A workflow for large-scale computer-aided cytology and its applications

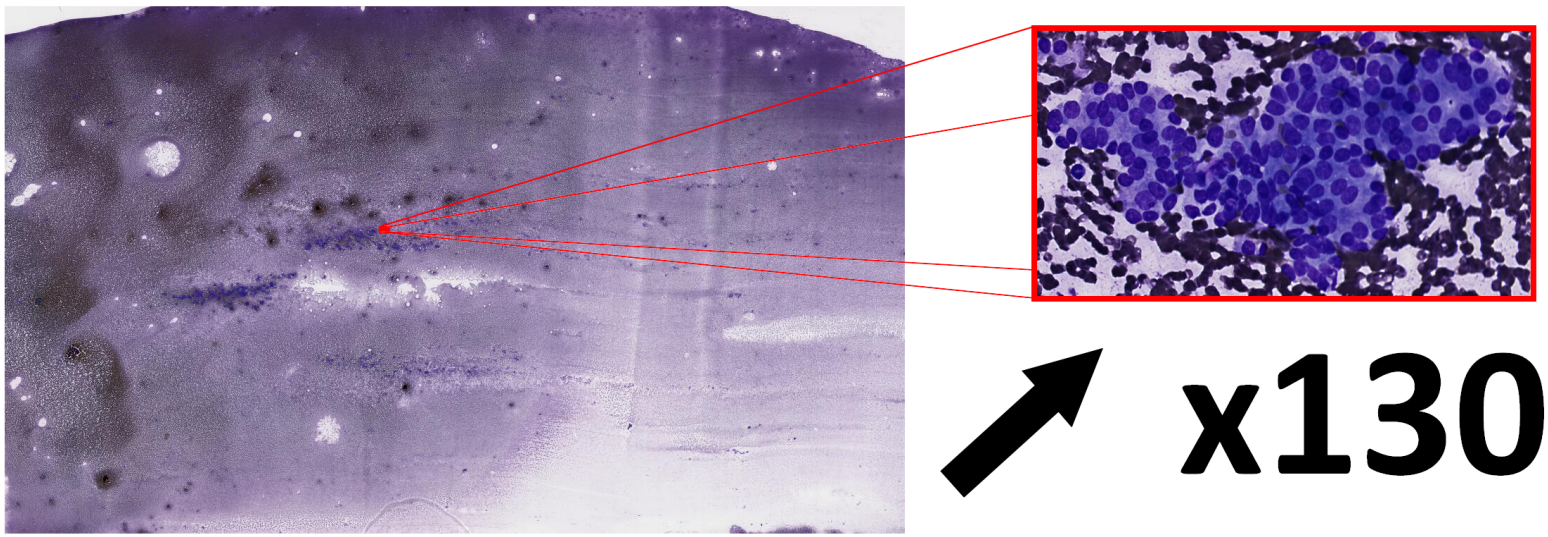
Romain Mormont

Université de Liège, Belgium

June 22, 2016

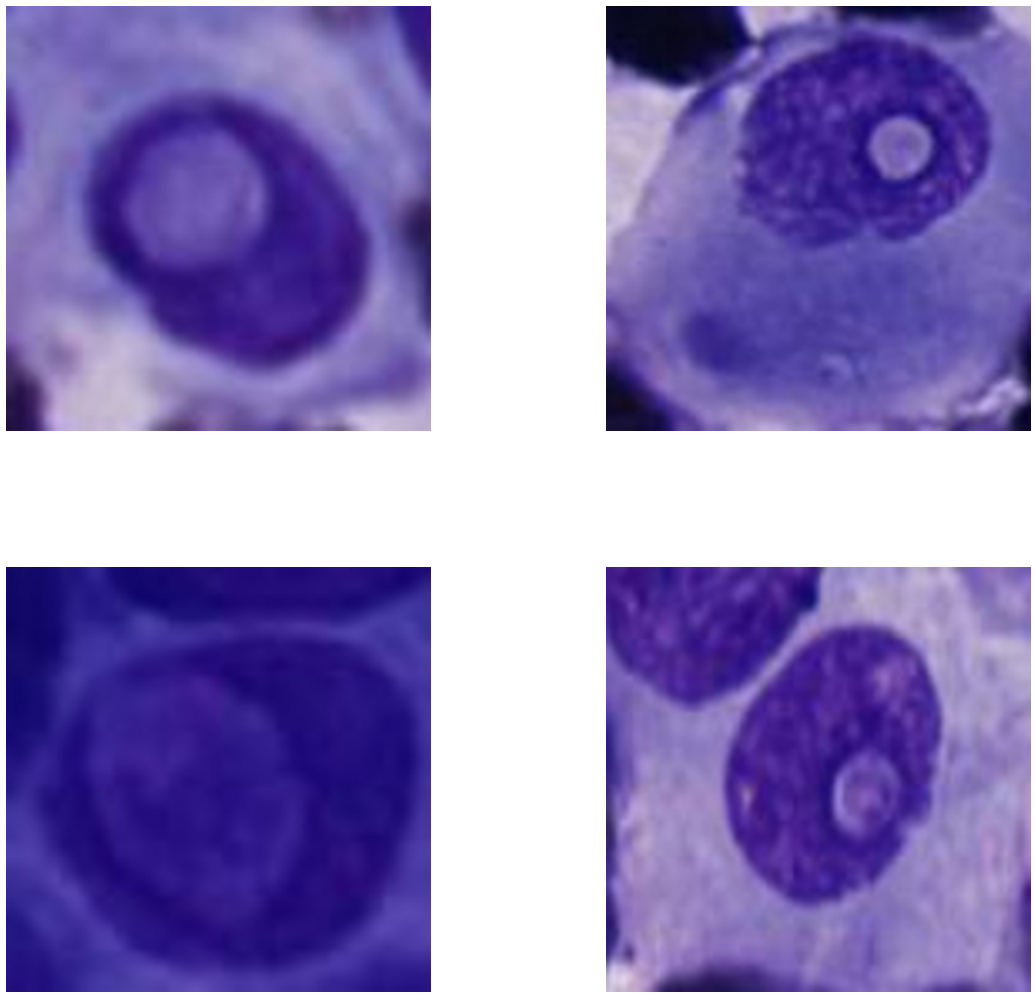
Supervisor : Dr. Raphaël Marée
Academic : Pr. Pierre Geurts

Context and application



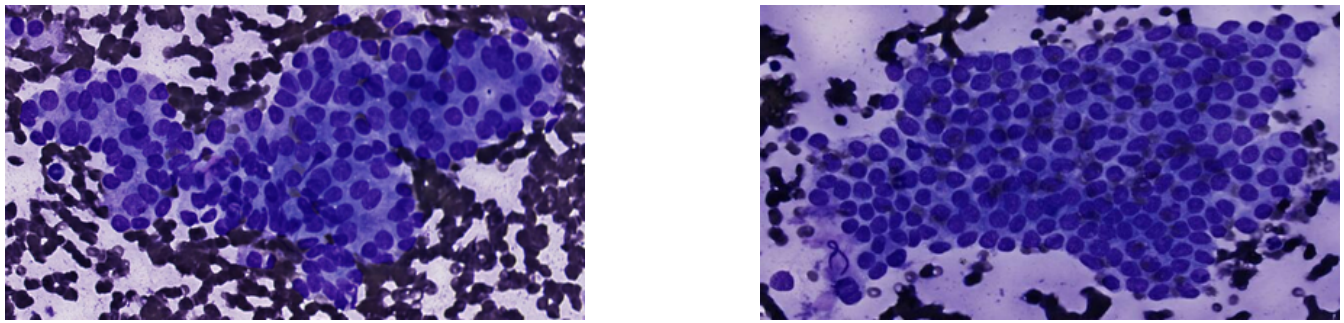
Microscope slide smeared with cell samples (15 gigapixels).

Application : thyroid nodule malignancy

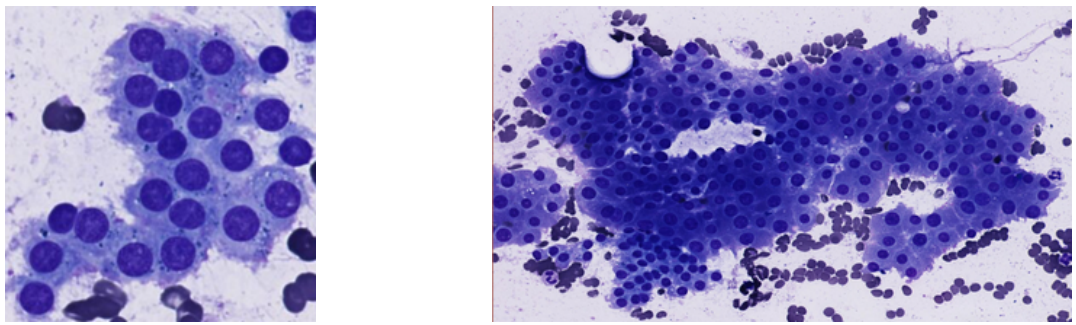


Cells with inclusion

Application : thyroid nodule malignancy



(a) Proliferative



(b) Non-proliferative

Architectural patterns

Application : thyroid nodule malignancy

Can be expressed as a problem of **object detection and classification** !

Objectives

1. **Developing a framework** for performing object detection and classification in multi-gigapixel images

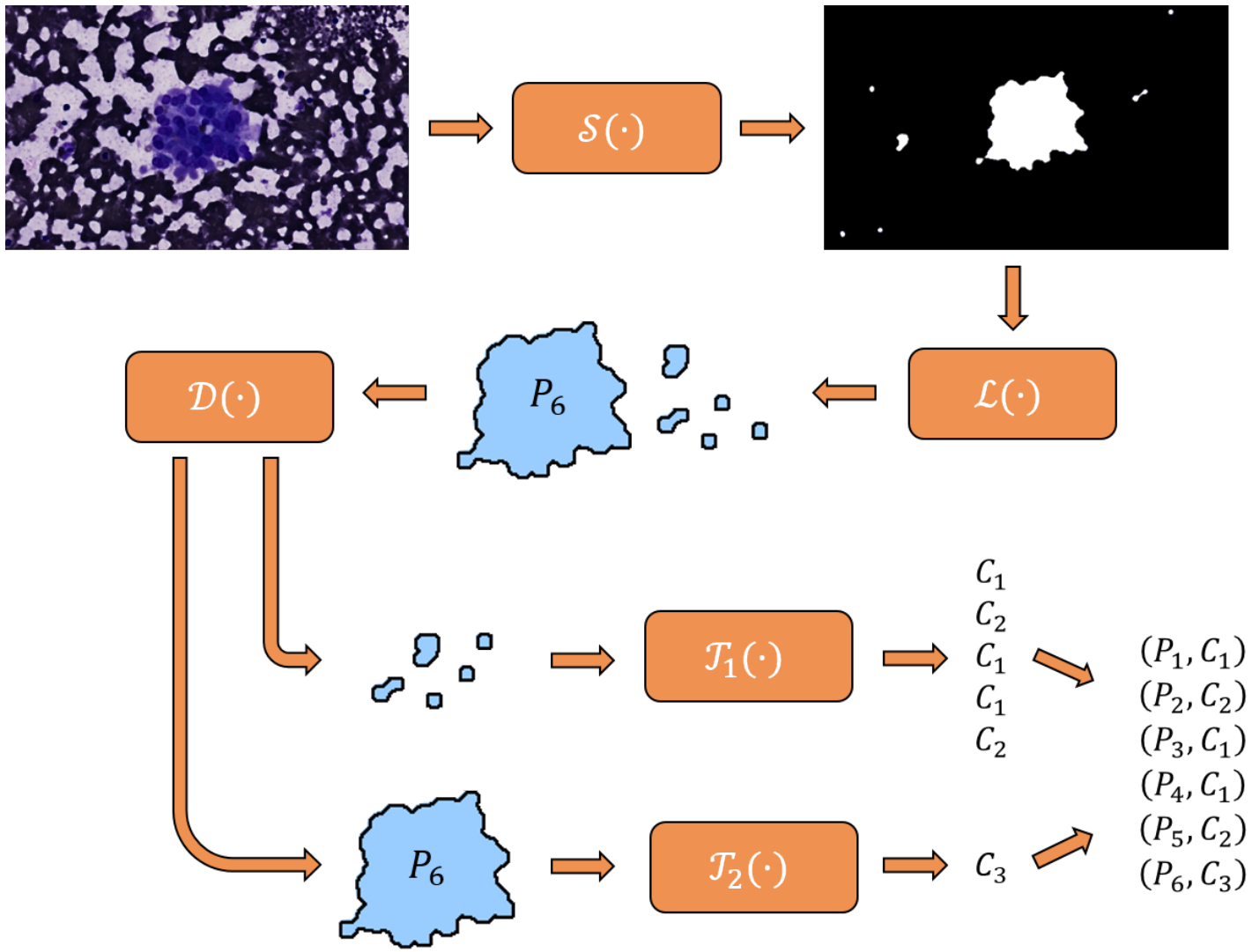
2. **Applying this framework** to the problem of thyroid malignancy diagnosis

SLDC framework

Technical challenges :

- Memory constraint (images do not fit into memory)
- Efficiency
- Parallelism
- Genericity

SLDC framework : algorithms (workflow)



SLDC framework : implementation

Key features :

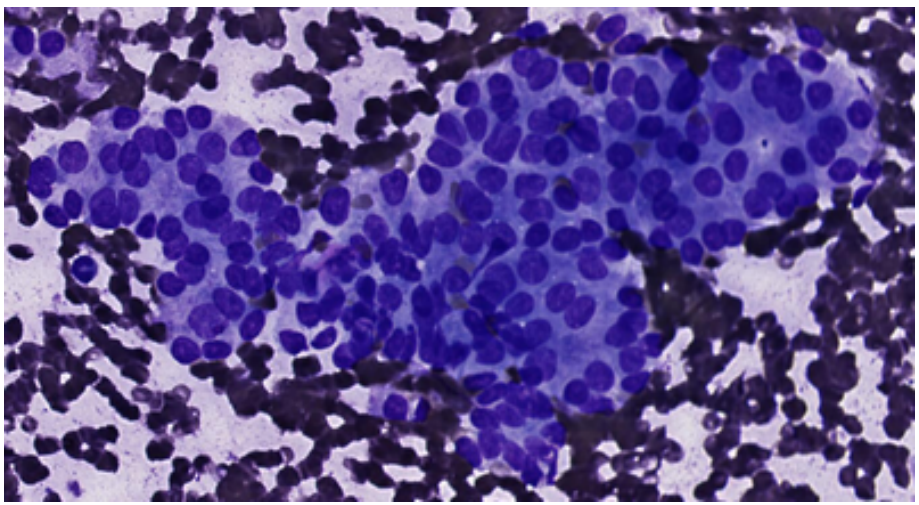
- Memory constraint handled by **splitting images into tiles**
- Customizable **logging system** allowing user to keep track of execution progress
- **Several levels of parallelism** available
- **Builder components** providing an easy way of building complex workflows

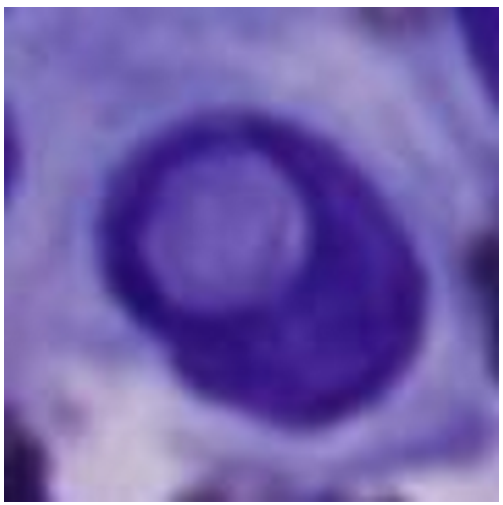
About the implementation :

- Implemented as a Python library
- Available on GitHub at `https://github.com/waliens/sldc`
- Unit-tested (coverage of 85 %)

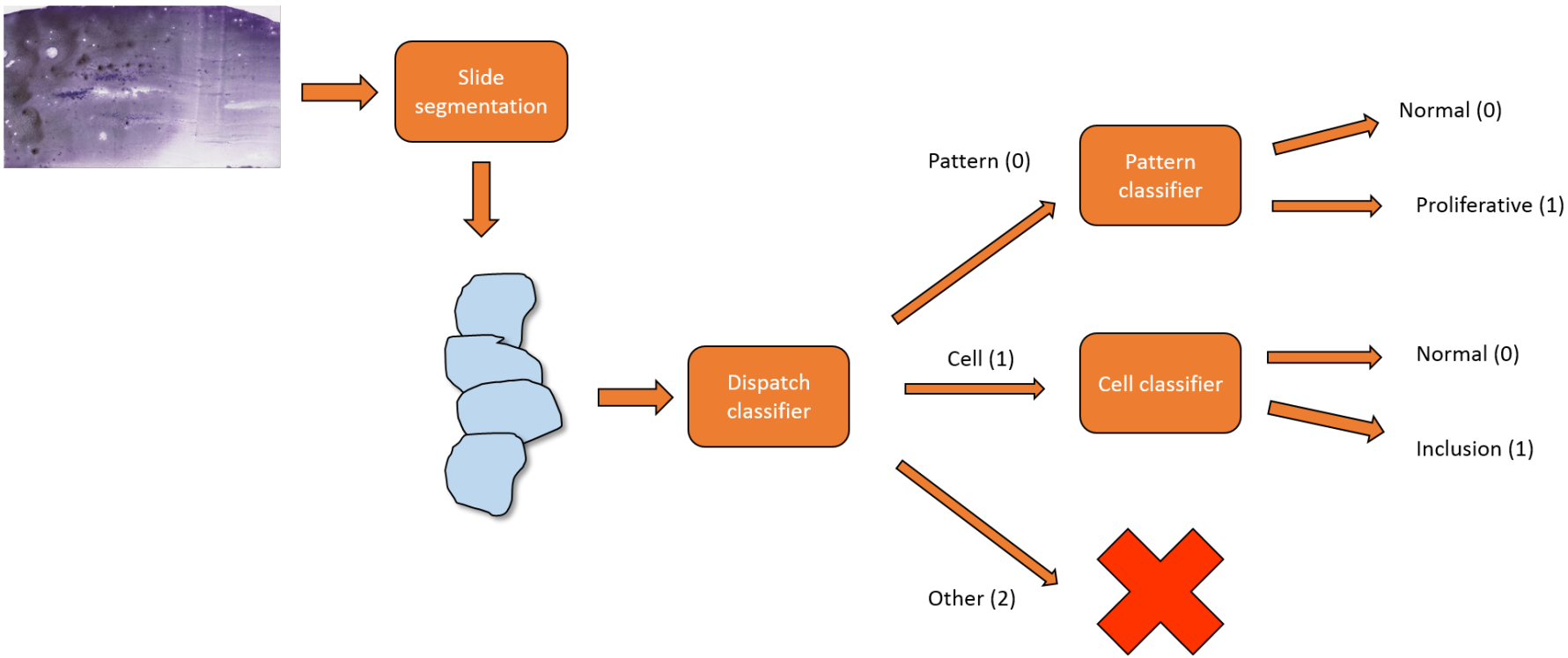
SLDC at work

Aim : detect **cells with inclusion** and **proliferative architectural patterns**

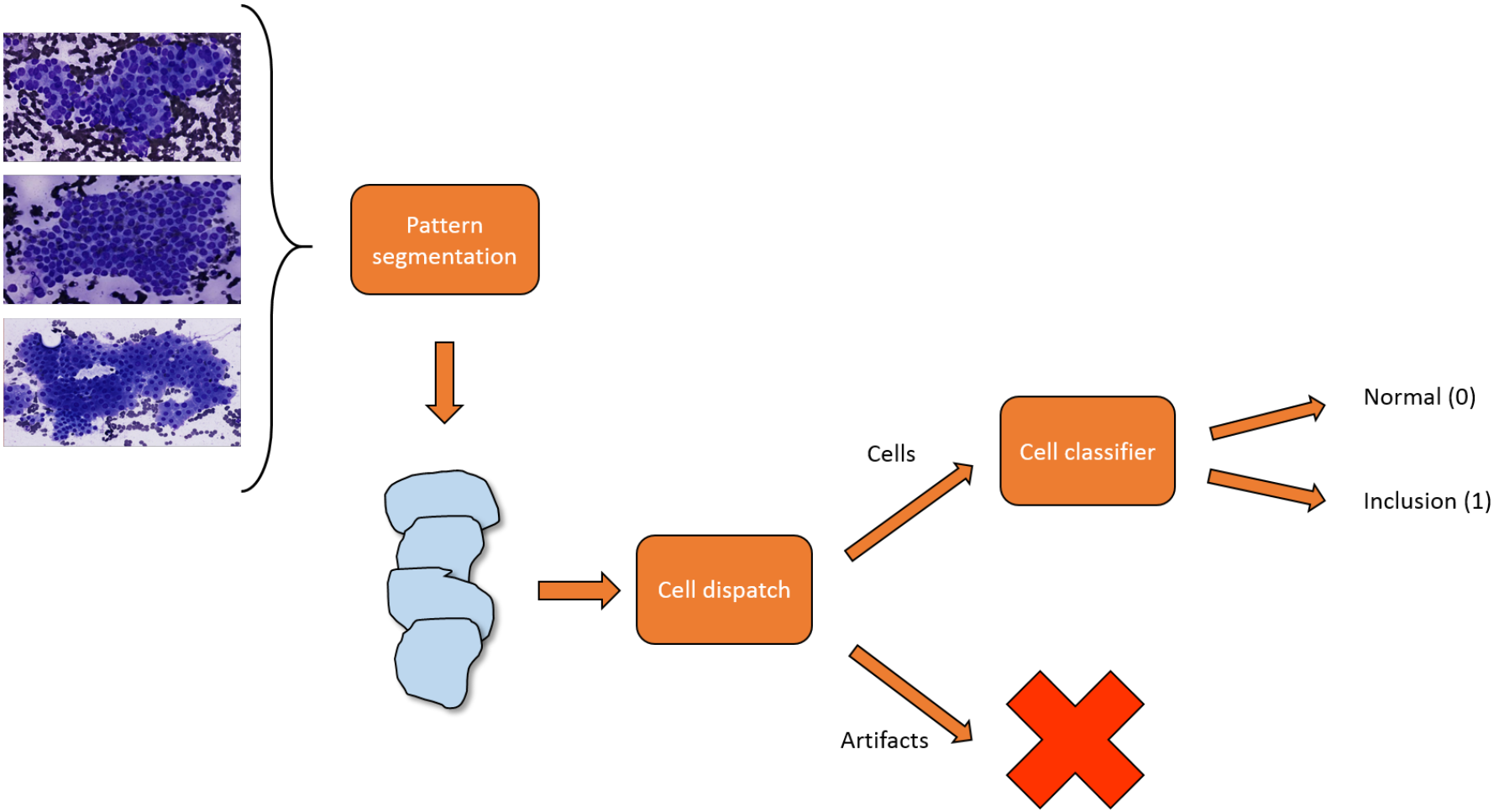




SLDC at work : slide processing

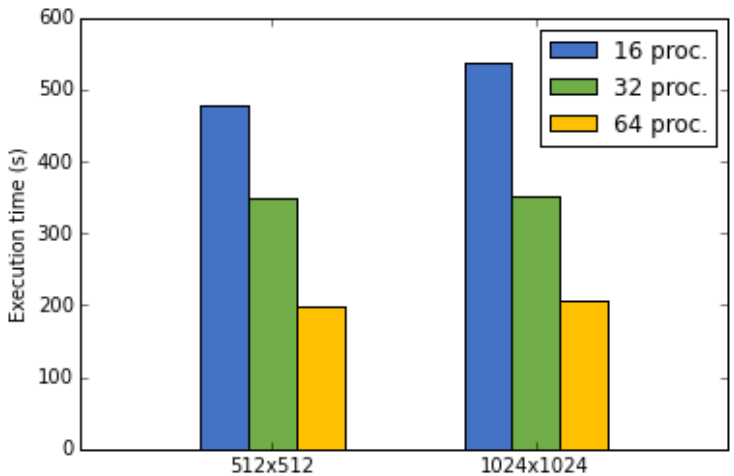


SLDC at work : pattern processing

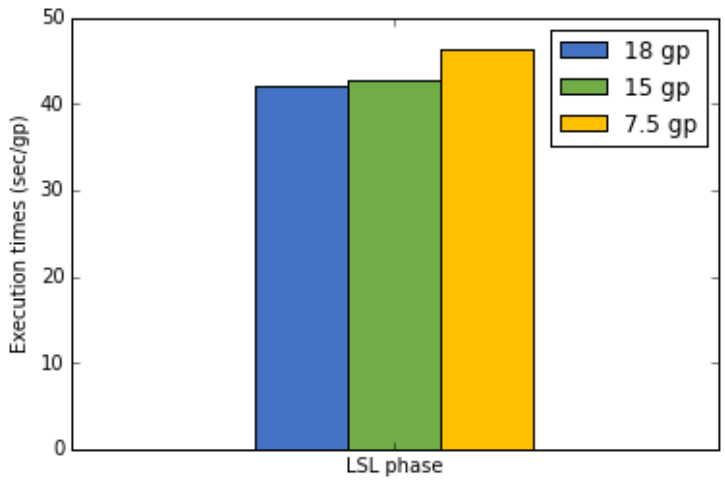


SLDC at work : execution times

- Execution times mostly dominated by network communication
- Effective time for 16 giga-pixels images is 18 minutes
- **Effective time for 8 giga-pixels images is 8 minutes**

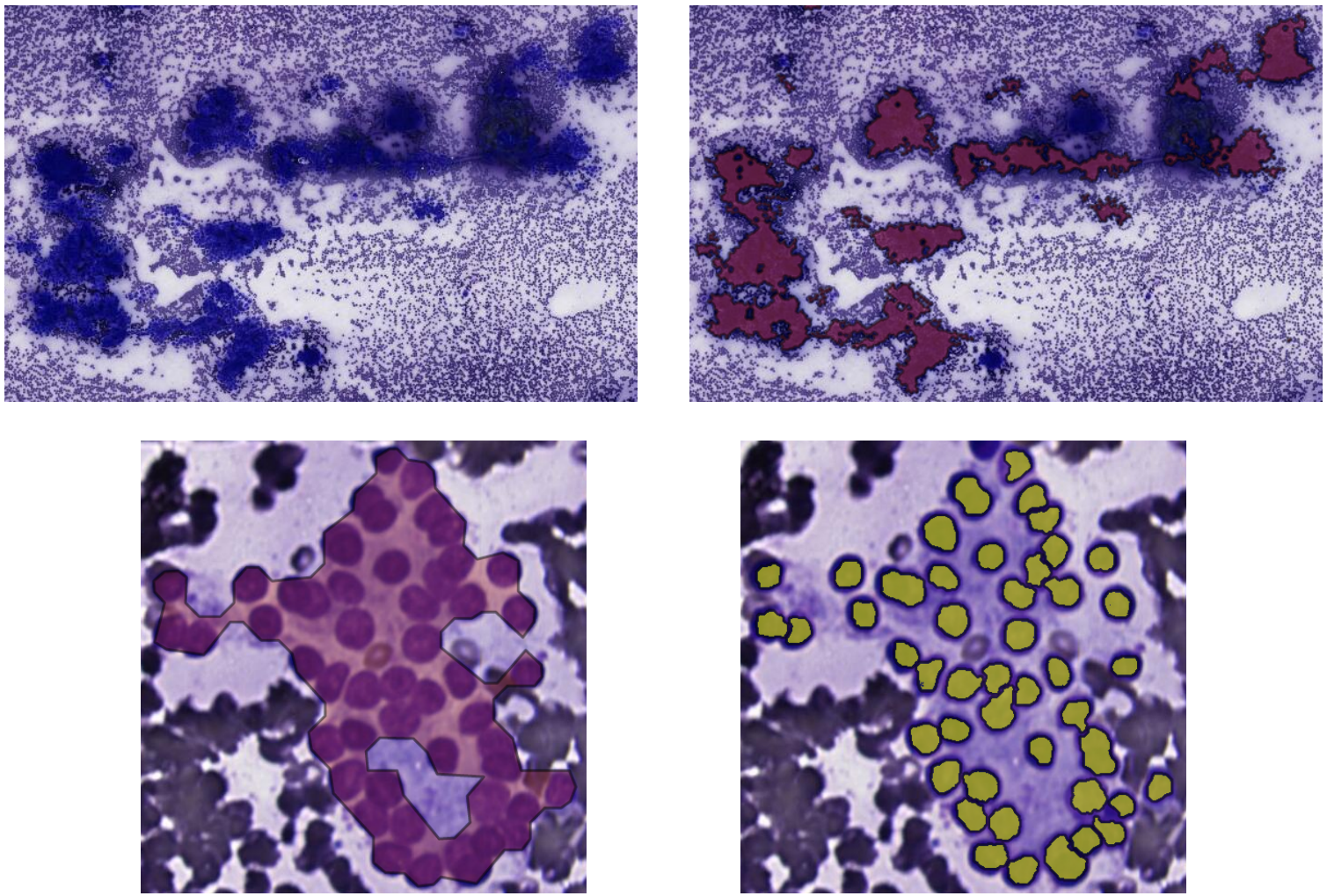


(a) Acceleration brought by using more processes.



(b) Scalability of the LSL phase.

SLDC at work : detection



Conclusions

1. The **framework is production-ready** and available on GitHub.

2. The implemented workflow provides **promising results** but **still requires some improvements**.

Future works

SLDC framework

- Was **improved since the master thesis submission** (parallelization, ease of use,...).
- But some minor improvements must still be performed.

Thyroid workflow

- Improving the segmentation procedures
- Improving the classification models

Thank you for your attention !
Any question ?