AUTOMATIC SEGMENTATION OF NUCLEI IN HISTOPATHOLOGY IMAGES USING ENCODING-DECODING CONVOLUTIONAL NEURAL NETWORKS

Deniz Sayin Mercadier[⋆] *Beril Besbinar*[†] *Pascal Frossard*[†]

[⋆] Department of Computer Engineering
Middle East Technical University (METU)
[†] Signal Processing Laboratory (LTS4)
École Polytechnique Fédérale de Lausanne (EPFL)

ABSTRACT

Accurate and fast segmentation of nuclei in histopathological images plays a crucial role in cancer research for detection and grading, as well as personal treatment. Despite the important efforts, current algorithms are still suboptimal in terms of speed, adaptivity and generalizability. Popular Deep Convolutional Neural Networks (DCNNs) have recently been utilized for nuclei segmentation, outperforming *traditional* approaches that exploit color and texture features in combination with shallow classifiers or segmentation algorithms. However, DCNNs need large annotated datasets that require extensive amount of time and expert knowledge. In addition, segmentation results obtained by either traditional or DCNN approaches often require a post-processing step to separate cluttered nuclei. In this paper, we propose a computationally efficient nuclei segmentation framework based on DCNNs exhibiting an encoding-decoding structure. We use a partially-annotated dataset and develop an effective training solution. We also use a weighted background model for network to give more importance to borders of nuclei to overcome the problem of clutters. The abolition of any pre-processing or post-processing step without any compromise on the performance leads to a fast and parameter-free system, which presents important advantages with respect to state-of-the-art.

Index Terms— Nuclei segmentation, digital pathology, histopathology, convolutional neural networks

1. INTRODUCTION

Detection and segmentation of nuclei in widely used histopathology images is a key step for cancer diagnosis, grading and prognosis. As it provides more information about nuclei features, such as size, texture and shape, segmentation, in particular, can be regarded as a key yet challenging task in pathology images, which is either achieved manually [1] or automatically [2]. One main reason in this challenge is the variation of nuclear size and shape depending on the tissue type, the existence and even the severity of a disease. Another difficulty relates to the variations in staining and scanning procedures. Although many normalization techniques have been proposed for the latter [3], an efficient and robust segmentation algorithm that can generalize well over different tissues and staining techniques is still of utmost importance for digital pathology.

The early work in this field addressed the problem of nuclei segmentation with *traditional* approaches, where the algorithms rely on color, texture or shape based features followed by *shallow* classifiers or segmentation algorithms such as graph-cut or watershed [4, 5, 6, 2]. On the other hand, favorable results of deep learning algorithms in computer vision tasks have led to a new class of approaches in biomedical image processing. One well-known example is the U-Net architecture [7] that has been successful in segmentation of electron microscopy images. An early adaptation of DCNNs to nuclei segmentation is proposed in [8], following a strategy to classify every pixel individually using a small patch around it. That work also introduced the first labeled dataset with partially annotated nuclei. Many other works address this problem in a similar fashion, which results in computationally-expensive sliding-window approach in inference time [8, 9, 10, 11, 12]. The work of [12] formulates the problem as a 3-way classification rather than two, third class being the nuclei boundary, which ameliorates the performance in case of cluttered nuclei. Another group of works within the deep learning umbrella employs architectures that make use of so-called deconvolution layers on top of the representations obtained with convolution layers, in order to obtain a high resolution output serving as initial segmentation map [13, 14]. Presenting a comparative study, the concurrent work of [14] found U-Net to be the most convenient architecture for segmentation. Nonetheless, regardless of the architecture utilized, all of the aforementioned works employ a sophisticated and computationally expensive post-processing

The first author performed the work while at LTS4, EPFL

This research was supported in part by Swiss Data Science Center through iLearn project

step to split cluttered or overlapping nuclei, which have been an established problem due to its challenging nature.

In this paper, we propose a framework that aims at solving the above limitations by providing means to effectively detect nuclei in adverse settings. In particular, we present a new algorithm that attains background modelling with the help of unsupervised algorithms and gives higher importance to pixels at nuclei borders to effectively handle clutters. We believe that the proposed training scheme that exploits unsupervised clustering techniques can be employed in many different applications in biomedical imaging with regard to unavailability of fully annotated datasets. Furthermore, the encoding-decoding structure for segmentation leads to competent inference time, which is essential in large scale processing as the nuclei segmentation mostly serves as a fundamental step for further examination. Our comparative experimental results indeed confirm that the proposed framework exploit the limited labels very well without a compromise in inference time or performance.

2. METHOD

Known with their preferable performance, DCNNs require labeled data for problems like classification, detection and segmentation. On the contrary, manual annotation of data for segmentation is an exhausting task and need expert knowledge, especially for biomedical data [1]. Partial annotations can be considered as an opportunity to make DCNNs more widely applicable to problems in medical images provided that efficient algorithms to make use of partial annotations exist. Although it is more straightforward to use partial annotations for patch-classification based algorithms, the ambiguity in nonanotated pixels, either being foreground or background in a two-class segmentation problem, makes the problem inherently challenging. Our method exploits the unsupervised clustering algorithms to provide a solution to this problem and uses the dataset introduced in [8] as a showcase. In this section, after providing the architecture details, we explain our background modelling and present the new loss function.

2.1. Deep Network Architecture

In this work, we employ a modified version of U-Net architecture [7], which is composed of five encoding and five decoding layers. We replace *valid* convolutions with zero-padded counterparts to provide more flexibility in terms of input image size. In addition, a batch normalization layer is added after each non-linearity operation to account for the variations in color distributions of images as a result of different staining and scanning procedures. Employing batch normalization allows us to eliminate the need for pre-processing steps, such as histogram normalization or matching. The empirical evaluation of different architectures has led to the current design of the architecture, which is illustrated in Figure 1.

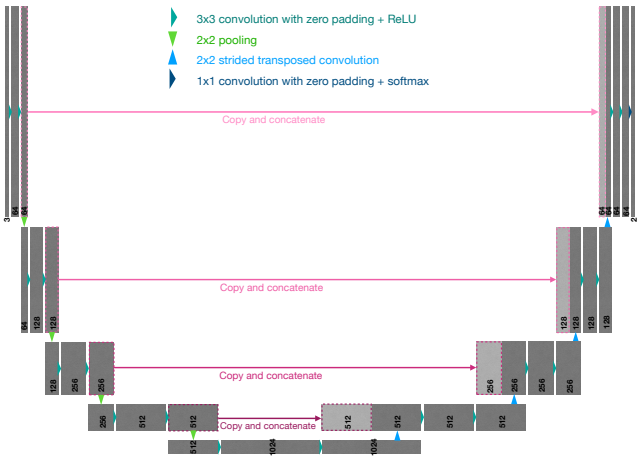


Fig. 1: The proposed architecture where each grey box represents a multi-channel feature map, the number written at the bottom of each box indicating the channel number. The light grey boxes represent the copied feature maps from the corresponding encoding branch.

2.2. Background Modelling

The dataset in use is partially annotated, which translates to the fact that only *some* of the nuclei are labeled for the groundtruth mask. Such labeling does not serve well for segmentation architectures as the pixels with label '0' do not always represent background. However, as a first step, the very close neighborhood of annotated nuclei can be regarded as background. In addition, the prior knowledge of a dataset, such as the fact that most of the cytoplasm is known to be eosinophilic in histopathology images, can improve the modelling. Hence, we perform a color deconvolution to obtain hematoxylin and eosin channels. The background model is then reinforced by applying k-means algorithm to the deconvolved image with k=2, where the cluster with larger area is supposed to be background with less certainty.

2.3. Loss Function

The widely-used cross-entropy loss penalizes the output of a model as it diverges from the true label; however, the ambiguity in true background label in partial annotation case provoked the modification of the cross-entropy loss. Let $y_{ij} \in 0, 1$ represent the groundtruth value of a pixel at location (i, j), where 1 symbolizes the nuclei and 0 indicates non-annotated nuclei or non-nuclei regions. Let $\hat{y}_{ij}^{\text{fg}}, \hat{y}_{ij}^{\text{bg}} \in [0, 1]$ represent the output of the network layer for foreground and background masks, respectively. In order to compensate for nuclei that are not annotated, we present a new masking term $m_{ij} \in [0, 1]$ as given in Eq.(1), where W, H indicate the height and width of an image. As the term $(1 - y_{ij})$ corresponds to groundtruth pixels with value 0, m_{ij} can be

interpreted as weighting term to model background. Not only does it let us avoid pixels with high uncertainty for training by equating its value to zero, but it also enables us to give more importance to some specific pixels, such as the close neighborhood of annotated nuclei.

$$\mathcal{L} = \sum_{(i,j)\in WxH} y_{ij} \log \hat{y}_{ij}^{\text{fg}} + m_{ij}(1 - y_{ij}) \log \hat{y}_{ij}^{\text{bg}} \quad (1)$$

To obtain m_{ij}, we first take the distance transform of the groundtruth mask, d_{ij}, using Euclidean distance. We then apply a soft thresholding mechanism, as given by Eq.(3).

$$d_{ij} = \min_{\substack{(i,j),(u,v)\in WxH \\ \{(i,j) \mid y_{ij}=1\}}} (\sqrt{(i-u)^2 + (j-v)^2}) \quad (2)$$

$$m_{ij} = \begin{cases} 1, & \text{if} \quad d_{ij} \leq \alpha \\ e^{-\beta(d_{ij}-\alpha)}, & \text{if} \quad d_{ij} > \alpha \end{cases} \quad (3)$$

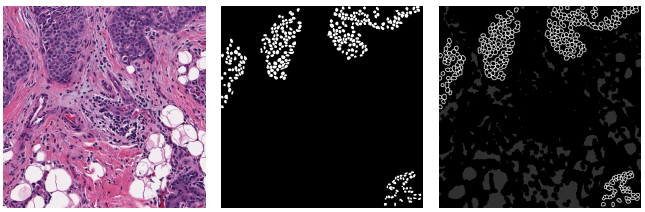
However, when m_{ij} depends only on the close neighborhood of nuclei boundary, it does not contain any information about stroma, cytoplasm or lumens. Thus, we modify the mask by employing the background cluster obtained with k-means to provide more information about the background. The modified weight mask is denoted as $\hat{m}_{ij}$ and given in Eq.(4), where $\mathbf{b}$ represents the background cluster, $\ominus$ denotes the erosion operation by a circular structuring element c and λ is the weighing coefficient. Due to higher uncertainty, the background cluster is given less importance than the nuclei boundary that is represented by m_{ij}, via a λ value less than 1.

$$\hat{m}_{ij} = m_{ij} \uplus \lambda(\mathbf{b} \ominus c) \quad (4)$$

As an example, Figure 2 illustrates an image from the training set, as well as the partial annotations and the proposed background mask with $\lambda = 0.5$.

As the final modification, we employ L2 regularization on the weights w of convolutional layers in order to avoid overfitting. Eq.(5) summarizes the resulting loss function that we minimize to train our neural network.

$$\mathcal{L} = \sum_{(i,j)\in WxH} y_{ij} \log \hat{y}_{ij}^{\text{fg}} + \hat{m}_{ij}(1 - y_{ij}) \log \hat{y}_{ij}^{\text{bg}} + \gamma \|w\|_2^2 \quad (5)$$



(a) Image (b) Annotation (c) Weighted Background Mask

Fig. 2: Use of partially annotated dataset for segmentation

3. EXPERIMENTS

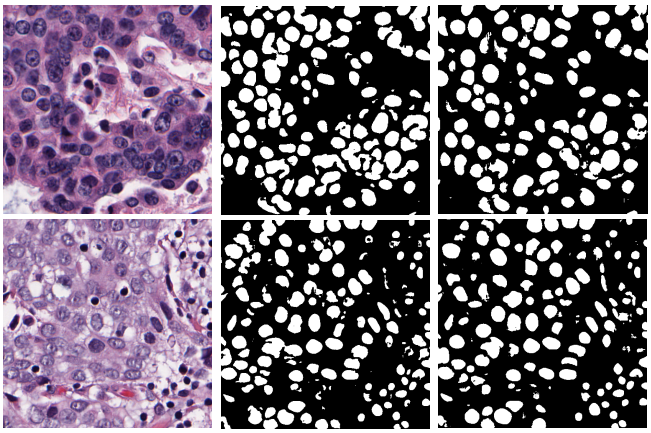
In this section, after providing the implementation details, we first qualitatively demonstrate the performance of the trained architecture on the test set. Then, we present the comparative analysis of our method and the method in [8] on two other nuclei segmentation datasets, introduced in [14] and [12], both of which are fully annotated. Neither of the algorithms have used any examples from these two datasets for training.

3.1. Implementation Details

The dataset introduced in [8] is composed of 141 hematoxylin and eosin (H&E) stained images of size 2000x2000 at 40x magnification. We use a split of 121/5/15 for training, validation and testing. Smaller patches of size 256x256 are extracted from the training set, where the morphologically dilated foreground and background masks have more nonzero pixels than a certain threshold, 10% and 40% of total number of pixels, respectively. Although the data augmentation is not applicable to every problem, biomedical image analysis benefit from the algorithms that are equivariant to rotations and translations. Thus, to avoid overfitting and improve generelizability, we use data augmentation by 90 degree rotations, flipping and elastic deformations. In the end, 71842 patches are extracted for training by sliding a window of size 256x256 over the training images with a stride of 128.

Training is performed on Tensorflow by using Adam optimizer for 100K steps of update, with a batch size of 6, which took less than 15 hours using an old generation of NVIDIA Titan X GPU. The learning rate is initialized as 10^{-3} and multiplied by a factor of 0.95 once every 1,000 iterations with a lower limit of 10^{-5}. The coefficient for weight regularization is chosen to be $\gamma = 10^{-5}$.

For inference, once the input image of an arbitrary size is provided, patches of size 256x256 are extracted by a slid-



(a) Image (b) [8] (c) Proposed method

Fig. 3: Samples from the test set and output masks generated by [8] and the proposed method

Table 1: Performance of the proposed algorithm in comparison to [8] for datasets presented in [13] (DS1) and [12] (DS2)

			Recall	Precision	F1 score	Accuracy	JI (IoU)	AJI	Time (sec)
DS 1	pp	[8]	0.35 (0.21)	**0.91 (0.06)**	0.45 (0.21)	0.92 (0.02)	0.32 (0.18)	0.84 (0.04)	467.88 (13.67)
		Proposed	**0.60 (0.14)**	0.89 (0.04)	**0.70 (0.10)**	**0.94 (0.02)**	**0.55 (0.11)**	**0.88 (0.03)**	**0.47 (0.01)**
	all	[8]	0.33 (0.23)	**0.92 (0.08)**	0.44 (0.23)	0.92 (0.06)	0.31 (0.20)	0.85 (0.10)	467.88 (13.67)
		Proposed	**0.60 (0.17)**	0.90 (0.06)	**0.70 (0.13)**	**0.94 (0.04)**	**0.55 (0.14)**	**0.89 (0.08)**	**0.47 (0.01)**
DS 2		[8]	0.59 (0.23)	0.81 (0.15)	0.63 (0.18)	0.85 (0.08)	0.49 (0.17)	0.70 (0.14)	1642.99 (39.91)
		Proposed	**0.73 (0.16)**	**0.82 (0.09)**	**0.76 (0.12)**	**0.89 (0.06)**	**0.62 (0.13)**	**0.78 (0.10)**	**1.84 (0.06)**

ing window with a stride of 128. Batching all the extracted patches, we efficiently obtain foreground and background probability maps by forward propagation. The binary segmentation map is then generated via maximum likelihood approach for each patch. In order to obtain the segmentation mask for the whole image, we merge the masks of these overlapping patches by considering the central part of size 128x128 to be valid, in order to avoid the artifacts due to convolutions with zero padding.

3.2. Qualitative Analysis

Figure 3 demonstrates some patches extracted from the test set, accompanied by the outcome of our network, and of the network in [8]. Although both algorithms perform comparably good, generation of the output mask takes 10 seconds on average using our framework for an image of size 2,000x2,000, whereas the method in [8] needs an average of 3.5 hours with the Python-Caffe toolkit provided with [8]. Both algorithms are tested on a single NVIDIA Titan X GPU.

Furthermore, the performance of a method on different datasets resembling different acquisition techniques is of vital importance as it is a very realistic scenario. Figure 4 illustrates some patches from datasets in [14] and [12], accompanied by the outcomes of our method. It is important to emphasize that we do not employ any pre-processing or post-processing steps while testing our method on different datasets. Hence, uncluttered segmentation results that can generalize well over different datasets are a clear benefit of our architecture and loss design.

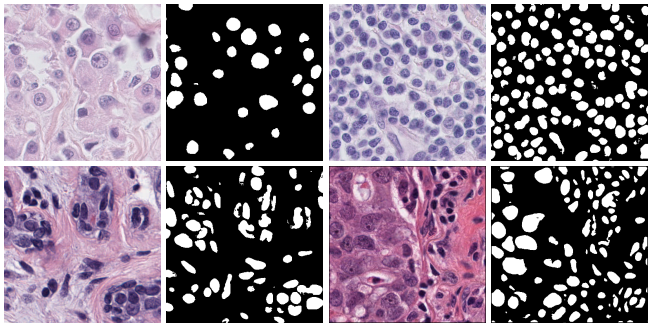


Fig. 4: Sample outputs from datasets in [14] (first row) and [12] (second row)

3.3. Quantitative Analysis

For a quantitative analysis, we test both algorithms on fully annotated datasets of [14] (**DS1**) and [12] (**DS2**). DS1 corresponds to 50 H&E images of size 500x500, taken from 11 different patients. On the other hand, DS2 is composed of 30 H&E images of size 1,000x1,000 with groundtruth that are obtained from 7 different organs. As proposed in [13], which uses a preliminary version of DS1 for training and testing, evaluation metrics are averaged over patients (per-patient: **pp**), yet, we also compute the average values over the whole dataset (**all**). For generating the output masks, we apply the exact same procedure described in Section 3.1 without any further training or fine tuning the model.

For evaluation, we use two groups of metrics to establish the performance of algorithms both regarding per-pixel classification success and object detection outcome. As per pixel metrics, we use recall, precision and accuracy values as well as F1 score and Jaccard Index (**JI**), also known as Intersection over Union (**IoU**). On the other hand, we use aggregated Jaccard index (**AJI**), which is proposed in [12], as per-object metric. Table 1 summarizes the performance evaluation and articulates the consistently superior performance of the proposed method with an emphasis on the computation time, which is orders of magnitude smaller compared to the method in [8].

Both qualitative and quantitative analyses reveal the efficacy of the proposed method regarding speed and accuracy.

4. CONCLUSION

In this paper, we presented an approach that employs DCNNs for automated nuclei segmentation in histopathology images. Our loss function enables the deconvolutional segmentation architecture to benefit from sparsely annotated datasets for efficient training and shorter inference time. The masking term not only exploits the unsupervised clustering algorithms to alleviate partial annotations, but also lets us give higher importance to boundary pixels to overcome the clutter problem without a costly post-processing step. Our superior performance in different datasets points out the ability of the proposed framework to generalize well, due to the proper normalization layers introduced in the architecture.

https://github.com/denizmsayin/nuclei-segmentation

5. REFERENCES

[1] Humayun Irshad, Laleh Montaser-Kouhsari, Gail Waltz, Octavian Bucur, JA Nowak, Fei Dong, Nicholas W Knoblauch, and Andrew H Beck, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd," in *Pacific Symposium on Biocomputing Co-Chairs*. World Scientific, 2014, pp. 294–305.

[2] Humayun Irshad, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE reviews in biomedical engineering*, vol. 7, pp. 97–114, 2014.

[3] Derek Magee, Darren Treanor, Doreen Crellin, Mike Shires, Katherine Smith, Kevin Mohee, and Philip Quirke, "Colour normalisation in digital histopathology images," in *Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*. Daniel Elson, 2009, vol. 100.

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

[5] Mitko Veta, Paul J Van Diest, Robert Kornegoor, André Huisman, Max A Viergever, and Josien PW Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PloS one*, vol. 8, no. 7, pp. e70221, 2013.

[6] Pin Wang, Xianling Hu, Yongming Li, Qianqian Liu, and Xinjian Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Processing*, vol. 122, pp. 1–13, 2016.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[8] Andrew Janowczyk and Anant Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of pathology informatics*, vol. 7, 2016.

[9] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A Cree, and Nasir M Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[10] Fuyong Xing, Yuanpu Xie, and Lin Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 550–566, 2016.

[11] David A Van Valen, Takamasa Kudo, Keara M Lane, Derek N Macklin, Nicolas T Quach, Mialy M DeFelice, Inbal Maayan, Yu Tanouchi, Euan A Ashley, and Markus W Covert, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS computational biology*, vol. 12, no. 11, pp. e1005177, 2016.

[12] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[13] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 933–936.

[14] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Transactions on Medical Imaging*, 2018.