TRAIL TRUSTED AI LABS
BY DIGITALWALLONIA / SPW-RECHERCHE

Sparsely-annotated dataset segmentation with self-training

Romain Mormont

03/11/2021

[FR] Soutenu par le Service public de Wallonie - Recherche (Convention de recherche n° 2010235 « ARIAC BY DIGITALWALLONIA4.AI »).
[EN] Supported by Service public de Wallonie – Recherche under grant n°2010235 « ARIAC BY DIGITALWALLONIA4.AI »

Who am I ?

- PhD student from the University of Liège
- Supervisors: **Pierre Geurts** and **Raphaël Marée**
- Research topics:
  - **Deep learning** applied to **digital pathology (DP)**
  - How to cope with **data scarcity** ?
  - **Transfer learning (TL)** for classification
  - **Self-training** for image segmentation

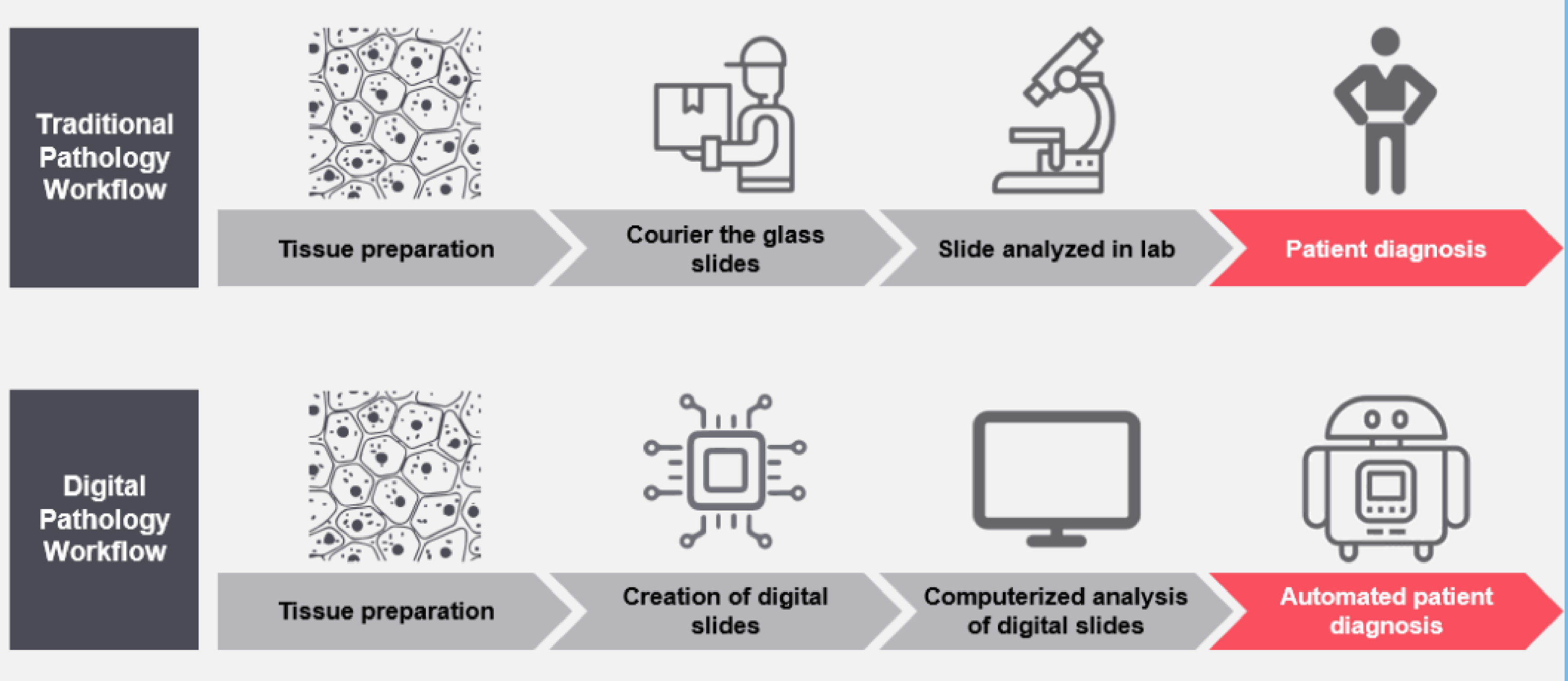
LIÈGE université

cytomine

Digital pathology

*"**Digital pathology** incorporates the acquisition, management, sharing and **interpretation** of pathology information – including slide and data"*

(https://www.leicabiosystems.com)

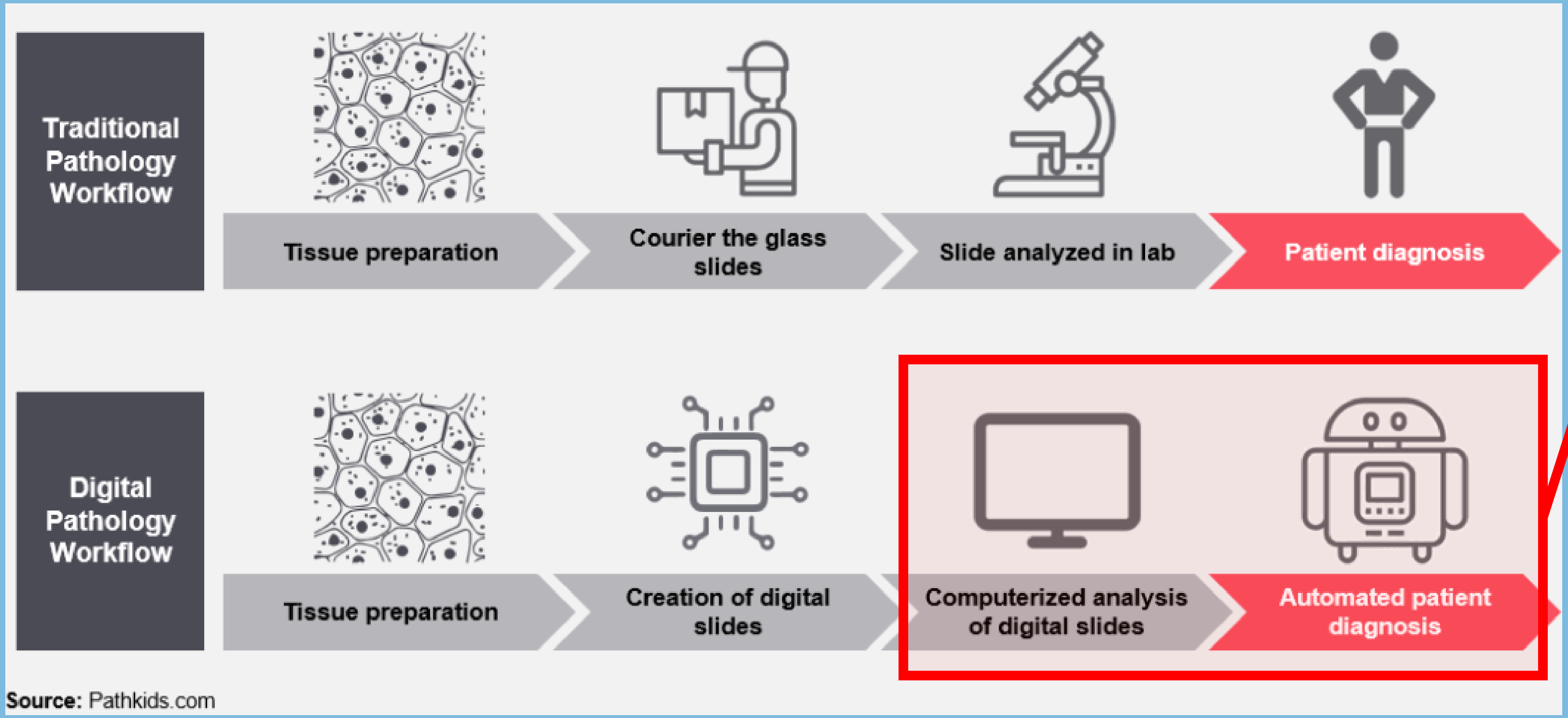


Source: Pathkids.com

Digital pathology

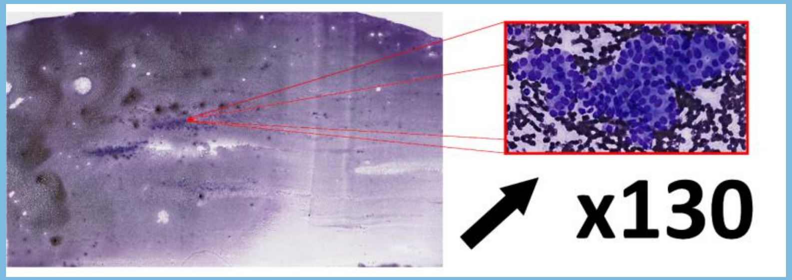
*"**Digital pathology** incorporates the acquisition, management, sharing and **interpretation** of pathology information – including slide and data"*

(https://www.leicabiosystems.com)



… challenging !

- Big data but data scarcity
- Image variability
- Many possible kinds of tasks



Past research – transfer learning

Contribution 1

Research question: **how should one use deep transfer learning in digital pathology ?**

How?

- Empirically evaluate feature extraction and fine-tuning from ImageNet using 8 DP classification datasets

Main takeaways:

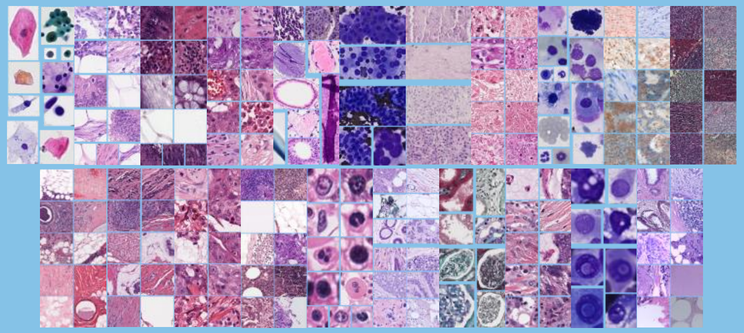
- Fine-tuning > feature extraction
- Feature extraction is a strong baseline

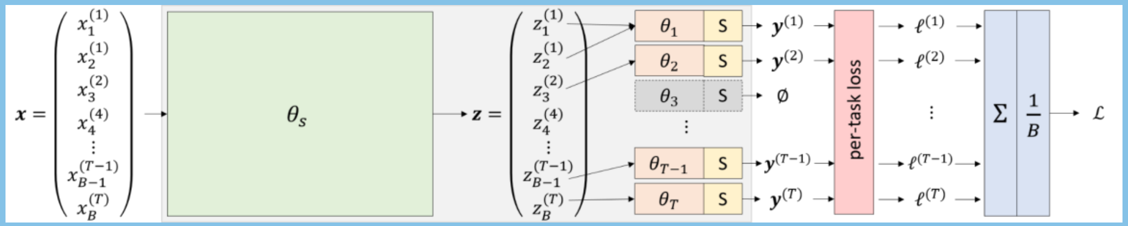
Contribution 2

Research question: **can we pretrain a model on pathology data ?**

Problem: the source task should be large, no ImageNet equivalent in DP

Idea: collect as many DP datasets as possible and pretrain the model in a multi-task fashion





22 classification tasks,
~900 000 images
81 classes

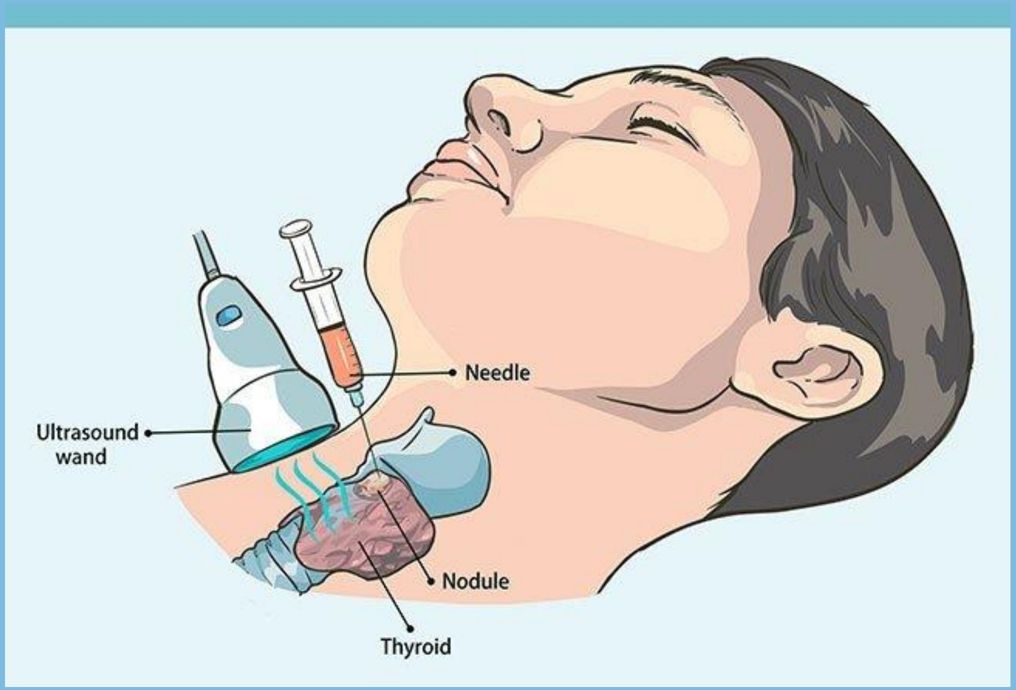
Main takeaway: our pre-trained models either improve significantly over ImageNet models or provide comparable performance

Mormont, Geurts, Marée (2018). "**Comparison of deep transfer learning strategies for digital pathology**". In: Proceedings of the IEEE CVPR Workshops, pp. 2262–2271
Mormont *et al.* (2020). "**Multi-task pre-training of deep neural networks for digital pathology**". In: IEEE journal of biomedical and health informatics 25.2, pp. 412–421

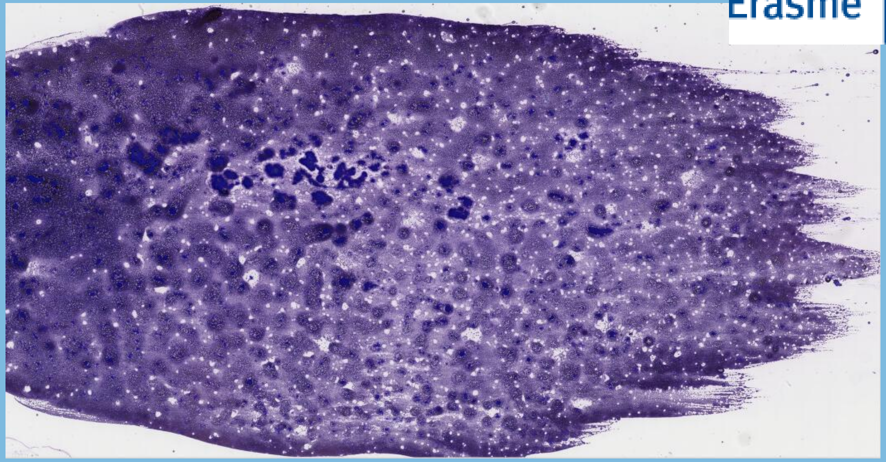
Code and models available online:
https://github.com/waliens/multitask-dipath

Thyroid cancer diagnosis

Fine-needle aspiration biopsy

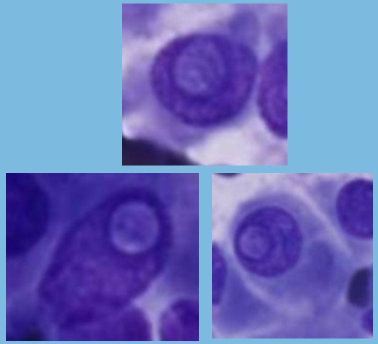


Source: https://images.medicinenet.com

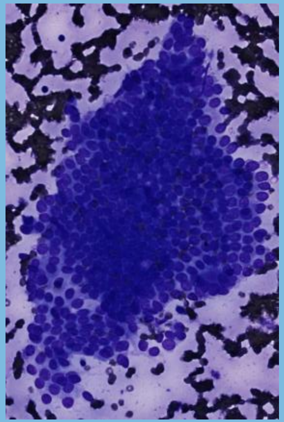


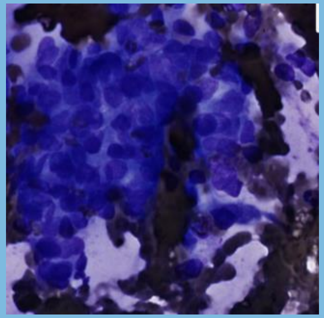
Malignant if presence of…

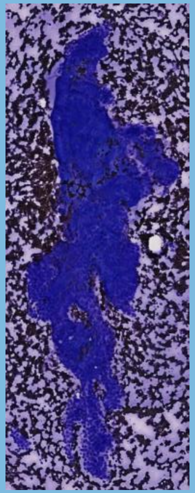
… nuclei with inclusion



… proliferative architectural patterns

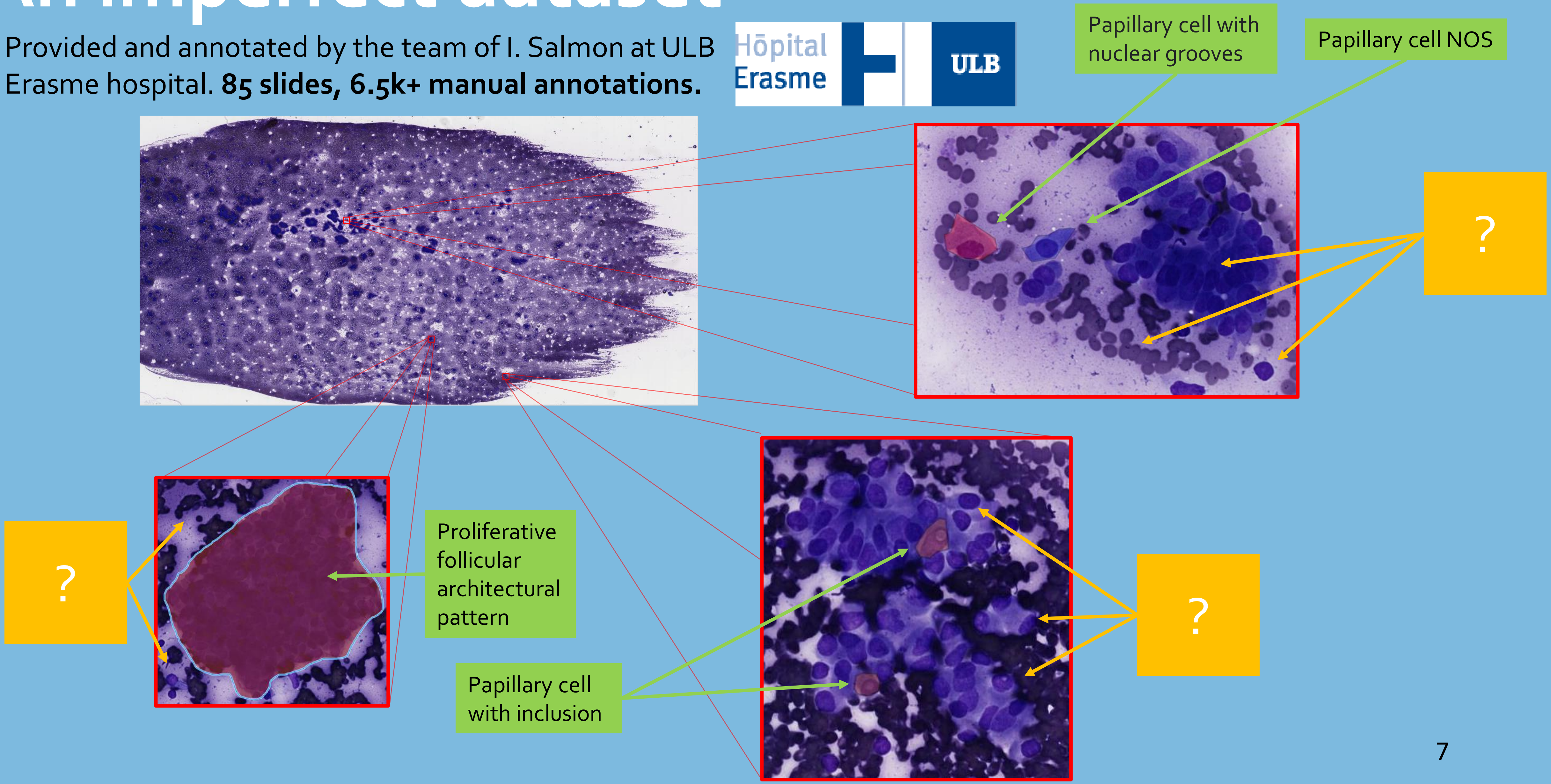






An imperfect dataset

Provided and annotated by the team of I. Salmon at ULB Erasme hospital. **85 slides, 6.5k+ manual annotations.**



Segmentation with self-training (I)

Goal: binary segmentation of nuclei (and patterns)

Dataset: crops of pathologist annotations

Problem: sparse annotations prevent the use of usual segmentation networks like U-Net

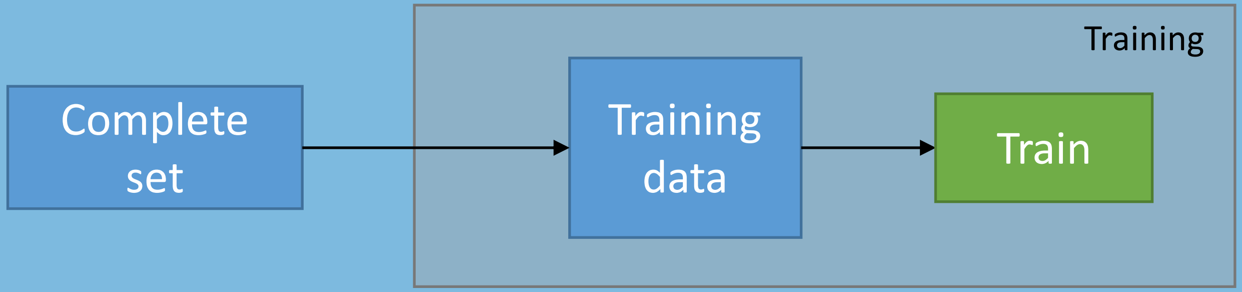
Hypothesis:

1) Pattern annotations are likely less sparse than cell annotations
2) Convolution can deal with « *a bit of noise* » in the ground truth

Ideas:

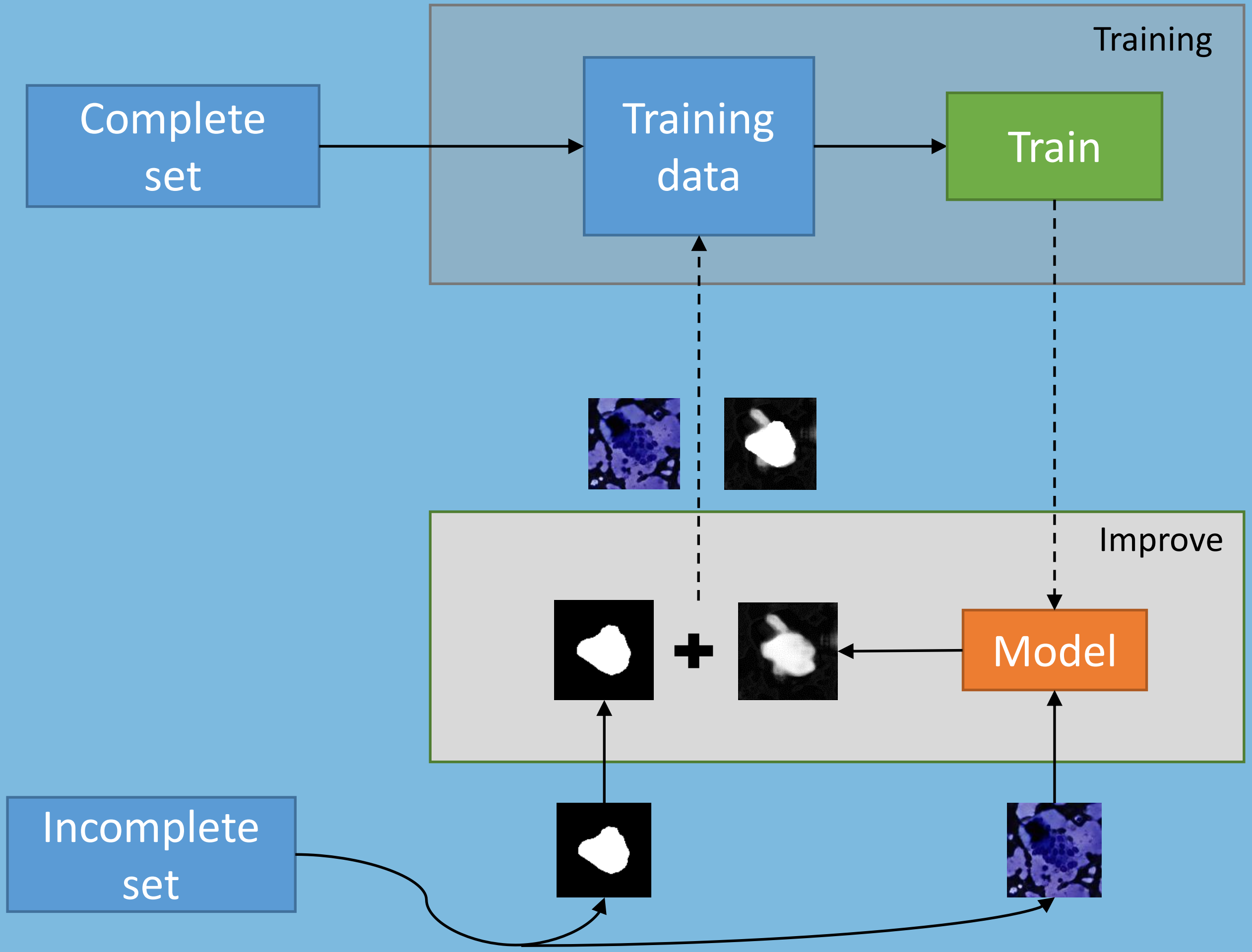
1) Split our dataset in 2 subsets: **complete** (pattern crops) vs. **incomplete** (cell crops)
2) Train a U-Net with our two subsets
   - *Complete set*: use as-is, no ground truth = background
   - *Incomplete set*: **use self-training to help filling the gaps in the ground truth**

Segmentation with self-training (II)



For few epochs, we only use the complete set for training the model. Then…

Segmentation with self-training (III)



… we start including an « improved » incomplete set to the data.

Improved ?

At the end of **each** epoch:

1. Extract the model being trained
2. Forward all samples from the incomplete set into the model
3. Re-create a new ground-truth by combining the expert ground truth with the predictions

How to combine ?

Expert ground truth is kept as-is. Pixels where there is no ground truth are assigned the probability predicted by the network.

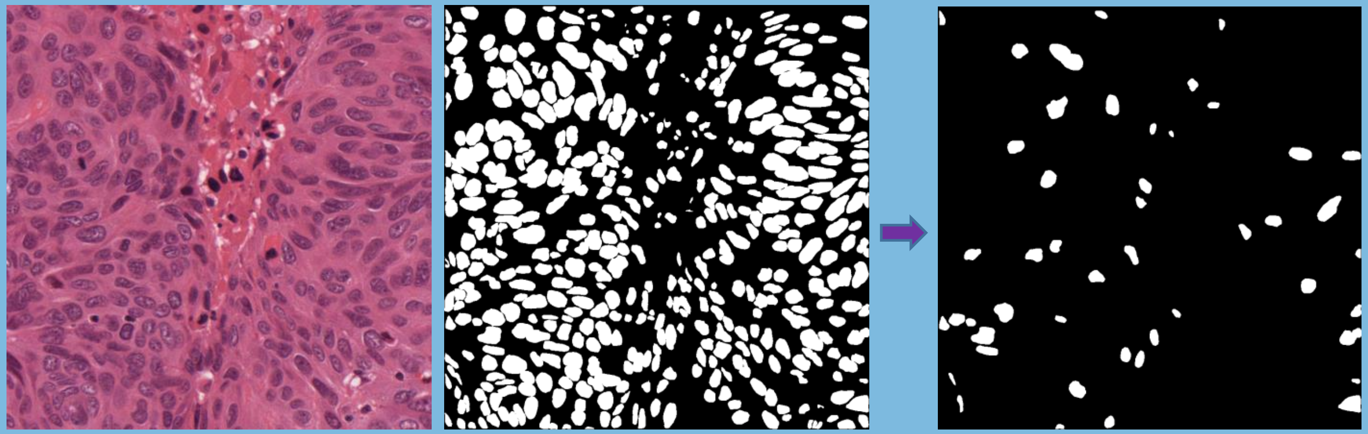
Segmentation with self-training (IV)

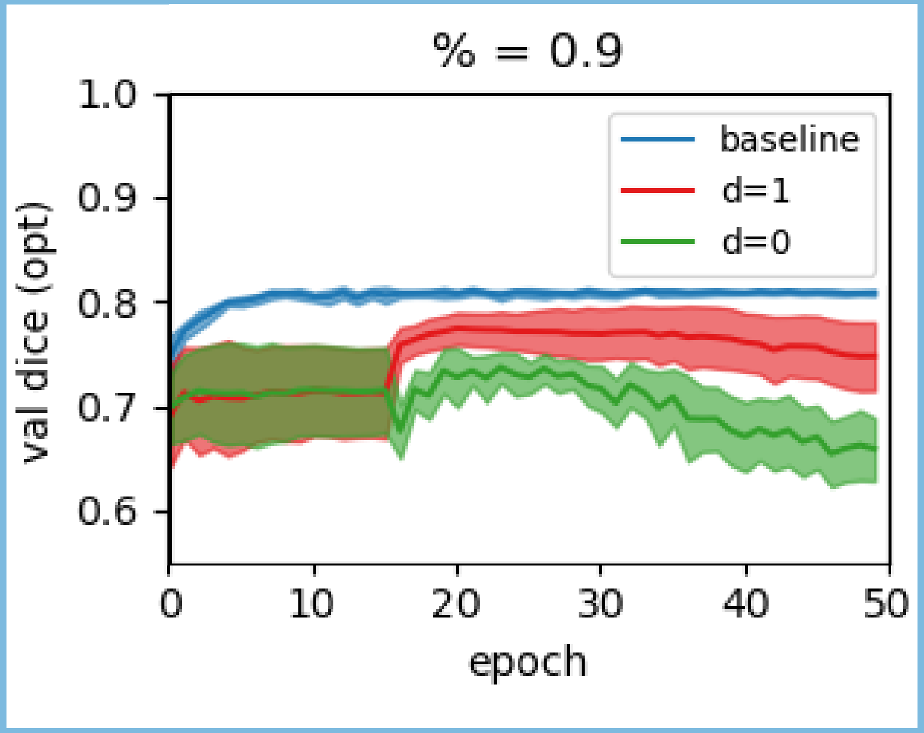
Preliminary results on another dataset (Monuseg).

Simulating data scarcity:

- Complete set: 1 image (1000x1000)
- Incomplete set: 29 images

Randomly removing 90% of nuclei annotation in incomplete set.





Our approach yield better performance compared to using the dataset without self-training, even with a large amount missing data.

Wrapping up

In our first experiments, we have used only **~13% of the annotated data** to train our model.

Exhaustive annotation might not be needed to successfully train a segmentation algorithm (?)

Next steps:

- **Validating** the observation on other datasets (including thyroid)
- **Weighting** the contribution of certain pixels when computing the training loss
- Using **thresholded prediction** instead of raw prediction to complement ground truth in the incomplete set
- Actually **apply the model to a entire whole-slide image** efficiently

Thank you !

Acknowledgements

Pierre Geurts and Raphaël Marée are both part of the ARIAC project. Raphaël Marée is funded by BigPicture.

Data contributors: ULB Erasme Hospital, Isabelle Salmon and Caroline Degand