Neurocomputing 191 (2016) 214–223

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images

Jun Xu [a,*], Xiaofei Luo [a], Guanhao Wang [a], Hannah Gilmore [b], Anant Madabhushi [c]

[a] Jiangsu Key Laboratory of Big Data Analysis Technique, Nanjing University of Information Science and Technology, Nanjing 210044, China
[b] Institute for Pathology, University Hospitals Case Medical Center, Case Western Reserve University, OH 44106-7207, USA
[c] Department of Biomedical Engineering, Case Western Reserve University, OH 44106, USA

ARTICLE INFO

Article history:
Received 14 September 2015
Received in revised form
21 December 2015
Accepted 2 January 2016
Communicated by Shaoting Zhang
Available online 17 February 2016

Keywords:
Deep Convolutional Neural Networks
Feature representation
The classification of epithelial and stromal regions
Breast histopathology
Colorectal cancer

ABSTRACT

Epithelial (EP) and stromal (ST) are two types of tissues in histological images. Automated segmentation or classification of EP and ST tissues is important when developing computerized system for analyzing the tumor microenvironment. In this paper, a Deep Convolutional Neural Networks (DCNN) based feature learning is presented to automatically segment or classify EP and ST regions from digitized tumor tissue microarrays (TMAs). Current approaches are based on handcraft feature representation, such as color, texture, and Local Binary Patterns (LBP) in classifying two regions. Compared to handcrafted feature based approaches, which involve task dependent representation, DCNN is an end-to-end feature extractor that may be directly learned from the raw pixel intensity value of EP and ST tissues in a data driven fashion. These high-level features contribute to the construction of a supervised classifier for discriminating the two types of tissues. In this work we compare DCNN based models with three handcraft feature extraction based approaches on two different datasets which consist of 157 Hematoxylin and Eosin (H&E) stained images of breast cancer and 1376 immunohistological (IHC) stained images of colorectal cancer, respectively. The DCNN based feature learning approach was shown to have a F1 classification score of 85%, 89%, and 100%, accuracy (ACC) of 84%, 88%, and 100%, and Matthews Correlation Coefficient (MCC) of 86%, 77%, and 100% on two H&E stained (NKI and VGH) and IHC stained data, respectively. Our DNN based approach was shown to outperform three handcraft feature extraction based approaches in terms of the classification of EP and ST regions.

1. Introduction

Stromal (ST) tissue includes the fatty and fibrous connective tissues surrounding the ducts and lobules, blood vessels, and lymphatic vessels, which are supportive framework of an organ. Epithelial (EP) tissue is the cellular tissue lining and found in the ductal and lobular system of the breast milk ducts. About 80% breast tumors originate in the breast EP cells. Although ST tissue is typically considered as not being part of malignant tissue, the changes in the stroma tend to drive tumor invasion and metastasis [11]. Therefore, tumor-stroma ratio in histological tissues is being recognized as an important prognostic value [12], since cancer growth and progression is dependent on the microenvironment of EP and ST tissues. Yuan et al. in [31] found that the spatial arrangement of stromal cell in tumors is a prognostic factor in breast cancer. Consequently a critical initial step in developing automated computerized algorithms for risk assessment and prognosis determination is to be able to distinguish stromal from epithelial tissue compartments on digital pathology images. This is however extremely challenging due to the high data density, the complexity of the tissue structures, and the inconsistencies in tissue preparation. Therefore, it is crucial to develop intelligent algorithms for the segmentation of different tissue structures in an accurate, fast, practical and robust manner [25,32–34].

2. Previous works

There has been substantial interest recently in developing approaches for automated classification of stromal and epithelial regions within H&E tissue images. In [19], local binary pattern (LBP) and contrast measure based texture features were used for discriminating epithelium and stroma from immunohistochemistry (IHC) stained tumor tissue microarrays (TMAs) of colorectal cancer. Five perception-based features (coarseness, contrast,

* Corresponding author.
E-mail address: xujung@gmail.com (J. Xu).

http://dx.doi.org/10.1016/j.neucom.2016.01.034
0925-2312/

directionality, line-likeness and roughness), features related to human perception, were presented in [6] to differentiate EP and ST patches [19]. In [14], color based texture features extracted from square image blocks for automated segmentation of stromal tissue from IHC images of breast cancer. A binary graph cuts approach where the graph weights were determined based on the color histogram of two regions, was used for segmenting EP and ST regions from odontogenic cysts images in [13]. In [17], a cell graph feature describing the topological distribution of the tissue cell nuclei was used for discriminating tumor and stromal areas on immunofluorescence histological images. In [3], IHC stained TMA cores were automatically stratified as tumor or non-tumor cores based on a visual word dictionary learning approach. As LBP based approaches can only deal with gray scale images, in [19], prior to feature extraction, each color image is converted into gray scale images by computing a weighted sum of R, G, and B components. However since the conversion assumes that each pixel in the gray scale image is a linear combination of three color components, an assumption that is not always true, LBP features could be derived off sub-optimal image representations.

The fixed-size window or pixel-grid is one of the traditional ways to select patches from bigger images prior to feature extraction. Recently, superpixel based approaches [23] are being employed to group pixels into meaningful atomic regions based on similarity. Two popular superpixel algorithms are Normalized Cut (Ncut)-based [23,24] and Simple Linear Iterative Clustering (SLIC)-based [1]. Ncut-based superpixel algorithm essentially employs graph theory to explore the pixel-wise similarity among the pixels being interrogated and their neighbourhood pixels. The SLIC-based superpixel algorithm is based on clustering and employs the similarity of each pixel's color and Euclidean distance. Ncut-based superpixel algorithm is more accurate but is more computationally intensive. Compared to Ncut-based algorithm, the SLIC-based approach is simple and faster, but is less accurate. Compared to traditional pixel-grids, the atomic regions generated via a superpixel algorithm represent a natural partitioning of visual scenes. As different tissue structures are mutually present in histologic images, superpixel based approaches are often employed as a pre-processing step to mitigate the issue of possible over-segmentation of the tissue images into atomic regions. The atomic regions are then subsequently segmented into epithelial and stromal regions. In [4], a superpixel based algorithm was used to over-segment breast tissue Hematoxylin and Eosin (H & E) images into small compartments. Subsequently the cell nuclei and cytoplasm within each smaller subcompartment were further classified into epithelial and stromal regions by a Support Vector Machine (SVM) classifier. Similarly, a superpixel based SVM was employed to separate EP from ST areas in tissue regions of oropharyngeal squamous cell carcinoma in [2].

All the previously proposed methods were based off handcrafted features such as color and texture which aim to simulate the visual perception of human pathologist in interpreting the tissue samples [30]. Recently, however, there has been interested "deep learning" (DL) strategies for classification and analysis of big data. Histopathology, given the data complexity and density, is ideally aligned with deep learning approaches that attempt to use deep architectures to learn complex features from data. DL approaches unlike handcrafted feature approaches represent end-to-end feature learning approach which attempt to learn high-level structural features from a large amount of training data to best discriminate between the classes of interest. The DL approach can thus serve as a good feature extractor for better data representation [18]. In [9], a deep max-pooling convolutional neural network was presented for detecting mitosis in breast histological images. The approach comprised a deep neural network involving a convolutional and a max-pooling layer which were employed to learn the representation of high-level features. Then, a supervised softmax classifier was trained to classify each pixel within a square window as containing a mitotic nucleus or not. In [10], a convolutional neural networks (CNN) and autoencoder were combined for histopathological image representation based learning. Then a softmax classification approach was employed for distinguishing cancerous and non-cancerous tissue. The approach in [10] used a one-layer autoencoder for high-level feature representation. In [28,29], we presented a Stacked Sparse Autoencoder (SSAE) framework for automated nuclear detection from high resolution breast histopathological images. Handcrafted features were combined with CNN features in [26] for mitosis detection in breast cancer pathology. DCNN is a hierarchical neural network which mimics the network structure of neural systems. It is a multi-layer network of interconnected simple "neurons" by connecting links characterized by a weight.

Building on these approaches, in this work, we present a patch based DCNN approach for distinguishing epithelial and stromal compartments within H&E images of breast cancers [8]. Each histologic image is first represented by thousands of cropped sub-images. Two different approaches involving the use of superpixel (SP) and a fixed-size square window is used to generate sub-images from H&E and IHC stained images, respectively. Different from color or intensity based features, such as LBP [19] and texture [6], our approach employs architectural features of atomic regions in the tumor and stroma for tissue classification. The DCNN based feature learning is applied to two classifications of EP and ST patches on (1) IHC stained histologic images of colorectal cancer and (2) on H&E stained images of breast cancer. For simplicity, throughout this paper, we use two different terms "Classification" and "Segmentation" to represent the two different applications, respectively. The classification of EP and ST patches of IHC stained images is an easier task which aims to assign a single label to the respective patch. Segmentation of EP and ST regions is more difficult since it aims to detect the regions of interest (ROIs) and then assign a label to each corresponding ROI. For the classification task, we employed a fixed-size SW to extract candidate sub-images defined via a sliding window scheme. These are then fed to the DCNN for training the network. The flowchart for the classification framework with DCNN is shown in Fig. 2(g)–(k). As the separation of the epithelial and stromal regions from H&E images is a more difficult task, we firstly employ a superpixel based scheme to over-segment the image into atomic regions. Then the atomic regions are resized into fixed-size square images, prior to feeding them to a DCNN for feature learning.

The rest of this paper is organized as follows. A detailed description of DCNN is presented in Section 3. The experimental setup and comparative strategies are presented in Section 4. The experiment results and a discussion of the results are reported in Section 5. Concluding remarks are presented in Section 6.

3. Methods

3.1. The deep convolutional neural networks (DCNN)

The DCNN approach employed in this paper comprises two alternating convolutional layers (or $\mathcal{C}$ layers, see Fig. 1(b)), max-pooling (or $\mathcal{P}$ layers, see Fig. 1(c)), two full connection layers, and a final classification layer. The $\mathcal{C}$ and $\mathcal{P}$ layers produce a convolution and a max-pooling feature map via successive convolution and max-pooling operations, respectively. These feature maps then enable the extraction and combination of a set of appropriate image features from the training exemplars.