A web-based framework for visualization, annotation, and automatic exploitation of high-resolution bioimages using tree-based machine learning methods

Benjamin Stevens¹,², Loïc Rollus¹,², Louis Wehenkel¹,², and Raphaël Marée¹,²,³

¹GIGA-Systems Biology & Chemical Biology, GIGA-R, University of Liège, Belgium
²Systems and Modeling Research Unit, Montefiore Institute, University of Liège, Belgium
³GIGA Bioinformatics Core Facility, GIGA, University of Liège, Belgium

*These two authors contributed equally to this work.

Introduction

Human interpretation of large-scale bioimaging datasets is time-consuming and somehow subjective. This stresses the need for replicable and generic computational methods to automate the extraction of useful, quantitative, information from these images. Towards that goal, there is a strong need for intuitive and efficient tools to foster collaboration between life science researchers working with high-throughput and high-resolution imaging equipments (such as whole-slide virtual microscopy, high-content screening, etc.), and computer scientists who need large and realistic training datasets to develop effective machine learning and computer vision methods.

This poster presents the ongoing development of a web-based software platform for remote visualization, collaborative annotation, and automated analysis of high-resolution biological images. The potential of such a framework is illustrated in the context of ongoing studies within the CYTOMINE research project (2010-2013) about lung tissue recognition and quantification in whole-slide Hematoxylin&Eosin stained images of experimental mice, and cell counting in whole-slide images of LBA cytology, from the Laboratory of Tumor and Development Biology (Prof. D. Cataldo, GIGA-Cancer, ULg), and from the Department of Pathology at Erasme University Hospital (Prof. I. Salmon, ULB).

Visualization and annotation of high-resolution images on the web

We rely on Java/JavaScript/REST technologies and a spatial, relational database to allow collaborative annotations of images with user-defined ontology terms.

Integration of image processing and machine learning algorithms

Ongoing integration of generic machine learning algorithms using tree-based approaches [1,2,3] is expected to allow scientists to explore more efficiently their large-scale cellular and tissular data through ready-to-use content-based image retrieval, image segmentation and classification techniques exploiting training sets of manual annotations.

Acknowledgments

This poster presents results of the CYTOMINE project (research grant n° 1017072 of the Walloon Region, DGO6). Raphaël Marée is also supported by the GIGA with the help of the Walloon Region and the European Regional Development Fund. We thank Didier Cataldo, Natacha Rocks, Fabienne Perin, and Christine Fink for feature suggestions, and continuous beta testing.

References

2 Dumont et al., Fast Multi-Class Image Annotation with Random Subwindows and Multiple Output Randomized Trees, VISAPP 2009.
3 Marée et al., Incremental Indexing and Distributed Image Search using Shared Randomized Vocabularies, MIR 2010.