Original citation:
Sirinukunwattana, Korsuk, Raza, Shan-e-Ahmed, Tsang, Yee-Wah, Snead, David, Cree, Ian A. and Rajpoot, Nasir M. (Nasir Mahmood). (2016) Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Transactions on Medical Imaging. doi: 10.1109/TMI.2016.2525803

Permanent WRAP url:
http://wrap.warwick.ac.uk/77351

For more information, please contact the WRAP Team at: publications@warwick.ac.uk

http://wrap.warwick.ac.uk

Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images

Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David R. J. Snead, Ian A. Cree
& Nasir M. Rajpoot, *Senior Member, IEEE*

Abstract—Detection and classification of cell nuclei in histopathology images of cancerous tissue stained with the standard hematoxylin and eosin stain is a challenging task due to cellular heterogeneity. Deep learning approaches have been shown to produce encouraging results on histopathology images in various studies. In this paper, we propose a Spatially Constrained Convolutional Neural Network (SC-CNN) to perform nucleus detection. SC-CNN regresses the likelihood of a pixel being the center of a nucleus, where high probability values are spatially constrained to locate in the vicinity of the center of nuclei. For classification of nuclei, we propose a novel Neighboring Ensemble Predictor (NEP) coupled with CNN to more accurately predict the class label of detected cell nuclei. The proposed approaches for detection and classification do not require segmentation of nuclei. We have evaluated them on a large dataset of colorectal adenocarcinoma images, consisting of more than 20,000 annotated nuclei belonging to four different classes. Our results show that the joint detection and classification of the proposed SC-CNN and NEP produces the highest average F1 score as compared to other recently published approaches. Prospectively, the proposed methods could offer benefit to pathology practice in terms of quantitative analysis of tissue constituents in whole-slide images, and could potentially lead to a better understanding of cancer.

Index Terms—Nucleus Detection, Histology Image Analysis, Deep Learning, Convolutional Neural Network.

I. INTRODUCTION

TUMORS contain a high degree of cellular heterogeneity due to their ability to elicit varying levels of host inflammatory response, angiogenesis and tumor necrosis amongst other factors involved in tumor development [1], [2]. The spatial arrangement of these heterogeneous cell types has also been shown to be related to cancer grades [3], [4]. Therefore, the qualitative and quantitative analysis of different types of tumors at cellular level can not only help us in better understanding of tumor but also explore various options for cancer treatment. One way to explore these cell types is to use multiple protein markers which can mark different cells in cancer tissues. That will, however, require deep biological understanding of tumors to identify informative markers and is costly in terms of laboratory time and the use of tissue to study cell types [5]. An alternative and more efficient approach which is taken in this paper is to use morphological clues in local neighborhoods to develop automated cellular recognition via image analysis which can then be deployed for sophisticated tissue morphometry in future [6].

There are several factors that hinder automatic approaches for detection and classification of cell nuclei. First of all, the inferior image quality may arise due to poor fixation and poor staining during the tissue preparation process or autofocus failure during the digitization of slide. On the other hand, complex tissue architecture (Fig 1a), clutter of nuclei, and diversity of nuclear morphology pose a challenging problem. Particularly, in case of colorectal adenocarcinoma, dysplastic and malignant epithelial cells often have irregular chromatin texture and appear highly cluttered together with no clear boundary, which makes detection of individual nuclei a challenging task (Fig 1b). Variability in the appearance of the same type of nuclei, both within and across different sample, is another factor that makes classification of individual nucleus equivalently difficult.

In this paper, we present novel *locality sensitive* deep learning approaches to detect and classify nuclei in routine hematoxylin and eosin (H&E) stained histopathology images of colorectal adenocarcinoma, based on convolutional neural networks (CNNs). Standard CNN based methods follow a sliding window approach whereby the sliding window is centered at the pixel to be labeled or regressed. Our locality sensitive deep learning approach is based on two premises: (*a*) distance from the center of an object (nucleus, in this case) should be incorporated into the calculation of probability map for detecting that object, and (*b*) a weighted ensemble of local predictions for a class label can yield more accurate labeling of an object. For detection, we propose a spatially constrained CNN (SC-CNN), a new variant of CNN that includes parameter estimation layer and spatially constrained layer for spatial regression. SC-CNN can be trained to predict the probability of a pixel being the center of a nucleus. As opposed to other approaches [7], [8] that do not enforce the pixels

Supplementary material is available in the supplementary files /multimedia tab.

K. Sirinukunwattana, S.E.A. Raza are with the Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK. e-mail: K.Sirinukunwattana@warwick.ac.uk

Y.W. Tsang, D.R.J. Snead and I.A. Cree are with the Department of Pathology, University Hospitals Coventry and Warwickshire, Walsgrave, Coventry, CV2 2DX, UK.

N.M. Rajpoot is with the Department of Computer Science and Engineering, Qatar University, Qatar and the Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK. email: Nasir.Rajpoot@ieee.org
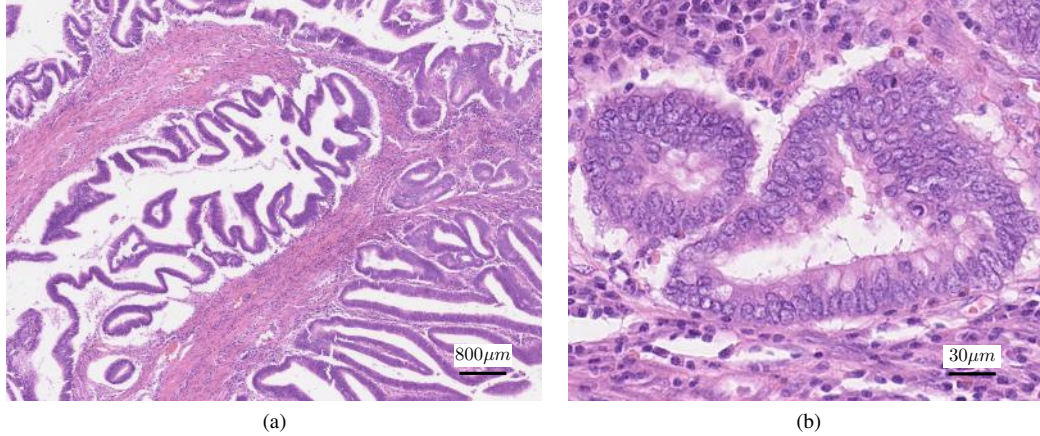
| (a) | (b) |

Fig. 1. An example of the cancer region in colorectal adenocarcinoma. (a) disarray of intestinal glandular architecture. (b) clutter of malignant epithelial nuclei due to high proliferation rate.

close to the center of a nucleus to have a higher probability value than those further away, the predicted probability values produced by SC-CNN are topologically constrained such that high probability values are concentrated in the vicinity of the center of nuclei. For classification, we introduce neighboring ensemble predictor (NEP) to be used in conjunction with a standard softmax CNN. This predictor based on spatial ensembling leverages all relevant patch-based predictions in the local neighborhood of the nucleus to be classified, which in turn produces more accurate classification results than its single-patch based counterpart. Moreover, the proposed approaches for detection and classification do not require the difficult step of nucleus segmentation [9] which can be fairly challenging due to the reasons mentioned above. The proposed approach for detection and classification uses a sliding window to train the networks on small patches instead of the whole image. The use of small patches not only increases the amount of training data which is crucial for CNNs, but essentially also localizes our analysis to small nuclei in images [10], [11].

The organization of the paper is as follows. The review of recent literature on cell and nucleus detection and classification is given in Section II. We briefly introduce CNNs in Section III. Sections IV-VI describe the proposed approach in detail, and experimental results are presented in Section VII. Finally, we discuss some potential applications of the work, and conclude in Sections VIII and IX, respectively.

II. RELATED WORK

Most existing methods for cell and nucleus detection can be classified into one of the following categories: thresholding followed by morphological operations, region growing, level sets, k-means, and graph-cuts. Recently proposed techniques for nucleus detection in routine H&E histology images rely on morphological features such as symmetry and stability of the nuclear region to identify nuclei [12]. Cosatto et al. [13] proposed detection of cell nuclei using difference of Gaussian (DoG) followed by Hough transform to find radially symmetrical shapes. Al Kofahi et al. [14] proposed a graph cut-

based method that is initialized using response of the image to Laplacian of Gaussian (LoG) filter. Kuse et al. [15] employed local phase symmetry to detect bilaterally symmetrical nuclei. Similarly, Veta et al. [16] relied on the direction of gradient to identify the center of nuclei. These methods may fail to detect spindle-like nuclei and irregular-shaped malignant epithelial nuclei. Arteta et al. [17] employed maximally stable extremal regions for detection, which is likely to fall victim to weakly stained nuclei or nuclei with irregular chromatin texture. Ali et al. [18] proposed an active contour-based approach to detect and segment overlapping nuclei based on shape models, which is highly variable in the case of tumor nuclei. In a recent study, Vink et al. [19] employed AdaBoost classifier to train two detectors, one using pixel-based features and the other based on Haar-like features, and merged the results of two detectors to detect the nuclei in immunohistochemistry stained breast tissue images. The performance of the method was found to be limited when detecting thin fibroblasts and small nuclei.

Cell and nucleus classification have been applied to diverse histopathology related applications. Dalle et al. [20] and Cosatto et al. [13] used shape, texture and size of nuclei for nuclear pleomorphism grading in breast cancer images. Malon et al. [21] trained CNN classifier to classify mitotic and non-mitotic cells using color, texture and shape information. Nguyen et al. [22] classified nuclei on the basis of their appearance into cancer and normal nuclei and used the location of detected nuclei to find cancer glands in prostate cancer. Yuan et al. [9] classified nuclei into cancer, lymphocyte or stromal based on the morphological features in H&E stained breast cancer images. This requires accurate segmentation of all the nuclei in the tissue including the tumor nuclei which is not straight forward due to the complex micro-architecture of tumor. Shape features have also been used in an unsupervised manifold learning framework [23] to discriminate between normal and malignant nuclei in prostate histology images. Recently, Sharma et al. [24] proposed nuclei segmentation and classification using AdaBoost classifier using intensity, morphological and texture features. The work was focused on nuclei segmentation, whereas evaluation on classification

performance was limited.

Recent studies have shown that deep learning methods produce promising results on a large number of histopathological image datasets. Wang *et al.* [25] proposed a cascaded classifier which uses a combination of *hand-crafted* features and features learned through CNN to detect mitotic cells. Cruz *et al.* [26] showed that for detection of basal-cell carcinoma, features learned using deep learning approaches produce superior and stable results compared to pre-defined bag of feature and canonical (discrete cosine transform/wavelet transform based) representations. For nucleus detection, Xu *et al.* [7] used stacked sparse autoencoder to learn a high-level representation of nuclear and non-nuclear objects in an unsupervised fashion. The higher level features are classified as nuclear or non-nuclear regions using a softmax classifier. Ciresan *et al.* [8] utilized CNNs for mitotic figure detection in breast cancer images, where CNNs were trained to regress the probability of belonging to a mitotic figure for each pixel, taking a patch centered at the pixel as context. Xie *et al.* [27] recently proposed structural regression CNN capable of learning a proximity map of cell nuclei and was shown by the authors to provide more accurate detection results. Another closely related deep learning work is by Xie *et al.* [28], which localizes nucleus centroids through a voting scheme.

In summary, most of the cell detection methods rely on shape of nuclei and stability of features for cell detection and on texture of the nuclei for classification. Deep learning approaches have been successfully used in the past for identification of nuclei in histopathology images, mostly for binary classification using pixel based approaches. In this paper, we propose a novel deep learning approach which is sensitive to local neighborhood and is capable of detecting and assigning class labels to multiple types of nuclei.

In terms of methodology for nucleus detection, our SC-CNN is closely related to the methods proposed by Xie *et al.* [27], [28]. SC-CNN contains a layer that is specifically designed to predict the centroid locations of nuclei as well as the confidence of the locations whether they correspond to true centroids. These parameters are in essence similar to the voting offset vectors and voting confidence in Xie *et al.* [28]. However, SC-CNN generates a probability mask for spatial regression based on these parameters rather than employing them directly for detection as done by Xie *et al.* [28]. Although, both Xie *et al.* [27] and our SC-CNN share the idea of using spatial regression as a means of localizing the nucleus centers, the regression in SC-CNN is model-based, which explicitly constrains the output form of the network. Thus, high probability values are likely to be assigned to the pixel locations in the vicinity of nucleus centers.

III. Convolutional Neural Network

A convolutional neural network (CNN) f is a composition of a sequence of L functions or layers $(f_1, .., f_L)$ that maps an input vector $\mathbf{x}$ to an output vector $\mathbf{y}$, i.e.,

$$
\begin{aligned}
\mathbf{y} &= f(\mathbf{x}; \mathbf{w}_1, ..., \mathbf{w}_L) \\
&= f_L(\,\cdot\,; \mathbf{w}_L) \circ f_{L-1}(\,\cdot\,; \mathbf{w}_{L-1}) \circ ... \\
&\quad \circ f_2(\,\cdot\,; \mathbf{w}_2) \circ f_1(\mathbf{x}; \mathbf{w}_1),
\end{aligned} \tag{1}
$$

where $\mathbf{w}_l$ is the weight and bias vector for the lth layer f_l. Conventionally, f_l is defined to perform one of the following operations: a) convolution with a bank of filters; b) spatial pooling; and c) non-linear activation. Given a set of N training data $\{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^N$, we can estimate the vectors $\mathbf{w}_1, .., \mathbf{w}_L$ by solving the optimization problem

$$
\underset{\mathbf{w}_1, ..., \mathbf{w}_L}{\mathrm{argmin}} \; \frac{1}{N} \sum_{i=1}^N \ell(f(\mathbf{x}^{(i)}; \mathbf{w}_1, ..., \mathbf{w}_L), \mathbf{y}^{(i)}), \tag{2}
$$

where ℓ is an appropriately defined loss function. Numerical optimization of (2) is often performed via backpropagation and stochastic gradient descent methods.

IV. Nucleus Detection

A. Spatially Constrained Regression

In regression analysis, given a pair of input $\mathbf{x}$ and output $\mathbf{y}$, the task is to estimate a function g that represents the relationship between both variables. The output $\mathbf{y}$, however, may not only depend on the input $\mathbf{x}$ alone, but also on the topological domain (time, spatial domain, etc.) on which it is residing.

Let Ω be the spatial domain of $\mathbf{y}$, and suppose that the spatially constrained regression model g is known *a priori*, and is of the form $\mathbf{y} = g(\Omega; \boldsymbol{\theta}(\mathbf{x}))$ where $\boldsymbol{\theta}(\mathbf{x})$ is an unknown parameter vector. We can employ CNN to estimate $\boldsymbol{\theta}(\mathbf{x})$ by extending the standard CNN such that the last two layers (f_{L-1}, f_L) of the network are defined as

$$
\boldsymbol{\theta}(\mathbf{x}) = f_{L-1}(\mathbf{x}_{L-2}; \mathbf{w}_{L-1}), \tag{3}
$$

$$
\mathbf{y} = f_L(\Omega; \boldsymbol{\theta}(\mathbf{x})) \equiv g(\Omega; \boldsymbol{\theta}(\mathbf{x})), \tag{4}
$$

where $\mathbf{x}_{L-2}$ is an output of the $(L-2)$th layer of the network, (3) is the new parameter estimation layer and (4) is the layer imposing the spatial constraints.

B. Nucleus Detection Using a Spatially Constrained CNN

Given an image patch $\mathbf{x} \in \mathbb{R}^{H \times W \times D}$ with its height H, width W, and the number of its features D, our aim is to detect the center of each nucleus contained in $\mathbf{x}$ (Fig. 2a). To tackle this problem, we first define the training output $\mathbf{y} \in [0, 1]^{H' \times W'}$ as a probability map of size $H' \times W'$ (Fig. 2b). Let $\Omega = \{1, ..., H'\} \times \{1, ..., W'\}$ be the spatial domain of $\mathbf{y}$. The jth element of $\mathbf{y}$, $j = 1, ..., |\Omega|$, is defined as

$$
y_j = \begin{cases} \frac{1}{1+(\|\mathbf{z}_j - \mathbf{z}_m^0\|_2^2)/2} & \text{if } \begin{aligned} &\forall m \neq m', \|\mathbf{z}_j - \mathbf{z}_m^0\|_2 \\ &\leq \|\mathbf{z}_j - \mathbf{z}_{m'}^0\|_2 \leq d, \end{aligned} \\ 0 & \text{otherwise,} \end{cases} \tag{5}
$$

where $\mathbf{z}_j$ and $\mathbf{z}_m^0$ denote the coordinates of y_j and the coordinates of the center of the mth nucleus on Ω, respectively,
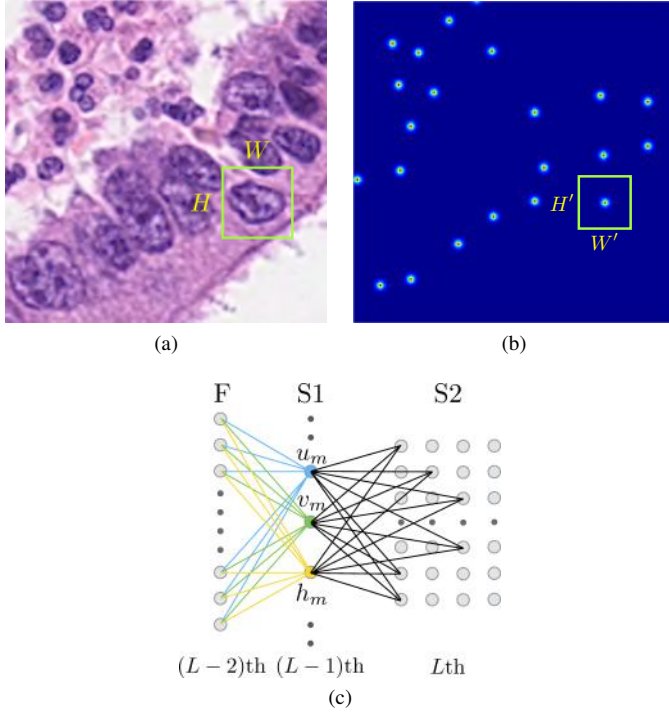
Fig. 2. An illustration of the proposed spatially constrained CNN. (a) An input patch $\mathbf{x}$ of size $H \times W$ from an image. (b) A training output patch $\mathbf{y}$ of size $H' \times W'$ from a probability map showing the probability of being the center of nuclei. (c) An illustration of the last three layers of the proposed spatially constraint CNN. Here, F is the fully connected layer, S1 is the new parameter estimation layer, S2 is the spatially constrained layer, and L is the total number of layers in the network.

and d is a constant radius. Pictorially, the probability map defined by (5) has a high peak in the vicinity of the center of each nucleus $\mathbf{z}_m^0$ and flat elsewhere.

Next, we define the predicted output $\widehat{\mathbf{y}}$, generated from the spatially constrained layer (S2, the Lth layer) of the network (Fig. 2c). Following the known structure of the probability map in the training output described in (5), we define the jth element of the predicted output $\widehat{\mathbf{y}}$ as

$$\widehat{y}_j = g(\mathbf{z}_j; \hat{\mathbf{z}}_1^0, ..., \hat{\mathbf{z}}_M^0, h_1, ..., h_M)$$
$$= \begin{cases} \left(\frac{1}{1+(\|\mathbf{z}_j - \hat{\mathbf{z}}_m^0\|_2^2)/2} \right) h_m & \text{if } \begin{array}{l} \forall\, m \neq m', \|\mathbf{z}_j - \hat{\mathbf{z}}_m^0\|_2 \\ \leq \|\mathbf{z}_j - \hat{\mathbf{z}}_{m'}^0\|_2 \leq d, \end{array} \\ 0 & \text{otherwise,} \end{cases}$$
$$(6)$$

where $\hat{\mathbf{z}}_m^0 \in \Omega$ is the estimated center and $h_m \in [0,1]$ is the height of the mth probability mask, and M denotes the maximum number of the probability masks on $\widehat{\mathbf{y}}$. Note that $\widehat{\mathbf{y}}$ defined in this way allows the number of predicted nuclei to vary from 0 to M because of the redundancy provided by $h_m = 0$ or $\hat{\mathbf{z}}_m^0 = \hat{\mathbf{z}}_{m'}^0$ for $m \neq m'$. In our experiments, we set d in (5) and (6) to 4 pixels as to provide enough support area to the probability mask.

The parameters $\hat{\mathbf{z}}_m^0 = (u_m, v_m)$ and h_m are estimated in the parameter estimation layer (S1, the $(L-1)$th layer, as shown in Fig. 2c). Let $\mathbf{x}_{L-2}$ be the output of the $(L-2)$th layer of

the network. We define u_m, v_m, h_m as

$$u_m = (H' - 1) \cdot sigm(\mathbf{W}_{L-1,u_m} \cdot \mathbf{x}_{L-2} + b_{u_m}) + 1, \quad (7)$$
$$v_m = (W' - 1) \cdot sigm(\mathbf{W}_{L-1,v_m} \cdot \mathbf{x}_{L-2} + b_{v_m}) + 1, \quad (8)$$
$$h_m = sigm(\mathbf{W}_{L-1,h_m} \cdot \mathbf{x}_{L-2} + b_{h_m}), \quad (9)$$

where $\mathbf{W}_{L-1,u_m}, \mathbf{W}_{L-1,v_m}, \mathbf{W}_{L-1,h_m}$ denote the weight vectors and $b_{u_m}, b_{v_m}, b_{h_m}$ denote the bias variables, and $sigm(\cdot)$ denotes the sigmoid function.

To learn all the variables (i.e., weight vectors and bias values) in the network, we solve (2) using the following loss function:

$$\ell(\mathbf{y}, \widehat{\mathbf{y}}) = \sum_j (y_j + \epsilon) H(y_j, \widehat{y}_j), \quad (10)$$

where $H(y_j, \widehat{y}_j)$ is the cross-entropy loss [29]–[31] defined by

$$H(y_j, \widehat{y}_j) = -\Big[y_j \log(\widehat{y}_j) - (1 - y_j) \log(1 - \widehat{y}_j) \Big], \quad (11)$$

and ϵ is a small constant, which is set to be the ratio of the total number of non-zero probability pixels and zero probability pixels in the training output data. The first term of the product in (10) is a weight term that penalizes the loss contributed by the output pixels with small probability. This is crucial as, in the training output data, there are a large number of pixels with zero probability as compared to the non-zero probability ones.

Finally, to detect the center of nuclei from a big image, we use the sliding window strategy with overlapping windows. Since we use full-patch regression, the predicted probability of being the center of a nucleus is generated for each of the extracted patches using (6). These results are then aggregated to form a probability map. That is, for each pixel location, we average the probability values from all the patches containing that pixel. The final detection is obtained from the local maxima found in the probability map. In order to avoid over-detection, a threshold defined as a fraction of the maximum probability value found on the probability map is introduced. All local maxima whose probability values are less than the threshold are not considered in the detection. In our experiments, the threshold was empirically determined from the training data set.

V. NUCLEUS CLASSIFICATION

We treat the problem of nucleus classification as patch-based multiclass classification. Let $\mathbf{x} \in \mathbb{R}^{H \times W \times D}$ be an input patch containing a nucleus to be classified in the proximity of its center, and let $c \in \{1, ..., C\}$ denote the corresponding label of patch $\mathbf{x}$, in which C is the total number of classes.

We train a CNN to produce an output vector $\widehat{\mathbf{y}}(\mathbf{x}) \in \mathbb{R}^C$ in the last layer of the network such that

$$c = \operatorname*{argmax}_j \widehat{y}_j(\mathbf{x}), \quad (12)$$

where $\widehat{y}_j(\mathbf{x})$ denotes the jth element of $\widehat{\mathbf{y}}(\mathbf{x})$. The following logarithmic loss is employed to train the network via solving (2):

$$\ell(c, \widehat{\mathbf{y}}(\mathbf{x})) = -\log(p_c(\widehat{\mathbf{y}}(\mathbf{x}))), \quad (13)$$
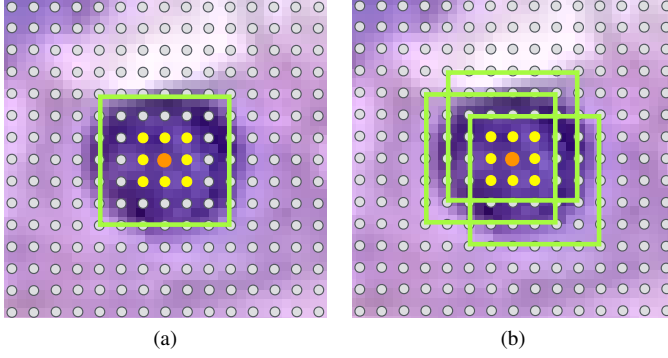
Fig. 3. An illustration of the prediction strategies used in conjunction with softmax CNN: (a) standard single-patch predictor; (b) neighboring ensemble predictor. An orange dot represents the center of the detected nucleus. The center of neighboring patches is represented as yellow dots.

and $p_j(\widehat{\mathbf{y}}(\mathbf{x}))$ is a softmax function [29]–[31] defined by

$$p_j(\widehat{\mathbf{y}}(\mathbf{x})) = \frac{\exp(\widehat{y}_j(\mathbf{x}))}{\sum_k \exp(\widehat{y}_k(\mathbf{x}))}. \qquad (14)$$

We refer to this classification CNN as softmax CNN. In this work, we employ two strategies for predicting the class label as described below.

A. Standard Single-Patch Predictor (SSPP)

Predicted class $\widehat{c}$ for input patch $\mathbf{x}$ with corresponding output $\widehat{\mathbf{y}}(\mathbf{x})$ from the network is given by

$$\widehat{c} = \operatorname*{argmax}_j p_j(\widehat{\mathbf{y}}(\mathbf{x})), \qquad (15)$$

where $\widehat{y}_j(\mathbf{x})$ denotes the jth element of $\widehat{\mathbf{y}}(\mathbf{x})$. See Fig. 3a for illustration.

B. Neighboring Ensemble Predictor (NEP)

Let $\mathcal{X}_{\mathrm{I}} \subset \mathbb{R}^{W \times H \times D}$ denote a set of all $W \times H \times D$ patches of image I, and let Ω_{I} denote a spatial domain of image I. For each patch $\mathbf{x} \in \mathcal{X}_{\mathrm{I}}$, denote $\mathbf{z}(\mathbf{x}) \in \Omega_{\mathrm{I}}$ as the location at which $\mathbf{x}$ is centered. We now define a set of neighboring patches of $\mathbf{x}$ as

$$\mathcal{B}(\mathbf{x}) = \left\{ \mathbf{x}^{(i)} \in \mathcal{X}_{\mathrm{I}} : \|\mathbf{z}(\mathbf{x}^{(i)}) - \mathbf{z}(\mathbf{x})\|_2 \leq d_\beta \right\}, \qquad (16)$$

which contains all patches centered in the ball of radius d_β centered at $\mathbf{z}(\mathbf{x})$.

Predicted class $\widehat{c}$ for an input patch $\mathbf{x}$ with corresponding network output $\widehat{\mathbf{y}}$ is given by

$$\widehat{c} = \operatorname*{argmax}_j \sum_{\mathbf{x}^{(i)} \in \mathcal{B}(\mathbf{x})} w(\mathbf{z}(\mathbf{x}^{(i)})) p_j(\widehat{\mathbf{y}}(\mathbf{x}^{(i)})), \qquad (17)$$

where $w : \Omega_{\mathrm{I}} \to \mathbb{R}$ is a function assigning weight to patch $\mathbf{x}^{(i)}$ based on its center position $\mathbf{z}(\mathbf{x}^{(i)})$. Ensemble predictor (17), defined in this way, is essentially a weighted sum of all relevant predictors (Fig. 3b). It takes into account uncertainty in detection of the center location, as well as, variability in the appearance of nuclei. In our experiments (Section VII-C), the

NEP (17) shows higher classification performance than that of the single-patch predictor (15).

In all experiments, we set $d_\beta = 4$ pixels in (16) so as to allow patches in $\mathcal{B}(\mathbf{x})$ to cover the majority area of a nucleus to be classified, and we set a uniform weight $w(\mathbf{z}(\mathbf{x}^{(i)})) = 1/|\mathcal{B}(\mathbf{x})|$ for all $\mathbf{x}^{(i)} \in \mathcal{B}(\mathbf{x})$ in (17).

VI. Network Architecture and Training Details

A. Architectures

The detailed architectures of the spatially constrained CNN (SC-CNN) for nucleus detection and the softmax CNN for nucleus classification are shown in Table I. These architectures are inspired by [8] and [32]. The networks consisted of conventional layers, including input, convolution, non-overlapping spatial max-pooling, and fully-connected layers. SC-CNN, in addition, included parameter estimation and spatially constrained layers. In both networks, a rectified linear unit (ReLU) activation function [32] was used after each convolution layer and the first two fully-connected layers (1st, 3rd, 5th and 6th layer). To avoid over-fitting, dropout [33] was implemented in the first-two fully-connected layers (5th and 6th layer, after ReLU is applied) with a dropout rate of 0.2. We implemented both networks using MatConvNet [34].

The input features were selected with respect to the task. In nucleus detection, we selected hematoxylin intensity as an input feature to SC-CNN for each patch. Because nucleic acids inside the nucleus is stained by hematoxylin, this feature is a reasonably good representation of localization of the nuclei. The hematoxylin intensity was obtained using a recently proposed color deconvolution method [35]. In classification, morphology of nuclei (shape, size, color, and texture) is necessary to distinguish between different types of them. Raw RGB color intensities which constitute the overall visual appearance of nuclei were, thus, chosen as input features to softmax CNN for each patch. For both networks, we set input patch size to $W \times H = 27 \times 27$ since the majority of the nuclei in a dataset used in our experiments (Section VII-A) have their size within this limit. We set the output patch size for SC-CNN to $W' \times H' = 11 \times 11$. Based on this output patch size, we found that the number of nuclei contained in the training output patches is mostly less than or equal to 2. Thus, the maximum number of predicted nuclei allowed in the S1 layer of SC-CNN was set to $M = 1$ or $M = 2$, accordingly.

B. Training Data Augmentation

For both networks, we arbitrary rotated patches $(0°, 90°, 180°, 270°)$ and flipped them along vertical or horizontal axis to alleviate the rotation-variant problem of the input features. To make softmax CNN persist to variability in color distribution, which is commonly found in histology images, we also arbitrary perturbed the color distribution of training patches. This was accomplished in HSV space, where hue (H), saturation (S), and value (V) variables were separately multiplied by random numbers $r_H \in [0.95, 1.05]$,

TABLE I
ARCHITECTURES OF THE SPATIALLY CONSTRAINED CNN (SC-CNN) FOR NUCLEUS DETECTION AND SOFTMAX CNN FOR NUCLEUS CLASSIFICATION.
THE NETWORKS CONSIST OF INPUT (I), CONVOLUTION (C), MAX-POOLING (M), FULLY-CONNECTED (F), PARAMETER ESTIMATION (S1), AND
SPATIALLY CONSTRAINED (S2) LAYERS.

Layer	SC-CNN for detection			softmax CNN for classification		
	Type	Filter Dimensions	Input/Output Dimensions	Type	Filter Dimensions	Input/Output Dimensions
0	I		$27 \times 27 \times 1$	I		$27 \times 27 \times 1$
1	C	$4 \times 4 \times 1 \times 36$	$24 \times 24 \times 36$	C	$4 \times 4 \times 1 \times 36$	$24 \times 24 \times 36$
2	M	2×2	$12 \times 12 \times 36$	M	2×2	$12 \times 12 \times 36$
3	C	$3 \times 3 \times 36 \times 48$	$10 \times 10 \times 48$	C	$3 \times 3 \times 36 \times 48$	$10 \times 10 \times 48$
4	M	2×2	$5 \times 5 \times 48$	M	2×2	$5 \times 5 \times 48$
5	F	$5 \times 5 \times 48 \times 512$	1×512	F	$5 \times 5 \times 48 \times 512$	1×512
6	F	$1 \times 1 \times 512 \times 512$	1×512	F	$1 \times 1 \times 512 \times 512$	1×512
7	S1	$1 \times 1 \times 512 \times 3$	1×3	F	$1 \times 1 \times 512 \times 4$	1×4
8	S2		11×11			

$r_S, r_V \in [0.9, 1.1]$, respectively. In addition, we extracted multiple patches of the same nucleus at different locations to account for location-variant that could negatively affect the classification performance of softmax CNN. This over-sample also allowed us to account for the class unbalance problem inherited in the dataset.

C. Initialization and Training of the Networks

We initialized all weights with 0 mean and 10^{-2} standard deviation Gaussian random numbers. All biases were set to 0. The networks were trained using stochastic gradient descent with momentum 0.9 and weight decay 5×10^{-4} for 120 epochs. We annealed the learning rate, starting from 10^{-2} for the first 60 epochs, then 10^{-3} for the next 40 epochs, and 10^{-4} for the last 20 epochs. We used 20% of training data for validation. The optimal networks for SC-CNN and softmax CNN were selected based on the root mean square error and the classification error on the validation set, respectively.

VII. EXPERIMENTS AND RESULTS

A. The Dataset[1]

This study involves 100 H&E stained histology images of colorectal adenocarcinomas. All images have a common size of 500×500 pixels, and were cropped from non-overlapping areas of 10 whole-slide images from 9 patients, at a pixel resolution of $0.55 \mu m$/pixel ($20 \times$ optical magnification). The whole-slide images were obtained using an Omnyx VL120 scanner. The cropping areas were selected to represent a variety of tissue appearance from both normal and malignant regions of the slides. This also comprises areas with artifacts, over-staining, and failed autofocussing to represent outliers normally found in real scenarios.

Manual annotation of nuclei were conducted mostly by an experienced pathologist (YT) and partly by a graduate student under supervision of and validation by the same pathologist. A total number of $29,756$ nuclei were marked at the center

[1]The dataset is available at http://www.warwick.ac.uk/BIAlab/data/CRChistoLabeledNucleiHE.

Epithelial Nuclei

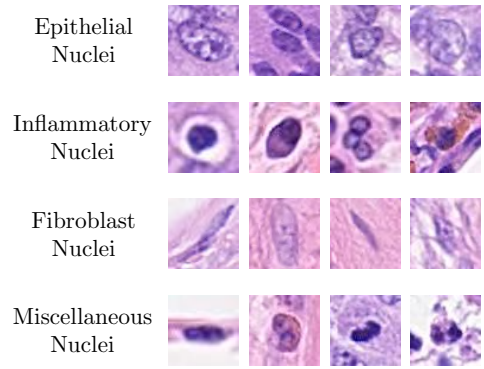Inflammatory Nuclei

Fibroblast Nuclei

Miscellaneous Nuclei

Fig. 4. Example patches of different types of nuclei found in the dataset: 1st Row - Epithelial nuclei, 2nd Row - Inflammatory nuclei (from left to right, lymphocyte, plasma nucleus, neutrophil, and eosinophil), 3rd Row - Fibroblasts, 4th Row - Miscellaneous nuclei (from left to right, adipocyte, endothelial nucleus, mitotic figure, and necrotic nucleus).

for detection purposes. Out of those, there were $22,444$ nuclei that also have an associated class label, i.e. epithelial, inflammatory, fibroblast, and miscellaneous. The remaining $7,312$ nuclei were unlabeled. The types of nuclei that were labeled as inflammatory include lymphocyte plasma, neutrophil and eosinophil. The nuclei that do not fall into the first three categories (i.e., epithelial, inflammatory, and fibroblast) such as adipocyte, endothelium, mitotic figure, nucleus of necrotic (i.e., dead) cell, etc. are labeled as miscellaneous. In total, there are $7,722$ epithelial, $5,712$ fibroblast, $6,971$ inflammatory, and $2,039$ miscellaneous nuclei. Fig. 4 shows some examples of the nuclei in the dataset.

B. Detection

The objective of this experiment is to detect all nuclei in an image by locating their center positions, regardless of their class labels. Details of the dataset used in this experiment are as described above in Section VII-A.

1) Evaluation: Precision, Recall, and F1 score were used to quantitatively assess the detection performance. Here, we define the region within the radius of 6 pixels from the annotated center of the nucleus as ground truth. If there are

multiple detected points within the same ground truth region, only the one closest to the annotated center is considered as true positive. The statistics, including median, 1st quartile, and 3rd quartile were also calculated to summarize the positively skewed distribution of the Euclidean distance between the detected points and their nearest annotated center of nucleus.

2) Other Approaches: The following nucleus detection algorithms were selected for comparison. Firstly, center-of-the-pixel CNN (CP-CNN) follows conventional method for object detection, where given a patch, it calculates the probability of being the center of a nucleus for a pixel at the center of the patch. Secondly, structural regression (SR-CNN) [27] extends the concept of a single-pixel regression of CP-CNN into a full-patch regression. The architectures of both CP-CNN and SR-CNN were set to resemble that of SC-CNN, except that the parameter estimation and the spatial constrained layers were replaced by a regression layer with a single node for CP-CNN and a regression layer with the number of nodes equal to the output patch size for SR-CNN. The training output patches for SR-CNN were the same as those for SC-CNN (see Fig. 2b for an example), whereas only the center pixel of those patches were used to train CP-CNN. Thirdly, stacked sparse autoencoder (SSAE) [7] consists of two sparse autoencoder layers followed by a softmax classifier which is trained to distinguish between nuclear and non-nuclear patches. If classified as a nucleus, all pixels inside the output patch are assigned the value of 1, or 0 otherwise. We set input feature, sizes of input and output patches of SSAE to be the same as those of SC-CNN. Fourthly, local isotropic phase symmetry measure (LIPSyM) yields high response values near the center of symmetric nuclei which can be used for detection. Lastly, CRImage [9] segments all nuclei with help of thresholding, followed by morphological operation, distance transform, and watershed. A centroid of individual segmented nucleus was used as the detected point. Note that for a fair comparison the detection of the center of nuclei for CP-CNN, SR-CNN, SSAE, and LIPSyM was done in the same fashion as that of SC-CNN (see last paragraph of Section IV-B for details). We implemented CP-CNN using MatConvNet [34], while CRImage is publicly available as an R package[2]. The implementations of SR-CNN and LIPSyM were provided by the authors, and we used a set of build-in functions in Matlab to implement SSAE.

3) Comparative Results: We employed 2-fold cross-validation (50 images/fold) in the experiment. Fig. 5 shows the precision-recall curve. The curve was generated by varying the threshold value applied to the predicted probability map before locating local maxima to avoid false positive detection. Table II reports the detection results when the threshold value is empirically chosen to optimize F1 score on the training fold. Detailed results are shown in Fig. SF1 and Fig. SF2 of the supplementary material. Here, we include two variants of SC-CNN; one with the maximum number of predicted nuclei in an output patch equal to 1 ($M = 1$) and the other equal to 2 ($M = 2$).
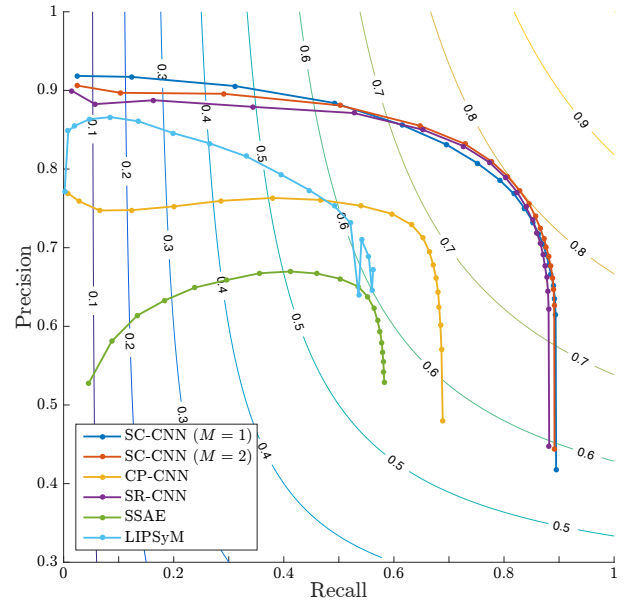
[2]http://www.bioconductor.org/packages/CRImage/

Fig. 5. Precision-recall curve for nucleus detection. Isolines indicate regions of different F1 scores. The curve is generated by varying the value of threshold applied to the predicted probability map before locating local maxima. Note that this thresholding scheme does not apply to CRImage [9].

TABLE II
COMPARATIVE RESULTS FOR NUCLEUS DETECTION.

Method	Precision	Recall	F1 score	Median Distance (Q1, Q3)
SC-CNN ($M = 1$)	0.758	**0.827**	0.791	**2.236** (1.414, 5.099)
SC-CNN ($M = 2$)	0.781	0.823	**0.802**	**2.236** (1.414, 5)
CP-CNN	0.697	0.687	0.692	3.606 (2.236, 7.616)
SR-CNN [27]	**0.783**	0.804	0.793	**2.236** (1.414, 5)
SSAE [7]	0.617	0.644	0.630	4.123 (2.236, 10)
LIPSyM [15]	0.725	0.517	0.604	**2.236** (1.414 , 7.211)
CRImage [9]	0.657	0.461	0.542	3.071 (1.377, 9.022)

Median, Q1 and Q3 refer to the median, 1st quartile and 3rd quartile of the distribution of the Euclidean distance between the detected points and their nearest annotated center of nucleus, respectively.

Overall, the comparison is in favor of the algorithms that learn to predict the probability of being the center of a nucleus, based on spatial context of the whole patch, i.e. SC-CNN and SR-CNN. This is consistent with the quantitative results shown in Fig. 6c and Fig. SF2. Visual inspection of the probability generated by SC-CNN (Fig. 6b and Fig. SF1a) and SR-CNN (Fig. SF1c) revealed that the pixels with high probability values are mostly located in the vicinity of the center of nuclei. However, SC-CNN, as imposed by spatial constrained in (6), has a narrower spread of pixels with high probability values. Probability maps generated by the CP-CNN (Fig. SF1b) and SSAE (Fig. SF1d), on the other hand, exhibit a wider spread of pixels with high probabilities away from the center of nuclei. This results in SC-CNN and SR-CNN yielding a lower value of median distance between the detected points and the annotated ground truth, as compared to those values for CP-CNN and SSAE. LIPSyM heavily relies on bilateral symmetry of nuclei for detection. For this reason, it could not precisely detect spindle-like and other irregularly-shaped nuclei such as fibroblasts and malignant epithelial nuclei (see Fig. SF2e). Due
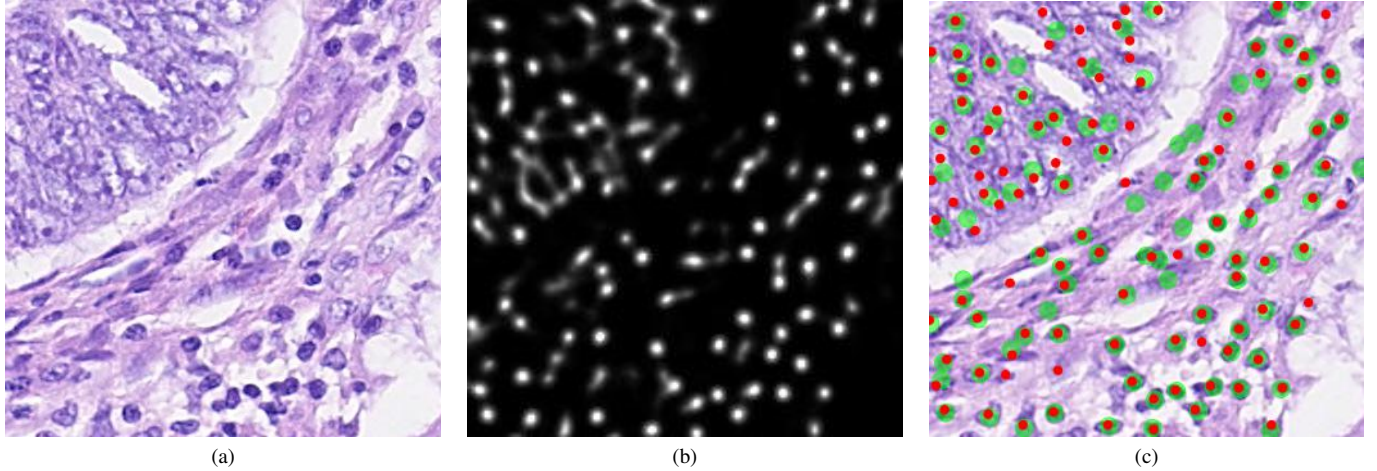
Fig. 6. Qualitative results for nucleus detection. (a) An example image. (b) Probability maps generated by SC-CNN ($M = 2$). The probability value in the probability map indicates the likelihood of a pixel being the center of a nucleus. Thus, one can detect the center of an individual nucleus based on the location of local maxima found in the probability map. (c) Detection results of SC-CNN ($M = 2$). Here, detected centers of the nuclei are shown as red dots and the ground truth areas are shown as green shaded circles. Probability maps and detection results of other methods can be seen in Fig. SF1 and Fig. SF2, respectively in supplementary material.

to the nature of nuclei in the dataset which often appear to have a weak boundary and/or overlapping boundaries, algorithms which require segmentation to detect nuclei such as CRImage are thus likely to fail for those cases. Fig. SF2f reflects this problem, where CRImage failed to detect malignant epithelial nuclei (top-left corner).

Given that a single output patch can contain multiple nucleus centers (mostly less than or equal to 2 in our experiment), SC-CNN ($M = 2$) thus gives a better performance than SC-CNN ($M = 1$). Nonetheless, the improvement mainly comes from the reduction in the false positive rate rather than the false negative rate. One of the possible explanations is that SC-CNN ($M = 1$) tries to compensate the loss incurred from not being able to regress the probability of multiple nucleus centers during the training process by giving a high confidence, i.e. h_m close to 1, to the estimated nuclear center. Thus, it tends to produce false detection as compared to SC-CNN ($M = 2$).

SC-CNN and SR-CNN share a closely resembling idea of using spatial regression to generate the probability map of nucleus center. In essence, SC-CNN uses a known spatial structure for regression, whereas SR-CNN directly learns the structure from the training output data. SR-CNN, hence, is more flexible in general. However, when the structure for regression is known and governed by a small number of parameters such as in this problem, SC-CNN can provide more advantages. Firstly, it simplifies the learning process of CNN by imposing the output functional form of the network. Secondly, the regressed structure is always consistent with that of the training data. This results in SC-CNN ($M = 2$) yielding better performance in terms of F1 score than SR-CNN. For further discussion about the detection performance when nuclei are stratified by their class label, see Section VII-D2.

C. Classification

The setting of this experiment is to classify patches of size 27×27 pixels, containing a nucleus at the center, into 4 classes: epithelial, inflammatory, fibroblast, and miscellaneous. Full details of the dataset can be found in Section VII-A.

1) Evaluation: We calculated the F1 score for each class of nucleus and their average weighted by the number of nucleus samples (see Section VII-A for details) to summarize the overall classification performance. We also considered an area under the receiver operating characteristic curve for multiclass classification (multiclass AUC) [36]. Multiclass AUC measures the probability that given a pair of samples with different class labels, a classifier will assign a high prediction score for class, say c, to the sample from class c, as compared to the sample from the other class. Here, the prediction score for softmax CNN is given by softmax function (14).

2) Other Approaches: First, superpixel descriptor [37] combines color and texture information, cumulated in superpixels, for classifying areas with different histologic pattern. This descriptor is used in conjunction with a random forest classifier. In the experiment, we treated a patch as a single superpixel. We implemented this method in Matlab according to the details outlined in [37]. Second, CRImage[3] [9] calculates a list of statistics[4] for each segmented nucleus and uses the support vector machine with radial basis kernel as a classifier. A successive spatial density smoothing is used to correct false classification. We implemented this method in Matlab.

3) Comparative Results: Fig. 7 and Table III show the comparative classification performance on 2-fold cross-validation

---

[3]CRImage is available as an R package (http://www.bioconductor.org/packages/CRImage/). As of Sep 1st, 2015, the package version 1.16.0 has a certain compatibility issue with package EBImage 4.10.1 and cannot reproduce the results on the test samples given in the manual of the package. We did our best to implement the method, but disclaim a perfect replication.
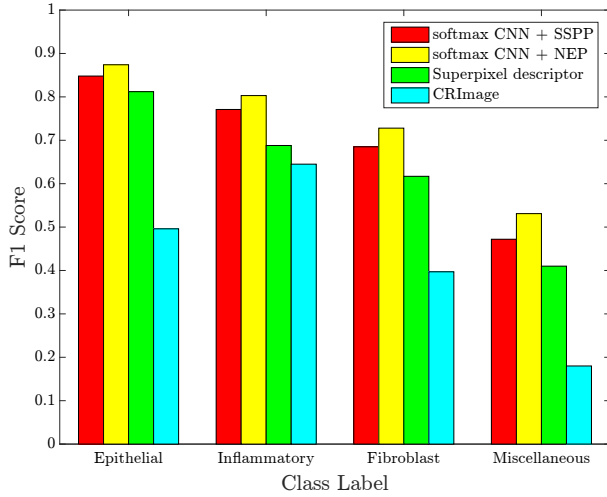
[4]See sweave file of [9] for details.

Fig. 7. Comparative results for nucleus classification stratified with respect to class label.

Fig. 8. Combined performance on nucleus detection and classification stratified according to class label.

TABLE III
COMPARATIVE RESULTS FOR NUCLEUS CLASSIFICATION.

Method	Weighted Average F1 score	Multiclass AUC
softmax CNN + SSPP	0.748	0.893
softmax CNN + NEP	**0.784**	**0.917**
superpixel descriptor [37]	0.687	0.853
CRImage [9]	0.488	0.684

TABLE IV
COMBINED PERFORMANCE ON NUCLEUS DETECTION AND
CLASSIFICATION.

Method		Weighted
Detection	Classification	Average F1 score
SC-CNN ($M = 1$)	softmax CNN + SSPP	0.664
	softmax CNN + NEP	0.688
SC-CNN ($M = 2$)	softmax CNN + SSPP	0.670
	softmax CNN + NEP	**0.692**
SR-CNN [27]	softmax CNN + SSPP	0.662
	softmax CNN + NEP	0.683

experiment. For detailed results, see Table ST1 of the supplementary material. The comparison is in favor of softmax CNN with the NEP in every nucleus class. The better performance of NEP than that of SSPP allows us to hypothesize that NEP is more resilient to variability in the appearance of nuclei. The superpixel descriptor [37] is originally devised to distinguish area with different histologic pattern. This descriptor does not directly contain features related to visual appearance of the nucleus, and thus yielded low classification performance as compared to softmax CNN. CRImage calculates features based the segmentation of the nuclei. As previously discussed, a reliable segmentation is difficult to be achieved when the nuclear boundary is weakly stained or there are overlapping boundaries. This prevents CRImage to perform well in this dataset. Another interesting observation is that all considered approaches including ours suffer from class imbalance problem. The classification performance declines as the number of the samples decreases (see Fig. 7 and Table ST1).

D. Combined Detection and Classification

In this experiment, we combine detection and classification as a single work flow. Nuclei are first detected and then classified. We consider the combinations of the top-performing approaches in each task for comparison. These are SC-CNN and SR-CNN for detection and softmax CNN together with SSPP and NEP for classification.

1) Evaluation: We calculated F1 score for combined detection and classification separately for each class label. For in-
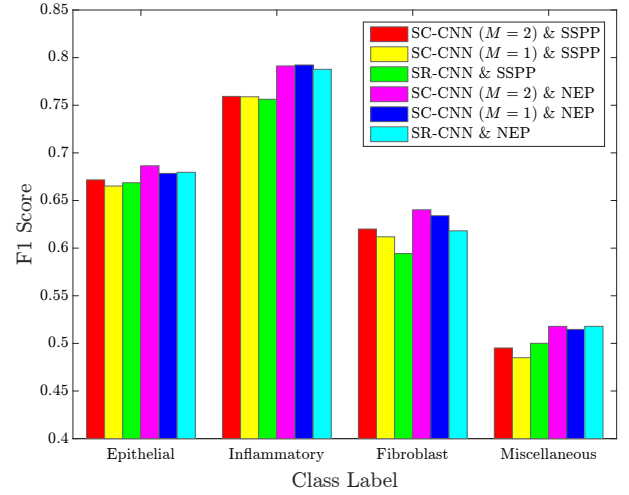
stance, consider class c. Precision is defined as the proportion between the number of correctly detected and classified class c nuclei and the total number of class c nuclei in the ground truth. The definition of Recall is the proportion between the number of correctly detected and classified class c nuclei and the total number of all detected objects classified as class c nuclei. The criterion for true positive for nucleus detection is outlined in Section VII-B1. Note that, in the dataset, the number of nuclei that have a class label associated with is smaller than the total number of annotated nuclei. Therefore, we only considered the labeled nuclei as ground truth and restrict the evaluation to the area covering by patches of size 41×41 pixels, centered at each ground truth nucleus in the image. To summarize the overall performance, we also calculated the weighted average F1 score, where the weight term for each nucleus class is defined by the number of data samples in that class (see Section VII-A for details).

2) Comparative Results: Fig. 8 and Table IV report the combined performance on detection and classification on a 2-fold cross validation experiment. See also Table ST2 of the supplementary material for the detailed results. As expected from the results in Section VII-C, the combinations that employ softmax CNN with SSPP performs better than their counterparts. SC-CNN and SR-CNN when combined with softmax CNN with SSPP yield similar F1 scores across different classes of nuclei, except for epithelial (irregular-shaped)

and fibroblast (elongated-shaped) where SC-CNN ($M = 2$) performs better. The same trend of performance can be seen for the combinations that employ softmax CNN with NEP.

VIII. DISCUSSION

Histopathological data are often incomplete and contaminated with subjectivity, which is also the case for the dataset used in this study. This is unavoidable due to the sheer number of cell nuclei and the enormous variation in morphology, making it difficult to identify all cells with certainty. In fact most pathologist rely on low power architecture to build the main picture of what is going on, using high power cellular morphological details to confirm or reject the initial impressions. Identifying individual cells on high power features alone without architectural clues will increase their misclassification. As previously mentioned, immunohistochemistry that stains a specific type of cell would provide a decisive judgement, yet it is costly and practically difficult to deal with in the laboratory, compared with H&E staining. That, nonetheless, could provide a stronger objective validation for the future work.

The architectures for CNN presented in this work were empirically chosen on the basis of resources available at hand. Finding theoretical justification for choosing an optimal network architecture is still an ongoing search. A large network architecture would allow more variation in high-level representations of objects, at the expense of training time and other resources. Yet, with random initialization of large number of network parameters, a gradient-based optimization may get stuck in a poor local minima. One could explore different strategies for training the network as described in [31]. There is also an issue of choosing network input features. For nucleus detection, we found that hematoxylin intensity could provide better results than standard RGB intensities. On the other hand, for nucleus classification, RGB did better than hematoxylin, but there is no significant difference between the results obtained using RGB and other standard color spaces such as LAB and HSV. Selecting a set of input features is task dependent, and suitable features should reduce the complexity of the task and allow better results.

Inspection of the results from all nucleus classification methods in Section VII-C revealed that the majority of misclassified nuclei often appeared isolated and biologically implausible, considering their spatial positions on the original images. Yuan *et al.* [9] proposed the use of hierarchical spatial smoothing to correct misclassification. However, it should be pointed out that this type of correction should be used with caution, as it may falsely eliminate biologically important phenomena such as tumor budding which consists of a small number of tumor nuclei [38] or isolated islands of tumor nuclei appearing at the invasive front of the tumor. In our experiments, we did not employ any spatial correction.

Automatic approaches for combined nucleus detection and classification could offer benefits to pathology practice in a number of ways. One of the potential applications is to locate and identify all tissue-constituent nuclei in whole-slide images. Fig. 9 shows the detection and classification results produced by SC-CNN and softmax CNN with NEP on a whole-slide image. This could facilitate quantitative analysis, and at the same time, remove tediousness and reduce subjectivity in pathological routine. This is an interesting prospect for future research and yet to be validated in a large-scale study.

Existing distributed computing technologies such as parallel computing and graphical processing unit (GPU) are important key factors to scale up the proposed framework for whole-slide histology images. In our experiment, a whole-slide image is first divided into small tiles of size $1,000 \times 1,000$ pixels. On a single 2.5 GHz CPU, the average execution time on an individual image tile is 47.6s (preprocessing 27.8s, detection 18.4s, and classification 1.4s). For a given whole-slide image captured at $20\times$ optical magnification and consisting of $60,000 \times 50,000$ pixel dimensions, there are 750 tiles of size $1,000 \times 1,000$ pixels to be processed assuming only 25% of the slide contains tissue. Theoretically speaking, by using a 12-core processor machine, the average execution time of the proposed detection and classification framework is around 50 mins per slide. However, it should be noted that the execution time reported here is for a research-grade implementation of the framework which has not yet been fully optimized to increase time efficiency, nor did we employ the computational power of GPUs which can significantly speed up the execution time for CNNs.

IX. CONCLUSIONS

In this study, we presented deep learning approaches sensitive to the local neighborhood for nucleus detection and classification in routine stained histology images of colorectal adenocarcinomas. The evaluation was conducted on a large dataset with more than $20,000$ annotated nuclei from samples of different histologic grades. The comparison is in favor of the proposed spatially-constrained CNN for nucleus detection and the softmax CNN with the proposed neighboring ensemble predictor for nucleus classification. The combination of the two could potentially offer a systematic quantitative analysis of tissue morphology, and tissue constituents, lending itself to be a useful tool for better understanding of the tumor microenvironment.

REFERENCES

[1] P. Dalerba, T. Kalisky, D. Sahoo, P. S. Rajendran, M. E. Rothenberg, A. A. Leyrat, S. Sim, J. Okamoto, D. M. Johnston, D. Qian *et al.*, "Single-cell dissection of transcriptional heterogeneity in human colon tumors," *Nature biotechnology*, vol. 29, no. 12, pp. 1120–1127, 2011.
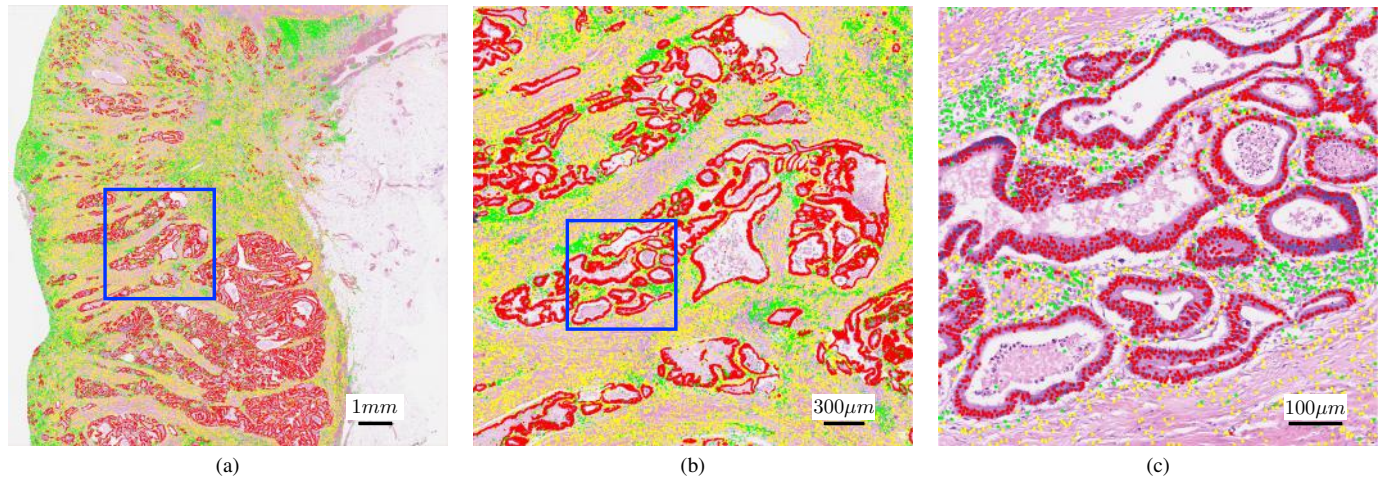
Fig. 9. Nucleus detection and classification on a whole-slide image. Detected epithelial, inflammatory and fibroblast nuclei are represented as red, green, and yellow dots, respectively. (a), (b) and (c) show the results overlaid on the image at $1\times$, $5\times$, and $20\times$, respectively. The blue rectangle in (a) contains the region shown in (b), and the blue rectangle in (b) contains the region shown in (c). The detection and classification were conducted at $20\times$ magnification using SC-CNN and softmax CNN with NEP. This figure is best viewed on screen with magnification 400%.

[2] C. A. OBrien, A. Pollett, S. Gallinger, and J. E. Dick, "A human colon cancer cell capable of initiating tumour growth in immunodeficient mice," *Nature*, vol. 445, no. 7123, pp. 106–110, 2007.

[3] A. Basavanhally, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, S. Ganesan, and A. Madabhushi, "Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: comparison to oncotype dx," *Journal of pathology informatics*, vol. 2, 2011.

[4] J. S. Lewis Jr, S. Ali, J. Luo, W. L. Thorstad, and A. Madabhushi, "A quantitative histomorphometric classifier (quhbic) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma," *The American journal of surgical pathology*, vol. 38, no. 1, pp. 128–137, 2014.

[5] G. N. van Muijen, D. J. Ruiter, W. W. Franke, T. Achtsttter, W. H. Haasnoot, M. Ponec, and S. O. Warnaar, "Cell type heterogeneity of cytokeratin expression in complex epithelia and carcinomas as demonstrated by monoclonal antibodies specific for cytokeratins nos. 4 and 13," *Experimental Cell Research*, vol. 162, no. 1, pp. 97 – 113, 1986.

[6] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review. current status and future potential," *Biomedical Engineering, IEEE Reviews in*, vol. 7, pp. 97–114, 2014.

[7] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *Medical Imaging, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[8] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 411–418.

[9] Y. Yuan, H. Failmezger, O. M. Rueda, H. R. Ali, S. Gräf, S.-F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell *et al.*, "Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling," *Science translational medicine*, vol. 4, no. 157, pp. 157ra143–157ra143, 2012.

[10] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *arXiv preprint arXiv:1505.04597*, 2015.

[12] M. Veta, J. Pluim, P. van Diest, and M. Viergever, "Breast cancer histopathology image analysis: A review," *Biomedical Engineering, IEEE Transactions on*, vol. 61, no. 5, pp. 1400–1411, May 2014.

[13] E. Cosatto, M. Miller, H. P. Graf, and J. S. Meyer, "Grading nuclear pleomorphism on histological micrographs," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[14] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 4, pp. 841–852, 2010.

[15] M. Kuse, Y.-F. Wang, V. Kalasannavar, M. Khan, and N. Rajpoot, "Local isotropic phase symmetry measure for detection of beta cells and lymphocytes," *Journal of Pathology Informatics*, vol. 2, no. 2, p. 2, 2011.

[16] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PLoS ONE*, vol. 8, no. 7, p. e70221, 07 2013.

[17] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, "Learning to detect cells using non-overlapping extremal regions," in *Medical image computing and computer-assisted intervention–MICCAI 2012*. Springer, 2012, pp. 348–356.

[18] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *Medical Imaging, IEEE Transactions On*, vol. 31, no. 7, pp. 1448–1460, 2012.

[19] J. P. Vink, M. Van Leeuwen, C. Van Deurzen, and G. De Haan, "Efficient nucleus detector in histopathology images," *Journal of microscopy*, vol. 249, no. 2, pp. 124–135, 2013.

[20] J.-R. Dalle, H. Li, C.-H. Huang, W. K. Leow, D. Racoceanu, and T. C. Putti, "Nuclear pleomorphism scoring by selective cell nuclei detection." in *WACV*, 2009.

[21] C. D. Malon and E. Cosatto, "Classification of mitotic figures with convolutional neural networks and seeded blob features," *Journal of pathology informatics*, vol. 4, 2013.

[22] K. Nguyen, A. K. Jain, and B. Sabata, "Prostate cancer detection: Fusion of cytological and textural features," *Journal of pathology informatics*, vol. 2, 2011.

[23] M. Arif and N. Rajpoot, "Classification of potential nuclei in prostate histology images using shape manifold learning," in *International Conference on Machine Vision (ICMV)*. IEEE, 2007, pp. 113–118.

[24] H. Sharma, N. Zerbe, D. Heim, S. Wienert, H.-M. Behrens, O. Hellwich, and P. Hufnagl, "A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, 2015, pp. 37–46.

[25] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection," vol. 9041, 2014, pp. 90 410B–90 410B–10.

[26] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 403–410.

[27] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *Medical Image Computing and Computer-Assisted InterventionMICCAI 2015*. Springer, 2015, pp. 358–365.

[28] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, and L. Yang, "Deep voting: A robust approach toward nucleus localization in microscopy images," in *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015*. Springer, 2015, pp. 374–382.

[29] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[30] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[31] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *The Journal of Machine Learning Research*, vol. 10, pp. 1–40, 2009.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[34] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for matlab," *CoRR*, vol. abs/1412.4564, 2014.

[35] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *Biomedical Engineering, IEEE Transactions on*, vol. 61, no. 6, pp. 1729–1738, 2014.

[36] D. J. Hand and R. J. Till, "A simple generalisation of the area under the roc curve for multiple class classification problems," *Machine learning*, vol. 45, no. 2, pp. 171–186, 2001.

[37] K. Sirinukunwattana, D. R. Snead, and N. M. Rajpoot, "A novel texture descriptor for detection of glandular structures in colon histology images," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2015, pp. 94 200S–94 200S.

[38] B. Mitrovic, D. F. Schaeffer, R. H. Riddell, and R. Kirsch, "Tumor budding in colorectal carcinoma: time to take notice," *Modern Pathology*, vol. 25, no. 10, pp. 1315–1325, 2012.