Pattern Recognition Letters 000 (2017) 1–8

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images

Zahraa Al-Milaji[a], Ilker Ersoy[b], Adel Hafiane[c], Kannappan Palaniappan[a], Filiz Bunyak[a,*]

[a] Department of Computer Science, University of Missouri, Columbia, MO 65211, United States
[b] Department of Pathology & Anatomical Sciences, University of Missouri, Columbia, MO 65212, United States
[c] INSA Centre Val de Loire, Laboratoire PRISME EA 4229, Bourges F-18000, France

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Epithelium and stroma
H&E images
Convolutional neural networks
Segmentation
Classification

ABSTRACT

Initiation, progression, and therapeutic response in cancer are largely influenced by tumor microenvironment. Segmentation of tumor into epithelial vs. stromal regions constitutes the first step for the study of tumor microenvironment and its effects on disease progression. This paper proposes a new method for stromal vs. epithelial tissue identification from images of H&E stained specimens. The proposed method integrates convolutional neural networks (CNN) based supervised classification with unsupervised image segmentation. The scheme combines the strengths of deep learning (feature learning and classification) with the boundary localization accuracy of image segmentation for enhanced performance. Our experimental results on Stanford Tissue Microarray Database show that integration of CNN classification with explicit image segmentation leads to better adherence of identified class boundaries to actual tissue boundaries and improves the classification accuracy.

1. Introduction

Initiation, progression, and therapeutic response in cancer is largely influenced by tumor microenvironment [19,21,31,36]. Quantitative analysis of tumor microenvironment in histopathology images has potential to predict or measure response to therapy. In cancer, changes in the stroma tend to drive tumor invasion and metastasis. The stroma is essential for the maintenance of epithelial tissues. When the epithelium changes, the stroma inevitably changes also [12]. Therefore, computational analysis can assist researchers and clinicians with quantification of the features that they already assess qualitatively or may lead to discovery of new informative features. In [4], Beck et al. combined image derived morphometric features with image classification and machine learning techniques to study prognostic information from stromal and epithelial regions of breast tumors. Since epithelial and stromal regions have different significance for prognosis, segmentation of tumor into epithelial vs. stromal regions constitutes a first step for many histopathology image analysis tasks [41]. Qu et al. [30] and Linder et al. [23] extract texture features at pixel or block levels and use support vector machine (SVM) classifier to segment H&E stained histopathology images into epithelial and stromal regions. In [8], we proposed an epithelium-stroma classification system that operates on superpixels rather than pixels or blocks. These systems use carefully selected features as input to their epithelium-stroma classifiers. Performance of pattern recognition and classification tasks greatly depends on the features they use [7]. For complex data (i.e. images of H&E stained specimens) manually selected features may not always capture the best representation to delineate the underlying structure.

Deep learning methods [5] enable automatic learning of complex features required for visual pattern recognition. Very recently, deep learning methods have shown outstanding performance in computer vision and pattern recognition tasks, not only in classical computer vision but also in biomedical applications : Ciresan et al. won ISBI brain image segmentation challenge [10], and MICCAI mitosis detection challenge [11]. This deep learning method not only won against hand-crafted pipelines but also got a score comparable to the inter-observer agreement among pathologists [40].

In this paper, we present our image processing and machine learning pipeline developed for stromal vs. epithelial tissue identification from images of H&E stained specimens. We propose a hybrid system that combines strengths of deep learning feature learning and classification capabilities (specifically using deep convolutional neural networks) with boundary localization accuracy of image segmentation. Each histologic image is first sub-divided into patches. Deep convolutional neural networks (CNNs) are trained on these patches to extract hierarchical features from raw pixels of H&E stain images and to perform epithelium vs. stroma classifica-

* Corresponding author.
E-mail address: bunyak@missouri.edu (F. Bunyak).

http://dx.doi.org/10.1016/j.patrec.2017.09.015
0167-8655/© 2017 Published by Elsevier B.V.

tion. The classification results from deep convolutional neural networks are further fused with an explicit segmentation results for improved performance. Also, instead of using raw intensity/color values to segment the images into coherent regions, learned features from CNN networks are used as pixel-wise raw features. Feature maps used in segmentation were obtained by convolving the input image with the filters learned at the first convolutional layer. The resulted feature maps are considered as the feature vector for SLIC and HFCM segmentation. This is the first step of a broader multidisciplinary study to analyze predictive and prognostic value of morphological features of tumor microenvironment.

Very recently, few other studies have used deep learning approaches for epithelium stroma classification [20,43]. Huang et al. [20] use CNNs with unsupervised domain adaptation. However, rather than tissue segmentation classification of image patches is addressed. Xu et al. [43] use CNNs for tissue segmentation. While our paper and Xu et al. [43] rely on CNN classification and superpixels, the two systems differ in (1) CNN architecture, (2) segmentation method, and (3) particularly in fusion scheme used. Xu et al. [43] pre-segment the images, resize the obtained superpixels into fixed-sized square images, and feed them to a CNN. Pre-segmentation makes the whole system too sensitive to segmentation accuracy as well as shape of epithelial and stromal regions. Our post-segmentation approach overcomes both problems, making the system more robust against segmentation method and region shape, while not artificially altering texture of a region. Post-segmentation also allows identification of mixed regions, which can then be further sub-divided to improve the results.

This paper is organized as follows. Section 2 describes details of deep learning based stroma-epithelium classification. Section 3 describes integration of segmentation to classification. Our experimental results on Stanford Tissue Microarray Database [4] are given in Section 4. Section 5 concludes the paper and gives future directions.

2. Stroma-epithelium classification using CNN

Image analysis pipelines traditionally involve a series of steps including pre-processing, image segmentation, training classifiers with carefully selected features, and classification. The performance of these systems is highly dependent on the selected features and the accuracies of the preceding steps. In recent years, deep artificial neural network approaches have shown outstanding performance in computer vision and pattern recognition tasks. Deep learning provides methods that enable automated learning of feature sets for particular problems as opposed to designing and/or selecting features. Convolutional neural network (CNN) is one of the most popular types of deep learning models used in image analysis. CNNs have been mostly used in various image or image block classification tasks including cell or nuclei classification [11,14,25,29,42]. Only very recently, CNNs have started to be adapted to semantic image segmentation tasks [13,26,28,32].

A convolutional neural network is a function g mapping data x (i.e. an image), to an output vector y. The function g is the composition of a sequence of simpler functions f_l, which are called computational blocks or layers; $g = f_L \circ \ldots \circ f_1$. Assuming the network input is $x_0 = x$, and the network outputs are, $x_1, x_2, \ldots, x_L$. Each output $x_l = f_l(x_{l-1}; w_l)$ is computed from the previous output x_{l-1} by applying the function f_l with parameters w_l [39]. The network is called *convolutional* network, because the functions f_l act as a local and translation invariant operator. Stochastic gradient descent (SGD) and backpropagation algorithms are used to fine-tune or learn CNN parameters. CNNs output a vector of class probabilities $\hat{y} = f(x)$ for all image classes or labels.

We have developed a convolution neural network architecture to identify epithelium and stroma regions in H&E stained tissue images. The proposed CNN architecture is implemented using MatConvNet [39] toolbox and consists of eleven layers as described in Table 1. The architecture uses four types of layers: (1) convolution, (2) pooling, (3) fully connected, and (4) dropout. Convolutional layers convolve the output of their previous layers with a set of learnable filters. Most of the existing CNN architectures or pre-trained networks are initially designed for natural or general images which are considerably different than histopathology images. It may be possible to repurpose some of those architectures or pre-trained networks with appropriate transfer learning steps. However, for this particular task where the characteristics of the subject matter and resulting images are far too different than general natural images, we opted to use a specific CNN architecture. Custom architectures are widely used in biomedical classification tasks including the earlier epithelium-stroma classification network in Xu et al. [43]. The proposed architecture was mostly guided by the size of our input patch, which we determined empirically by balancing region purity (Fig. 2) and information support. As CNN input, we have tested various patch sizes and empirically selected 32×32. Input patch size determined number of polling layers in the architecture. Since the number of filters and hidden layers depend on the distribution and complexity of the input patch, as we go in depth down the network we increase the number of filters [34]. We set the filter sizes of conv layers into 5×5 and 4×4 since they are big enough to capture the scale of most edges and salient features in the patch. Larger filters may result in skipping relevant information specially in the first conv layers [22,44]. Dropout layers were added to address the problem of overfitting. The last layer, a fully connected layer, performs the classification task. We initialized the weights of the neurons to small random numbers, known as symmetry breaking. We initialized the neuron biases in the convolutional layers and in the fully-connected hidden layers with the constant 0 since the symmetry breaking is provided by the small random numbers in the weights [6].

Table 1
Architecture of the CNN training model.

Layer number	Layer type	Parameters
Layer 1	Batch normalization	–
Layer 2	Convolution	Kernel number: 32 Kernel size: $5 \times 5 \times 3$
Layer 3	Pooling	Pooling region size: 3×3 Pooling method: max-pooling Activation function: ReLU
Layer 4	Convolution	Kernel number: 32 Kernel size: $5 \times 5 \times 32$ Activation function: ReLU
Layer 5	Pooling	Pooling region size: 3×3 Pooling method: average-pooling
Layer 6	Dropout	Dropout with learning rate: 0.5
Layer 7	Convolution	Kernel number: 64 Kernel size: $5 \times 5 \times 32$ Activation function: ReLU
Layer 8	Pooling	Pooling region size: 3×3 Pooling method: average-pooling
Layer 9	Convolution	Kernel number: 64 Kernel size: $4 \times 4 \times 64$ Activation function: ReLU
Layer 10	Dropout	Dropout with learning rate: 0.5
Layer 11	Fully-connected	Kernel number: 2 Kernel size: $1 \times 1 \times 64$ Activation function: Softmax with log-loss

2.1. Data set

Two (TMA) data sets were evaluated on this work. Netherlands Cancer Institute (NKI) and Vancouver General Hospital (VGH). The data sets consist of 157 images (106 NKI, 51 VGH) of size

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015

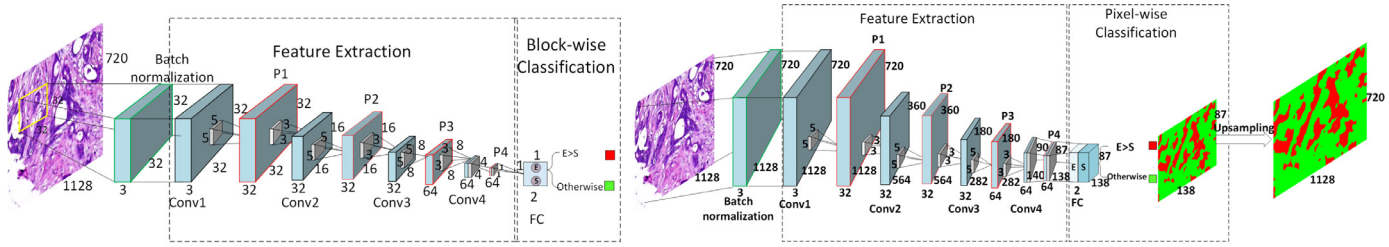


Fig. 1. Architecture of the CNN labeling model using (a) non-overlapping rectangular image blocks; (b) sliding window on whole input image. Conv, P, and FC represent convolutional, pooling, and fully connected layers, respectively.

1128×720 pixels in which Epithelial and Stromal regions were manually annotated by pathologists. For the NKI, five fold cross validation is performed. We used 80% of the images for training and 20% of the images for tesring. For comparison, we used 69 NKI images for training and 37 NKI images for testing. For the VGH, we used 36 images for training and 15 images for testing.

2.2. Training

Original tissue images and corresponding expert labeled ground-truth segmentation masks are partitioned into non-overlapping image patches (Fig. 2). Each patch is normalized to have zero mean and unit variance [27]. Patch normalization is a common practice for CNN [3,15,33,37]. Normalization is particularly important for histopathology images because of possible large variations in staining. The network is trained by comparing extracted labels from corresponding expert labeled images to the sampled patches. Mixed-class patches and patches with large undecided (by pathologist) classes or background regions are discarded in training. We have labeled blocks that have 80% or more of their pixels from one class as pure, remaining blocks as mixed. For our 32×32 block size the percentages of mixed class blocks (not used in training) were 31.19% for NKI dataset and 28.28% for VGH dataset. After a specific number of epochs, the training is stopped. Each convolution layer in those networks learns filter coefficients which represent the discriminative features. CNN classification layer produces an output of two channels. These channels represent confidence scores for epithelium and stroma classes. Class label for a block or pixel is determined by picking the corresponding output channel with the maximum confidence score.

2.3. Testing

The CNN architecture described so far classifies an image *block* as epithelium or stroma. Unlike the usual application of a CNN that involves classifying whole images into one of the classes or detecting the existence of certain objects (faces, vehicles, etc.) within approximate bounding boxes, the goal of the proposed system is to detect epithelium and stroma regions and accurately delineate their boundaries within the images. In order to achieve this goal, we implemented and compared three different labeling schemes. We then integrated the CNN classification output with region segmentation results to improve epithelium/stroma region boundary localization. Here, we will first describe the three CNN labeling schemes. In all cases, training is performed with non-overlapping regular rectangular image patches as described above.

(a) Non-overlapping blocks: This is the simplest form of labeling, where we partition the input image into a set of non-overlapping rectangular image blocks, and each block is fed to the trained CNN independently. A single class label corresponding to epithelium or stroma is obtained for each block. Class labels for the input blocks are stitched together to form a coarse segmentation map for the input image. The resolution of the output image is $1/(bsize \times bsize)$, where *bsize* is block size ($bsize = 32$ for our experiments). Resolution of the output label image can be increased by decreasing block size. However, very small blocks may not capture enough information to reliably classify the local tissue region. Fig. 1(a) shows this scheme.

(b) Overlapping blocks and voting: In order to increase the resolution of the output label map without decreasing block size, we subdivide the input image into blocks overlapped by half block size. Each block is processed as described above. Each pixel in the input image gets covered by four overlapping blocks. Class label for each pixel is then determined by classification with highest confidence score.

(c) Full image labeling: Ideally, a brute force sliding window approach where a block is extracted around each pixel would give the highest resolution for the label map. Yet this technique is computationally very expensive. In order to avoid repeated convolutions for the overlapping regions, Su et al. [35] proposed a CNN labeling scheme where the whole image is directly fed to the network without dividing into regular blocks and equivalent of sliding window class labels are obtained all at once. Due to the pooling downsampling, the output label map is slightly smaller than the input image. Therefore, to retain the original tested image size, this map is upsampled. Fig. 1(b) shows this scheme.

3. Integration of segmentation with CNN

While some recent works on stroma-epithelium classification operate on pixels or blocks [23,30], our pipeline includes an explicit segmentation module. Interactions between supervised classification and unsupervised segmentation occurs at two different levels: (1) learned features by CNN are used as input to segmentation to improve segmentation performance; (2) region boundaries produced by segmentation are used to improve tissue class boundary adherence for classification. This mutual interaction between supervised classification with unsupervised segmentation (Fig. 3) takes full advantage of the strengths of these two classes of approaches. In this study, we used two segmentation approaches: Simple Linear Iterative Clustering (SLIC superpixels) described in [1,38], and HFCM segmentation [8], developed by our group described below.

3.1. Hierarchical fuzzy C-means with spatial constraint (HFCM)

This work extends our works in [9,17,18]. The classical Fuzzy C-means (FCM) algorithm minimizes an objective function defined by the sum of similarity measures. It is an iterative process that minimizes the distance between each point and the prototypes. It does not incorporate any spatial information. Spatial correlation and multiresolution bring robustness and efficiency to the fuzzy c-means algorithm [18]. HFCM extends FCM by incorporating spatial

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015

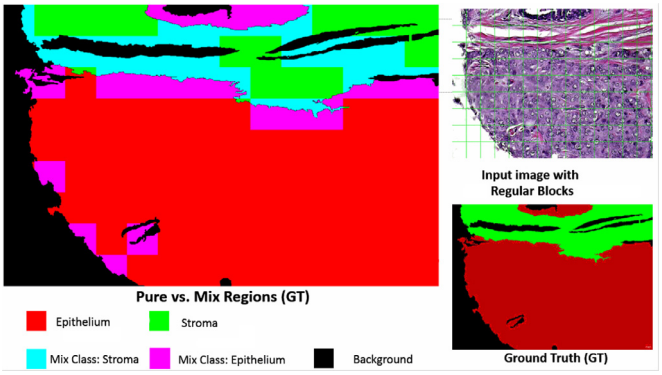


Fig. 2. Ground-truth class label assignment to image patches.

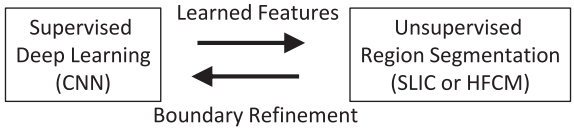


Fig. 3. Mutual interaction between supervised deep learning and unsupervised region segmentation.

information to its objective function:

$$J_{SCM}(U,V) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} \| x_j - v_i \|^2 + \frac{n}{2}\alpha \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} e^{-\sum_{k\in\Omega} u_{ik}^{m}} + \beta \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} f_i^{(n-1)}(x_j) \quad (1)$$

where $X = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N}\}$ denotes data (pixel feature vectors). $V = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_C}\}$ represents the prototypes (clusters centers). Ω is a set of neighbors ($k \neq j$). $f_i^{(n-1)}(x_j)$ is the point x_j's ancestor membership function to the ith cluster in lower layer $(n-1)$. Parameters α and β control the influence of the associated terms. α is multiplied by scale factor $(\frac{n}{2})$ to reduce the effect of spatial constraint at lower levels. The HFCM objective function (1) contains three terms. The first term is the same as in regular FCM. The second term is a spatial penalty that forces neighboring pixels to belong to the same class. It reaches a minimum when the membership value of neighbors for a particular cluster is large. The third term incorporates the relationship between classes of elements at different resolutions for more feature consistency. Optimization of (1) with respect to U is done in a classical way by a Lagrange multiplier technique.

$$J_{SCM}(U,V) = \sum_{j=1}^{N}\lambda_j\Big(1 - \sum_{i=1}^{C} u_{ij}\Big) + \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} \times \left(\| x_j - v_i \|^2 + \frac{n}{2}\alpha e^{-\sum_{k\in\Omega} u_{ik}^{m}} + \beta f_i^{(n-1)}(x_j) \right) \quad (2)$$

After taking the derivative of Eq. (2) vs. u_{ij}, solving for u_{ij}, and solving for λ_j with respect to the constraint eventually leads to the following membership update equation:

$$u_{ij} = \frac{1}{\sum_{p=1}^{C}\left(\frac{\|x_j - v_i\|^2 + \frac{n}{2}\alpha e^{-\sum_{k\in\Omega} u_{ik}^{m}} + \beta f_i^{(n-1)}(x_j)}{\|x_j - v_p\|^2 + \frac{n}{2}\alpha e^{-\sum_{k\in\Omega} u_{pk}^{m}} + \beta f_p^{(n-1)}(x_j)} \right)^{\frac{1}{m-1}}} \quad (3)$$

As in (3), u_{ij}, the membership value of a point j to cluster i, depends on membership values of its neighbors and ancestor in the pyramidal representation. Regularization is controlled by weights α and β. The prototype update equation is the same as in standard FCM, since the second component of (1) does not depend on v_i. Centroids update obeys the equation:

$$v_i = \left(\sum_{j=1}^{N} u_{ij}^{m} x_j\right) \Big/ \left(\sum_{j=1}^{N} u_{ij}^{m}\right) \quad (4)$$

Instead of using raw intensity/color values to segment the images into coherent regions, feature maps from CNN networks were used as pixel-wise raw features. Feature maps used in segmentation were obtained by convolving the input image with the filters learned at the first convolutional layer. The resulted feature maps were considered as the feature vector for SLIC and HFCM segmentation algorithms as shown in Fig. 4. The new features have improved the performance of the segmentation results once used with CNN classification results.

3.2. Integration of region segmentation with CNN classification

Convolutional neural networks (CNNs) are often used as pixel-wise classification tools where a small image patch centered on a pixel is used as input to the classifier and a single class label is obtained. Efficient segmentation using CNNs has only started to be explored very recently [16,24,35].

In this paper, we propose a hybrid system that combines strengths of CNN feature learning and classification capabilities with boundary localization accuracy of image segmentation. The

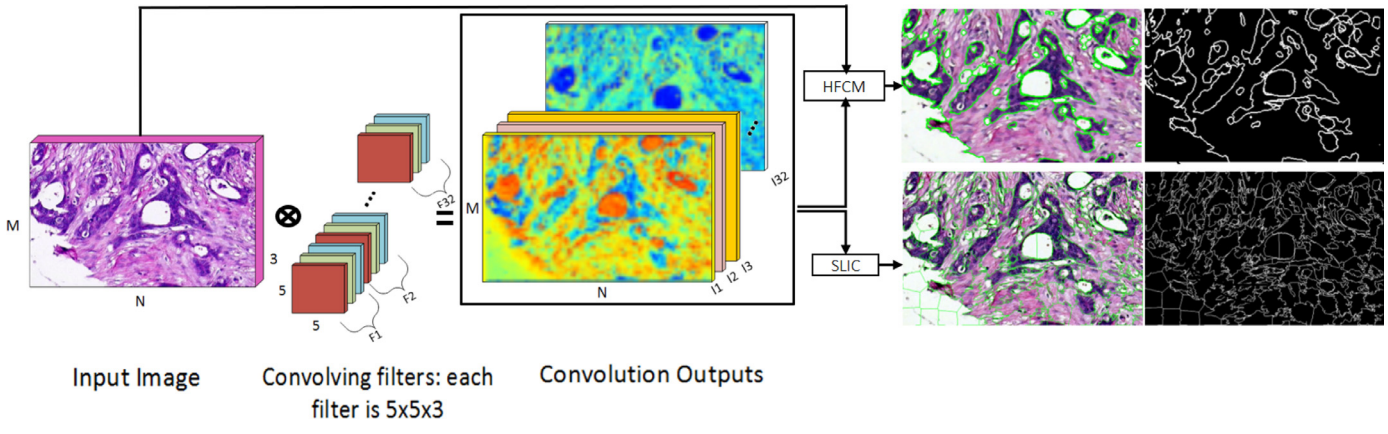


Fig. 4. SLIC/HFCM segmentation on CNN learned feature maps. Feature maps used in segmentation were obtained by convolving the input image with the filters learned at the first convolutional layer of the proposed and trained CNN architecture.

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015

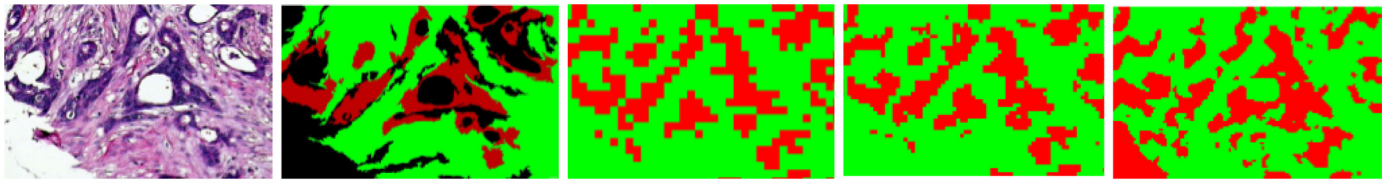


Fig. 5. Sample epithelial vs. stromal tissue classification results using different CNN labeling schemes. (a) original image, (b) manual ground truth (red: epithelium, green: stroma, black: unknown or ignore), (c) non-overlapping blocks, (d) overlapping blocks and voting, (e) full image processing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

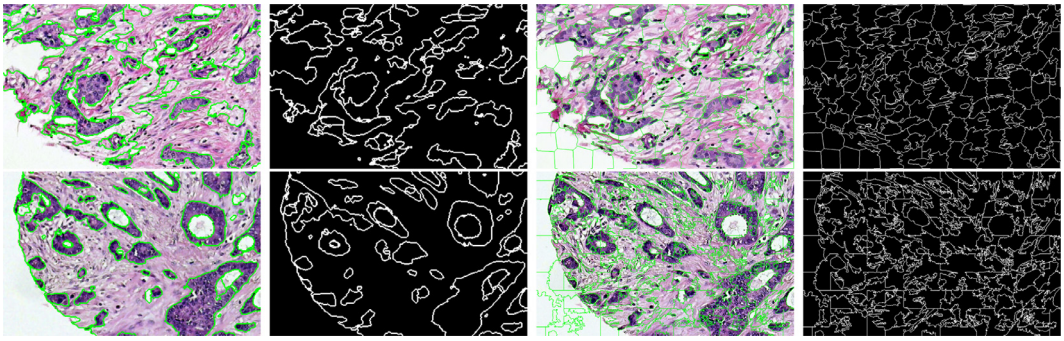


Fig. 6. Sample segmentation results for two images of Stanford TMA database [4]. The results correspond to HFCM partitions and SLIC superpixels [1]. (a) HFCM segmentation boundaries overlaid on original image, (b) HFCM partitions, (c) SLIC segmentation boundaries overlaid on original image, (d) SLIC superpixels.

Table 2
Epithelial vs. stromal tissue identification accuracies for CNN-only and CNN-hybrid systems for three CNN labeling schemes and two region segmentation methods.

Classification	CNN	CNN	CNN	CNN	CNN
Segmentation	/	$SLIC_{RGB}$	$SLIC_{CNN}$	$HFCM_{RGB}$	$HFCM_{CNN+RGB}$
NonOverlappingBlocks	77.73	81.66	83.52	86.22	**88.03**
OverlappingBlocks+Voting	80.14	82.12	84.12	86.16	**87.73**
Full image	81.39	82.79	84.33	86.85	**89.32**

superpixels produced by SLIC or HFCM are labeled by using the label maps produced by the CNN. This scheme has an added advantage of fusing global and local image information. Through training, CNN learns tissue characteristics across multiple images whereas image segmentation relies on information from a single image and better adapts to batch variations between specimen images. Fusion is implemented as follows:

1. Image $\mathcal{P}$ is partitioned into superpixels $\mathcal{P}_i$ using SLIC or HFCM segmentation methods ($\mathcal{P} = \cup\mathcal{P}_i$).
2. CNN classification output is resized to original image size and assigned to a matrix $\mathcal{C}$ where $\mathcal{C}(i, j) \in \{0, 1\}$ corresponds to epithelium or stroma classes.
3. For each partition $\mathcal{P}_i$
   (a) Pixel counts for each class k is computed:
   $$Count(\mathrm{P}_i, k) = \sum_{p_i \in \mathrm{P}_i} \delta(\mathcal{C}(p_i) - k)$$
   (b) Refined class for the partition $\mathcal{C}_R(\mathrm{P}_i)$ is determined as the majority class in region P_i
   $$C_R(\mathrm{P}_i) = \arg\max_{k \in \{0,1\}} (Count(\mathrm{P}_i, k))$$

There is a tradeoff between information content and purity of an image patch. As expected, when the block size becomes larger, blocks get more heterogeneous (mixed class), and assigning the majority class to the center pixel introduces errors (i.e. the center pixel does not necessarily belong to the majority class within the block at the boundaries of the classes). Even if the ground truth is used to assign classes to the blocks, this "quantization error" still happens. On the other hand, smaller blocks do not capture enough texture or structure information and fail to reliably determine tissue type (e.g. tissue class of a single pixel cannot be determined by just its color). While larger blocks contain more context information that can help classification, the likelihood of having both tissue classes in the block also increases with the block size con-

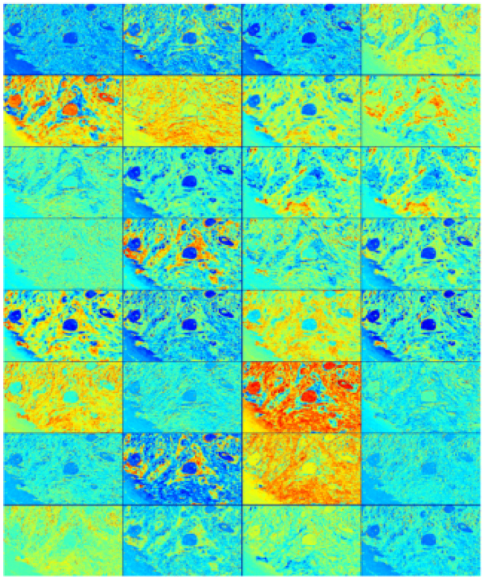
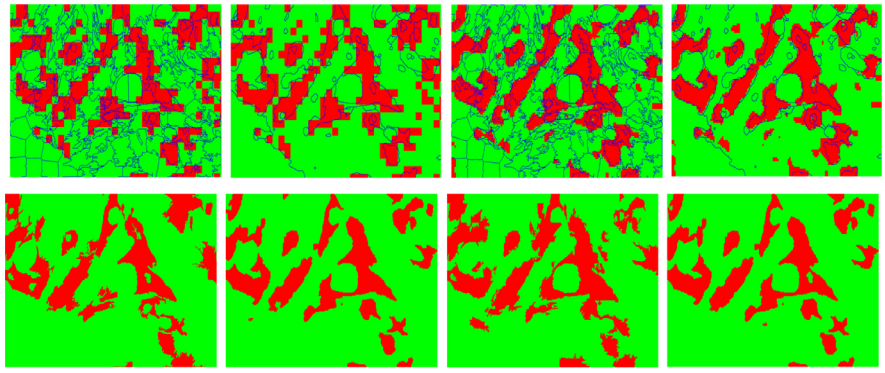


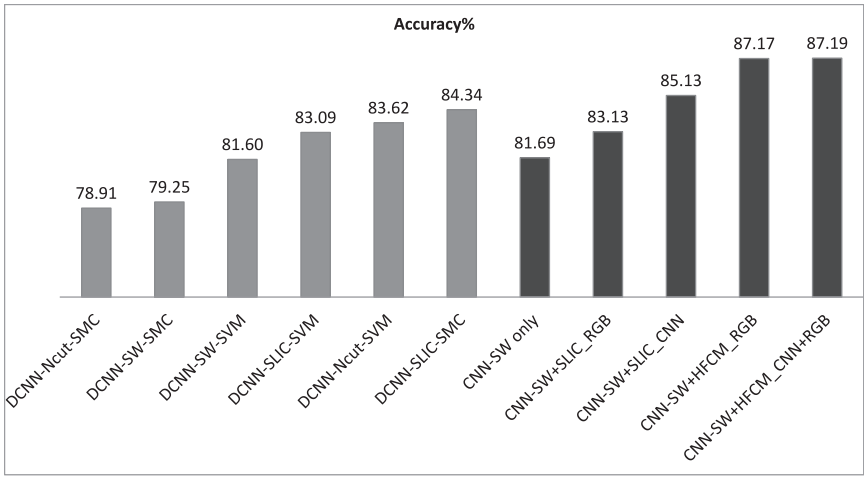
Fig. 7. CNN intermediate representation: $360 \times 564 \times 32$ feature set from the fourth layer when using full image labeling instead of image blocks. These automatically extracted features are also used as inputs to image segmentation.

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015

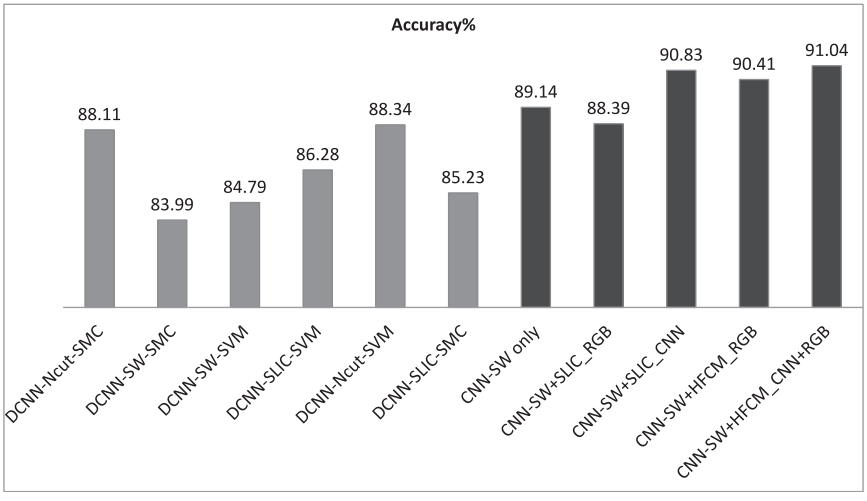


(a) CNN-block + S-LIC (b) CNN-block + H-FCM (c) CNN-full +S LIC (d) CNN-full + H-FCM

Fig. 8. Example of fusing CNN classification with region segmentation. Row 1: boundaries of segmentation overlaid on CNN class labels. Row 2: fused output. Fusion increases CNN-only classification accuracy from 81% to (a) 84% for SLIC, and to (b) 90% for HFCM segmentation in non-overlapping block labeling. Fusion increases CNN-only classification accuracy from 86% to (c) 90% for SLIC, and to (d) 94% for HFCM segmentation in full image labeling.



(a)



(b)

Fig. 9. Experimental results comparing our proposed methods (dark bars) with methods proposed in [43] (gray bars). Accuracies for different combinations of segmentation (none, SLIC Ncut, HFCM), feature sets (RGB only, RGB+CNN learned), and classification (CNN, SVM): (a) on NKI dataset [4], (b) on VGH dataset [4].

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015

Table 3
The quantitative evaluation of classification results on NKI and VGH data sets.

Data sets	Models	TPR	TNR	PPV	NPV	FPR	FDR	FNR	ACC	F1	MCC
NKI	DCNN-SLIC-SMC [43]	86.31	82.15	84.11	84.60	17.85	15.89	13.66	84.34	85.21	68.60
	CNN_{SW} only	81.43	82.89	84.11	80.05	17.11	15.89	18.57	81.69	82.75	64.24
	CNN_{SW}+$HFCM_{CNN+RGB}$	89.48	**85.96**	85.94	89.50	**14.04**	14.06	10.52	**87.19**	87.68	75.44
	CNN_{SW}+$HFCM_{RGB}$	**89.66**	85.92	85.86	**89.71**	14.08	14.14	**10.34**	87.17	**87.72**	**75.58**
	CNN_{SW}+$SLIC_{CNN}$	85.02	85.83	**86.64**	84.13	14.17	**13.36**	14.98	85.13	85.82	70.81
	CNN_{SW}+$SLIC_{RGB}$	83.62	84.46	85.38	82.61	15.54	14.62	16.38	83.83	84.49	68.04
VGH	DCNN-Ncut-SVM [43]	88.29	88.40	89.93	86.55	11.60	10.07	11.71	88.34	89.10	76.59
	CNN_{SW} only	90.32	88.15	92.98	83.97	11.85	7.02	9.68	89.14	91.63	77.70
	CNN_{SW}+$HFCM_{CNN+RGB}$	**91.96**	**92.21**	**95.45**	**86.59**	**7.79**	**4.55**	**8.04**	**91.04**	**93.67**	**83.10**
	CNN_{SW}+$HFCM_{RGB}$	91.16	91.70	95.20	85.15	8.30	4.80	8.84	90.41	93.14	81.60
	CNN_{SW}+$SLIC_{CNN}$	91.66	90.71	94.51	86.17	9.29	5.49	8.34	90.83	93.06	81.52
	CNN_{SW}+$SLIC_{RGB}$	89.03	89.52	94.08	81.37	10.48	5.92	10.97	88.39	91.49	76.99

fusing the classification process. CNN output suffers from boundary localization inaccuracies. On the other hand, image segmentation methods such as SLIC [1,2] or HFCM [8,9] lack the capability to determine tissue class. However, these segmentation methods group pixels together with similar features into superpixels that adhere well to actual tissue boundaries.

4. Experimental results

We have implemented the proposed deep convolutional neural network architecture using MatConvNet toolbox [39]. The network is trained and tested on Stanford Tissue Microarray Database [4] for classification of epithelial vs. stromal tissue. Expert labeled ground truth labels presented in [4] are used for supervised training and evaluation with a five fold cross validation.

Sample CNN classification results obtained by using different labeling schemes are shown in Fig. 5. As expected, full image labeling provides the highest resolution and hence the highest accuracy. Our framework integrates supervised classification with unsupervised segmentation. Two methods of image segmentation are considered SLIC superpixels [1] and our HFCM (Hierarchical Fuzzy C-means with Spatial Constraint) clustering [8,9] (Fig. 6).

We have modified both SLIC and HFCM segmentation schemes to enable use of learned features from CNN in addition to color features. Fig. 7 shows sample multi-scale features learned at intermediate layers of CNN and used as input to segmentation. CNN classification results are further fused with unsupervised region segmentation results to better reflect irregular tissue boundaries as described above. Fig. 8shows sample integrated results. For the sample image in Fig. 8, CNN produces classification results with accuracies of 81% and 86% for non-overlapping block vs. full image labeling schemes. Fusion with SLIC superpixels increases these accuracies to 84% and 90%, respectively. Fusion with HFCM partitions further increases these accuracies to 90% and 94%, respectively.

Table 2 presents a summary of our quantitative performance evaluation results for CNN-only and CNN-hybrid systems for three different labeling schemes, two different region segmentation methods, and two different feature sets (raw color vs. learned CNN features). The highest accuracy is reached by combining CNN classification with HFCM segmentation on RGB color plus CNN learned features.

Very Recently, another epithelium-stroma classification system combining information from CNN networks with region segmentation has been proposed [43]. In Fig. 9, we compare the performances of different configurations of our system (black bars) against different configurations of the system presented in [43] (gray bars). Our results outperformed results presented in [43], in terms of best results 87.19% vs. 84.34% for NKI dataset and 91.04% vs. 88.34% for VGH dataset; in terms of CNN classification only results 83.99% vs. 79.25% for NKI dataset and 89.14% vs. 81.69% for VGH dataset; and in terms of CNN classification combined with SLIC segmentation (SLIC on RGB image for [43] vs. SLIC on CNN learned features for ours). 85.13% vs. 84.34% for NKI dataset and 90.83% vs. 85.23% for VGH dataset; Our best results for both NKI and VGH datasets are obtained from CNN classification integrated with HFCM segmentation using CNN learned features ($CNN + HFCM_{CNN+RGB}$). Best results for [43] are obtained using CNN classification with SLIC segmentation (DCNN-SLIC-SMC) for NKI and using CNN features with SVM classification and Normalized Cut segmentation (DCNN-Ncut-SVM) for VGH dataset.

The quantitative performance for epithelium-stroma classification on NKI and VGH data sets compared to the state of the art is shown in Table 3. True Positive Rate (TPR), True Negative Rate (TNR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Positive Rate (FPR), False Discovery Rate (FDR), False Negative Rate (FNR), Accuracy (ACC), F1 Score (F1), and Matthews Correlation Coefficient (MCC) outperform the approaches described in [43]. Our improved performance are due to three factors:

1. Deeper CNN architecture that better capture complex multi-scale tissues features. Our CNN architecture consists of 4 convolution, 3 pooling, 1 fully connected, 2 dropout layers compared to 2 convolution, 2 pooling, and 2 fully connected layers in [43].
2. HFCM segmentation vs. SLIC or Ncut segmentation. The hierarchical nature of our HFCM segmentation scheme produces coherent partitions with less mixed-class regions, which in turn translate to better classification accuracy.
3. Use of CNN learned features in segmentation (for both SLIC superpixels and HFCM clustering) improves segmentation and translates to better integrated system output.

5. Conclusion and future work

We have presented a hybrid image processing and machine learning system for identification of stromal vs. epithelial tissue regions from images of H&E stained specimens. The proposed system combines strengths of deep learning (feature learning across multiple images and classification capabilities) with those of image segmentation (accuracy of tissue boundary localization). The proposed framework was systemically evaluated on a set of expert annotated images for three different labeling schemes, two different segmentation approaches, and two different feature sets. Combining deep learning with region segmentation shows promising results in capturing complex features of stromal and epithelial tissues. Utilizing automatically extracted features of CNN along with the regular color features improves classification accuracy even further. We are currently working on using the presented approach as a first step in quantitative analysis of morphological features of tumor microenvironment.

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels, Technical Report 149300, EPFL, Lausanne, Switzerland, 2010.
[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[3] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopez, Convolutional neural networks for mammography mass lesion classification, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 797–800.
[4] A.H. Beck, A.R. Sangoi, S. Leung, R.J. Marinelli, T.O. Nielsen, M.J. van de Vijver, R.B. West, M. van de Rijn, D. Koller, Systematic analysis of breast cancer morphology uncovers stromal features associated with survival, Sci. Transl. Med. 3 (108) (2011).
[5] Y. Bengio, Learning deep architectures for AI, Found. Trends® Mach. Learn. 2 (1) (2009) 1–127.
[6] Y. Bengio, Practical recommendations for gradient-based training of deep architectures, in: Neural Networks: Tricks of the Trade, 2012, pp. 437–478.
[7] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.
[8] F. Bunyak, A. Hafiane, Z. Al-Milaji, I. Ersoy, A. Haridas, K. Palaniappan, A segmentation-based multi-scale framework for the classification of epithelial and stromal tissues in h&e images, in: IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM), IEEE, 2015, pp. 450–453.
[9] F. Bunyak, A. Hafiane, K. Palaniappan, Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets, in: Software Tools and Algorithms for Biological Systems, Springer, 2011, pp. 413–424.
[10] D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.
[11] D.C. Cireşan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2013, pp. 411–418.
[12] O. De Wever, M. Mareel, Role of tissue stroma in cancer cell invasion, J. Pathol. 200 (4) (2003) 429–447.
[13] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1915–1929.
[14] Z. Gao, J. Zhang, L. Zhou, L. Wang, HEp-2 cell image classification with convolutional neural networks, in: IEEE Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014, pp. 24–28.
[15] L. Ge, H. Liang, J. Yuan, D. Thalmann, Robust 3d hand pose estimation in single depth images: from single-view CNN to multi-view CNNs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3593–3601.
[16] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587.
[17] A. Hafiane, F. Bunyak, K. Palaniappan, Fuzzy clustering and active contours for histopathology image segmentation and nuclei detection, in: Lecture Notes in Computer Science (ACIVS), vol. 5259, 2008, pp. 903–914, doi:10.1007/978-3-540-88458-3_82.
[18] A. Hafiane, B. Zavidovique, FCM with spatial and multiresolution constraints for image segmentation, Image Analysis and Recognition, Springer, 2005.
[19] D. Hanahan, L.M. Coussens, Accessories to the crime: functions of cells recruited to the tumor microenvironment, Cancer Cell 21 (3) (2012) 309–322.
[20] Y. Huang, H. ZHENG, C. LIU, X. Ding, G. Rohde, Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images, IEEE J. Biomed. Health Inf. (2017).
[21] M.R. Junttila, F.J. de Sauvage, Influence of tumour micro-environment heterogeneity on therapeutic response, Nature 501 (7467) (2013) 346–354.
[22] A. Krizhevsky, G. Hinton, Convolutional deep belief networks on CIFAR-10, Unpublished manuscript 40(2010).
[23] N. Linder, J. Konsti, R. Turkki, E. Rahtu, M. Lundin, S. Nordling, C. Haglund, T. Ahonen, M. Pietikäinen, J. Lundin, Identification of tumor epithelium and stroma in tissue microarrays using texture analysis, Diagn. Pathol. 7 (1) (2012) 22.
[24] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: IEEE Conf. on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[25] C.D. Malon, E. Cosatto, Classification of mitotic figures with convolutional neural networks and seeded blob features, J. Pathol. Inf. 4 (2013).
[26] M. Mostajabi, P. Yadollahpour, G. Shakhnarovich, Feedforward semantic segmentation with zoom-out features, in: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition, 2015, pp. 3376–3385.
[27] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proceedings of the 27th Int. Conf. on Machine Learning (ICML-10), 2010, pp. 807–814.
[28] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE Int. Conf. on Computer Vision, 2015, pp. 1520–1528.
[29] B. Pang, Y. Zhang, Q. Chen, Z. Gao, Q. Peng, X. You, Cell nucleus segmentation in color histopathological imagery using convolutional networks, in: Chinese Conf. on Pattern Recognition (CCPR), 2010, pp. 1–5.
[30] A. Qu, J. Chen, L. Wang, J. Yuan, F. Yang, Q. Xiang, N. Maskey, G. Yang, J. Liu, Y. Li, Two-step segmentation of hematoxylin-eosin stained histopathological images for prognosis of breast cancer, in: 2014 IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM), IEEE, 2014, pp. 218–223.
[31] D.F. Quail, J.A. Joyce, Microenvironmental regulation of tumor progression and metastasis, Nat. Med. 19 (11) (2013) 1423–1437.
[32] H.R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E.B. Turkbey, R.M. Summers, DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015, pp. 556–564.
[33] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, F. Moreno-Noguer, Discriminative learning of deep convolutional feature point descriptors, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 118–126.
[34] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Preprint arXiv:1409.1556 (2014).
[35] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, L. Yang, Region segmentation in histopathological breast cancer images using deep convolutional neural network, in: IEEE Int. Symposium on Biomedical Imaging (ISBI), 2015, pp. 55–58.
[36] T.D. Tlsty, L.M. Coussens, Tumor stroma and regulation of cancer development, Annu. Rev. Pathol. Mech. Dis. 1 (2006) 119–150.
[37] E.S. Varnousfaderani, S. Yousefi, A. Belghith, M.H. Goldbaum, Luminosity and contrast normalization in color retinal images based on standard reference image., in: Medical Imaging: Image Processing, 2016, p. 97843N.
[38] A. Vedaldi, B. Fulkerson, VLFeat: an open and portable library of computer vision algorithms, in: Proc. Int. Conf. Multimedia, 2010.
[39] A. Vedaldi, K. Lenc, MatConvNet – convolutional neural networks for MATLAB, in: Proceeding of the ACM Int. Conf. on Multimedia, 2015.
[40] M. Veta, P.J. van Diest, S.M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A.B. Larsen, J.S. Vestergaard, A.B. Dahl, et al., Assessment of algorithms for mitosis detection in breast cancer histopathology images, Med. Image Anal. 20 (1) (2015) 237–248.
[41] M. Veta, J.P. Pluim, P.J. van Diest, M. Viergever, et al., Breast cancer histopathology image analysis: a review, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1400–1411.
[42] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, A. Madabhushi, Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features, J. Med. Imaging 1 (3) (2014) 034003.
[43] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223.
[44] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, Preprint arXiv:1311.2901 (2013).

Please cite this article as: Z. Al-Milaji et al., Integrating segmentation with deep learning for enhanced classification of epithelial and stromal tissues in H&E images, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.09.015