NUCLEI SEGMENTATION IN HISTOPATHOLOGY IMAGES USING DEEP NEURAL NETWORKS

Peter Naylor[1,2,3*] *Marick Laé*[4] *Fabien Reyal*[5,6,7] *Thomas Walter*[1,2,3]

[1]MINES ParisTech, PSL Research University, CBIO - Centre de Bioinformatique, 77300 Fontainebleau, France;
[2]Institut Curie, [3]INSERM U900, Paris, [4]Service de Pathologie, Institut Curie, Paris, F-75248, France
[5]Residual Tumor & Response to Treatment Laboratory, RT2Lab, Translational Research Department, Institut Curie,
[6]U932, Immunity and Cancer, INSERM, Institut Curie, [7]Department of Surgery, Institut Curie, Paris, F-75248, France.

ABSTRACT

Analysis and interpretation of stained tumor sections is one of the main tools in cancer diagnosis and prognosis, which is mainly carried out manually by pathologists. The avent of digital pathology provides us with the challenging opportunity to automatically analyze large amounts of these complex image data in order to draw biological conclusions from them and to study cellular and tissular phenotypes at a large scale. One of the bottlenecks for such approaches is the automatic segmentation of cell nuclei from this type of image data. Here, we present a fully automated workflow to segment nuclei from histopathology image data by using deep neural networks trained from a set of manually annotated images and by processing the posterior probability maps in order to split jointly segmented nuclei. Further, we provide the image data set that has been generated for this study as a benchmark set to the scientific community.

Index Terms— Deep Learning, Convolutional Neural Networks, Nuclei Segmentation, Histopathology, Digital Pathology, Breast Cancer, Cellular Phenotyping

1. INTRODUCTION

Today, large sequencing approaches build the main body of cancer research programs and they have revolutionized our understanding of the molecular basis of cancer. In clinical practice however, molecular profiling is paralleled with the more traditional (and mostly manual) analysis of stained histological tumor sections. With the avent of digital pathology, i.e. the scanning and digital storage of diseased tissue sections, it is now possible to build tools for the quantitative and automatic analysis of these complex and informative image data, which are complementary to genomic and expression data.

For these reasons, analysis of histopathology data has received much attention over the last years. In particular for approaches aiming at relating biologically relevant features (such as phenotype distributions or heterogeneity measures) to clinical variables, segmentation of nuclei from tissue images is essential.

However, segmentation of nuclei is a complicated task : tissue type, staining differences and cell type convey them different visual characteristics, which makes it very difficult to design traditional image segmentation algorithms that work satisfactorily for all of these different cases. On the other hand, deep learning algorithms have been used recently with great success to complex segmentation tasks in biology[1, 2].

The contribution of this paper is three-fold : (1) we generated a set of manually annotated, representative images containing segmentation results for more than 2000 cells. We believe that such datasets can trigger methodological developments and are important for the field. (2) We apply a deep neural network end-to-end strategy and demonstrate the superiority of this approach with respect to previously proposed simpler architectures. (3) We propose a post-processing strategy of the posterior probabilities provided by the network. This approach is based on mathematical morphology and unlike many adhoc procedures allows a clearly defined and intuitive criterion for the object splits.

2. RELATED WORK

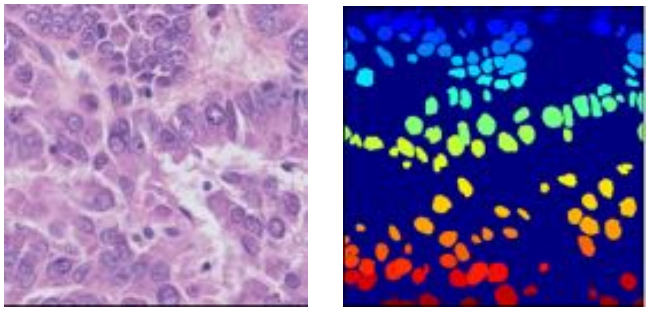
Many traditional techniques have been proposed for the segmentation of nuclei in histopathology images, ranging from simple background subtraction and color threshold techniques to much more sophisticated approaches, such as marked point processes. Many of these methods have been recently reviewed [3].

More recently, advances in deep convolutional neural networks (CNN), and in particular in their optimization have made them become the state-of-the- art model for object recognition. Indeed, deep neural networks have been used for the localization and classification of nuclei in histopathology data from breast and colon cancer [4, 5], but these previously published approaches do no allow for segmentation.

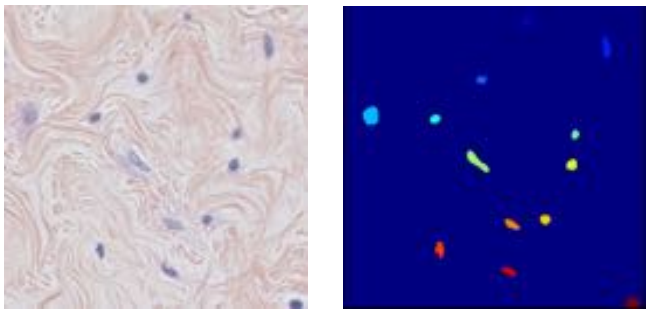
∗. Peter Naylor has received a PhD fellowship from the Ligue contre le Cancer

On the other hand, deep CNN have also been used for semantic segmentation in other fields, where [6] use "deconvolution layers" and up-sampling in order to directly output binary images. The rationale of our proposed method was therefore to learn segmentation directly from annotated samples with this or similar architectures.

Learning based methods rely on annotated data sets. But for the problem of nuclei segmentation in histopathology images, there are only few manually segmented images freely available. In [7], a partially annotated data set has been released. However, this data set does not cover the morphological variability typically encountered in real histopathology data.



Sample from patient 1 Associated ground truth



Sample from patient 2 Associated ground truth

Fig. 1. Random annotated samples from the dataset

3. DATA SET

One of the contributions of this paper is the now publicly available nuclei detection dataset within HE stained histopathology images which can be found at `http://cbio.mines-paristech.fr/~pnaylor/BNS.zip`. This annotated dataset provides images clustered by patients. Each patient has at least 3 annotated 512×512 HE histopathology images with their associated ground truth. Each ground truth image is a 512×512 label image where each pixel value above 0 is considered as the label of the corresponding nucleus. See figure 1 for an example of three annotated images. This annotation was conducted via the help of the software ITK-snap, [8] `http://www.itksnap.org` by the authors of this paper.

The patients were randomly picked from an unpublished study on triple negative breast cancer (TNBC). For each of the patients we had access to a biopsy and a whole slide image (WSI). WSI are very large images and can be up to 60 GB (uncompressed). As they cannot be stored in the RAM of a standard computer, we randomly cropped 512×512 samples from the WSI. 3 to 7 images were choosen from the randomly sampled images to try and give the most diverse dataset among these patients. Once the samples were choosen, we fully annotated each nucleus via the software ITK-snap; touching nuclei were differentiated by the different label value.

In this data set we have annotated a considerable amount 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). For the moment, we did not annotate the different cell types, however.

In total, we have 33 images with a total of 2754 annotated cells, the maximum number of cells in one sample is 293 and the minimum number of cells in one sample is 5. We also have on average 83 cells per sample with a high standard deviation of 63.

4. METHODOLOGY

Let A be the space of RGB images, A can typically be $\mathbb{R}^{n \times p \times 3}$ and let B the space B the space of annotation images, in our case $\{0,1\}^{n \times p}$. We have a set of $(A_l, B_l)_{l \in [1,N]}$ for a supervised learning approach. Our goal is a prediction task named as semantic segmentation, we wish to maximize the prediction of an unseen element belonging to B given an new element in A. We maximize thus prediction by modelizing our prediction function as the softmax output of a deep neural network. We find the model parameters by minimizing a log loss function defined as : $\frac{1}{\sum_{i,j} w_{ij}} \sum_{i,j} \sum_k w_{i,j} t_{i,j,k} \log(\widehat{p_{i,j,k}})$, where k designates a certain label, $w_{i,j}$ is a certain weight given to pixel i, j, $t_{i,j,k}$ is equal to 1 if pixel i, j is of class k and $\widehat{p_{i,j,k}}$ designates the estimated probability of pixel i, j of being k via the softmax output of the neural network. We minimize the loss function via stochastic gradient descent.

We have our training set $(A_l, B_l)_{l \in [1,N]}$ where N is equal to 33, however each element A_l belongs to a certain patient and several elements A_l can belong to the same patient. As we are dealing with histopathology images, it is known that samples can widely vary from one patient to the other. We thus validate our model by a leave-one-patient-out scheme : for a given set of hyper parameters, we train our model on every patient except one that is used for validation. Our final score is averaged over all patients. Several metrics give a detailed assessment of the model quality : accuracy, intersection over union (IU), F1 score and a performance score which is the mean between true positive rate and true negative rate.

To train our models, as the number of available annotated is scarse, we used a great number of transformation for the data augmentation. From an original size of 33 annotated images, adding flips, rotations, bluriness and random elastic deformations enabled us to have more then 400000 training images. We also try out several hyperparameter configurations : the learning rate and momentum for the stochastic gradient descent, the weight decay value. In practice, we found that hyperparameters tuning didn't influence the scores much, the exepction being the learning rate. If the learning rate was not of the right magnitude the given network did not seem to learn. We therefore fixed the momentum to be 0.9 and the weight decay to be 5.10^{-5} for all of the experiments, the learning rate was tuned according to the model chosen. We also experimented with different initialization values and if possible, we also considered pretrained layers. The use of pretrained layers made learning more efficient and made scores more robust.

The results obtained by the Deep Neural Network were encouraging, but touching nuclei were often segmented as one single object. Solving this issue by weighting the error term for pixels between objects did not solve the issue for our case (data not shown). However, we did observe that for touching and even partially overlapping nuclei, the posterior probability at the nucleus border is systematically lower than in the putative center of the nucleus, but may still be relatively large. As in the center of nuclei, the posterior probability is maximal, we can readily assume that the local maxima of the posterior correspond to putative nuclei, which we call candidates. Let $\mathcal{X} = (x_1, x_2, \ldots, x_N)$ be a path that joins two candidates (i.e. $\forall i,\ x_i$ and x_{i+1} are neighbor pixels) and $\mathcal{P} = (p_1, p_2, \ldots, p_N)$ the corresponding posterior probabilities. Without loss of generality, we assume $p_1 \leq P_N$. We define for each path $\mathcal{X}$ a cost $C(\mathcal{P}) = \max_{i=2\ldots N} p_1 - p_i$, which is the maximal decrease in posterior probability along the path, when starting from the candidate with lower probability. Considering all paths joining two candidates, we can now state a criterion that allows us to decide whether to perform a split (and thus accept the candidates as being different nuclei) or not : if all the paths involve a decrease in probability larger than a parameter λ, we will perform a split. Conversely, if we can find at least one path that joins the two candidates with a probability decrease smaller or equal to λ, no split will be performed : we argue that in this case, it is probably one single object. Hence, the split is performed if :

$$\min_{\mathcal{P}} C(\mathcal{P}) = \min_{\mathcal{P}} \{ \max_{i=2\ldots N} p_1 - p_i \} > \lambda$$

where λ is a free parameter. This is actually nothing else than the morphological dynamics[9]. The actual split locations are then obtained by applying the watershed algorithm to the inverted posterior map, starting from the local minima with morphological dynamics larger than a free parameter λ. To speed up this procedure, we have developed an algorithm that performs selection and Watershed algorithm in a single pass by selectively building the Watershed line or fusing the regions.

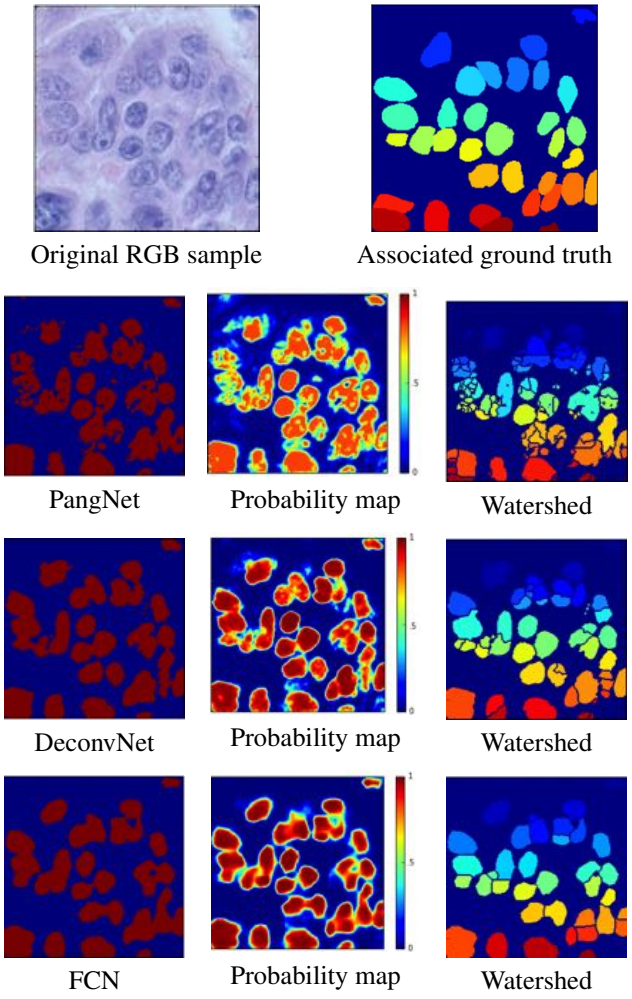


Fig. 2. Prediction via different classifiers of a random sample on the (same) left out patient

5. DIFFERENT ARCHITECTURES

We experience with 3 known arhitectures in semantic segmentation, named PangNet, Fully Convolutionnal Net (FCN) and DeconvNet. The most basic architecture, PangNet, consists of 4 convolutionnal layer where each convolutionnal layer has 8 feature map, for more information please refer to [10]. This net, being not deep, has the advantage of being less computationnally intensive. FCN is a first attempt of applying "deep feature" representations to the task of semantic segmentation. This architecture has the advantage of re-using a classical deep learning architecture with additional upsampling and skip layers. The upsampling layers enable the network to learn a pixel level classification and the skip layers enable the network to fuse different levels of abstraction to the fi-

	PangNet	DeconvNet	FCN	Ensemble
Accuracy	0.924	0.954	0.944	0.944
IU	0.722	0.814	0.782	0.804
Recall	0.655	0.773	0.752	0.900
Precision	0.814	0.864	0.823	0.741
F1	0.676	0.805	0.763	0.802
Performance	0.803	0.875	0.862	0.922

Table 1. Results

nal prediction. This model can be fine tuned with a set of pretrained-weights extracted from the classical deep learning architecture[6]. DeconvNet is also based on a classical architecture, however, in this network their are no skip layers as one intends to learn the proper upsampling through repeated deconvolution and convolution layers. This model can also be fine tuned[11]. We also did an ensemble classifier of these two previous nets as suggested in [11] but did not include any illustration. Processing one image of size 224×224 pixels took on average 0.91 secondes for PangNet, which is the smallest network, and it took on average 3.38 secondes for DeconvNet, the largest network. These times do not depend on the number of cells. These computation times were estimated on a K620 Quadro graphics card.

6. RESULTS AND DISCUSSION

We conducted our experiments on GPUs via the caffe framework, [12]. We present our results in table 1 where we averaged the metrics over all left out patients. We also added, for illustration purposes 2, image predictions and associated probability maps to evaluate some differences between the classifiers. We notice that the simpler net, PangNet seems to have learnt simple rules such as color information. As soon as the nuclei are not homogeneous and dark inside, the method fails ("holes" in the cells). On the contrary, deeper nets as FCN and DeconvNet have learnt deeper features and seem capable of recognizing whole cells. On these probability maps we applied our postprocessing scheme and we can picture the results in the third column. By construction, the postprocessing relies on a smooth posterior probability as we can see in the PangNet prediction : many unrelated minimas will lead to oversegmentation. This phenomenom also appears with the DeconvNet where undesired cuts were performed, like for the cell at the bottom left of the image. We picked $\lambda = 7$ which seemed to give a nice partitionning of the image, this value can also be found by cross validation.

7. CONCLUSION

We presented a fully automized workflow for segmenting nuclei in histopathology images based on deep learning and mathematical morphology. We have also generated a manually curated segmentation data base, which we make publicly available. We believe that such a method can become very useful for the investigation of cellular phenotypes at a tissue level and their relation to disease.

8. REFERENCES

[1] Ciresan et al, "Deep neural networks segment neuronal membranes in electron microscopy images," in *NIPS*, 2012, pp. 2852–2860.

[2] Ronneberger et al, "U-net : Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pp. 234–241. Springer, 2015.

[3] Irshad et al, "Methods for nuclei detection, segmentation, and classification in digital histopathology : A review—current status and future potential," *Biomedical Engineering, IEEE Reviews in*, vol. 7, pp. 97–114, 2014.

[4] Xu et al, "Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2016.

[5] Sirinukunwattana et al, "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.

[6] Long et al, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[7] Gelasca et al, "Evaluation and benchmark for biological image segmentation," in *IEEE International Conference on Image Processing*, Oct 2008.

[8] 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.

[9] Michel Grimaud, "New measure of contrast : the dynamics," in *Proc. SPIE*, 1992, vol. 1769, pp. 292–305.

[10] Pang et Al., "Cell nucleus segmentation in color histopathological imagery using convolutional networks," in *Pattern Recognition (CCPR), 2010 Chinese Conference on*. IEEE, 2010, pp. 1–5.

[11] Noh et al, "Learning deconvolution network for semantic segmentation," *arXiv preprint arXiv :1505.04366*, 2015.

[12] Yangqing et al, "Caffe : Convolutional architecture for fast feature embedding," *arXiv preprint arXiv :1408.5093*, 2014.