Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map

Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter

Abstract—The advent of digital pathology provides us with the challenging opportunity to automatically analyze whole slides of diseased tissue in order to derive quantitative profiles that can be used for diagnosis and prognosis tasks. In particular, for the development of interpretable models, the detection and segmentation of cell nuclei is of the utmost importance. In this paper, we describe a new method to automatically segment nuclei from Haematoxylin and Eosin (H&E) stained histopathology data with fully convolutional networks. In particular, we address the problem of segmenting touching nuclei by formulating the segmentation problem as a regression task of the distance map. We demonstrate superior performance of this approach as compared to other approaches using Convolutional Neural Networks.

Index Terms—Cancer research, deep learning, digital pathology, histopathology, nuclei segmentation.

I. Introduction

THE advent of next generation sequencing methods has revolutionized cancer research by providing unprecedented insights into the molecular basis of cancer [1] and by establishing big data approaches as a main strategy of this domain. Yet, while massive sequencing approaches and the analysis of genomic, epigenomic and transcriptomic features build the main body of current bioinformatics programs related to cancer, diagnosis and prognosis in clinical practice are not solely based on sequence and gene expression data. Indeed, histopathology data is routinely used in clinical practice for cancer diagnosis and prognosis. These data correspond to tissue slides encompassing the tumor and the surrounding tissue, stained with agents in order to highlight specific structures, such as cells, cell nuclei or collagen. A pathologist will routinely check a patient's histopathology image data in order to determine the next step in the patient's treatment. Tissue slides are informative on many aspects of the disease, such as cancer subtype, the grade or the reaction of the patient's immune system. They give access to morphological properties of cells and tissues and their spatial organization and are thus complementary to information revealed by *omics* measurements. However, the interpretation of such data is not straightforward and requires special training and a deep understanding of the underlying physiology. In addition, while some criteria have proven to be useful for the interpretation of stained tissue data and correlate well with clinically relevant variables, it is highly unlikely that all relevant and exploitable aspects contained in images are known as of today.

Digital pathology has risen in popularity in the last decade thanks to improvements in terms of hardware and software. Today, Whole Slide Images (WSI) can be acquired using highly automated scanners, and stored and accessed using dedicated software. The advent of digital pathology does not only simplify storage and sharing of stained tumor slides, it also paves the way for the automatic analysis of very large cohorts with image analysis and machine learning methods in order to quantitatively model these challenging and rich data and to provide tools for Computer Aided Diagnosis (CAD).

However, it is still an open question which information is to be extracted from tissue slides in order to provide useful and predictive quantitative profiles. There are two main strategies, one consisting in learning general image features with the best discriminative power with respect to some clinical variable, the other in detecting the important elements in a histopathology image (such as particular tissue types or cellular phenotypes) and quantitatively describing the abundance and spatial organization of these elements, leading to a physiologically interpretable quantitative profile. In the context of this second kind of approach, the most important biological elements to be detected in tissue slides are nuclei, because they are indicative of many cellular phenotypes [2]. Their morphology is currently used by pathologists in order to identify the mitotic index and the level of nuclear pleomorphism [3] and the standard staining procedure makes them distinguishable.

In this paper, we propose a new method for the segmentation of nuclei in histopathology data based on a CNN, which is an essential step for cellular and tissular phenotyping. In particular, we develop in this paper a new idea to segment touching or overlapping nuclei by formulating the problem as a regression task. Segmenting touching objects is a recurrent problem in the field of biology, and in particular for the segmentation of nuclei across many imaging modalities.

The paper is organized as follows: in section II we describe related work. In section III, we present the data sets that will be used in this study. We explain the core of our method in section IV. Sections V and VI are devoted to experiments on simulated and real data, respectively. In section VII, we discuss the results and impact of this study and conclude in section VIII.

II. Related Work

In this section we give an overview of related work. After a short presentation of the major challenges in the field of digital pathology in section II-A, we review recently developed approaches for the detection of nuclei in II-B. Given the increasing importance of annotated data sets in this field, we give a short description of different available data sets in section II-C.

A. Challenges in Digital Pathology

Analysis of histopathology data has interested the image analysis and computer vision communities for many years. The first works go back to the late eighties [4]. More recently, the advent of digital pathology and the availability of large amounts of digitized scans have triggered a deluge of new methodological developments in this field, mostly targeting systems for CAD. Many of these approaches are developed for tissue sections stained with H&E.

Popular topics in this field include color normalization [5]–[7], detection of particular cellular states, such as mitoses [8], [9] or the classification of tissue regions, such as glands [10] and necrotic or metastatic regions [11], [12] and prediction of tumor type and grade [13]–[15]. Such predictions can be made directly from the raw images, but they can also be based on the detection of nuclei [13], [15]–[18] allowing for a biological interpretation of the predictive features.

B. Nuclei Segmentation

Due to its essential role in the automatic interpretation of stained tissue sections, segmentation of nuclei has been addressed by many authors with a variety of traditional approaches, including mathematical morphology [13], [19]–[21], pixel classification [22], level sets [23] and graph-based segmentation methods [24] (see also [25], [26] for extensive reviews).

Today, neural networks are considered to be among the most powerful methods in the field of computer vision. Initially, deep neural networks were developed in the context of large-scale image classification [27], [28], they have rapidly been adapted to object detection and image segmentation. Indeed, the most straightforward way of applying classification networks to detection and segmentation tasks is a sliding window approach, where the center pixel of the window is classified into 2 (object / background) or more classes. This approach has been successfully applied to the detection [8], [17], [29], [30] and segmentation [31], [32] of nuclei in H&E images and other imaging modalities [33], [34]; see [35] for a comprehensive review. As the corresponding image patch needs to be classified for every pixel, the number of operations to perform is relatively high. Instead of using architectures optimized for image classification and applying them for each individual pixel, fully convolutional neural networks [36]–[38] directly predict binary maps from images. These architectures are thus optimized for segmentation in terms of both accuracy and speed. They have also been applied to biomedical data for various segmentation tasks, such as cell segmentation in microscopy images [38].

While these networks are certainly the most powerful methods for segmentation today, they face one major issue when applied to biological data: biological samples are made of individual components that are often in direct physical contact to each other, such as cells in culture or in tissue. In traditional image analysis, this problem can be addressed by imposing shape priors implicitly by the design of dedicated algorithms, such as deformable models [26], [31]. In principle, CNNs are supposed to overcome this kind of handcrafted design by finding the optimal rules themselves by generalizing from the set of annotated samples.

In practice however, standard CNNs without any kind of postprocessing are often not capable of correctly handling the issue of touching objects, because the loss function minimized by CNNs is usually defined at the pixel level. In order to address this problem, one would need to turn to instance segmentation models. For this, there are several strategies, such as giving larger weights to pixels in close proximity of the object contours [38], predicting not only pixel labels, but also object properties [30], predicting both the objects and their contours by the CNNs [32], [34] or learning a notion of the object by combining object region prediction with pixel-level segmentation [39].

C. Annotated Nuclei Datasets

Annotated histopathology data at the pixel level are scarce. Indeed, precisely annotating object boundaries in this type of data is time consuming and error-prone. For instance [40] compare results obtained by different groups of annotators, including crowd annotators, for both nuclei detection and segmentation, obtaining optimal F1 scores of 0.94 for detection and 0.66 for segmentation of nuclei.

[35] and [41] have both released large annotated data sets for nuclei segmentation in digital pathology. [35] have annotated 142 breast cancer images containing around 12 000 nuclei, [41] have annotated 58 images (the number of nuclei is not indicated). While these datasets are certainly of high value for the scientific community, annotation is not
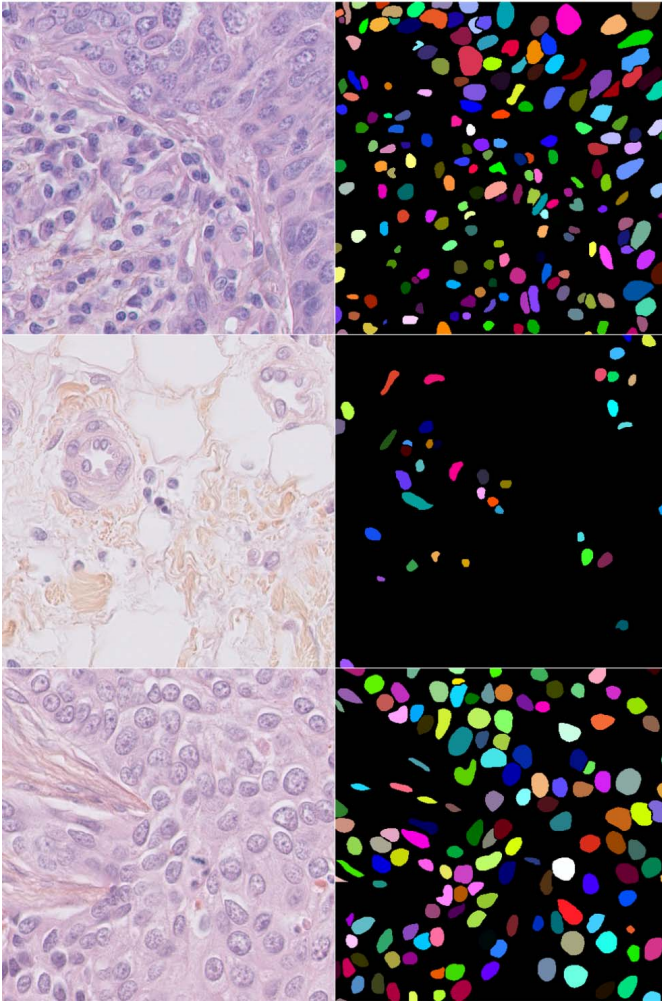
Fig. 1. Data set 1. Left column: 3 samples from 2 patients. Right column: associated ground truth.

exhaustive, i.e. only a subset of nuclei has been annotated. This leads to several problems when training CNNs on these data: first, it is challenging to find representative negative examples; second, it is difficult to focus on touching objects, which is the main challenge in nuclei segmentation; and third, it precludes fully convolutional neural networks for segmentation to be used, as this family of networks requires exhaustive annotation.

The data set published in [32] contains exhaustive segmentation of 30 images from different organs, containing more than 21 000 nuclei. At the same time, we have published a data set of 33 exhaustively annotated breast cancer images with 2 700 annotated nuclei [42]. To the best of our knowledge, these two data sets as of today represent the only H&E stained image sets with exhaustive annotation of nuclei.

III. DATA SETS

We have considerably extended dataset [42] and also make it publicly available. In this paper, we also use the annotated dataset provided by [32]. The other data sets listed in II-C were not exhaustively annotated and did therefore not meet our criteria to be used in this study.

A. Data Set 1 - DS1

We have generated a dataset at the Curie Institute consisting of annotated H&E stained histology images at 40× magnification. All slides are taken from a cohort of Triple Negative Breast Cancer (TNBC) patients and were scanned with Philips Ultra Fast Scanner 1.6RA.

For eleven patients, we extracted three to eight 512×512 patches from different areas of the tissue. The patches were selected in order to represent the heterogeneity of the image data including low cellularity regions which can be stromal areas or adipose tissue, but also high cellularity areas consisting of invasive breast carcinoma cells. This data set thus represents both intra- and inter-patient variability for the same cancer type.

We have annotated a large number of cells, including normal epithelial and myoepithelial breast cells (localized in ducts and lobules), invasive carcinomatous cells, fibroblasts, endothelial cells, adipocytes, macrophages and inflammatory cells (lymphocytes and plasmocytes). In total, our data set consists of 50 images with a total of 4022 annotated cells, the maximum number of cells in one sample is 293 and the minimum number of cells in one sample is 5, with an average of 80 cells per sample and a high standard deviation of 58. The annotation was performed by three experts: an expert pathologist and two trained research fellows. Each sample was annotated by one of the annotators, checked by another one and in case of disagreement, a consensus was established by discussion among the 3 experts. The annotations were performed on ITK-SNAP [43] where samples could be enlarged so that each image pixel occupied several pixels on the monitor. This allowed for a precise detection of the nuclei boundaries.

B. Data Set 2 - DS2

The second dataset, made available by the Indian Institute of Technology Guwahati consists of annotated H&E stained histology images captured at 40× magnification. The patients were selected from the TCGA website and originate in several hospitals [32] and represent different cancer types. For each WSI, one 1000×1000 patch was extracted. In total, there were 30 images from 7 different organs with a total of 21 623 annotated nuclei. The variability observed in this data set, as illustrated in figure 2, reflects the heterogeneity between organs, cancer types, different regions inside the organs and the technical variability between different pathology departments.

We used the data published in [32] for benchmarking against their method, where we made sure to use exactly the same subset of images their results were based on.

IV. NUCLEI SEGMENTATION FRAMEWORK

The main difficulties in the segmentation of nuclei in H&E stained images are (1) the variability of their appearance depending on cell type, phenotype and tissue and (2) the close proximity or even overlap of nuclei, making their recognition as individual objects difficult. While traditional
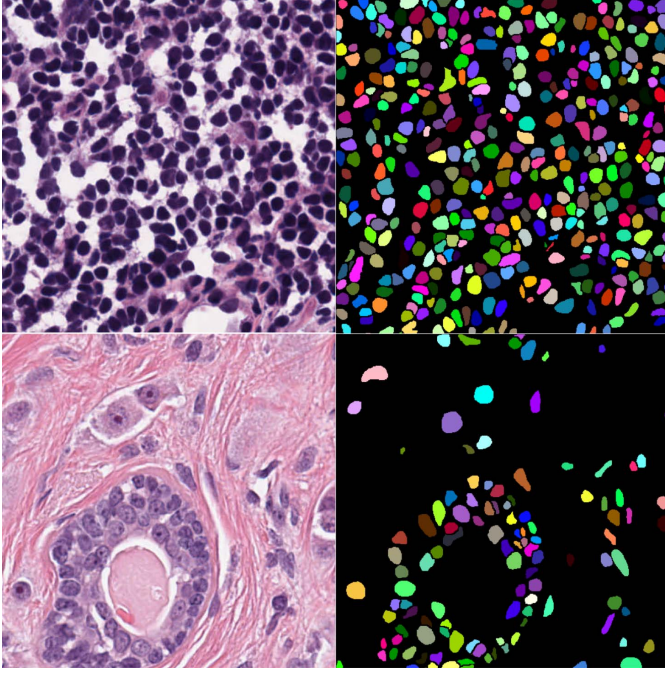
Fig. 2. Data set 2. Left column: samples from stomach (top) and breast (bottom). Right column: associated ground truth.

methods applied to this problem may work reasonably well on a relatively homogeneous subset of well contrasted cells (such as lymphocytes), it is very challenging to design a single algorithm to segment nuclei with varying appearance with satisfactory accuracy. On the other hand, they are distinguishable by human experts. This suggests the use of supervised learning approaches, and in particular the use of neural networks for segmentation.

For the next two sections, let $\mathcal{A}$ be the space of RGB images $\mathcal{A} = \mathbb{R}^{n \times p \times 3}$. Let $\mathcal{B}$ be the space of annotation images. Here, $\mathcal{B}$ can be $\{0,1\}^{n \times p}$, $[0;1]^{n \times p}$ or $\mathbb{R}^{n \times p}$, depending on the formulation of the segmentation problem. Let (A, B) be a tuple of an image $A \in \mathcal{A}$ and its annotation $B \in \mathcal{B}$. Our goal is to find a prediction function f that allows us to predict an annotation image B from an unseen image A. The prediction function f is found by minimizing a designated loss, which can be a standard classification loss or a regression loss. Finally, we will denote by $(A_l, B_l)_{l \in [1,N]}$ the annotated data set. We denote by N the size of the dataset.

A. Deep Neural Network

In our previous work, [42], we showed that fully convolutional networks were highly efficient for segmenting nuclei in histopathology images. In particular, rather than classifying individual pixels in a sliding window approach by classical CNNs, the fully convolutional neural networks allows to upscale the classical CNN feature representation and thereby to segment images in a single pass [36].

We therefore decided to use the U-Net [38] and the Fully Convolutional Network for segmentation (FCN)[1] [36] for

[1] In order to avoid any ambiguity, we will designate by FCN the fully convolutional network as defined in the original paper.

comparison. We denote by f_k a Deep Neural Network (DNN) that maps its final output to $\mathbb{R}^{n \times p \times k}$. This can be easily set up with any fully convolutional network by modifying the final 1×1 convolutional layer so that it maps each pixel to k channels. We define the global loss as

$$\mathcal{L} = \frac{1}{N} \sum_{l=1}^{N} loss(B_l, f_k(A_l)) + \lambda ||w||_2^2 \qquad (1)$$

where w are the model parameters and λ is a weight decay hyperparameter. We optimize the model parameters with Adam optimization [44]. Another crucial hyperparameter is the learning rate l_r, which sets the speed at which we update the model parameters. Finally, we introduce a scale parameter n_{feat} that allows us to scale the complexity of the DNN: n_{feat} is the number of filters of the first convolutional layer. The number of filters for all other layers are derived from n_{feat} as described in [38]. This parameter is only used for the U-Net, as the other architectures use pre-trained networks and therefore have a fixed architecture.

B. Classification Loss

In the case of the standard classification loss for a binary problem, we set k to 2 so that the network outputs two channels (object and background). Hence, the global loss $\mathcal{L}$ for these networks is

$$\mathcal{L} = \frac{1}{N} \sum_{l=1}^{N} loss(B_l, f_2(A_l)) + \lambda ||w||_2^2$$

$$loss(B_l, f_2(A_l)) = \frac{1}{np} \sum_{i,j} \sum_k t_{i,j,k} \log(\widehat{p_{i,j,k}}) \qquad (2)$$

where k designates a certain label. $t_{i,j,k}$ is equal to 1 if pixel $[i,j]$ is of class k, i.e. $t_{i,j,k} = 1 \Leftrightarrow B_l[i,j] = k$. $\widehat{p_{i,j,k}}$ designates the estimated probability of pixel $[i,j]$ of being k via the softmax output of the neural network, i.e. $\widehat{p_{i,j,k}} = \frac{\exp(-f_2(A_l)[i,j,k])}{\sum_c \exp(-f_2(A_l)[i,j,c])}$. In other words, the loss is defined as the cross-entropy of the logistic function of these two channels. We expect each channel to represent an optimal representation for object and background, respectively. The logistic function of these two channels will result in a probability map where the value of each pixel corresponds to the probability of belonging to the object class.

C. Regression Loss

1) Motivation: The main problem in segmentation of nuclei is that close or overlapping nuclei tend to be segmented as one object. The problem is particularly striking for learning based approaches, where some pixel-based loss is usually minimized. This is easy to understand, as the contribution of a single wrongly classified pixel to the global loss is minor, yet it systematically leads to joint segmentation of individual objects. For this reason, several authors propose to predict both the objects and their contours [32], [34], thereby finding an efficient internal representation of the object border.
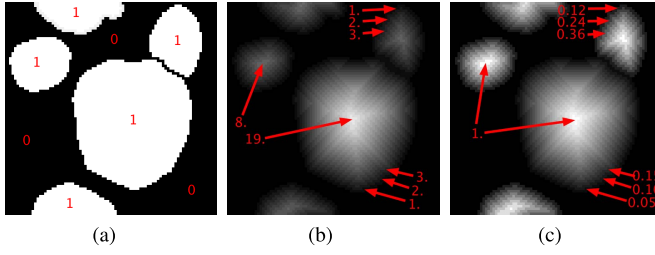
Fig. 3. Different versions of the distance map. Images have been rescaled to [0,255] for visualization. In red: gray level values at the specified locations. (a) $B \in \mathcal{B}$. (b) $B_D \in \mathcal{B}_D$. (c) $B_{D,N} \in \mathcal{B}_{D,N}$.

We felt that this formulation was not necessarily well adapted to the nature of histopathology data, as nuclei boundaries are often fuzzy and ill defined due to the thickness of the slice. With this limitation in mind, we wanted to invert the attention, i.e focus the attention of the model on the core of the nuclei by predicting an eroded version of the annotation. The extreme case of this reasoning leads to the detection of nuclear centers, as proposed in [17] and [30], as centers correspond to the ultimate erosion of the ground truth. We then argued that instead of predicting one eroded ground truth or the nuclear centers, we might as well predict different ground truth versions resulting from successive erosions. This then led to the rationale of predicting directly the distance transform of the nuclei images instead of predicting several binary maps: unlike [17], the segmentation task is thus formulated as a regression problem. We also note that this formulation is ideally suited for the post-processing step we discuss in section IV-D.

2) Framework: We define the space of distance maps as $\mathcal{B}_D = \{B_D; B_D = Dist(B), B \in \mathcal{B}\}$ where $Dist$ is some distance transform that assigns to each pixel $x = [i, j]$ with $B(x) > 0$ the distance to the closest background pixel, i.e. $Dist : \mathcal{B} \rightarrow \mathcal{B}_{\mathcal{D}}, Dist(x) = \min_{y, B(y)=0} \{d(x, y)\}$. Here, we use the chessboard distance. We denote by $\widehat{B_D}$ the prediction of the distance map from the input image A, i.e $\widehat{B_D} = f_k(A)$. The loss is defined as:

$$loss(B_{D,l}, \widehat{B_{D,l}}) = \frac{1}{np} \sum_{i,j} (B_{D,l}[i, j] - \widehat{B_{D,l}}[i, j])^2 \quad (3)$$

There are several design choices to be made for this regression task. First, the maximal value of $Dist(B)$ for one connected component is a proxy for the radius of the component (or more precisely the radius of the maximal ball inside the component). Consequently, the distance map will feature larger values for larger nuclei. It is not evident a priori how the neural network will cope with different value spectra corresponding to differently sized nuclei. In order to correct for these inter-nuclei differences, we can thus normalize the distance map, such that for each individual connected component its values are in the range [0; 1]. This can be achieved by simply dividing the distance map by its maximal value for each connected component. We denote this space $\mathcal{B}_{D,N}$, see figure 3.

Given this normalized distance map $B_{D,N}$, where all values are in the range $[0, 1]$, we further argued that it might

be advantageous to use an architecture whose output is in the same range. In addition, we knew that in principle the classification networks that learn two outputs, one for the object and the other for the background tend to provide very accurate results at the pixel-level. We therefore hypothesized that f_2 with a softmax layer would be a suitable architecture to perform regression of the normalized distance map.

To summarize, we tested the following settings:

- to predict the raw distance map $B_D \in \mathcal{B}_D$ by f_1.
- to predict the normalized distance map $B_{D,N} \in \mathcal{B}_{D,N}$ by f_1.
- to predict the normalized distance map in $B_{D,N} \in \mathcal{B}_{D,N}$ by $softmax \circ f_2$.

D. Post-Processing

The DNN provides us with an image in which the pixel values are indicative of the presence of nuclei: for the distance map, the pixel value is expected to be an estimation of the distance to the closest background pixel, for the traditional classification DNN it is the posterior probability of belonging to a nucleus. Object pixels are identified by a simple thresholding operation. Furthermore, we hypothesize that in either case, nuclei centers correspond to local maxima of the image and that between nuclei there is a significant drop in the signal. This drop in the signal is defined by morphological dynamics [45]: let M be a local maximum of the DNN output y. We say that M represents a nucleus if along all paths $\mathcal{P}$ connecting M with some higher maximum M', the decrease in y is at least p_1:

$$\min_{\substack{\mathcal{P}=(M,...,M'), \\ y(M')>y(M)}} \left\{ \max_{x \in \mathcal{P}} [y(M) - y(x)] \right\} > p_1 \quad (4)$$

where p_1 is a parameter that can be freely chosen. The nuclear regions are then found by the watershed transformation seeded from the maxima that fulfill this criterion. In addition, we require that all object pixels have a higher valuer than p_2 (which is the usual threshold operation).

In [42], we applied this strategy to the posterior probability map. While the strategy gave satisfying results in many cases, we also observed that — by construction — the method did not perform well in cases where the DNN was uncertain leading to bumpy probability maps and consequently to fragmented segmentation results. In addition, this method did not provide any mechanism to include prior information on the nuclear shape. Here, we try to overcome these limitations by applying the same strategy to the predicted distance map.

V. DESIGN CHOICES FOR THE DISTANCE MAP PREDICTION

To validate the regression model, we perform a simple experiment where we want to investigate which setup for the prediction of the distance map is most promising to be applied in this context. We compare the three formulations of the regression problem introduced in section IV-C: regression with f_1 on $\mathcal{B}_D$, regression with f_1 on $\mathcal{B}_{D,N}$ and regression with $softmax \circ f_2$ on $\mathcal{B}_{D,N}$. To compare these methods we constructed a dummy data set.

(a) $A \in \mathcal{A}$ (b) $B_{D,N} \in \mathcal{B}_{D,N}$ (c) $B_D \in \mathcal{B}_D$ (d) $\widehat{B_{D,N}} = f_1(A)$ (e) $\widehat{B_{D,N}} = \text{softmax} \circ f_2(A)$ (f) $\widehat{B_D} = f_1(A)$
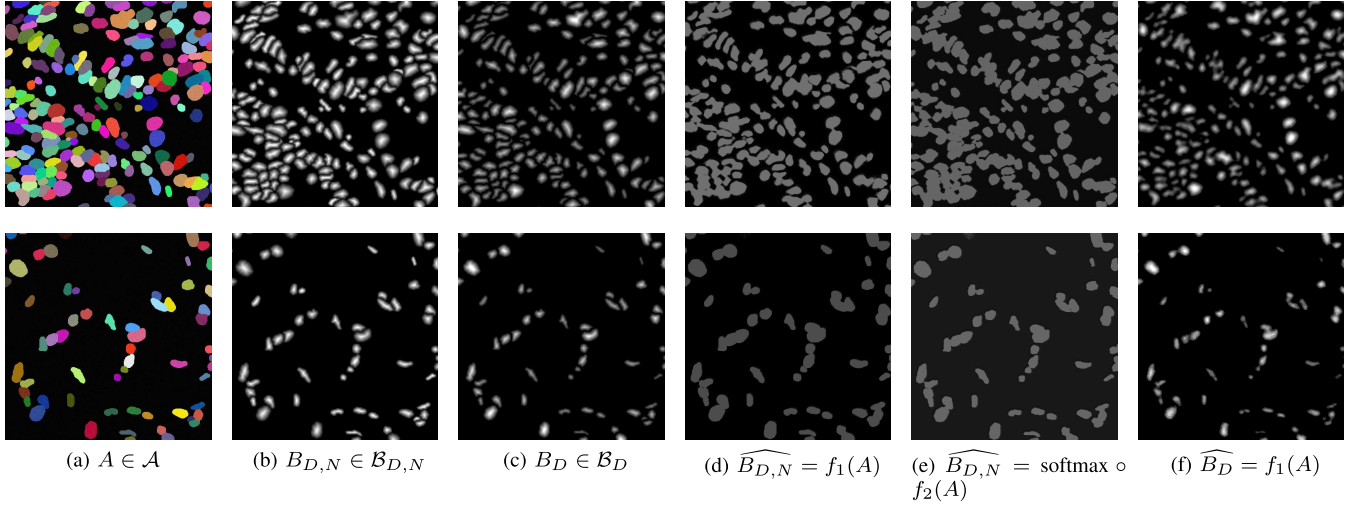
Fig. 4. Results of the distance map regression on the synthetic data set for different settings (regression of the raw distance map, regression of the normalized distance map and softmax regression for the normalized distance map). Contrast and brightness are modified for better visibility. The actual maximal values are: (b) max: 1, (c) max: 14, (d) max: 0.298, (e) max: 0.4 and (f) max: 12.

A. Synthetic Data Set

To design a very simple dataset we used the ground truth of data set $DS1$ introduced in III-A. We dilated the labeled ground truth images in order to increase the proportion of overlapping cells. For each nucleus in each image, we sample a color and assign this color to each pixel of the nucleus. We then add random Gaussian noise to each pixel in this nucleus. Finally, we add noise to the entire image that is drawn from a half-normal distribution.

Algorithm 1 Generating the Dummy Data Set

Input: Ground truth (binary) image
Output: Simulated sample
1: Label each connected component.
2: Dilate the image.
3: **for** each connected component (CC) **do**
4:     Draw $c_i \sim \mathcal{N}(\mu_1, \sigma_1 \mathbf{I}_3)$
5:     Assign c_i to CC
6:     Corrupt each pixel in CC by adding $\varepsilon \sim \mathcal{N}(0, \sigma_2)$
7: **end for**
8: Corrupt each pixel by $|\varepsilon|$ with $\varepsilon \sim \mathcal{N}(0, \sigma_3)$.
9: **return** Simulated image

For our experimental setup, we choose $\mu = 127$, $\sigma_1 = 100$, $\sigma_2 = 10$ and $\sigma_3 = 5$.

In this simple data set, each nucleus takes thus a constant color, corrupted by noise. With this data set, we can now ask the question if and under which conditions it is in principle possible to learn a distance map by regression with a DNN.

B. Results on Synthetic Data Set

Here, we apply a DNN in order to predict from the generated synthetic data the raw and normalized distance maps described in section IV-C. The objective is to understand which design allows us to predict the distance map easily, i.e. in a few epochs. In order to obtain unbiased results, we split the data set into two and retain 10 samples from the training procedure.

These 10 samples are kept to check how the models perform on unseen data.

As model, we use a simplified U-Net architecture with $n_{feat} = 2$, we take $l_r = 0.01$ and $\lambda = 0.0005$ and allow training for 10 epochs with a batch size of 4. Figure 4 shows typical examples obtained for the retained images. We notice that only the unnormalized f_1 regression seems to learn the distance map reasonably well (figure 4f). Figures 4d and 4e illustrate that the values inside the nucleus do not get smaller close to the border of the object. This suggests that predicting the normalized distance map is probably much more complicated than predicting the raw distance map, even though smaller and bigger cells are better balanced in terms of their value spectra.

To illustrate this, we can consider a large and a small cell. The small cell will have very few pixels and consequently, at equal distance to the border, the values of the normalized distance map are much larger than in the normalized distance map of a large cell. It is conceptually much harder to learn a rule that would assign different values in proximity of the border depending on the total size of the nucleus, as a more global property of the nucleus (of which the neural network has no prior knowledge) would influence a local property (the value of the distance map).

In both normalized settings, the network seems to assign a bi-value prediction. The normalized f_1 regression model assigns the value of 0.298 if it believes that the pixel belongs to a nucleus and 0 otherwise. The normalized distance however is not predicted, the neural net has learned a classification rather than a regression rule. On the contrary, regression of the unnormalized distance map converges in less then 5 epochs and produces the expected output.

We concluded from this experiment that the prediction of the unnormalized distance map is the most promising of the three approaches.

C. Conclusion

In conclusion of this section, we propose to predict the unnormalized distance map of nuclei with a U-Net architecture

and a regression loss. The output of this regression network is ideally suited for the postprocessing scheme. In the following, we call this workflow consisting in distance map regression and postprocessing DIST.

VI. APPLICATION TO REAL DATA

A. Metrics

We will use two types of metrics in this paper. A per pixel metric, the F1 score and a per object metric, the Aggregated Jaccard Index (AJI) as presented in [32]. The F1 measure is defined as the harmonic mean between recall and precision at the pixel level and is equivalent to the dice coefficient:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

The AJI is an extension of the global Jaccard index, where every ground truth component is first matched to one detected component by maximizing the Jaccard index. The AJI corresponds then to the ratio of the sums of the cardinals of intersection and union of these matched components respectively. In addition, all detected components that are not matched are added to the denominator. More formally, with $G = \bigcup_{i=1...L} G_i$ the ground truth set of object pixels, G_i its L connected components, $S = \bigcup_{k=1...M} S_k$ the detection results and S_k its M connected components, the AJI can be written as:

$$AJI = \frac{\sum_{i=1}^{L} |G_i \cap S_k^*(i)|}{\sum_{i=1}^{L} |G_i \cup S_k^*(i)| + \sum_{l \in U} |S_l|} \quad (6)$$

where $S_k^*(i)$ is the connected component from the segmentation result that maximizes the Jaccard Index with ground truth component G_i, i.e. $S_k^*(i) = \arg\max_{S_v} \frac{|G_i \cap S_v|}{|G_i \cup S_v|}$ (with $S_k^*(i) \cap S_l^*(j) = \emptyset$ for $i \neq j$) and U the set of indices of detected components that have not been assigned to any ground truth component.

B. Data Augmentation

We perform standard data augmentation techniques such as rotation, mirroring and blurring to deal with focus variations during acquisition. As suggested in [38], we also randomly augment our data by deforming the image content. Finally, we also modify the color content of the images in the following way: we perform standard color deconvolution, as described in [5], perturb the deconvolved channels by affine histogram modifications and invert the deconvolution to obtain the final image. This color augmentation technique aims at making the method more robust with respect to differences in staining and tissue properties.

C. Training, Validation and Test Set

In order to compare the performance of our method with previously published methods, we use the same dataset (DS2) for benchmarking. For this, we split DS2 into three sets: training, validation and test set. The test set is the same as the one used in [32], so that the reported results are directly comparable to the results we obtain. From the remaining images, we build a validation set containing one image from each organ and a training set containing the rest of the images.

The overall procedure is as follows: given a set of network hyperparameters, we train a DNN with the images in the training set and evaluate the performance on the validation set. This is repeated for all sets of network hyperparameters tested, and we finally choose the set of network hyperparameters that optimizes the F1 score.

To set the hyperparameters for the post-processing scheme, we use the selected model from the previous step and perform the post-processing procedure with different values of p_1 and p_2. We finally choose p_1 and p_2 that maximize the AJI score on the validation set.

The accuracy metrics we report are calculated on the test set containing the same images as in [32], none of which has been used neither for training nor for validation.

D. Models and Training Procedure

In this study, we studied three different architectures: FCN, U-Net and mask R-CNN.

1) Training FCN: As described in the original paper, we imported the pre-trained weights from *vgg16* trained on ImageNet and then fine-tuned with the training data described above. Fine-tuning FCN implies importing as initialization the values for the feature extraction part of the network, all other parameters are initialized with Glorot initialization [46]. The network hyperparameter l_r is optimized as described above. For the postprocessing, p_2 is set to 0.5, as the output is a probability map. p_1 is optimized as described above.

2) U-Net: In contrast to the other methods, this network has to be learned from scratch. We set the batch size to 10 for computational reasons and we decay the learning every 10 epochs exponentially. We tune a total of 5 parameters: 3 network hyperparameters n_{feat}, the weight decay λ and the learning rate l_r and the 2 postprocessing hyperparameters p_1 and p_2. We adjust p_2 (threshold) only in the case of regression and set it to 0.5 in case the U-Net outputs a probability map.

3) Mask R-CNN: has many hyperparameters most of which we set to recommended values (taken from https://github.com/matterport/Mask_RCNN). We modified the following parameters: we set the backbone to ResNet 101 [47] pre-trained on the COCO dataset [28] and we set the number of training anchors per image for the region proposal network part to 256. We set the number of regions of interest per image to train with to 256. Furthermore, we use a batch size of 4 and decay the learning rate by 10 every 20 epochs. We tuned several hyperparameters: the learning rate, the detection minimum confidence rate for regions of interest (DMC) and the non-maximum suppression threshold (NMST) for the region proposal network. The last two parameters are tuned during training and during validation. Indeed, during training it is better to have a high DMC as this means more relevant examples are passed through the network and during inference it is better to lower this value to have more proposals.

For each architecture and each set of hyperparameters we train for a maximum of 80 epochs (a couple of training hours) and monitor the train and validation error in order to
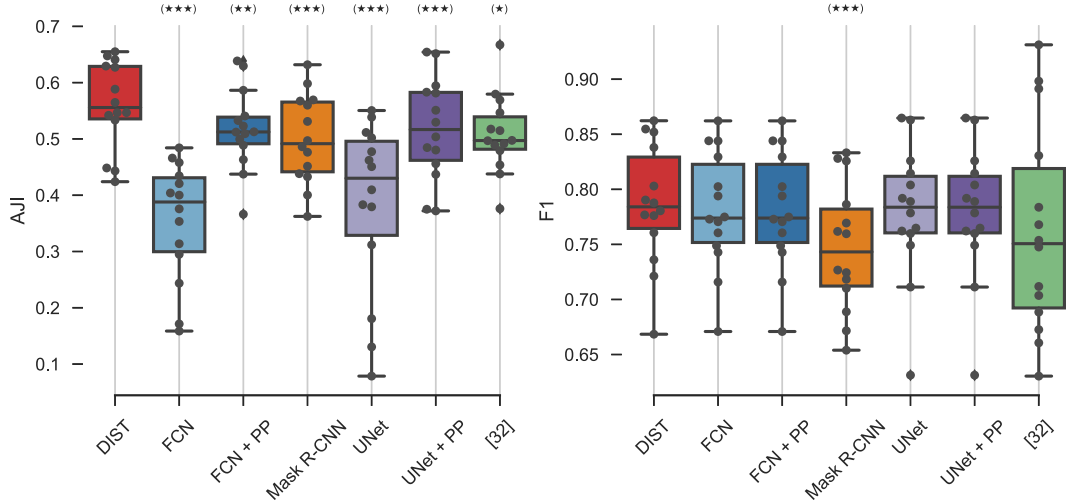
Fig. 5. Box plots for comparative analysis between the different methods on the test data set. Left-panel: AJI scores. Right-panel: F1 scores. Each black point corresponds to the score obtained for one image by the respective method. An image score is considered as outlier if it resides outside $[q_{25\%} - 1.5 \times iqr; q_{75\%} + 1.5 \times iqr]$, where q_α denotes the quantile of level α and $iqr = q_{75\%} - q_{25\%}$ denotes the inter-quantile range. We use ($\star\star\star$) if the p-value is below 0.01, ($\star\star$) if the p-value is below 0.05 and ($\star$) if the p-value is below 0.10. The test is a one-sided Wilcoxon signed-rank test between our method (DIST) and the other methods.

TABLE I
MODELS AND HYPER PARAMETERS

Hyperparameter	Range	U-Net	FCN	Mask R-CNN
l_r	$[10^{-i}$ for $i \in [2:6]]$	✓	✓	✓
n_{feat}	$[16, 32, 64]$	✓		
λ	$[5.10^{-i}$ for $i \in [3:6]]$	✓		
p_1 (Prob)	$[6:12]$	✓	✓	
p_2 (Prob)	0.5	✓	✓	
p_1 (Dist)	$[0:2]$	✓		
p_2 (Dist)	$[i.10^{-1}$ for $i \in [5:10]]$	✓		
DMC	$[i.10^{-1}$ for $i \in [4:9]]$			✓
NMST	$[i.10^{-1}$ for $i \in [5:9]]$			✓

perform early stopping with a view of 10 epochs, i.e. training is stopped when the validation error has not decreased for the last 10 epochs. We take the set of hyperparameters that minimizes the validation error and measure the performance on the independent test set. Table I shows all hyperparameters with their range.

E. Results - Comparison of Methods

With the training and model selection and testing procedure described in section VI-D, we quantitatively compared a panel of methods: in figure 5 we provide an overall comparison of both AJI and F1 score, in table II and figure 6 we provide a detailed image by image comparison of the method's performances. Furthermore, we show example images in figure 8.

Figure 5 and table II show that on average, the regression approach we propose (DIST) outperforms competing methods on the AJI score. On the F1 score, the regression approach is roughly at the same level as the FCN and the U-Net (with or without post-processing), and slightly better than the other methods. In order to assess the statistical significance

of the improvement brought by our method, we performed a one-sided exact Wilcoxon signed-rank test comparing DIST with all competing methods. This non-parametric test was chosen, as the paired differences cannot be assumed to be normally distributed, see figure 6. We report the corresponding levels of significance in figure 5 and conclude all differences are statistically significant for the AJI score ($p < 0.1$ for all methods tested).

We observe that on the pixel-wise accuracy (F1-score), except for the mask R-CNN, the fully convolutional networks show a higher score on average than the sliding window approach [32], which is the standard in the field [8], [31], [35]. While this difference is not statistically significant, this result is still interesting in the field of histopathology with its extremely large data sets, because fully convolutional networks are computationally much more efficient than sliding window approaches. Better scores can be reached by increasing the size of the training set (data not shown), but for the sake of comparability, we preferred not to include more data in the training and model selection procedure.

In terms of AJI, our method outperforms the others. Furthermore, we observe that in general, the most competitive AJI scores are obtained with a post-processing scheme. While the basic FCN and U-Net perform very well at the pixel level (F1 score), they perform poorly on the AJI score, indicating that an excellent pixel-level performance does not imply excellent performance at the object level. In contrast to this, mask R-CNN has similar AJI scores to [32] even though the F1 score is relatively low. Indeed, due to the nature of its architecture, in particular due to the region proposal part of the network, the mask R-CNN model will only label pixels if it has found a potential cell object first. This is different to the other approaches that maximize a per-pixel probability: object-level characteristics are implicitly taken into account.

TABLE II
COMPARATIVE ANALYSIS FOR INDIVIDUAL IMAGES FROM THE TEST DATA SET

Organ	Image	AJI (Object based)							F1 (Pixel based)				
		DIST	FCN	FCN + PP	Mask R-CNN	U-Net	U-Net + PP	[32]	DIST	FCN	Mask R-CNN	U-Net	[32]
Bladder	1	0.6475	0.4841	0.5864	0.5693	0.5115	0.5510	0.5465	0.8623	0.8439	0.8332	0.8258	0.9312
	2	0.5467	0.4001	0.4887	0.4328	0.4621	0.5296	0.4968	0.7768	0.7729	0.6888	0.7648	0.6304
Colorectal	1	0.4240	0.1713	0.3662	0.3624	0.0786	0.3751	0.4891	0.7212	0.6709	0.6539	0.7121	0.7679
	2	0.4484	0.2949	0.4374	0.4003	0.1305	0.4371	0.5692	0.7360	0.7158	0.7101	0.7599	0.7118
Stomach	1	0.6408	0.3757	0.5230	0.5983	0.4096	0.6515	0.4538	0.8547	0.7939	0.8257	0.8647	0.8913
	2	0.6550	0.4040	0.5327	0.6318	0.4507	0.6543	0.4378	0.8520	0.8025	0.8280	0.8629	0.8982
Breast	1	0.5334	0.2436	0.4632	0.4515	0.3831	0.4559	0.4974	0.7761	0.7605	0.7267	0.7492	0.6885
	2	0.5884	0.4345	0.6385	0.5311	0.5505	0.4803	0.5796	0.8380	0.8621	0.7694	0.8144	0.7476
Kidney	1	0.5648	0.4580	0.5126	0.5600	0.5386	0.5815	0.4792	0.7805	0.7488	0.7863	0.8039	0.6606
	2	0.5420	0.4660	0.5407	0.4764	0.5020	0.5037	0.6672	0.7606	0.7705	0.7244	0.7786	0.7837
Liver	1	0.5466	0.4206	0.4996	0.4861	0.4773	0.4846	0.5175	0.7877	0.7749	0.7597	0.7648	0.6726
	2	0.4432	0.3138	0.5093	0.4383	0.3794	0.3721	0.5148	0.6684	0.7428	0.6716	0.6313	0.7036
Prostate	1	0.6273	0.1586	0.5121	0.4970	0.1807	0.5831	0.4914	0.8030	0.8440	0.7183	0.7889	0.8306
	2	0.6294	0.3534	0.6296	0.5673	0.3118	0.5945	0.3761	0.7903	0.8294	0.7618	0.7919	0.7537
Overall		0.5598	0.3556	0.5172	0.5002	0.3833	0.5182	0.5083	0.7863	0.7809	0.7470	0.7793	0.7623

We also see from the results presented in figure 5 and table II that there is not one method that outperforms systematically all other methods on all images. This is not surprising: there are for instance images with relatively few and well separated nuclei, where the separation of touching nuclei does not bring any advantage. For these images, it is expected that using post-processing strategies will not be beneficial. For other images, the choice of hyperparameters might not have been optimal, because they are too different from the validation set we have used. While there is variability in the performance of the tested methods and while we understand the origin of this variability, our results still demonstrate that fully convolutional networks in combination with the proposed post-processing strategies outperform the other methods on these data.

Finally, we compare the two workflows with post-processing, i.e. our previously published method where we apply maximum selection and watershed transform on the posterior probability maps [42], and DIST described in section V-C where we apply maximum selection and watershed transform on the predicted distance map. We observe a clear advantage for the regression strategy. Indeed, the results reported in figure 5, table II and figure 6 suggest that our formulation pushes the neural network to learn a mapping that incorporates some notion of objects.

In terms of visual results, as illustrated in figure 8, our formulation of the problem seems to work like a shape prior. The reason for this is that by regressing the distance map on the original image, the network tries to mimic the typical shape of the distance map. This implies also more roundish or elliptical shapes, as can be seen in figure 8b. In comparison to our previously published post-processing scheme that relied on application of morphological dynamics and watershed to the probability maps, the advantage is that the reconstruction of the distance map seems to work like a regularization of the signal and is therefore not confronted to the fragmentation typically observed in cases of high uncertainty.

Interestingly, we observed that we achieve the highest AJI score with p_1 set to 0. This means that any local maximum in the predicted distance map suggests the presence of a nucleus. However, higher values of p_1 help to compensate

TABLE III
SELECTED HYPERPARAMETERS

Models	l_r	λ	n_{feat}	p_1	p_2
U-Net	10^{-4}	5.10^{-7}	32	11	0.5
DIST	10^{-3}	5.10^{-6}	32	0	0.5
FCN	10^{-4}	x	x	10	0.5

Models	l_r	Training		Inference	
		DMC	$NMST$	DMC	$NMST$
Mask R-CNN	10^{-2}	0.9	0.5	0.5	0.5

noisy images and to aggregate local maxima to form a single cell. Higher values of p_1 should be used if one is predicting a set where the domain is different to normal and the classifier is less confident. This is also reflected by the size distributions of the detected nuclei. In figure 7, we show the kernel density estimations of the cell size distributions for the ground truth (black) and a selection of other methods. We observe that the distribution of nuclei as predicted by DIST is very similar to the true size distribution, in contrast to the results obtained by the other methods. These results suggest that shape descriptors calculated on DIST results more faithfully represent the true feature representations.

F. Results on TNBC Slides

The comparison of different methods and approaches showed so far indicates that predicting the distance map with DNN is a competitive approach for nuclei detection and excels in particular in the separation of clustered nuclei. Our main interest is the analysis of breast cancer samples and so we want to ask the question whether we can still increase performance on breast cancer WSI by specifically training with slides from the same tissue.

For this, we designed an experiment where we compare the performance of three models, all trained with the same hyperparameters (i.e. there is no further model selection). For all three models, we evaluated the results on the breast samples of $DS2$. The first model was trained solely on $DS2$. The second model was trained on $DS1 \cup DS2$ (training for 80 epochs without early stopping). The third model was trained on the
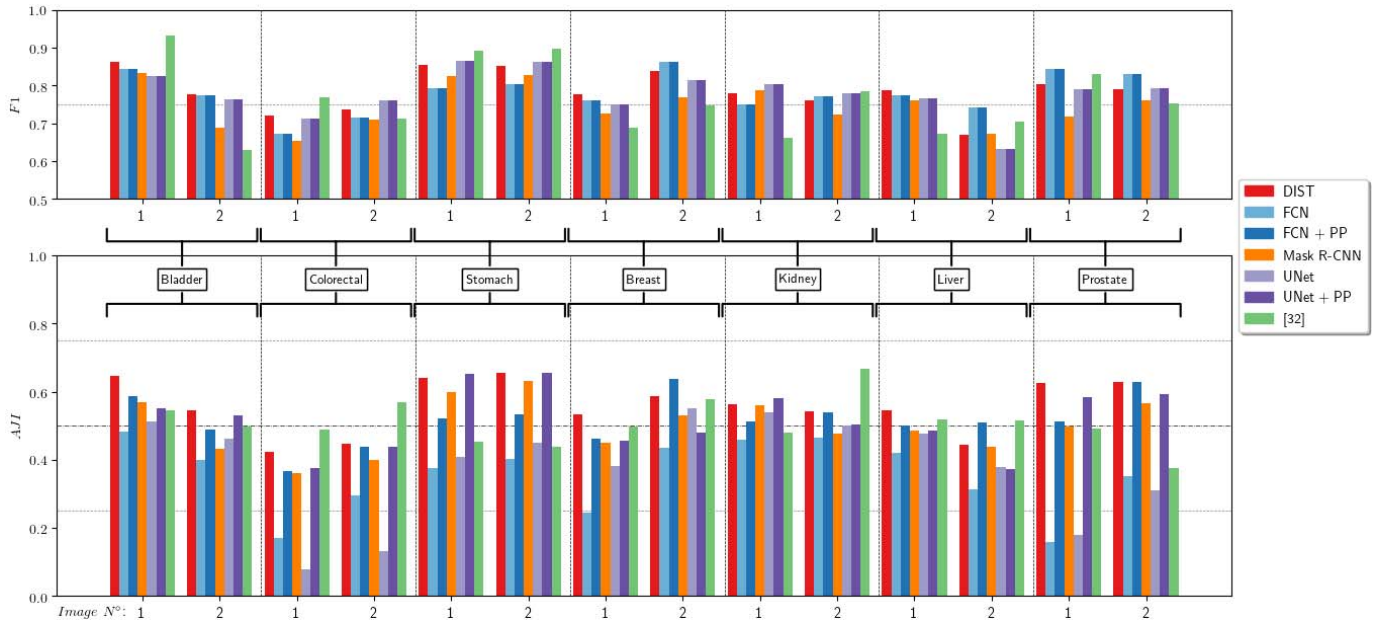
Fig. 6. Comparative analysis for individual images from the test data set. Top-panel: F1 score for each individual image and all methods. Bottom-panel: AJI score for each image and all methods. Color code is given alongside the graphs.
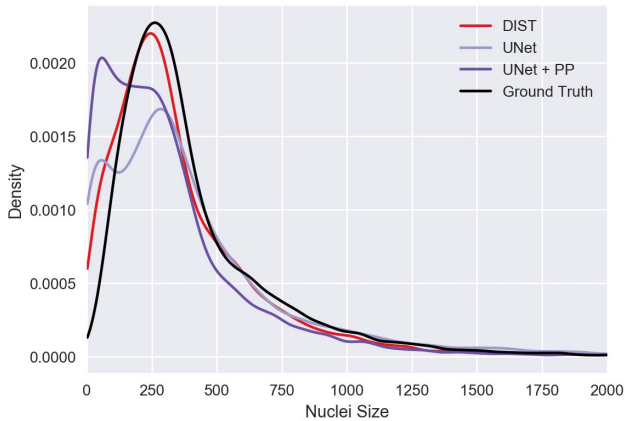

Fig. 7. Distribution of cell size for different methods (in pixels). Curves represent the density as obtained by kernel density estimation. The curves were cut at 2000 for better visibility.

TABLE IV
RESULTS FOR BREAST CANCER SLIDES AND
DIFFERENT TRAINING SETS

	F1	AJI
[32]	0.7180	0.5385
Trained with *DS2*	0.8070	0.5609
Trained with *DS1* and *DS2*	0.8110	0.5733
Trained only on Breast	0.8236	0.5853

breast samples of $DS1$ and $DS2$. The overall results are shown in table IV.

We observe that the model trained solely on breast cancer images performed best. Apparently, training with samples obtained from different organs slightly deteriorates the performance of the model. Conversely, it seems advantageous to focus the training on data sets that are most representative for the data the method is going to be applied to. On the other hand, the results are not fundamentally different, indicating that it is also possible to develop more generally applicable segmentation networks, if losing a few percent accuracy is acceptable.

VII. DISCUSSION

Automatic analysis of histopathology data bears great promise for both clinical practice and research related to cancer. The segmentation of nuclei is of great importance if one wants to build interpretable models for the prediction of clinical variables, such as the risk of relapse or the efficiency of a treatment. Interpretable models are one major line of research in this field, and will be even more popular when the field moves towards data integration including genomic, epigenomic, expression and phenotypic data. We therefore believe that the segmentation of nuclei is a cornerstone for such integrative approaches, as they provide the building blocks of tissues: their detection allows for a detailed analysis of both morphological and spatial signatures of the tissue. Furthermore, they represent the natural intermediate level between the molecular information typically revealed by sequencing approaches and the macroscopic level of a diseased tissue.

In this work, we show that fully convolutional networks are well suited to the task of nuclei segmentation. In addition to increased accuracy at the pixel-level, they have also the benefit of being faster than traditional sliding window approaches. This is an essential aspect, as the images in histopathology are extremely large and — once automatic analysis becomes a standard in pathology departments — numerous. Furthermore, we focus on the separation of touching nuclei, which is clearly one of the major problems in segmentation of nuclei in general and in histopathology data in particular. We address this problem by formulating the segmentation task as a regression problem, where we aim at predicting the distance map of nuclei. This formulation is unusual, as naturally segmentation
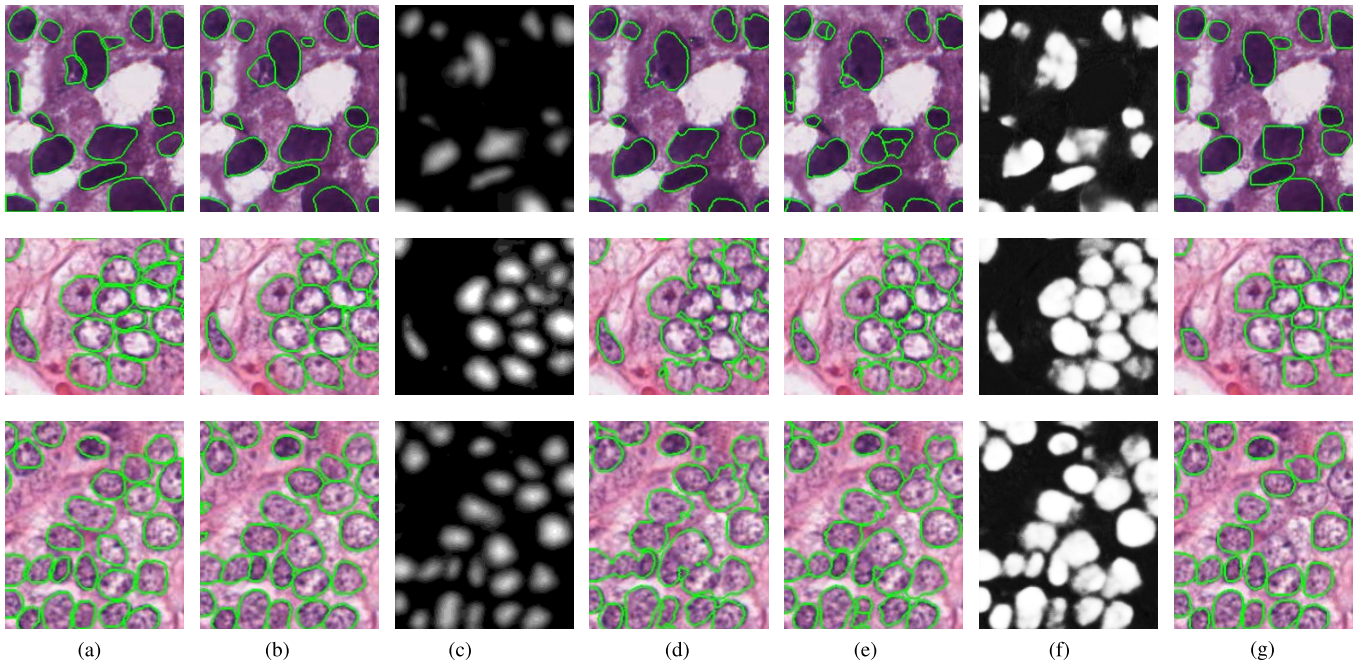
Fig. 8. Comparing segmentation results on cluttered cells. (a) Ground Truth. (b) DIST. (c) Distance regression output. (d) U-Net. (e) U-Net + PP. (f) U-Net probability map. (g) Mask R-CNN.

is cast as a classification problem. The underlying rationale is that this formulation implicitly imposes a shape prior on the network and makes the network learn some object-level notions. Furthermore, it also has some regularization capacity, as the network is forced to generate a spatially smoothed output, therefore leading to more regular structures. Other approaches in this field have focused on pixels at the intersection between objects or on the object contours. Our analysis suggests that it is advantageous to shift the focus on the interior of the object, in particular in case of the presence of blurred boundaries or contour pixels that might contain a less informative signature. Finally, our approach can also be seen as an ideal and particularly robust preparation to the classical and very popular separation by the watershed transformation.

From a larger perspective, the problem of segmenting touching entities in biological data is certainly not limited to nuclei detection in histopathology data. Indeed, in virtually all imaging modalities at the cellular level we are confronted with this problem, e.g. in fluorescence microscopy. We therefore believe that our formulation of the problem as a regression task that can be efficiently solved by fully convolutional neural networks may also have some impact on other fields of bioimaging.

VIII. CONCLUSION

In this article, we present a new method based on deep learning for the segmentation of nuclei in H&E stained tissue slides. The main problem we are currently facing in this field is the joint segmentation of close nuclei. The core idea of the paper is to formulate the segmentation task as a regression of the intra-nuclear distance maps. With this novel formulation, we implicitly impose a shape prior on the segmentation network. We demonstrate the power of the method by applying

it to a previously published data set, showing that the method outperforms previously published methods, including our own.

Furthermore, we also generated a carefully annotated data set for nuclei segmentation, which we made publicly available to the scientific community on the website https://peterjacknaylor.github.io/. This data set was published previously [42], but we extended it considerably in order to cover the intra- and inter-patient heterogeneity for a single type of cancer. We also showed that the use of a dedicated data set can increase the performance of segmentation algorithms.

In addition to the methodological developments and in the spirit of reproducible research, we also make the code for all experiments publicly available under https://peterjacknaylor.github.io/ for the synthetic data described in section V and for the real data described in section VI. This code is written for python 3 and is based on scikit-image [48] and tensorflow [49]. In addition, we provide nextflow scripts [50] which we use to unify the scripts in one pipeline file.

REFERENCES

[1] T. J. Hudson *et al.*, "International network of cancer genome projects," *Nature*, vol. 464, no. 7291, pp. 993–998, Apr. 2010.
[2] K.-H. Chow, R. E. Factor, and K. S. Ullman, "The nuclear envelope environment and its cancer connections," *Nature Rev. Cancer*, vol. 12, no. 3, pp. 196–209, Feb. 2012.

[3] C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up," *Histopathology*, vol. 19, pp. 403–410, Nov. 1991.

[4] P. H. Bartels, J. E. Weber, and L. Duckstein, "Machine learning in quantitative histopathology," *Anal. Quant. Cytol. Histol.*, vol. 10, no. 4, pp. 299–306, 1988.

[5] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.

[6] B. E. Bejnordi *et al.*, "Stain specific standardization of whole-slide histopathological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 404–415, Feb. 2016.

[7] H. Cho, S. Lim, G. Choi, and H. Min. (2017). "Neural stain-style transfer learning using GAN for histopathological images." [Online]. Available: https://arxiv.org/abs/1710.08543

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

[9] M. Veta *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Med. Image Anal.*, vol. 20, no. 1, pp. 237–248, 2015.

[10] S. Manivannan, W. Li, S. Akbar, J. Zhang, E. Trucco, and S. J. McKenna, "Local structure prediction for gland segmentation," in *Proc. 13th Int. Symp. Biomed. Imag. (ISBI)*, Prague, Czech Republic, Apr. 2016, pp. 799–802.

[11] Y. Liu *et al.*, "Detecting cancer metastases on gigapixel pathology images," in *Proc. Comput. Vis. Pattern Recognit.*, 2017, pp. 1–13.

[12] Y. Xu *et al.*, "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC Bioinf.*, vol. 18, no. 1, pp. 1–17, 2017.

[13] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," *Signal Process.*, vol. 122, pp. 1–13, May 2016.

[14] H. R. Ali *et al.*, "Lymphocyte density determined by computational pathology validated as a predictor of response to neoadjuvant chemotherapy in breast cancer: Secondary analysis of the ARTemis trial," *Ann. Oncol.*, vol. 28, no. 8, pp. 1832–1835, Aug. 2017.

[15] J. M. Chen *et al.*, "Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review," *Tumor Biol.*, vol. 39, no. 3, p. 1010428317694550, 2017.

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

[17] N. Kumar *et al.*, "Convolutional neural networks for prostate cancer recurrence prediction," *Proc. SPIE*, vol. 10140, Mar. 2017.

[18] H. R. Ali *et al.*, "Computational pathology of pre-treatment biopsies identifies lymphocyte density as a predictor of response to neoadjuvant chemotherapy in breast cancer," *Breast Cancer Res.*, vol. 18, no. 1, p. 21, 2016.

[19] J. Cheng and J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 741–748, Mar. 2009.

[20] M. Veta *et al.*, "Automatic nuclei segmentation in H&E stained breast cancer histopathology images," *PLoS ONE*, vol. 8, no. 7, p. e70221, 2013.

[21] P. Faridi, H. Danyali, M. S. Helfroush, and M. A. Jahromi. (2016). "Cancerous nuclei detection and scoring in breast cancer histopathological images." [Online]. Available: https://arxiv.org/abs/1612.01237

[22] Y. Gao *et al.*, "Hierarchical nucleus segmentation in digital pathology images.," *Proc. SPIE*, vol. 9791, p. 979117, Mar. 2016.

[23] P. Guo, A. Evans, and P. Bhattacharya, "Segmentation of nuclei in digital pathology images," in *Proc. IEEE 15th Int. Conf. Cognit. Inform. Cogn. Comput. (ICCI CC)*, Aug. 2016, pp. 547–550.

[24] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.

[25] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—Current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2013.

[26] 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.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[28] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," *CoRR*, vol. 8693, pp. 740–755, May 2014.

[29] E. F. Braz and R. D. A. Lotufo, "Nuclei detection using deep learning," in *Proc. Simpósio Brasileiro Telecomunicações Processamento Sinais*, 2017, pp. 1059–1063.

[30] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[31] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.

[32] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[33] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421–2433, Oct. 2015.

[34] D. A. Van Valen *et al.*, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Comput. Biol.*, vol. 12, no. 11, p. e1005177, 2016.

[35] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Inf.*, vol. 7, no. 1, p. 29, 2016.

[36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[37] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241.

[39] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[40] H. Irshad *et al.*, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: Evaluating experts, automated methods, and the crowd," in *Proc. Pacific Symp. Biocomput. Co-Chairs*. Singapore: World Scientific, 2014, pp. 294–305.

[41] E. D. Gelasca, B. Obara, D. Fedorov, K. Kvilekval, and B. Manjunath, "A biosegmentation benchmark for evaluation of bioimage analysis methods," *BMC Bioinform.*, vol. 10, no. 1, p. 368, 2009.

[42] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 933–936.

[43] P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[44] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[45] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. New York, NY, USA: Springer-Verlag, 2003.

[46] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[48] S. van der Walt *et al.*, "Scikit-image: Image processing in Python," *PeerJ*, vol. 2, p. e453, Jun. 2014.

[49] M. Abadi *et al.* (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org

[50] P. Di Tommaso, M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, and C. Notredame, "Nextflow enables reproducible computational workflows," *Nature Biotechnol.*, vol. 35, no. 4, pp. 316–319, 2017.