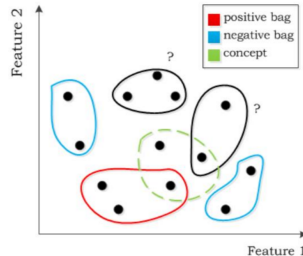


Image from : <http://www.miproblems.org/mi-learning/>

Relaxed MIL assumptions

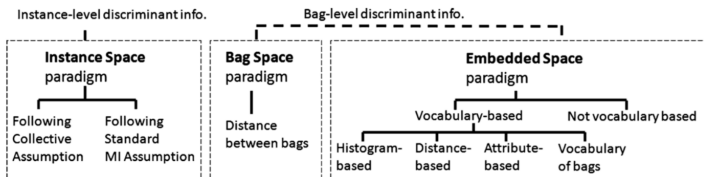
In many applications, the standard MIL assumption is too restrictive. MIL can alternatively formulated as:

- A bag is positive when it contains a sufficient number of positive instances.
- A bag is positive when it contains a certain combination of positive instances.
- Positive and negative bags differ by their instance distributions.

More on MIL assumption: J. Foulds and E. Frank, “A Review of MultilInstance Learning Assumptions,” Knowl. Eng. Rev., vol. 25, no. 1, pp. 1–25, Mar. 2010.

Taxonomy of MIL Methods

A generally accepted taxonomy divides MIL methods based on their reasoning space:



Taxonomy from: J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, Aug. 2013.

Last Details

- Performance is measured using the area under ROC curve (AUC).
- Leaderboard on 50% of the test set (try not to overfit :))
- In the report, we expect more metrics...

Good luck and most of all...

Have Fun !