
Master thesis : Development of cell counting algorithms within pathological tissues

Auteur : Rubens, Ulysse

Promoteur(s) : Geurts, Pierre; Maree, Raphael

Faculté : Faculté des Sciences appliquées

Diplôme : Master en ingénieur civil en informatique, à finalité spécialisée en "computer systems and networks"

Année académique : 2016-2017

URI/URL : <http://hdl.handle.net/2268.2/3237>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Development of cell counting algorithms within pathological tissues.

Author: Ulysse Rubens

Advisor: Prof. Pierre Geurts

Supervisor: Dr. Raphaël Marée

Master in Computer science and Engineering

Academic year 2016-2017

Counting objects in images is within the reach of everyone. When the number of objects to count becomes large, the task is more difficult and we tend to make errors because of some oversights. For a computer, counting or detecting objects in images is hard. However, the repetitive characteristic of this task does not worry it.

In histopathology, the study of pathology combined to histology (the study of cells and tissues), scientists frequently have to count cells from images of tissues to address hypotheses about pathological or developmental processes. The rise of digital imaging services and the power of computers and networks led to the emergence of digital histopathology, a new field where histopathologists analyze cells and tissues on computer screens. Therefore, it would be interesting to have automatic counting methods available to lighten the work of life scientists.

In this thesis, the problem of cell counting and the closely related problem of cell detection are studied. Supervised machine learning workflows are proposed in that purpose. Five methods are presented, re-implemented and tried on four databases of histopathological images, whose two are newly introduced for this kind of problem. We set up the software modules allowing to use these algorithms directly on databases from Cytomine, a web-based application for collaborative analysis of multi-gigapixel images.

Basically, two ways of counting are studied. Firstly, counting can be done by cell detection, the counting trivially being the number of detections. Secondly, counting can be estimated by density estimation, knowing the density for each image's pixel.

An important aspect of this work is the assessment and the comparison of these approaches. Performance results in machine learning literature are sometimes too optimistic due to weak model validation. Robust metrics are thus defined in order to make comparisons possible. Each machine learning algorithm has a set of parameters that can be tuned. They have a large influence on the performance. In order to choose the best combination of parameters, a systematic model selection protocol is employed. It is the first time that such a set of methods has been evaluated on several datasets of such a large size. Our study has helped to better understand the functioning of these methods and to establish their limits.

Results are encouraging. In general, at least one algorithm is able to reach a counting error less than 10% for each dataset.