Localization and Classification of Cell Nuclei in Post-Neoadjuvant Breast Cancer Surgical Specimen Using Fully Convolutional Networks

Rene Bidart[a], Mehrdad J. Gangeh[*b], Mohammad Peikari[b], Sherine Salama[c],
Sharon Nofech-Mozes[c], Anne L. Martel[b,d], Ali Ghodsi[a]

[a] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada
[b] Departments of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[c] Laboratory of Medicine and Pathobiology, University of Toronto, ON, Canada
[d] Department of Physical Sciences, Sunnybrook Research Institute, Toronto, ON, Canada

ABSTRACT

Neoadjuvant therapy (NAT) is an option for locally advanced breast cancer patients to downsize tumour allowing for less extensive surgical operation, better cosmetic outcomes, and lesser post-operative complications. The quality of NAT is assessed by pathologists after examining the tissue sections to reveal the efficacy of treatment, and also associate the outcome with the patient's prognosis. There are many factors involved with assessing the best treatment efficacy, including the amount of residual cancer within tumour bed. Currently, the process of assessing residual tumour burden is qualitative, which may be time-consuming and impaired by inter-observer variability. In this study, an automated method was developed to localize, and subsequently classify nuclei figures into three categories of lymphocyte (L), benign epithelial (BE), and malignant epithelial (ME) figures from post-NAT tissue slides of breast cancer. A fully convolutional network (FCN) was developed to perform both tasks in an efficient way. In order to find the cell nuclei in image patches (localization), the FCN was applied over the entire patch, generating four heatmaps corresponding to the probability of a pixel being the centre of an L, BE, ME, or non-cell nuclei. Non-maximum suppression algorithm was subsequently applied to the generated heatmaps to estimate the nuclei locations. Finally, the highest probability corresponding to each predicted cell nucleus in the heatmaps was used for the classification of the nucleus to one of the three classes (L, BE, or ME). The final classification accuracy on detected nuclei was 94.6%, surpassing previous machine learning methods based on handcrafted features on this dataset.

Keywords: Fully convolutional networks, pathology, localization, cell nuclei classification, heatmaps, non-maximum suppression, breast cancer

1. INTRODUCTION

Neoadjuvant therapy (NAT) is a treatment of choice for selected high-risk and/or locally advanced breast cancer (LABC) patients[1]. The goal of NAT is to downsize the tumour, allowing for less extensive surgical operation resulting in better cosmetic outcomes and reduced post-operative complications. Also, after NAT, the efficacy of the therapy can be assessed *in vivo* with the aim of associating the treatment response to the new therapeutic agents in drug trials and modifying the treatment for patients who show little to no response to the conventional therapies.

It is desirable to achieve pathologic complete response (pCR) after the course of NAT, which indicates the absence of residual cancer in the breast tissue and associated axillary lymph nodes. A pCR is associated with better patient prognosis[2]. Currently, residual tumour burden assessment is done manually by pathologists on hematoxylin and eosin (H&E) stained tissue sections through a qualitative and time-consuming process. Although there are different scoring factors reported for assessing tumour response to NAT, the cellularity fraction of cancer within tumour bed provides a better prognostic indicator of the efficacy of the treatment compared to other factors[3].

The objective of this paper was to present an automated method based on a fully convolutional network (FCN)[4] to locate and distinguish cancer nuclei from other nuclei figures within the tissue micro-environment. Major nuclei figures within tissue microenvironment are stromal, lymphocytic (L), benign epithelial (BE), and malignant epithelial (ME) nuclei.

Stromal nuclei are usually thinly elongated figures that could be identified by their much longer major axis. Therefore, the focus of this study was to distinguish ME figures from BE and L figures.

2. RELATED WORK

This problem can be approached using a variety of methods, including semantic segmentation[4–7], object detection[8,9], and regression[10,11] based on deep learning or feature-based[12] methods. In semantic segmentation, the goal is to predict a label for each pixel in the image, after which post processing can be used to improve the segmentation. Object detection methods aim to find a bounding box for each object, while regression-based methods have the goal of counting the number of occurrences of each class.

2.1 Semantic segmentation

Patchwise training and testing has been used for the segmentation of cells in pixel labeled images as far back as 2005[6]. Here, the authors used a simple convolutional neural network (CNN), followed by post-processing to refine the segmentation. This method was also used in a straightforward way to segment neuronal structures by training a CNN as a binary classifier to predict the label of each individual pixel[7]. The training data was submitted to the classifier as a subsection of size 95×95 pixels, centered at the pixel to be classified. In order to classify a test sample, the classifier was applied to all pixels in the image, generating a map of neuron probability. This map was then spatially smoothed by a 2-pixel-radius median filter, and thresholded to obtain the pixel classifications. The downside to this method is that at test time, the classifier must be applied to all pixels in the image, which is extremely computationally expensive. Fully convolutional architectures can eliminate this redundant computation.

Shelhamer *et al.*[4] have demonstrated that fully convolutional networks (FCN) are superior to previous methods for semantic segmentation. FCNs improve the speed at test time by recognizing that convolution layers can be applied over the entire image, regardless of the size of the image it was trained on. In addition to the speedup during inference, FCNs also improve the speed of training by training on whole images, rather than on patches. Training on patches is inefficient because when the input patches overlap, there is duplicated computation because the same convolutional features are computed multiple times. U-Nets[5], on the other hand, use a similar FCN approach to produce pixel level neuronal segmentations. These methods have been extended to include inputs of multiple resolutions and sizes.

For cell segmentation, one method[13] uses an FCN operating on a low resolution image to decide if it would be helpful to use a classifier on a higher resolution image. This leverages the benefits in accuracy of using a higher resolution image, but only uses it when absolutely necessary, so that the computational burden is kept as low as possible. Spatially constrained CNNs have also been proposed for nuclei detection and classification[14] without performing explicit segmentation. This method generates a fixed set of nuclei locations, along with a probability for each nucleus. FCNs have also been used for mitosis detection[15], with a two-stage approach. In the first stage, a coarse model has been used for rough localization, and subsequently, an Imagenet pre-trained model applied to each identified location to generate a better classification.

2.2 Regression-based methods

Regression based methods use a CNN to estimate the number of cells contained in a given area. By applying the CNN over an entire image, a cell probability heatmap is produced that can be integrated over to compute a count of the total number of cells[10,11]. In Count-ception[10], object counting and localization have been performed for one class of cells, using an inception[16] style network taking an input of 32×32 pixels, and generating a count of all the cells within that window. By applying this network across the entire input image, a count map is created.

2.3 Methods based on handcrafted features

On the particular problem addressed in this study, previous research has shown that traditional machine learning methods can be useful. For example, Peikari *et al.*[12] first segmented the images using an unsupervised approach and then extracted features describing shape and texture. Spatial features, which were based on the number of other nuclei surrounding the cell and their distances, were also generated from the segmented images. These features were then used to train a support vector machine (SVM) to distinguish between the three classes. The approach was validated on the same dataset as used in this study.
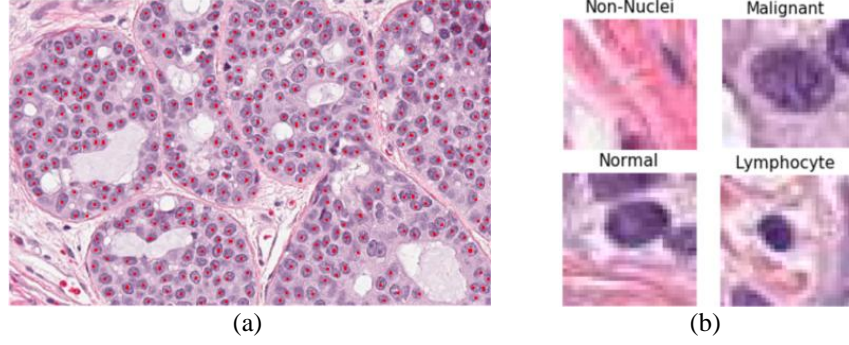
Figure 1- (a) Subsection of H&E stained slide, taken at 20× magnification, showing the malignant nuclei labeled in red (b) examples of extracted nuclei (tiles) from the image patch. Each tile is 32×32 pixels (16 µm × 16 µm).

3. METHODS

3.1 Patient data

The dataset was obtained from H&E stained whole slide images (WSIs) of breast tissue samples from 46 patients underwent post-neoadjuvant treatment, and was prepared and scanned at Sunnybrook Health Sciences Centre. There were 92 WSIs in the dataset that were scanned at 20× magnification (0.5 µm/pixel). Nuclei figures from three categories of L, BE, and ME were marked by an expert pathologist by clicking on the centroid of the individual nuclei within $n = 148$ rectangular regions of arbitrary size drawn by the pathologist, using the Sedeen viewer[17]. Selected regions were chosen in such a way that contained a mixture of nuclei categories. More than 27000 nuclei ($m_1 = 3034$ lymphocyte, $m_2 = 9964$ benign epithelial, and $m_3 = 14314$ malignant epithelial figures) were marked by the pathologist from all rectangular regions and used to assess the performance of the automated method presented in this paper.

We will refer to three different levels of data, as explained below:

1. **Slides** – Raw 20× magnification H&E stained slides.
2. **Patches** – Annotated subsections of the full slides, selected by the pathologist. There were in total 148 of these patches available in this study with the nuclei labeled.
3. **Tiles** – Generated by creating a box around each nucleus centered at the pathologist's annotation. In addition, tiles of non-nuclei tissues were created by randomly selecting points that were at least 5 µm (10 pixels) from any other cells. Consequently, each tile had a class of malignant, normal, lymphocyte, or non-nuclei. The input image was zero padded to ensure that the same size tile will be produced for all nuclei. Each tile may contain multiple nuclei, but was centered on the one of interest. Experiments were performed with various tile sizes from 32 to 192 pixels. In total, 183 non-nuclei tiles were selected from each slide, resulting in 26,600 nuclei and non-nuclei examples.

3.2 Outline of the methodology

The pipeline for the proposed scheme was consisting of three main steps, including: 1) training an FCN to recognize three classes of cell nuclei including BE, ME, and L nuclei, as well as no cell; 2) localization of cell nuclei by applying the trained FCN in previous step to generate four heatmaps of the probability of each nuclei class, including no nuclei class, followed by a non-maximum suppression algorithm to estimate nuclei locations ; and 3) classification of the cell nuclei at predicted cell nuclei locations.

3.3 Training an FCN model to recognize cell nuclei subtypes

In order to perform cell nuclei localization, a network needed to be trained first to recognize different cell classes in addition to no-cell class. To this end, a training dataset was created consisting of tiles that each included nuclei of one class, or no nucleus (see Figure 1 for examples of tiles on four classes). The tiles were created by taking a box of fixed size around each nucleus, centered at the pathologist's annotations. Since there could be more than one nucleus in each

tile, the FCN was trained based on the nucleus at the centre of a tile. Tiles of non-cell tissues were created by randomly selecting points that were at least 5 μm (10 pixels) from any other cell. Including no-cell class (tiles) in the training set would enable the trained FCN to perform localization as well.

3.4 Cell nuclei localization

The trained FCN in previous step was applied to the input images to generate the probability of a nucleus being centered at that location, in addition to the class probabilities of that nucleus. By applying the trained FCN across an entire image, four heatmaps were generated for each image corresponding to each of the cell probabilities, including the probability of no cell. An example for the heatmaps generated for an image is shown in Figure 2. By looking at $1 - p(\text{no cell}^{\dagger})$, the probability of there being a cell in any location could be explicitly obtained. This produced heatmaps that indicated the probability of a given pixel being the center of a nucleus. In order to convert the four heatmaps generated in the previous step into a set of cell nuclei locations, a non-maximum suppression algorithm was employed. This relied on the fact that the points on the cell probability heatmaps corresponding to the highest probability were most likely to be the centre of the cells. In this procedure, the pixels with the highest cell probability were iteratively found, and then all points within a radius r of these pixels were set to 0.

3.5 Classification of cell nuclei at predicted locations

After estimation of cell nuclei locations in previous step, the class labels (L, BE, or ME) for the predicted locations was assigned based on the corresponding highest probability from the heatmaps for those points.

3.6 Remarks on the implementations

Evaluation – The data was divided to three sets: 60% for training, 15% for validation (tuning the hyper-parameters and model selection), and 25% for testing the network. The test set remained unseen during the hyper-parameter tuning and model selection.

Network architecture – In order to choose the best network architecture for this problem, many different architectures were tested with varying number of filters in each layer. Inception module[16] was also investigated and found useful. We settled on using a fully convolutional network with two 3×3 convolutions, followed by two inception modules, as shown in Figure 2. For the final layer, instead of fully connected layers, a convolutional kernel that spanned the full width of the feature map was employed with no zero padding. The output from the network consisted of four convolutional feature maps, corresponding to the class probabilities.

Fully convolutional architecture was preferred over networks containing fully connected layers, mainly because it allowed for fast inference at test time. It also allowed for varying sized inputs, as the network was a non-linear filter instead of being a nonlinear function. When the network was tested on a larger image than what was used during training, it would output multiple probabilities corresponding to the spatial locations of the image. Some other standard architecture choices are included:

Batchnorm[18] – Was used to normalize the layer inputs by the mean and variance of the batch, making gradient flow better with less dependency on the weight initialization.

Dropout[19] – Regularization on the network was partially accomplished by randomly dropping nodes and their associated weights during training, helping to prevent overfitting.

ReLU – The ReLU non-linearity, max(0, x) was used instead of a sigmoid function to deal with saturating gradient problem.

Optimizer – Optimization was performed using the Adam optimizer with a cyclical learning rate schedule[20]. The implementation of the network was accomplished using the TensorFlow library[21], along with the Keras[22] package.

---

† This was the main reason that the non-cell data was added as a class for the training the FCN.
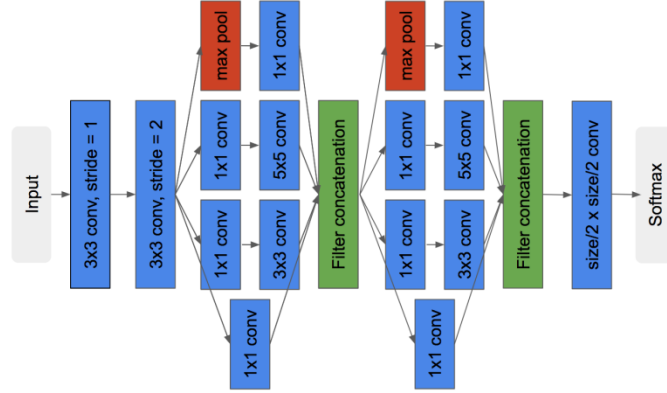
Figure 2 -The architecture of the best performing CNN. This was based on the inception[16] architecture, but with fewer layers and using a large convolution kernel instead of global average pooling at the last layer.

Data augmentation – Data augmentation was used to artificially increase the size of the training set by randomly applying transformations to each batch of data before submission to the network. If reasonable transformations are selected, this should help to reduce overfitting. Our data augmentation process was quite standard, but with a few adjustments to prevent translation invariance.

Rotations and flips – Since the data did not have a specific orientation, the classification of a nucleus was invariant to any rotation or flip. Therefore, the training data was augmented using random rotations and flips.

Translation – A small amount of translation was used. The intuition was that the doctor's annotation was probably not always in the exact center of the nucleus, and therefore it was possible to augment the data by slightly moving the nuclei off center in the image.

Choosing the radius for non-maximum suppression – The aim in non-maxima suppression (NMS) algorithm was to find the center of a nucleus, and then set all other points inside the nucleus to 0, to ensure only one location is returned per nucleus. If all nuclei were uniform size, this would be easy to implement. The NMS radius could be set equal to the radius of the nuclei, which would work provided the heatmap had given good predictions for the nuclei centers. However, the problem was that nuclei had different sizes, which was especially significant for the malignant nuclei.

When picking the radius of nuclei in the NMS algorithm, there were two opposing objectives: the radius should be wide enough to ensure that there was only one prediction per nucleus, but it also should be small enough that it would not suppress a neighboring cell's nucleus. To find the optimal value, we looked at the distribution of the distance from a nucleus to its closest neighbor. Almost all nuclei were more than 8 pixels apart, and therefore, if the radius is made significantly wider than 8 pixels, the nucleus's neighbors would be suppressed. Empirically, a radius of 10 gave the best results.

Moreover, since the nuclei in different classes have different average sizes, conditioning the radius on the predicted class of the nuclei was also investigated. Intuitively, it makes sense to use a larger radius for the larger malignant nuclei, but empirically, this did not lead to significant improvements.

4   RESULTS

Figure 3 depicts the heatmaps generated for a sample input image to the network. Table 1 presents the classification accuracy, sensitivity, and specificity achieved by the proposed method on the test set. The results were compared with a machine learning method based on handcrafted features. The rival method consisted of segmentation of nuclei figures using decorrelation stretching technique[23], feature extraction based on extracted textural, morphological, and spatial features to describe the nuclei figures properties, and a support vector machine (SVM) with radial basis function (RBF) kernel as the classifier[12]. The rival method was labeled as SVM in Table 1. As can be observed from these results, the

proposed method based on a fully convolutional network boosted the performance on the classification of the cell nuclei at predicted locations compared to the handcrafted features.

As shown in Table 2, increasing the size of the classifier's input improved the performance. Considering that a cell nucleus fits inside a 32×32 box, these results indicate that contextual information is important, and that the shape and appearance of the nuclei alone is not sufficient to generate a good classification accuracy.
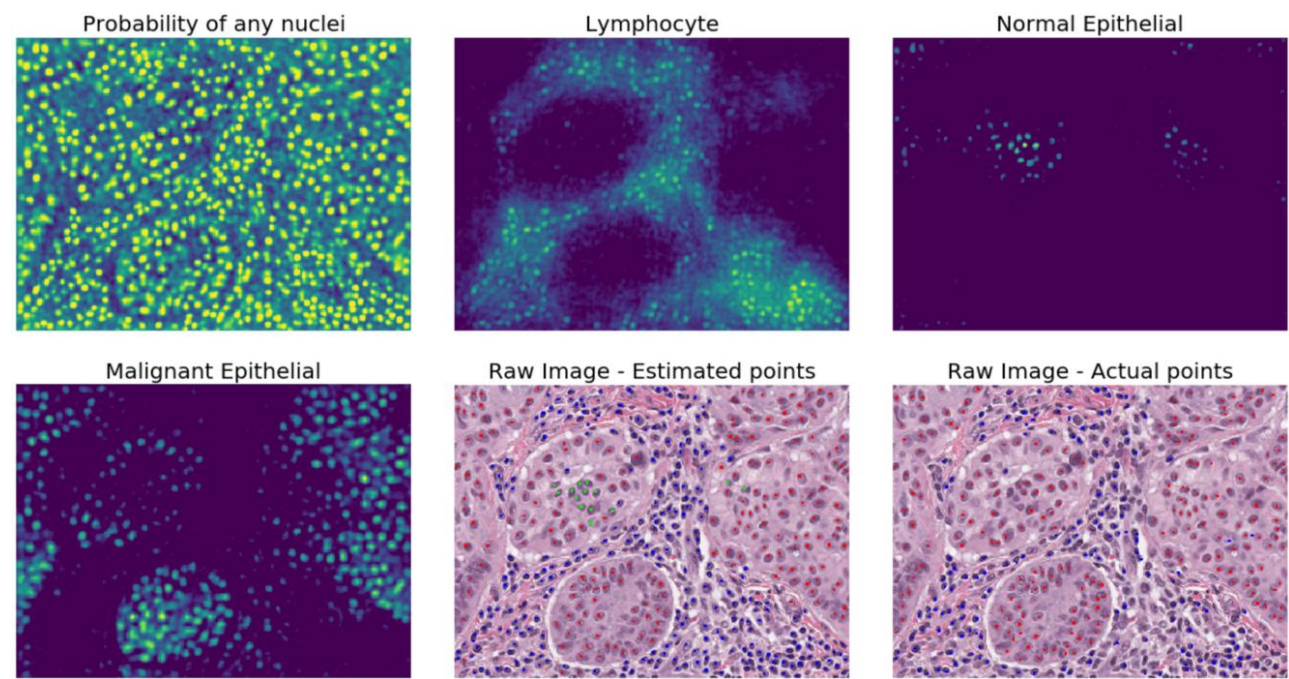

Figure 3- Representative heatmaps produced by the network for a typical input image, and the predictions on the raw image. L, BE, and ME nuclei are shown as blue, green and red dots, respectively (the images should be digitally enlarged for the best view).

Table 1-The performance of the proposed network in comparison with handcrafted features on the test set.

Class	Accuracy (%)		Sensitivity (%)		Specificity (%)	
	FCN	SVM	FCN	SVM	FCN	SVM
L	99	92	93	80	100	94
BE	95	75	99	50	94	92
ME	95	77	93	91	99	63

Abbreviations: L: Lymphocyte
                 BE: Benign Epithelial
                 ME: Malignant Epithelial
                 FCN: Fully Connected Network (proposed approach)
                 SVM: The rival method based on handcrafted features and SVM classifier

Table 2-Accuracy compared to the input size for the CNN.

Input size	Classification Accuracy
32x32	82.7%
64x64	93.7%
128x128	93.9%
192x192	94.6%

5   CONCLUSIONS

A fully convolutional network equipped with inception modules was designed to localize and classify cell nuclei in post-neoadjuvant therapy tissue slides of breast cancer. The proposed network accurately localized the cell nuclei using a pre-trained network, and based on the generation of heatmaps of probability of each nuclei class followed by a non-maximum suppression algorithm. The predicted cell nuclei locations were subsequently classified as lymphocyte, benign epithelial or malignant epithelial with high accuracy, sensitivity, and specificity. The results validated the viability of the proposed approach in this application for the first time, particularly in comparison with handcrafted features.

ACKNOWLEDGMENTS

REFERENCES

[1]   Faneyte, I.F., Schrama, J.G., Peterse, J.L., Remijnse, P.L., Rodenhuis, S., and Vijver, M.J. Van De, "Breast cancer response to neoadjuvant chemotherapy : predictive markers and relation with outcome," British Journal of Cancer 88(3), 406–412 (2003).

[2]   Nahleh, Z., Sivasubramaniam, D., Dhaliwal, S., Sundarajan, V., and Komrokji, R., "Residual cancer burden in locally advanced breast cancer : a superior tool," Current Oncology 15(6), 271–278 (2008).

[3]   Rajan, R., Poniecka, A., Smith, T.L., Yang, Y., Frye, D., Pusztai, L., Phil, D., Fiterman, D.J., Gal-gombos, E., et al., "Change in tumor cellularity of breast carcinoma after neoadjuvant chemotherapy as a variable in the pathologic assessment of response," Cancer 100(7), 1365–1373 (2004).

[4]   Shelhamer, E., Long, J., and Darrell, T., "Fully convolutional networks for semantic segmentation," Transactions on Pattern Analysis and Machine Intelligence 39(4), 640–651 (2017).

[5]   Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in Int. Conf. Med. Image Comput. Comput. Interv., 234–241 (2015).

[6]   Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., and Barbano, P.E., "Toward automatic phenotyping of developing embryos from videos," IEEE Transactions on Image Processing 14(9), 1360–1371 (2005).

[7]   Ciresan, D., Giusti, A., Gambardella, L.M., and Schmidhuber, J., "Deep neural networks segment neuronal membranes in electron microscopy images," in Adv. Neural Inf. Process. Syst., 2843–2851 (2012).

[8]   Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards real-time object detection with region proposal networks," in Adv. Neural Inf. Process. Syst., 91–99 (2015).

[9]   Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 779–788 (2016).

[10]   Cohen, J.P., Lo, H.Z., and Bengio, Y., "Count-ception: Counting by fully convolutional redundant counting," arXiv preprint arXiv:1703.08710 (2017).

[11]   Xie, W., Noble, J.A., and Zisserman, A., "Microscopy cell counting and detection with fully convolutional regression networks," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization1–10 (2016).

[12]   Peikari, M., Salama, S., Nofech-Mozes, S., and Martel, A.L., "Automatic cellularity assessment from post-treated breast surgical specimens," Cytometry Part A 91(11), 1078–1087 (2017).

[13]   Janowczyk, A., Doyle, S., Gilmore, H., and Madabhushi, A., "A resolution adaptive deep hierarchical (RADHicaL) learning scheme applied to nuclear segmentation of digital pathology images," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization1–7 (2016).

[14] Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.-W., Snead, D.R.J., Cree, I.A., and Rajpoot, N.M., "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," IEEE Transactions on Medical Imaging 35(5), 1196–1206 (2016).

[15] Chen, H., Dou, Q., Wang, X., Qin, J., Heng, P.-A., and others, "Mitosis detection in breast cancer histology images via deep cascaded networks.," in AAAI, 1160–1166 (2016).

[16] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wonja, Z., "Rethinking the inception architecture for computer vision," in IEEE Conf. Comput. Vis. Pattern Recognit., 2818–2826 (2016).

[17] Martel, A.L., Hosseinzadeh, D., Senaras, C., Zhou, Y., Yazdanpanah, A., Shojaii, R., Madabushi, A., Gurcan, M.N., "An image analysis resource for cancer research: PIIP—pathology image informatics platform for visualization, analysis, and management," Cancer Research 77(21), e83–e86 (2017).

[18] Ioffe, S., and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Int. Conf. Mach. Learn., 448–456 (2015).

[19] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research 15, 1929–1958 (2014).

[20] Smith, L.N., "Cyclical learning rates for training neural networks," in IEE Winter Conf. Appl. Comput. Vis., 464–472 (2017).

[21] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., et al., "TensorFlow: Large-scale machine learning on heterogeneous systems" (2015).

[22] Chollet, F., and others, "Keras" (2015).

[23] Peikari, M., and Martel, A.L., "Automatic cell detection and segmentation from H&E stained pathology slides using colorspace decorrelation stretching," in SPIE Med. Imaging, 979114 (2016).