Robust detection and segmentation of cell nuclei in biomedical images based on a computational topology framework

Rodrigo Rojas-Moraleda [a,e], Wei Xiong [b], Niels Halama [c], Katja Breitkopf-Heinlein [d], Steven Dooley [d], Luis Salinas [e], Dieter W. Heermann [b], Nektarios A. Valous [a,*]

[a] Applied Tumor Immunity Clinical Cooperation Unit, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany
[b] Statistical Physics and Theoretical Biophysics Group, Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany
[c] Department of Medical Oncology, National Center for Tumor Diseases, University Hospital Heidelberg, Heidelberg, Germany
[d] Department of Medicine II, Faculty of Medicine at Mannheim, Heidelberg University, Mannheim, Germany
[e] Department of Informatics, Technical University Federico Santa María, Valparaiso, Chile

ARTICLE INFO

ABSTRACT

The segmentation of cell nuclei is an important step towards the automated analysis of histological images. The presence of a large number of nuclei in whole-slide images necessitates methods that are computationally tractable in addition to being effective. In this work, a method is developed for the robust segmentation of cell nuclei in histological images based on the principles of persistent homology. More specifically, an abstract simplicial homology approach for image segmentation is established. Essentially, the approach deals with the persistence of disconnected sets in the image, thus identifying salient regions that express patterns of persistence. By introducing an image representation based on topological features, the task of segmentation is less dependent on variations of color or texture. This results in a novel approach that generalizes well and provides stable performance. The method conceptualizes regions of interest (cell nuclei) pertinent to their topological features in a successful manner. The time cost of the proposed approach is lower-bounded by an almost linear behavior and upper-bounded by $\mathcal{O}(n^2)$ in a worst-case scenario. Time complexity matches a quasilinear behavior which is $\mathcal{O}(n^{1+\varepsilon})$ for $\varepsilon < 1$. Images acquired from histological sections of liver tissue are used as a case study to demonstrate the effectiveness of the approach. The histological landscape consists of hepatocytes and non-parenchymal cells. The accuracy of the proposed methodology is verified against an automated workflow created by the output of a conventional filter bank (validated by experts) and the supervised training of a random forest classifier. The results are obtained on a per-object basis. The proposed workflow successfully detected both hepatocyte and non-parenchymal cell nuclei with an accuracy of 84.6%, and hepatocyte cell nuclei only with an accuracy of 86.2%. A public histological dataset with supplied ground-truth data is also used for evaluating the performance of the proposed approach (accuracy: 94.5%). Further validations are carried out with a publicly available dataset and ground-truth data from the Gland Segmentation in Colon Histology Images Challenge (GlaS) contest. The proposed method is useful for obtaining unsupervised robust initial segmentations that can be further integrated in image/data processing and management pipelines. The development of a fully automated system supporting a human expert provides tangible benefits in the context of clinical decision-making.

1. Introduction

Topology concerns the study of fundamental and intrinsic properties of spaces which are invariant to continuous deformations, extrinsic evaluations (Weatherson and Marshall, 2014) or changes in magnitude e.g. scale, shape, distance, intensity, etc. Bernhard

Riemann discovered the modern concept of topology by the second half of the 19th century. However, the foundations of topology and homology are due to the work of Henry Poincare, Enrico Betti, and Emmy Noether. Topology became an influential field of research by the middle of the 20th century. For most of its history topology has not been applied to the scientific domain due to its abstract and qualitative nature. However, during the last ten years it has emerged as one of the most effective representation models for exploring massive and complex data sets (Gyulassy et al., 2009)