The thyroid workflow can still be improved (classifiers and segmentations)

Second pass workflow

Cells with incl. (malignant)

Pattern annot. per term

Proliferative (malignant)

Normal cells (benign)

Cell annot. per group

Cell annot. per term

Machine learning and image processing

• Due to the lack of proper tool, those slides are usually analysed manually!

• Thanks to modern technologies, microscope slides are often digitized to be further analysed on computers. But slides are huge: ~100K × 100K pixels typically.

In cytology, cytopathologists analyse microscope slides in order to find malignant cells or artefacts to diagnose diseases like cancer (see Figure below).

In the images and to associate them with a field-related label.

In cytology, cytopathologists analyse microscope slides in order to find malignant cells or artefacts to diagnose diseases like cancer (see Figure below).

SLDC is an open-source Python framework created for accelerating development of large image analysis workflows.

How?

• It encapsulates problem-independent logic (parallelism, memory limitation due to large images handling, …).

• It provides a concise way of declaring problem dependent components (segmentation, object classification, …).

SLDC framework

Object detection and classification

• In some applications, multi-gigapixel images must analysed to gather information and/or take decisions.

• The analysis usually consists in finding some objects in the images and to associate them with a field-related label.

• In cytology, cytopathologists analyse microscope slides in order to find malignant cells or artefacts to diagnose diseases like cancer (see Figure below).

• In cytology, cytopathologists analyse microscope slides in order to find malignant cells or artefacts to diagnose diseases like cancer (see Figure below).

Features

• Tile-based: images are splitted into tiles which are loaded into memory one after another. A full image is never loaded into memory at once.

• Parallel: the framework can take advantage of parallelism to reduce its execution time. Parallelism is available at several levels (tiles, objects, images, …).

• Talkative: a customizable logging system provides a rich real-time feedback about the execution.

• Integrable: thanks to Python, integration with other libraries such as scikit-learn (ML), open-cv (IP), PyCuda (GPU), … is effortless.

• Convenient: building components provide an easy way of constructing complex workflows.

How SLDC works

Given an input image, the framework produces a set of polygons and classification labels representing the objects of interest. In order to do so, it executes the following steps:

• Segment $S$: a segmentation procedure is applied on the input image (top-left) and produces a binary mask (top-right) locating the objects of interest in the image.

• Detect $D$: polygons representing the objects found in the image are extracted from the binary mask.

• Classify $T$: a classifier produces classification labels for the polygons/objects it is passed.

All problem-independent concerns being encapsulated by the framework, developers only have to define the segmentation $S$, the dispatching rules $r$ and the classifiers $T$.

Workflow

To actually detect those objects, two passes are performed on the images. The first consists in extracting standalone cells and architectural patterns. The second consists in extracting the cells contained in the detected patterns.

Classification

Classification is performed based on the detected objects’ crop images using random subwindows and extremely randomized trees [MWG16].

Results

To diagnose thyroid cancer, cytopathologists screen microscope slides smeared with thyroid nodule cell samples. The illness is diagnosed when two types of objects are found on those slides:

SLDC at work: thyroid nodule malignancy diagnosis

Data

The dataset is stored on the Cytomine [Mar+16] web platform. It consists in:

• 84 images with size ranging from 4 to 18 gigapixels

• 68 annotated images

• 5921 labelled annotations made by cytopathologists from ULB (Team of Pr. Isabelle Salmon, Dept. of Pathology)

Conclusion

• The SLDC framework is production ready, feel free to use it: https://github.com/waliens/sldc

• The thyroid workflow can still be improved (classifiers and segmentations)

Bibliography
