Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 90 (2016) 119 – 124

Procedia Computer Science

International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6-8 July 2016, Loughborough, UK

Persistent Homology for Fast Tumor Segmentation in Whole Slide Histology Images

Talha Qaiser[a], Korsuk Sirinukunwattana[a], Kazuaki Nakane[b], Yee-Wah Tsang[c], David Epstein[d], Nasir Rajpoot[a,*]

[a]*Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom*
[b]*Osaka University, Suita, 565-0871, Japan*
[c]*University Hospital of Coventry and Warwickshire, Coventry, CV2 2DX, United Kingdom*
[d]*Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom*

Abstract

Automated tumor segmentation in Hematoxylin & Eosin stained histology images is an essential step towards a computer-aided diagnosis system. In this work we propose a novel tumor segmentation approach for a histology whole-slide image (WSI) by exploring the degree of connectivity among nuclei using the novel idea of *persistent homology profiles*. Our approach is based on 3 steps: 1) selection of exemplar patches from the training dataset using convolutional neural networks (CNNs); 2) construction of persistent homology profiles based on topological features; 3) classification using variant of k-nearest neighbors (k-NN). Extensive experimental results favor our algorithm over a conventional CNN.

Peer-review under responsibility of the Organizing Committee of MIUA 2016

Keywords: Digital Pathology; Tumor Segmentation; Histology Image Analysis; Persistent Homology; Colorectal Cancer;

1. Introduction

Colorectal cancer is the second most commonly diagnosed cancer in females and the third most in males, with an estimated 1.4 million cases and 693,000 deaths occurring in 2012[1]. Identification of tumor-rich areas from histopathology colon images is one of the primary tasks for a computer-aided diagnosis system. Automated tumor segmentation methods are therefore highly desirable and have received much research effort during recent years.

Recently published techniques for tumor segmentation mostly rely on morphological appearance and local texture features of tissue components[2]. However, the high variability of features in tumor regions and tissue samples pose the risk of over-emphasizing the spatial properties of a particular dataset; these methods do not produce the desired results. Some methods focus on nucleus segmentation or sub-cellular components of tissues[3,4]. However, cell segmentation is non-trivial due to the atypical characteristics and heterogeneous appearance and, often, clumped structure of cancerous cells, so these methods are also likely to fail. For these reasons, detection of tumor regions is a highly challenging problem.

* Corresponding author
E-mail address: N.M.Rajpoot@warwick.ac.uk

Peer-review under responsibility of the Organizing Committee of MIUA 2016
doi:10.1016/j.procs.2016.07.033

In tumor regions nuclei have different atypical characteristics, with non-uniform chromatin texture and irregularity in their shape and size. In tumor areas nuclei clump together, filling the inter-cellular regions, and the structure of individual nuclei becomes more difficult to discern. In contrast, nuclei remain relatively distinct in normal regions and maintain their structure and appearance (Fig 1). The arrangement of nuclear structures is a significant feature for tumor classification. In this paper, we propose a new approach to tumor segmentation in WSIs by characterizing the connectivity between cells, using persistent homology. Our approach works in the following ways : 1) a method for constructing persistent homology profiles based on topological features; 2) an algorithm for selection of exemplar patches from the training dataset based on highly activated nodes of a CNN; and 3) a variant of k-NN classifier.

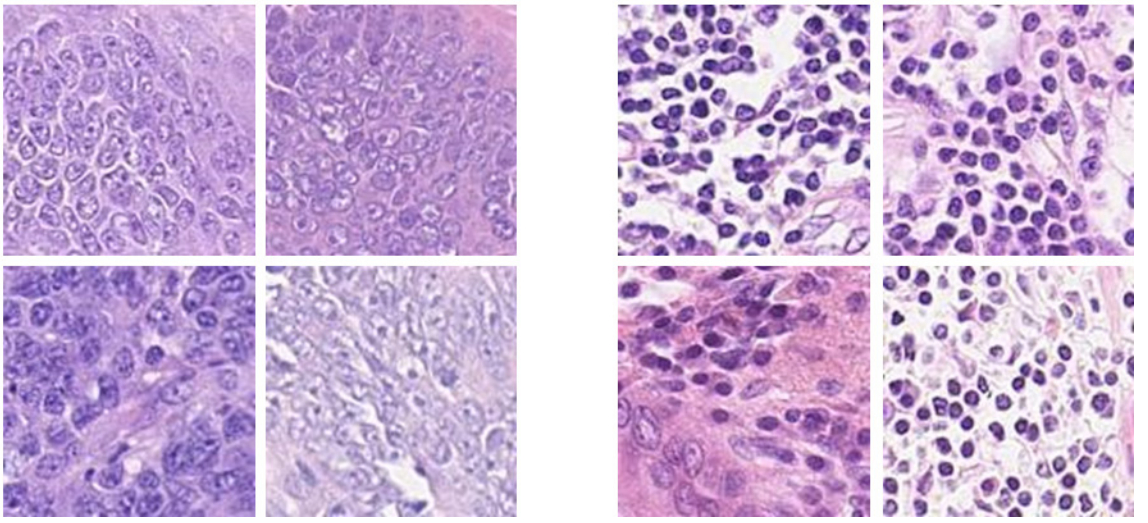

Fig. 1: (a) Tumor patches, exhibiting non-uniformity in chromatin texture and atypical characteristics. (b) Non-tumor patches showing the homogeneous structure of normal nuclei.

2. Proposed Algorithm

Given a colorectal WSI, we first divide it into patches. The problem is to then classify each patch as tumor or non-tumor. We approach this problem by exploring the connectivity between cells. We associate with each patch two statistical distributions, which we call *persistent homology profiles*, computed using certain topological invariants, as will be explained below. For simplicity, we will explain our method using a single persistent homology profile for each patch. In that case, given two patches, we have the persistent homology profile for each, and the symmetrised Kullback-Leibler divergence (KLD) between these two distributions gives a numerical estimate for how far the patches are from each other. An input patch Q is classified by a k-NN classifier, based on KLD distances between the persistent homology profiles of representative patches, denoted by P (Fig 2). We decide on our representative patches by training a CNN, and then selecting, separately for the tumor class and the non-tumor class, patches whose activation during the training is large. The essence of this approach is to use the subset of highly activated patches from convolutional layers as exemplars, rather than using the set of all the patches in the training dataset.

2.1. Persistent Homology

2.1.1. Overview.

Persistent homology is a fairly recent concept, surveyed in[5] and[6]. Let M denotes a graylevel image of size mxn, where gray intensities are integer values between 0 and 255 and B be an $m \times n$ closed rectangle. For each value of the threshold t, let $B(t) \subset B$ be the union of pixels with intensity less than or equal to t. We want $B(t)$ to be a closed subspace of B, and so, with each pixel in $B(t)$, we include its four corners and four sides. In effect, we binarize the intensity values in M, replacing any value less than or equal to t by 0 and replacing other values by 1. The resulting matrix $M(t)$ gives us a black and white image. The notation is supposed to remind us that $B(t)$ is the union of black pixels. We have $B(0) \subseteq B(1) \subseteq \cdots \subseteq B(255) = B$, so that B is a filtered space, persistent homology's essential ingredient. We will work with the zeroth Betti number $\beta_0 : [0, 255] \cap \mathbb{Z} \to \mathbb{Z}$ which maps a threshold value t to the