This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering

Pixel-to-pixel Learning with Weak Supervision for Single-stage Nucleus Recognition in Ki67 Images

Fuyong Xing, Toby C. Cornish, Tell Bennett, Debashis Ghosh, and Lin Yang, *Member, IEEE*

Abstract*—*Objective: **Nucleus recognition is a critical yet challenging step in histopathology image analysis, for example in Ki67 immunohistochemistry stained images. Although many automated methods have been proposed, most use a multi-stage processing pipeline to categorize nuclei, leading to cumbersome, low-throughput and error-prone assessments. To address this issue, we propose a novel deep fully convolutional network for single-stage nucleus recognition. *Methods:* Instead of conducting direct pixel-wise classification, we formulate nucleus identification as a deep structured regression model. For each input image, it produces multiple proximity maps, each of which corresponds to one nucleus category and exhibits strong responses in central regions of nuclei. In addition, by taking into consideration the nucleus distribution in histopathology images, we further introduce an auxiliary task, region of interest (ROI) extraction, to assist and boost the nucleus quantification with weak ROI annotation. The proposed network can be learned in an end-to-end, pixel-to-pixel manner for simultaneous nucleus detection and classification. *Results:* We have evaluated this network on a pancreatic neuroendocrine tumor Ki67 image dataset, and the experiments demonstrate that our method outperforms recent state-of-the-art approaches. *Conclusion:* We present a new, pixel-to-pixel deep neural network with two sibling branches for effective nucleus recognition and observe that learning with another relevant task, ROI extraction, can further boost individual nucleus localization and classification. *Significance:* Our method provides a clean, single-stage nucleus recognition pipeline for histopathology image analysis, especially a new perspective for Ki67 image quantification, which would potentially benefit individual object quantification in whole-slide images.**

***Index Terms*—Nucleus classification, nucleus detection, fully convolutional networks, Ki67, neuroendocrine tumor, microscopy images**

I. Introduction

Histopathology images usually exhibit different types of cells including lymphocytes, epithelial cells and stromal cells. Cellular morphology and spatial configuration of nuclei/cells are related to disease development and used as biomarkers to determine tumor grades and facilitate treatment decision-making [1], [2]. For example, in order to calculate the Ki67 labeling index, which is a cellular marker to measure growth fraction of tumors and used clinically to predict survival and tumor recurrence in several types of cancer such as gastrointestinal and pancreatic neuroendocrine tumors (NETs), it is necessary to correctly count immunopositive tumor and immunonegative tumor, while ignoring non-tumor cells [3]. Thus, it is very critical to recognize or categorize nuclei/cells for disease characterization. However, manually recognizing nuclei/cells is prohibitively expensive because there might be a large number of images in one single experiment, each of which can have tens of thousands of or even more nuclei/cells [4]. Additionally, manual image analysis assessment leads to potential inter- and intra-observer variations [5], [6]. Therefore, automated methods that can greatly improve the efficiency and objectiveness have drawn considerable attention in recent digital histopathology and microscopy image analysis [7], [8].

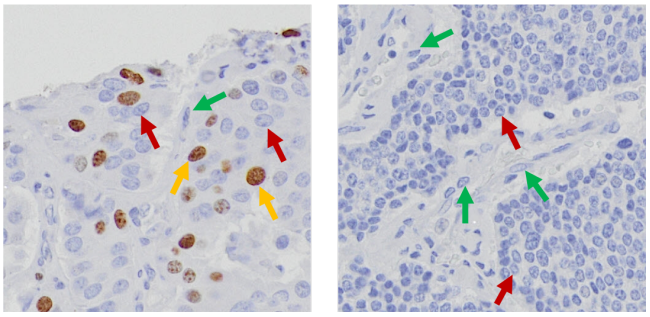


Fig. 1. Example Ki67 immunohistochemistry stained images of pancreatic neuroendocrine tumor. Color coding of arrows: yellow, red and green for immunopositive tumor nuclei, immunonegative tumor nuclei and non-tumor nuclei, respectively.

Automated nucleus recognition in histopathology images contains individual nucleus detection and classification, which are very challenging tasks. First, histopathology images can have strong background noise or image artifacts due to tissue sectioning, staining or imaging characteristics. Second, cellular characteristics including intensity, scale and shapes can exhibit significant intra-class but slight inter-class variations such that robust nucleus classification can be very difficult. For instance, in Ki67 immunohistochemistry (IHC) stained images, tumor (e.g., immunopositive or immunonegative) and non-tumor (e.g., glandular or inflammatory) nuclei may exhibit similar appearance in terms of color and shape. Finally, it is not unusual that nuclei are densely clustered and they

F. Xing and D. Ghosh are with the Department of Biostatistics and Informatics and the Data Science to Patient Value Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO 80045.

T. C. Cornish is with the Department of Pathology, University of Colorado Anschutz Medical Campus, Aurora, CO 80045.

T. Bennett is with the Department of Pediatrics and the Data Science to Patient Value Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO 80045.

L. Yang is with the Department of Electrical and Computer Engineering, the Department of Computer and Information Science and Engineering, and the J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering

might touch or even partially overlap each other, leading to ambiguous boundary cues for individual object detection and classification. The challenges are illustrated in Figure 1.

Early nucleus recognition approaches mainly rely on non-learning-based image processing algorithms; however, these methods might not precisely transfer domain knowledge into rules, and manual algorithm adaptation is necessary for different datasets or images. Machine learning is an alternative approach that infers rules from example data and has been widely applied to pathology and microscopy image analysis [9]. Nevertheless, classic machine learning relies heavily on data representations (features), which are often manually designed and require significant domain expertise. This feature engineering is a non-trivial task, especially for complex histopathological images. Deep neural networks, which automatically learn multi-level feature representations from raw data, have demonstrated state-of-the-art performance in various biomedical image computing applications [8], [10], [11]. Convolutional neural networks (CNNs) [12] and fully convolutional networks (FCNs) [13], in particular, have been applied to nucleus/cell localization and classification, leading to improved accuracy. However, almost all previous methods, particularly non-deep learning algorithms, adopt a multi-stage processing pipeline, which typically consists of nucleus/cell detection (and/or segmentation), cellular feature extraction and classification. The final object recognition significantly depends on the previous stages, which are themselves very challenging tasks. This multi-stage pipeline leads to a low-throughput image analysis, which prevents the methods from scaling up to large datasets.

In this paper, we propose KiNet, a novel FCN architecture (see Figure 2) for single-stage nucleus recognition in histopathology images. KiNet is capable of differentiating immunopositive tumor nuclei, immunonegative tumor nuclei and non-tumor nuclei in pancreatic NET Ki67 images. More specifically, we present a structured regression model to encode the position information of nucleus centers so that the network is encouraged to predict high values for pixels in central regions of nuclei. In order to enhance the discriminative power for nucleus classification, a correlation penalization is introduced into inter-class hidden representations and response maps. Furthermore, a region-of-interest (ROI) extraction task is incorporated into the network to assist the nucleus identification task. This joint learning network is trained in an end-to-end, pixel-to-pixel manner. For each input image, it predicts multiple class-aware proximity maps (i.e., one for each nucleus category), where local maxima indicate nucleus locations for each class. The proposed model is extensively evaluated on a set of pancreatic NET images and compared with several recent state-of-the-art deep models. Additionally, an experimental analysis of important parameter selection is also provided. In summary, our contributions are three-fold:

- We propose a novel pixel-to-pixel deep neural network with two sibling branches, which significantly outperforms multiple recent state-of-the-art deep models on nucleus recognition, including both individual object detection and classification.
- We take advantage of an auxiliary task, ROI extraction, to assist and further boost nucleus recognition. This ROI extraction task only requires very weak data annotation, which is much easier to achieve than fine-grained and expensive individual nucleus labeling.
- Our network allows for single-stage, simultaneous nucleus detection and classification, which can greatly facilitate Ki67 labeling index assessment in Ki67 IHC stained images. Compared with previous multi-stage processing approaches, it is more concise and efficient, and requires no disk storage for feature representation caching. More important, it does not need individual nucleus segmentation, which still remains very challenging.

II. Related Work

Nucleus or cell recognition is a fundamental task in the analysis of cytologic and histologic images. Early methods mainly use basic image processing techniques to identify individual nuclei/cells, but they usually produce inferior performance compared to machine learning-based approaches [9]. Therefore, supervised learning has been widely applied to object classification in microscopy images. Larsen *et al.* [14] have trained a support vector machine (SVM) with shape index histograms to classify different types of HEp-2 cell staining patterns in fluorescence images, Qi *et al.* [15] have learned an SVM with both texture and shape information for cell classification, and Xu *et al.* [16] have employed a linear SVM classifier with a co-occurrence differential texton feature to recognize HEp-2 cells. Other learning-based approaches, such as linear classifiers, Bayes classifiers, conventional artificial neural networks, and so on., have also been applied to nucleus/cell classification [17], [18], [19]. All of these methods require manual feature engineering, which is a challenging task for microscopy images, especially for histopathology images that exhibit significant variation in appearance.

Recently CNNs have achieved state-of-the-art object recognition performance in various digital pathology applications [8], [20]. Zhang *et al.* [21] have adopted a deep CNN to separate abnormal from normal cervical cells in Pap-stained and hematoxylin and eosin (H&E) stained images, Gao *et al.* [22] have trained a CNN to classify HEp-2 cells in fluorescence microscopy images, and Liu *et al.* [23] have combined a CNN with a deep autoencoder for individual HEp-2 cell classification. These papers show that deep CNNs outperform other traditional machine learning methods. In order to deal with limited training data, transfer learning of CNNs has been reported for nucleus/cell recognition. Bayramoglu and Heikkilä [24] have fine-tuned CNNs trained on natural image data towards colon cancer histopathology image data for nucleus classification, while Phan *et al.* [25] have used a pre-trained CNN as a fixed feature extractor and learn multiple SVMs with CNN features for HEp-2 cell recognition. On the other hand, a CNN with active learning is presented in [26] for nucleus classification in H&E stained pathology images, which dynamically selects the most informative nuclei for human annotation. All of these deep learning-based methods, as well as those aforementioned traditional machine learning approaches, assume that the positions of nuclei/cells are known

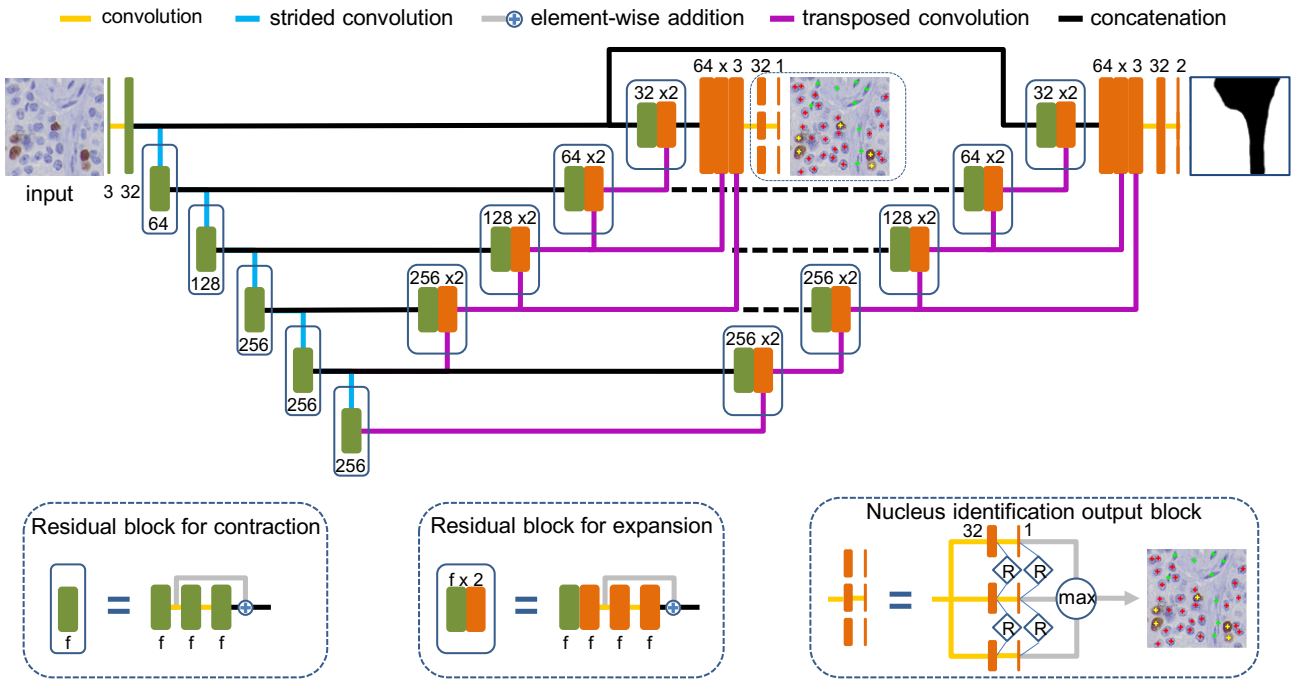


Fig. 2. Proposed network architecture. The green or orange boxes denote feature maps, and the number of feature maps in each layer is provided below or above the feature map. The connections with different colors represent distinct operations. To avoid cross connections for better illustration, we use dashed black lines to present concatenation operations from the contraction path to the ROI extraction expansion path. In the output block of nucleus identification, *R* denotes the regularization/penalization between feature maps of different types of nuclei and for clarity, some regularization notation is omitted; *max* is an element-wise operation among different feature maps.

a priori and that a single object is centered at each input image, but this assumption does not hold in real applications. In digital pathology, one single microscopy image usually has thousands of or more nuclei/cells and individual object localization itself is a very challenging problem [27].

Many automated methods have been proposed to localize nuclei or cells for object classification in different kinds of microscopy images. Xing *et al.* [3] have detected and segmented individual nuclei with a radial symmetry voting algorithm and a repulsive deformable model, and then train multiple SVM classifiers with morphometry and intensity features to recognize different types of nuclei in pancreatic NET images. A similar pipeline is presented in [28], which computes object representations with contextual information, for nucleus classification in H&E stained and IHC stained histology images. Wang *et al.* [29] have used graph cut and level set algorithms to segment nuclei and then extract nuclear structure information as object-level features for SVM-based nucleus classification in Feulgen stained images. Chen *et al.* [30] have adopted a semi-automatic method to segment individual Pap smear cells and then train an SVM with cellular morphology and texture features for cell classification. Some other methods learning with cellular features for cell or nucleus classification are reported in [31], [32], [33], [34], which provide promising performance in various microscopy images. However, all the approaches above use a multi-stage processing pipeline for nucleus/cell recognition, which typically consists of object detection and/or segmentation, feature representation computation, and classification, leading to image analysis inefficiency and great variability. In particular, those methods that require individual nucleus/cell segmentation might be error-prone, because object segmentation is very difficult and still remains unsolved in microscopy image analysis.

Deep neural networks, which learn feature representations from raw image data and are able to obviate individual object segmentation, provide an alternative approach to nucleus/cell recognition. Cireşan *et al.* [35] have introduced a CNN model to identify mitotic nuclei in H&E stained breast cancer images. Two partially shared CNNs [36] are trained independently for nucleus detection and classification in lung cancer images. These CNN-based methods take as input small image patches and conduct model inference in a sliding window manner, which is computationally expensive for pixel-wise predictions. To avoid dense predictions for nucleus classification, CNNs can be applied on only selected patch candidates [37], [38]. However, patch candidate extraction in these approaches still requires pixel-wise predictions; more important, it introduces an additional step of object candidate detection, adding another layer of inefficiency. Recently, a CNN with multiple instance learning [39] is proposed for cell localization and recognition, but it requires back-propagating class-specific feature maps to the input space. On the other hand, a deep neural network [40] is presented for nucleus detection and classification in colorectal adenocarcinoma histopathology images, and it is able to predict nucleus categories with one-pass forward propagation. This work differs from ours because it uses the detection as an object prior to constrain the classification. Our FCN model takes advantage of weak labels, i.e. ROI region annotation, to assist individual nucleus identification. This auxiliary task boosts nucleus identification by encouraging the network to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering

learn more general representations. More important, it can reduce human effort for fine-grained nucleus annotation, which is much more expensive. In addition, the method in [40] requires multi-stage model training, while our model allows for more efficient single-stage training.

III. Pixel-to-pixel Learning for Single-stage Nucleus Recognition

Given histopathology images with weak ROI (i.e., tumor and non-tumor regions) annotation as well as labeling of nucleus positions and categories, our goal is to design a deep FCN model for single-stage classification of individual nuclei. In order to take into account the topology in the label space, we formulate nucleus recognition as a deep structured regression problem. Meanwhile, we take advantage of data information from a related task, ROI extraction, and use it to assist the nucleus identification task. These two tasks are unified into a single FCN network that has two output branches, one per task. The network is jointly trained in an end-to-end, pixel-to-pixel manner and directly produces pixel-wise predictions for individual object localization and labeling.

A. Model Architecture

Our network architecture is built on the recent state-of-the-art neural network, U-Net [41], a variant of FCN. We replace the conventional convolution connections with residual learning [42] such that the risk of gradient vanishing can be alleviated. Additionally, in order to handle scale variation of nuclei, we apply a multi-context aggregation to fusion of multi-level features for object localization. We also introduce a sibling output branch, ROI prediction with weak supervision, to assist and boost nucleus recognition. Unlike previous patched-based CNN methods, our model takes as input an arbitrary-sized image and produces an identical-sized output without using a sliding window. In this scenario, it significantly reduces the computational complexity for dense prediction.

Pixel-to-pixel learning for nucleus identification. This branch is composed of contraction, expansion and multi-context aggregation paths, as shown in Figure 2. The contraction path mainly consists of multiple residual learning blocks, aiming to extract hierarchical feature representations from input images. Each residual block contains two 3×3 convolutions with padding 1, with each followed by a batch normalization [43] and an exponential linear unit (ELU) [44]. Additionally, a shortcut connection is used to conduct an identity mapping and its output is added to the stacked non-linear output. A strided convolution with stride 2 is exploited to downsample feature maps between two consecutive residual blocks and in total four blocks are stacked by following the rule [45]: double the number of feature maps when their dimensions are halved. The expansion path also consists of four residual blocks but uses transposed convolution [46] to upsample feature maps to a desired size for dense prediction. In order to facilitate object localization in the final output maps, high-resolution information from the contraction path are copied and concatenated to corresponding upsampled ouputs for successive learning. Note that the expansion and contraction paths might not be exactly symmetric such that their output feature maps do not have the same size and are not eligible for direct pixel-wise concatenation. To address this problem, we pad the upsampled feature maps with zeros to match corresponding downsampled ones.

A network with a single-sized receptive field might not localize objects with different scales effectively. In histopathology images, nuclei often exhibit different sizes such that learning with a single receptive field would be difficult to capture proper information for localization. Inspired by [47], we apply a multi-context aggregation to information fusion from the hierarchy of feature maps. Specifically, we perform transposed convolution on different levels of downsampled feature maps, which correspond to distinct sizes of receptive fields, and concatenate these transposed convolutional outputs to the last upsampled feature map. Then this concatenated map is fed into an output block, which consists of parallel two-convolutional (3×3) layers, for class-specific pixel-wise predictions, which have the same size as the input image and each corresponds to one nucleus category.

Sibling outputs for ROI prediction. Non-tumor nuclei typically appear in glandular regions, stroma and other non-tumor areas, and these non-tumor regions usually exhibit different spatial distributions of nuclei from tumor regions. In practice, pathologists identify tumor nuclei based on not only nuclear appearances, such as object size, shape, intensity, and so on., but also on spatial configuration of nuclei. Tumor nuclei are typically adjacent to tumor nuclei, and vice versa. By taking into consideration these histopathological image characteristics, we design a specific sibling branch for region classification, i.e. ROI extraction, which is used to assist nucleus classification. Because this sibling output branch only focuses on ROI extraction, it should have different higher layers from the nucleus classification branch, although the lower layers that learn generic feature representations can be shared. In order to capture spatial configuration of nuclei, which has a broader view, it is necessary for the neurons in this newly introduced branch to have larger receptive fields compared to the nucleus recognition branch. Therefore, we add another residual block after the shared components (see Figure 2). Then, an expansion path that consists of five upsampled residual blocks is linked to this newly introduced residual block for a pixel-to-pixel mapping. The output of this expansion path is fed to two successive 3×3 convolutions followed by a softmax function for region classification.

The two tasks, nucleus identification and ROI extraction, influence each other through the shared representations in the low layers. The interaction between these tasks benefits each other by encouraging each to pay attention to only its own task and thus making each easier to learn. It will allow the network to learn more general representations to boost the nucleus identification performance. Another benefit of this joint learning is that ROI annotation is much weaker than fine-grained nucleus labeling. With limited annotation of individual nucleus localization and categories, using weakly-annotated ROIs is implicitly augmenting training data such that it can lower the overfitting risk [48]. Note that we only need to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering

(a) (b) (c)

Fig. 3. Weak ROI annotation. (a) Original image, (b) tumor (white) and non-tumor (black) regions, and (c) weak ROI annotation used for model training. Here only a randomly selected subset of, instead of entire, tumor and non-tumor regions are annotated.

randomly annotate a subset of tumor and non-tumor regions, rather than fully label all the images pixels (see Figure 3). This weak annotation is very easy to achieve and adds only a small overhead compared with fine-grained nucleus annotation alone.

B. Loss Function

The overall loss function has two components corresponding to the two different tasks, individual nucleus identification and ROI extraction.

Loss for nucleus identification. In our modeling, nucleus identification is formulated as a class-aware structured regression problem. Compared with commonly used methods of classification in nucleus/cell detection, regression modeling is able to exploit additional context information for better prediction during the learning phase [49], [50]. This is achieved by designing a regression output layer of the FCN network, which for each input image produces a structured-label output, i.e. multiple continuous-valued proximity maps, one per nucleus category. This regression modeling naturally fits in the pixel-to-pixel FCN network, which takes as input arbitrary-sized raw images and directly outputs equal-sized mask predictions. Formally, for image $\mathbf{x}^i \in \mathbb{R}^{w\times h\times c}$ where w, h and c are the width, height and the number of channels, respectively, we calculate its structured label containing k gold-standard proximity maps $\mathbf{y}^i \in \mathbb{R}^{w\times h\times k}$ (see Figure 4), where k is the number of nucleus categories (k=3 in our application), as follows

$$\mathbf{y}^i(u,v,s) = \begin{cases} \frac{e^{\alpha(1-\frac{D^i(u,v,s)}{d})}-1}{e^{\alpha}-1} & \text{if } D^i(u,v,s) \le d \\ 0, & \text{ifotherwise,} \end{cases} \quad (1)$$

where $D^i(u,v,s)$ is the Euclidean distance between pixel $(u,v) \in \mathbb{R}^{w\times h}$ and its closet annotated nucleus for class $s \in \{1,2,\cdots,k\}$. d is a distance threshold and α is a parameter controlling the proximity value decay. This function is normalized such that its value is in $[0,\ 1]$. With the threshold d, the major portion of the proximity map are assigned zeros and only small regions around object centers exhibit positive continuous values. The closer a pixel is to the center, the higher value it has.

Given a set of training images and corresponding proximity maps, we aim to learn an FCN-based regression function. One straightforward method is to minimize the mean square error (MSE) of the predictions. However, this standard MSE loss uses equal weights for all pixels and might lead to a trivial solution that simply assigns zeros to all the pixel predictions, because a dominant portion of each gold-standard proximity map is zero [49]. Intuitively, the pixels in central regions of nuclei should make more contributions than the others. To this end, we extend a weighted MSE loss [50] to handle multi-class regression. Let $\hat{\mathbf{y}}^i \in \mathbb{R}^{w\times h\times k}$ denote the prediction for image $\mathbf{x}^i$. The loss of nucleus identification for image $\mathbf{x}^i$ is defined as

$$\mathcal{L}(\hat{\mathbf{y}}^i,\mathbf{y}^i) = \frac{1}{2}\sum_{(u,v,s)\in\mathbf{y}^i} (y^i(u,v,s) + \lambda\bar{y}^i_s) \cdot(\hat{y}^i(u,v,s) - y^i(u,v,s))^2, \quad (2)$$

where λ controls the weights for different image regions. $\bar{y}^i_s$ denotes the mean value of $\mathbf{y}^i$ for the s-th class of nuclei and enables the learning to automatically adjust the weight for each training image. Because of $y^i(u,v,s) = 0$ for pixels in non-central regions of nuclei, the prediction errors for those central regions are penalized more than the other regions such as image background. In this way, the model would tend to predict non-zero values at nucleus centers.

For each input image $\mathbf{x}^i$, non-zero valued pixels in its proximity maps are mutually exclusive in terms of nucleus classes, i.e., pixels with positive values for one type of nuclei should have zero values for any other types of nuclei. To facilitate class-specific proximity map generation, we introduce a class-wise regularization to further penalize the correlation between different hidden representations in the output block of nucleus identification branch (see Figure 2) during model training as follows

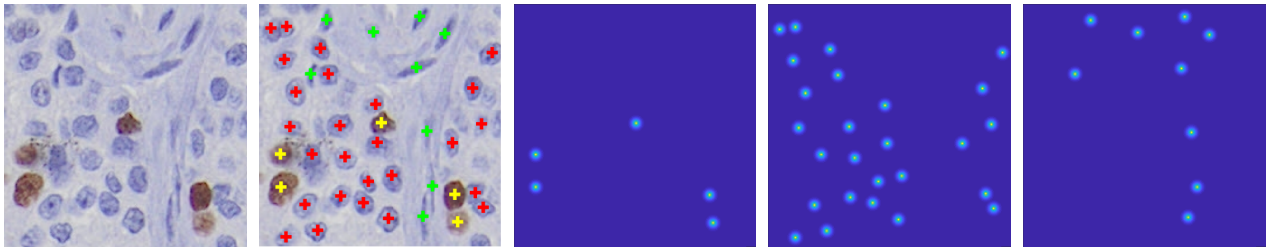
$$\mathcal{L}_{NI}(\hat{\mathbf{y}}^i,\mathbf{y}^i) = \mathcal{L}(\hat{\mathbf{y}}^i,\mathbf{y}^i) + \eta\sum_l\sum_{s=1}^{k-1}\sum_{t=s+1}^{k} \hat{\mathbf{h}}^{i,l}_s * \hat{\mathbf{h}}^{i,l}_t, \quad (3)$$

where l is the layer index in the nucleus identification output block and $\hat{\mathbf{h}}^{i,l}_s$ is the s-th channel of hidden representation at the l-th layer. For the last layer, hidden representations became proximity maps, $\hat{\mathbf{h}}^{i,l}_s = \hat{\mathbf{y}}^{i,l}_s$. The η is a weight parameter and $*$ denotes the dot product operation.

Loss for ROI extraction. ROI extraction is posed as a pixel-wise binary classification problem. For training image $\mathbf{x}^i$, we denote its gold-standard ROI label by 2 two-dimensional binary masks $\mathbf{z}^i \in \{0,1\}^{w\times h\times 2}$, one for tumor and the other for non-tumor regions. Each mask has value 1's for annotated regions and 0's for the others. We choose the cross-entry loss over pixels for this task. Because the ROIs are not fully annotated, we calculate the loss based on only annotated regions. Formally, denoting the ROI prediction for image $\mathbf{x}^i$ by $\hat{\mathbf{z}}^i \in \mathbb{R}^{w\times h\times 2}$, we define its loss as follows

$$\mathcal{L}_{ROI}(\hat{\mathbf{z}}^i,\mathbf{z}^i) = -\sum_{(u,v,s)\in\mathbf{z}^i} z^i(u,v,s)\log\hat{z}^i(u,v,s), \quad (4)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering



(a) (b) (c) (d) (e)

Fig. 4. Proximity map generation. (a) Original image and (b) manual annotation, where color coding is the same as that in Figure 1. (c-e) Proximity maps for immunopositive tumor, immunonegative tumor and non-tumor nuclei, respectively. Only central regions of nuclei in the proximity maps have continuous, non-zero values.

where $\hat{\mathbf{z}}^i \in [0,1]$ is the output of the softmax function.

Joint loss. Because fine-grained labeling of individual nuclei is more expensive than ROI annotation, it is very common that some training images only have ROI annotation available. Therefore, for a training dataset $\{\mathbf{x}^i, \mathbf{y}^i, \mathbf{z}^i\}_{i=1}^N$, where $\mathbf{y}^i$ is empty when nucleus localization and labeling are not available, the joint loss for nucleus identification with the assistance of ROI extraction is

$$\mathcal{L}_{joint} = \frac{1}{N}\sum_{i=1}^{N}\left[\gamma \mathbb{I}(\mathbf{y}^i \neq \emptyset)\mathcal{L}_{NI}(\hat{\mathbf{y}}^i, \mathbf{y}^i) + \mathcal{L}_{ROI}(\hat{\mathbf{z}}^i, \mathbf{z}^i)\right], \quad (5)$$

where $\mathbb{I}(\cdot)$ is an indicator function and γ is a factor balancing the contributions from nucleus identification and ROI extraction. In this scenario, the model training is not only strongly supervised with individual nucleus localization and label annotation if available, but also assisted by the region classification with weak supervision.

C. Training and Testing

The loss function in (5) is differentiable and the model can be trained with the standard backpropagation algorithm [12]. Because we might not have high-quality predicted proximity maps for different classes of nuclei at the beginning of the learning phase such that the dot product in (3) could be a very large value, we exclude the regularization term (i.e., set $\eta = 0$) when the number of training iterations is less than 10^4 to avoid the gradient explosion problem. In addition, because the number of training images with fine-grained nucleus annotation can be significantly less than that of images with only ROI annotation, we select a proper value of γ in (5) based on the model performance on the validation set.

In the testing stage, for an input image $\mathbf{x} \in \mathbb{R}^{w\times h\times c}$, the model outputs k proximity maps, $\hat{\mathbf{y}} \in \mathbb{R}^{w\times h\times k}$ for k different classes of nuclei, from the nucleus identification branch and one heatmap, $\hat{\mathbf{z}} \in \mathbb{R}^{w\times h\times 2}$ for tumor regions, from the ROI extraction branch. The pixels with small values in $\hat{\mathbf{y}}$, i.e. less than $\xi \cdot \max(\hat{\mathbf{y}})$ where $\xi \in [0,1]$, are suppressed. Then non-maximum suppression is applied to each channel of $\hat{\mathbf{y}}$ to localize individual nuclei and class labels are achieved by finding the highest value across the channels of $\hat{\mathbf{y}}$.

IV. Experiments and Discussion

A. Dataset

We evaluate the proposed model on 38 pancreatic NET cases with each corresponding to three images of size 500×500. The images are cropped to represent a variety of tissue appearances from whole slide scanned tissue microarray image data, which are Ki67 IHC stained and generated with a bright-field microscope at $20\times$ magnification (about $0.34\mu m$/pixel). Each image has both nucleus labeling (consisting of positions and categories) and weak ROI annotation available. We randomly split the data into two halves in at the case level (19 vs. 19): one for training and the other for testing. We further randomly select 20% (4 cases) of the training set as a validation set with the rest for model training. There is no overlap between the training, validation and testing sets. In the experiments, we aim to localize and classify nuclei into three categories, immunopositive tumor, immunonegative tumor and non-tumor, and this would facilitate the quantification of the Ki67 labeling index, which is a very critical biomarker for NET grading in clinical practice and research.

B. Implementation Details

We train the model from scratch using stochastic gradient descent with Nesterov momentum [51] and set the parameters as: learning rate=0.01, momentum=0.9, weight decay=10^{-6}, batch size=4 and number of iterations=10^5. We stop the training if the performance on the validation dataset does not improve for 2×10^4 iterations. We set $\alpha = 3$, $d = 15$ in Equation (1), $\lambda = 5$ in Equation (2) and $\gamma = 25$ in Equation (5). We penalize the correlation in the last layer and set $\eta = 10^{-6}$ in Equation (3). The hyperparameter of ELU is set as 1.0. Dropout [52] with a rate of 0.5 is used after the convolution operations in the last two residual blocks of the contraction path. For model inference, the suppression parameter ξ is set as 0.5. The model is implemented with PyTorch [53] and trained and tested on a machine with a 2.20 GHz Intel(R) Xeon(R) CPU and an Nvidia GeForce GTX 1080Ti GPU.

In the training phase, we randomly crop four $200 \times 200 \times 3$ patches (i.e., $w = h = 200$ and $c = 3$) from each training image. All patches are preprocessed with zero mean and standard deviation divided in each image channel. Data augmentation including random rotation, shifting and mirroring is exploited

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering

TABLE I
EVALUATION OF NUCLEUS DETECTION AND CLASSIFICATION IN TERMS OF PRECISION (P), RECALL (R), F_1 SCORE, MEAN (μ) WITH STANDARD DEVIATION (σ) OF THE EUCLIDEAN DISTANCE BETWEEN TRUE POSITIVES AND GOLD-STANDARD ANNOTATION, AND THE AREA UNDER THE PRECISION-RECALL CURVE (AUC).

Model	Detection				Classification			
	P	R	F_1	$\mu_{\pm\sigma}$	P	R	F_1	AUC
FCN-8s	0.825	0.615	0.705	$5.087_{\pm 2.846}$	0.779	0.560	0.649	0.525
U-Net	0.859	0.650	0.740	$6.009_{\pm 3.017}$	0.769	0.581	0.660	0.550
FCRNA	0.976	0.511	0.671	$2.137_{\pm 1.805}$	0.853	0.439	0.578	0.590
FCRNB	**0.979**	0.543	0.698	$2.037_{\pm 1.556}$	**0.869**	0.486	0.622	0.637
SFCNOPI	0.952	0.680	0.793	$6.023_{\pm 2.760}$	0.863	0.616	0.716	0.561
FRCN	0.880	0.841	0.860	$4.604_{\pm 2.293}$	0.787	0.750	0.768	0.680
KiNet	0.861	**0.938**	**0.898**	$\mathbf{2.018}_{\pm 1.821}$	0.771	**0.841**	**0.804**	**0.724**

to enlarge the training set. Instead of pre-computing all the training patches, we dynamically crop the patches and apply data augmentation within each iteration to save storage space.

C. Evaluation Metrics

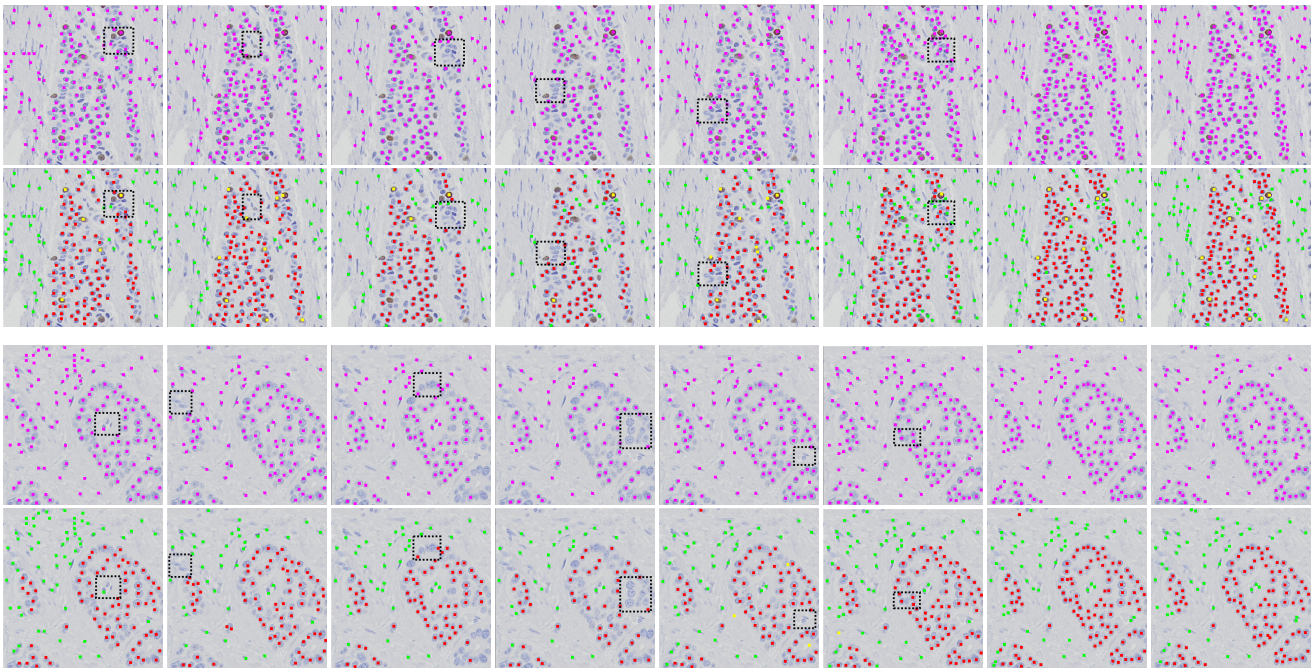
We evaluate our model in terms of two applications: nucleus detection and classification. We follow the metrics in [50] to evaluate detection performance. For each annotated nucleus center, we define its gold-standard region as the area centered at this annotation with a radius $r = 16$, which is roughly the average radius of nuclei in this dataset. The automatically detected centers within the gold-standard regions are matched with corresponding gold-standard annotations using the Hungarian algorithm [54]. Each detection corresponds to at most one gold-standard annotation, and vice versa. All matched detections are considered true positives (TP). Detections that are not matched with any gold-standard annotations are viewed as false positives (FP) and those gold-standard annotations that do not have matched detections are false negatives (FN). With these definitions, we report the precision (P), recall (R) and F_1 score for nucleus detection evaluation: $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$, $F_1 = \frac{2PR}{P+R}$. We use the metrics in [37] for nucleus classification. Specifically, we calculate the precision, recall and F_1 score for each class of nuclei and then their weighted average over all three categories. The weight is the percentage of each category in the entire testing dataset.

D. Model Evaluation

1) Comparison with state-of-the-art methods: We compare the proposed method, KiNet, with recent state-of-the-art deep models including FRCN [50], SFCNOPI [40], FCRNA [55], FCRNB [55], U-Net [41] and FCN-8s [13]. We select these pixel-to-pixel learning and inference models for a fair comparison. Table I shows the comparison results of nucleus detection and classification on the test set, where we also present the mean of Euclidean distances and their standard deviation ($\mu_{\pm\sigma}$) between true positives and corresponding matched gold-standard annotations, and the area under the precison-recall curve (AUC). We note that KiNet produces the best performance in the metrics of recall and F_1 score for both nucleus detection and classification. In particular, it outperforms the others by a large margin of around $4\% \sim 23\%$ in terms of F_1 score, which is the unary evaluation metric for nucleus detection and classification. Specifically, FCRNA and FCRNB produce the lowest F_1 scores, probably because they suffer from inaccurate nucleus localization due to high-resolution information loss in the downsampling operation, especially for those densely clustered nuclei. FCN-8s, U-Net, SFCNOPI and FRCN provide relatively better F_1 scores than FCRNA and FCRNB probably because they take into consideration the feature maps from low layers, but they are still outperformed by KiNet, which also produces significantly better performance in terms of the distance measurement ($\mu_{\pm\sigma}$) for detection and the AUC for classification. In addition, models formulated with direct pixel-wise classification such as FCN-8s, U-Net and SFCNOPI exhibit higher Euclidean distances, perhaps because all pixels inside nuclei contribute equally such that local maxima might not locate at real nucleus centers during testing. These three models also produce significantly lower AUC values than the others, especially compared to KiNet. We also observe that for all methods, precision, recall and F_1 score of nucleus detection are significantly higher than that of classification. This is expected because classification requires the labels of detected nuclei match the gold-standard annotations. Several representative qualitative results are shown in Figure 5.

2) Learning with limited and weak supervision: In this experiment, we evaluate learning with limited and weak supervision for nucleus detection and classification. To this end, we generate the training set by using all the weak ROI annotation in the entire training data and a randomly-selected subset of fine-grained annotation, i.e., $5\%, 10\%, 20\%, 40\%, 60\%, 80\%$ and 100% of training images with nucleus labeling. As shown in Figure 6, when significantly limited (e.g., 5% or 10%) fine-grained annotation is used, both nucleus recognition learning alone and joint learning exhibit very poor performance due to model overfitting. With the increase of fine-grained annotation training data, nucleus detection and classification accuracies grow accordingly for both learning strategies. It shows significant improvements at the beginning (i.e., very limited fine-grained annotation) and then slight accuracy increases when more than 60% fine-grained annotation is used. One potential reason is that the model approaches its capacity with more individual cell labeling. This suggests strong supervision, i.e., individual nucleus annotation, is necessary for nucleus localization and categorization. Without sufficient fine-grained

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering



(a) FCN-8s (b) U-Net (c) FCRNA (d) FCRNB (e) SFCNOPI (f) FRCN (g) KiNet (h) Gold standard

Fig. 5. Comparative nucleus detection (rows 1 and 3) and classification (rows 2 and 4) using different methods on two example images. The automatic results and manual annotation are marked with dots of different colors. Magenta denotes detection in rows 1 and 3. Yellow, red, and green in rows 2 and 4 represent immunopositive tumor, immunonegative tumor and non-tumor, respectively. Some missed or false detection/classification is highighted with black dashed rectangles.

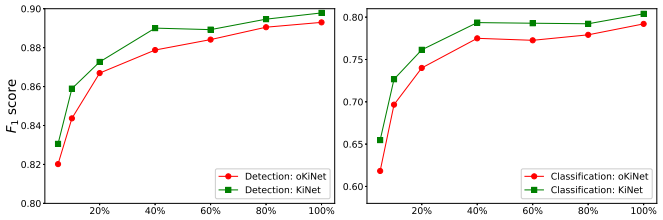


Fig. 6. Comparison between learning with (KiNet) and without (oKiNet) weak supervision, i.e., ROI annotation. The x and y axes represent the percentage of fine-grained annotation data and F_1 score, respectively.

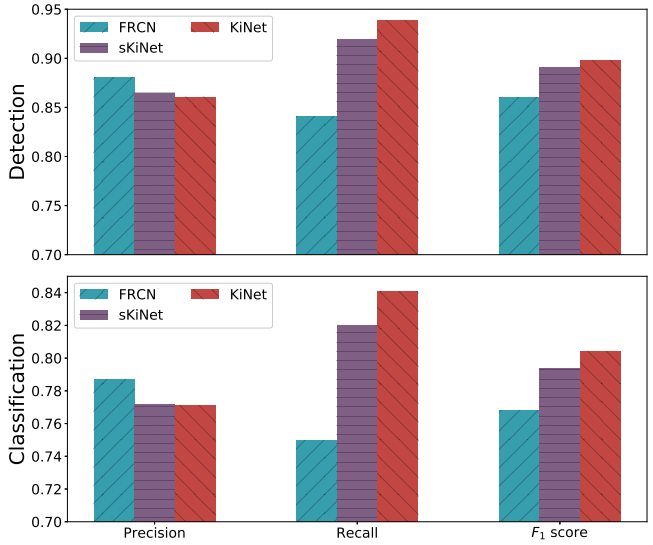


Fig. 7. Comparison between joint learning (KiNet) and stage-wise training (sKiNet) for nucleus detection and classification. We also present the performance of the baseline network, FRCN, for a comparison.

annotation training data, the model would fail to learn useful representations and thus have difficulty detecting and classifying individual nuclei, no matter whether it learns with weak ROI annotation or not.

Figure 6 also demonstrates that compared with learning nucleus recognition alone, learning with an auxiliary task consistently boosts both detection and classification for all the cases with different numbers of fine-grained annotation training images. This might be attributed to the fact that joint learning implicitly augments training data such that the network is able to learn more general representations.We also observe that for a fixed nucleus detection or classification accuracy, joint learning requires less fine-grained annotation data than learning nucleus recognition alone. This is very important because weak ROI annotation is much easier to achieve than individual nucleus labeling, thereby providing an efficient way to reduce human effort for training data annotation.

3) Joint learning versus stage-wise training: In order to evaluate whether joint learning indeed improves nucleus detection and classification in KiNet, we conduct sequential or stage-wise training of KiNet, referred to as sKiNet. Here we first train the ROI extraction branch with keeping the other branch fixed, and then train the nucleus identification branch while freezing the other. Figure 7 shows the comparative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2019.2900378, IEEE Transactions on Biomedical Engineering

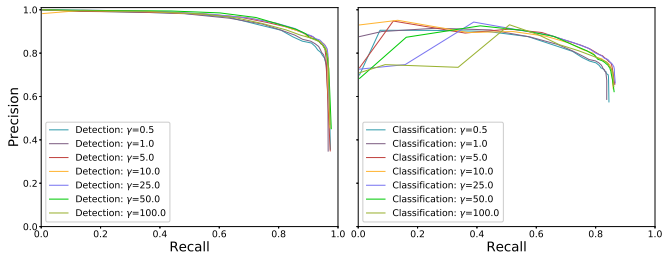


Fig. 8. Precision-recall curves of nucleus detection and classification with different γ values in Equation (5). Each curve is generated by varying the threshold ξ on the final predicted maps. The x/y axis represents recall/precision.

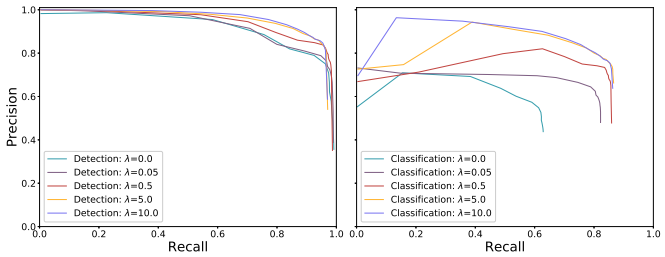


Fig. 9. Precision-recall curves of nucleus detection and classification with different λ values in Equation (2). Each curve is generated by varying the threshold ξ on the final predicted maps. The x/y axis represents recall/precision.

results, where we also present the performance of a baseline network, FRCN [50], which is a recent state-of-the-art FCN model for nucleus/cell quantification. Compared to sKiNet, KiNet improves by about 1% F_1 score for both nucleus detection and classification. This suggests joint learning, which allows both tasks to influence each other, has a consistent positive effect. More important, KiNet training is single-stage via the joint loss and thus avoids managing a sequential task learning pipeline. Finally, it is worth mentioning that both sKiNet and KiNet significantly outperform the baseline by a large margin, which demonstrates the effectiveness of the proposed network architecture.

4) Effects of Parameters: Our nucleus recognition modeling has a critical parameter γ in Equation (5), which controls the loss balance between two tasks, nucleus recognition and ROI extraction. Figure 8 displays the precision-recall curves of nucleus detection and classification with respect to γ. Each curve is generated by varying ξ from 0 to 1 with an interval of 0.05. As we can see, learning with a small γ value (e.g., $\gamma < 1$) leads to a low accuracy; when increasing γ, the performance improves accordingly. In general, our model can achieve a stable performance within a wide range of γ values from 5 to 50. Nucleus detection does not exhibit significant performance variation when $\gamma > 1$; however, nucleus classification might slightly decrease the accuracy for a much higher γ value (e.g., 100). The reason is that during model training, the ROI extraction task produces a higher loss than the nucleus recognition task and would dominate the joint loss due to the imbalanced contributions; with a proper weighting of the nucleus recognition task, it is able to balance their contributions to avoid bias predictions.

Another important parameter is the weight λ in Equation (2), which is used to weight the contributions of non-zero valued image regions, i.e., central regions of nuclei. Figure 9 shows the precision-recall curves of nucleus detection and classification for different λ values: $\lambda = 0,\ 0.05,\ 0.5,\ 5$ and 10. We also train models with $\lambda = 50$ and $\lambda = 500$, but observe the gradient explosion problem and thus do not report their corresponding results. We note that models with $\lambda > 1$ outperform those with $\lambda < 1$; in particular, the one without any additional penalty on nucleus central regions, i.e., $\lambda = 0$, produces very poor performance for both detection and classification. This is because for one single training image, the zero valued regions dominate its corresponding proximity maps (see Figure 4) and lack of a strong penalty on those non-zero-value regions would lead to a trivial solution. Therefore, it is necessary to emphasize the central regions of nuclei during model training. We set $\lambda = 5$ in all the experiments except this one for parameter analysis.

V. Conclusion

In this paper, we propose a novel FCN architecture for single-stage nucleus recognition, which can be efficiently trained in an end-to-end, pixel-to-pixel manner. Instead of conducting pixel-wise classification, we model nucleus recognition as a structured regression problem, which is able to take advantage of contextual information in the label space. Compared with many previous nucleus/cell recognition approaches, which usually rely on a multi-stage image processing pipeline, the proposed method allows for single-stage, simultaneous nucleus detection and classification. Meanwhile, it does not require individual nucleus segmentation for cellular feature extraction and can significantly improve the efficiency and effectiveness of image quantification, because nucleus segmentation is very challenging, especially for one single image with thousands of nuclei.

We observe that learning with the auxiliary task, i.e., ROI extraction, is helpful for individual nucleus recognition. This ROI extraction task needs only weak region annotation, which is much easier to achieve than individual nucleus labeling. Joint learning with the two tasks, nucleus recognition and ROI extraction, implicitly augments training data such that it can learn more general representations, which reduce the risk of overfitting. We also find that it is necessary to properly balance the contributions of the two tasks to obtain desirable performance. Finally, we demonstrate that learning with an emphasis on the central regions of nuclei is beneficial to nucleus recognition. Our method provides a clean, end-to-end nucleus recognition pipeline for histopathology image analysis, especially a new perspective for Ki67 image quantification, which would potentially benefit individual object quantification in whole-slide images. In particular, it would significantly facilitate the assessment of Ki67 labeling index, thereby enabling early detection and targeted treatments of the NET diseases.

REFERENCES

[1] S. Kothari *et al.*, "Pathology imaging informatics for quantitative analysis of whole-slide images," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 6, pp. 1099–1108, 2013.

[2] J. S. Lewis *et al.*, "A quantitative histomorphometric classifier (quhbic) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma," *Am. J. Surg. Pathol.*, vol. 38, no. 1, pp. 128–137, 2014.

[3] F. Xing *et al.*, "Automatic ki-67 counting using robust cell detection and online dictionary learning," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 859–870, 2014.

[4] M. Reid *et al.*, "Calculation of the ki67 index in pancreatic neuroendocrine tumors: a comparative analysis of four counting methodologies," *Mod. Pathol.*, vol. 28, pp. 686–694, 2015.

[5] M. N. Gurcan *et al.*, "Histopatological image analysis: a review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.

[6] M.-Y. C. Polley *et al.*, "An international study to increase concordance in ki67 scoring," *Mod. Pathol.*, vol. 28, pp. 778–786, 2015.

[7] H. Irshad *et al.*, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review – current status and fture potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.

[8] F. Xing *et al.*, "Deep learning in microscopy image analysis: A survey," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 20, no. 10, pp. 4550–4568, 2018.

[9] C. Sommer and D. W. Gerlich, "Machine learning in cell biology - teaching computers to recognize phenotypes," *J. Cell Sci.*, vol. 126, no. 24, pp. 5529–5539, 2013.

[10] H. Greenspan *et al.*, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[11] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60 – 88, 2017.

[12] Y. LeCun *et al.*, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, November 1998.

[13] E. Shelhamer *et al.*, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[14] A. B. L. Larsen *et al.*, "Hep-2 cell classification using shape index histograms with donut-shaped spatial pooling," *IEEE Trans. Med. Imaging*, vol. 33, no. 7, pp. 1573–1580, 2014.

[15] X. Qi *et al.*, "Hep-2 cell classification via combining multiresolution co-occurrence texture and large region shape information," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 2, pp. 429–440, 2017.

[16] X. Xu *et al.*, "Adaptive co-occurrence differential texton space for hep-2 cells classification," in *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 260–267.

[17] P. Foggia *et al.*, "Benchmarking hep-2 cells classification methods," *IEEE Trans. Med. Imaging*, vol. 32, no. 10, pp. 1878–1889, 2013.

[18] A. Taalimi *et al.*, "Multimodal dictionary learning and joint sparse representation for hep-2 cell classification," in *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 308–315.

[19] N. Theera-Umpon and S. Dhompongsa, "Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 353–359, 2007.

[20] D. Shen *et al.*, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[21] L. Zhang *et al.*, "Deeppap: Deep convolutional networks for cervical cell classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1633–1643, 2017.

[22] Z. Gao *et al.*, "Hep-2 cell image classification with deep convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 2, pp. 416–428, 2017.

[23] J. Liu *et al.*, "Hep-2 cell classification based on a deep autoencoding-classification convolutional neural network," in *IEEE Int. Symp. Biomed. Imaging*, 2017, pp. 1019–1023.

[24] N. Bayramoglu and J. Heikkilä, "Transfer learning for cell nuclei classification in histopathology images," in *Computer Vision – ECCV 2016 Workshops*, 2016, pp. 532–539.

[25] H. T. H. Phan *et al.*, "Transfer learning of a convolutional neural network for hep-2 cell image classification," in *IEEE Int. Symp. Biomed. Imaging*, 2016, pp. 1208–1211.

[26] W. Shao *et al.*, "Deep active learning for nucleus classification in pathology images," in *IEEE Int. Symp. Biomed. Imaging*, 2018, pp. 199–202.

[27] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234–263, 2016.

[28] K. Nguyen *et al.*, "Using contextual information to classify nuclei in histology images," in *IEEE Int. Symp. Biomed. Imaging*, 2015, pp. 995–998.

[29] W. Wang *et al.*, "An optimal transportation approach for nuclear structure-based pathology," *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 621–631, 2011.

[30] Y. Chen *et al.*, "Semi-automatic segmentation and classification of pap smear cells," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 1, pp. 94–108, 2014.

[31] J. Ge *et al.*, "A system for counting fetal and maternal red blood cells," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2823–2829, 2014.

[32] X. Chen *et al.*, "Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 762–766, 2006.

[33] Y. Yuan *et al.*, "Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling," *Sci. Transl. Med.*, vol. 4, no. 157, pp. 157ra143–157ra143, 2012.

[34] J. Kong *et al.*, "A comprehensive framework for classification of nuclei in digital microscopy imaging: An application to diffuse gliomas," in *IEEE Int. Symp. Biomed. Imaging*, 2011, pp. 2128–2131.

[35] D. C. Cireşan *et al.*, "Mitosis detection in breast cancer histology images with deep neural networks," in *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2013, pp. 411–418.

[36] S. Wang *et al.*, "Subtype cell detection with an accelerated deep convolution neural network," in *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2016, pp. 640–648.

[37] K. Sirinukunwattana *et al.*, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[38] M. Saha *et al.*, "An advanced deep learning approach for ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer," *Sci. Rep.*, vol. 7, no. 3213, pp. 1–14, 2017.

[39] O. Z. Kraus *et al.*, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 2016.

[40] Y. Zhou *et al.*, "SFCN-OPI: Detection and fine-grained classification of nuclei using sibling fcn with objectness prior interaction," in *AAAI Conf. Artif. Intell.*, 2018, pp. 2652–2659.

[41] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.

[42] K. He *et al.*, "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

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

[44] D.-A. Clevert *et al.*, "Fast and accurate deep network learning by exponential linear units (elus)," in *Int. Conf. Learn. Repres.*, 2016, pp. 1–14.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. Learn. Repres.*, 2015, pp. 1–14.

[46] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv:1603.07285 [stat.ML]*, pp. 1–31, 2016.

[47] H. Chen *et al.*, "Dcan: Deep contour-aware networks for object instance segmentation from histology images," in *Med. Image Anal.*, vol. 36, 2017, pp. 135–146.

[48] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv:1706.05098 [cs.LG]*, pp. 1–14, 2017.

[49] P. Kainz *et al.*, "You should use regression to detect cells," in *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 276–283.

[50] Y. Xie *et al.*, "Efficient and robust cell detection: A structured regression approach," *Med. Image Anal.*, vol. 44, pp. 245 – 254, 2018.

[51] I. Sutskever *et al.*, "On the importance of initialization and momentum in deep learning," in *Int. Conf. Mach. Learn.*, 2013, pp. III–1139–III–1147.

[52] N. Srivastava *et al.*, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[53] PyTorch, 2017. [Online]. Available: https://github.com/pytorch

[54] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

[55] W. Xie *et al.*, "Microscopy cell counting and detection with fully convolutional regression networks," *Comput. Methods Biomech. Biomed. Eng.: Imaging Visua.*, vol. 6, no. 3, pp. 283–292, 2018.