NUCLEI SEGMENTATION IN HISTOPATHOLOGY IMAGES USING DEEP NEURAL NETWORKS

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

[1]MINES ParisTech, PSL Research University, CBIO - Centre de Bioinformatique,
35 rue St Honoré 77300 Fontainebleau, France
[2]Institut Curie, 75248 Paris Cedex, France
[3]INSERM U900, 75248 Paris Cedex, France
[4]Service de Pathologie, Institut Curie, 75248 Paris Cedex, France

ABSTRACT

***Index Terms*—** Deep Learnnig, Convolutional Neural Networks, Nuclei Segmentation, Histpathology, 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. Indeed, this is a very promising approach for several reasons : image based approaches provide us with spatial information that is absent in genomic, transcriptomic and epigenetic studies of cancer. In addition they allow analysis of individual cells, such that it is possible to study cellular heterogeneity. Cellular heterogeneity has been acknowledged to be of major importance in cancer research and treatment. Also, it is possible to analyze cells in terms of morphology and phenotype, which brings in a complementary piece of functionally relevant information.

For these reasons, analysis of histopathology data has received much attention over the last years. Most of these works are in the frame of Computer Aided Diagnosis (CAD) and aim at automatically performing the task of a human pathologist or at assisting in the diagnosis. For such a task, the nature of the features according to which a certain classification or detection task is solved is mostly irrelevant. A conceptually different approach is to derive biologically relevant features, such as phenotype distributions, spatial cell type distributions or heterogeneity measures from the tissue data in order to reach some understanding as of which biological features can be predictive for clinical variables such as resistance to treatment or relapse probability.

In this latter case, one of the essential steps is to segment nuclei from tissue images in order to be able to assign them a cell type or phenotypic label and to evaluate their spatial distribution. We concentrate on nuclei, as they are indicative of many cellular phenotypes[1], and their morphology is currently used by pathologists in order to identify the mitotic index and the level of nuclear pleomorphism[2].

Yet, 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[3, 4].

The contribution of this paper is three-fold : (1) we have generated a ground truth of annotated images containing segmentation results for more than 2000 cells. The dataset can be used by the scientific community. (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 post-processing strategy is based on mathematical morphology and has a rigorous interpretation as to when objects are to be split.

2. RELATED WORK

In [5], they have a partially annotated data set for segmentation of histopathology images for nuclei segmentation. However their study does not sollely focus on tripple negativ breast cancer patients and not only segmentation.

Image segmentation and object recognition in medical imagining have been tackled for many years and many ap-

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

proaches have been found. Methods based on mathematical morphological operators have been widely used and applied for segmentation and feature extraction, however such methods rely on the user to manually define the information that is relevant for his task [6]. Segmentation methods based on a strong priori information are also possible and yield comparable results in terms of accuracy, in [7] they segment the picture using mathematical morphology with an ellispe fit condition which is considerably faster than other state of the art methods. Alternative methods have arisen from the achieving Convolutional neural networks [8], in particular strong incentive rose with the very impressive results that they achieved in the pascal voc challenge with [9]. Recent advances in deep neural network, and in particular in their optimization have made them become the state-of-the- art model for object recognition. Deep neural network have also been used for different task and have yielded very impressiv results, we can think of semantic segmentation problems, where [10] use "de- convolution layers" and up-sampling in order to identify and precisely locate objects within a picture. Different architecture arising from different intuition are also possible and have been applied in this paper.

3. DATASET

One of the main contributions of this paper is the now public available nuclei detection dataset within HE stainned histopathology images which can be found at website `http://cbio.mines-paristech.fr/~pnaylor/Downloads/HECellSegmentation.zip` (not a true address). This annotated dataset provides images clustered by patients. Each patient has at least 3 annotated 512×512 HE stainned histopathology images with their associated ground truth. Each ground truth image is a 512×512 image where each pixel value above 0 is considered as a pixel belonging to a nuclei. The differences in values of these pixels denote different nucleus, such an annotation can be used in several processing step which does or doesn't take into account clutered nuclei. See figure 1 for an example of three annotated images. This annotation was conducted via the help of the software ITK-snap, [11] `http://www.itksnap.org`, and were annotated by the authors of the paper.

These patients were randomly picked from an unpublished study on tripple negative breast cancer. For each of these patients we had access to their biopsy sample as a whole slide image (WSI). WSI enables a medical practitionner to digitize the huge amount of information that can be found in glass slides. WSI can be up to 60 GB big uncompressed and can't be stored in RAM on a standard computer. Given the WSI of a patient, 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 diversed dataset among these patients. Once the samples were choosen, we fully annotated each nuclei via the software ITK-snap and touching nuclei are differentiate via a different annotation value.

In this data set we have annotated a considerous 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, cell types have not been included, which means that in this dataset every annotation corresponds to a global type : cell. We can also give some simple statistics with respect to the dataset. 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, how ever each element A_l belongs to a certain patients and several elements A_l can belong to the same patient. As we are dealing with histopathology images, it is know that samples can widely vary from one patient to the other. We thus validate our model by a leave one patient out scheme. Our validation scheme is as followed, for a given set of hype 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 assess the quality of the model, the accuracy, the F1 score and a performance score which is the mean between the true positive rate and the 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 a 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 experiences, the learning rate was tunned according to the model choosen. We also experienced with different initialization value and if possible, we also considered pretrained layers. Using pretrained layers made learning more efficient and made score values more robust.

The results obtained by the Deep Neural Network were encouraging, but touching nuclei were often segmented as one single object. This is not surprising, as the error at a pixel level seems very minor, and no shape prior is used in the model. 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 larger than typically applied thresholds.This leads to the following rationale. 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 a 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 :

$$\min_{\mathcal{P}} C(\mathcal{P}) = \min_{\mathcal{P}} \{ \max_{i=2\ldots N} p_1 - p_i \} > \lambda \qquad (1)$$

where λ is a free parameter. This is actually nothing else than the morphological dynamics[12]. The actual split locations are thus obtained by applying the watershed algorithm, starting from the local maxima of the posterior probability map 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, depending on the morphological dynamic at the moment at which two regions get in contact. This algorithm will be made publicly available through the open-source software scikit-image[13], in which we implemented all the steps which were not related to deep learning.

5. DIFFERENT ARCHITECTURES/RESULTS

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 [14]. 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 archicture has the advantage of re-using a classical deep learning architecture with added on upsampling layer 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 final prediction. This model can be fine tuned with a set of pretrained-weights extracted from the classical deep learning architecture, for a more detailed explaination please refer to [10]. DeconvNet is also based on a classical architecture, however, in this network their are no skip layer as one intends to learn the proper upsampling threw repeated deconvolution and convolution layers. This model can also be fine tuned, for more information please refer to the original paper [15]. We did not ensemble classifier as suggested in [15] as our FCN and DeconvNet classifier achieve very similar scores on the test sets.

6. RESULT AND DISCUSSION

We conducted our experiments on GPUs via the caffe framework, [16]. We present our results in table 1 where we averaged the metrics other 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 we can not notice by the "holes" in the probability map when the interior of the cell is very bright. 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 watershed segmentation and we can picture the results in the third column. Our post-processing scheme is very sensitive to noise as we can see in the PangNet prediction, many unrelated minimas will lead to highly partitionned cell. 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 be found by cross validation or by simple visualation of the results.

7. CONCLUSION

	PangNet	DeconvNet	FCN
Accuracy	0.936	0.968	0.958
IU	0.759	0.856	0.857
Recall	0.744	0.856	0.855
Precision	0.741	0.858	0.878
F1	0.742	0.856	0.866
TP	0.744	0.856	0.855
TN	0.963	0.981	0.977
Performance	0.853	0.918	0.916
Pixel error	0.063	0.032	0.042

Table 1. Results

8. REFERENCES

[1] Kin-Hoe Chow, Rachel E Factor, and Katharine S Ullman, "The nuclear envelope environment and its cancer connections.," *Nature reviews. Cancer*, vol. 12, no. 3, pp. 196–209, mar 2012.

[2] C W Elston and 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, 1991.

[3] Dan Claudiu Ciresan, Alessandro Giusti, Luca Maria Gambardella, and Jürgen Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *NIPS*, 2012, pp. 2852–2860.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net : Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pp. 234–241. Springer, 2015.

[5] Elisa Drelie Gelasca, Jiyun Byun, Boguslaw Obara, and B.S. Manjunath, "Evaluation and benchmark for biological image segmentation," in *IEEE International Conference on Image Processing*, Oct 2008.

[6] 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," *Biomedical Engineering, IEEE Reviews in*, vol. 7, pp. 97–114, 2014.

[7] Petter Ranefall, Sajith Kecheril Sadanandan, and Carolina Wählby, "Fast adaptive local thresholding based on ellipse fit," in *International Symposium on Biomedical Imaging (ISBI'16), Prague, Czech Republic, April 13-16, 2016*, 2016.

[8] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

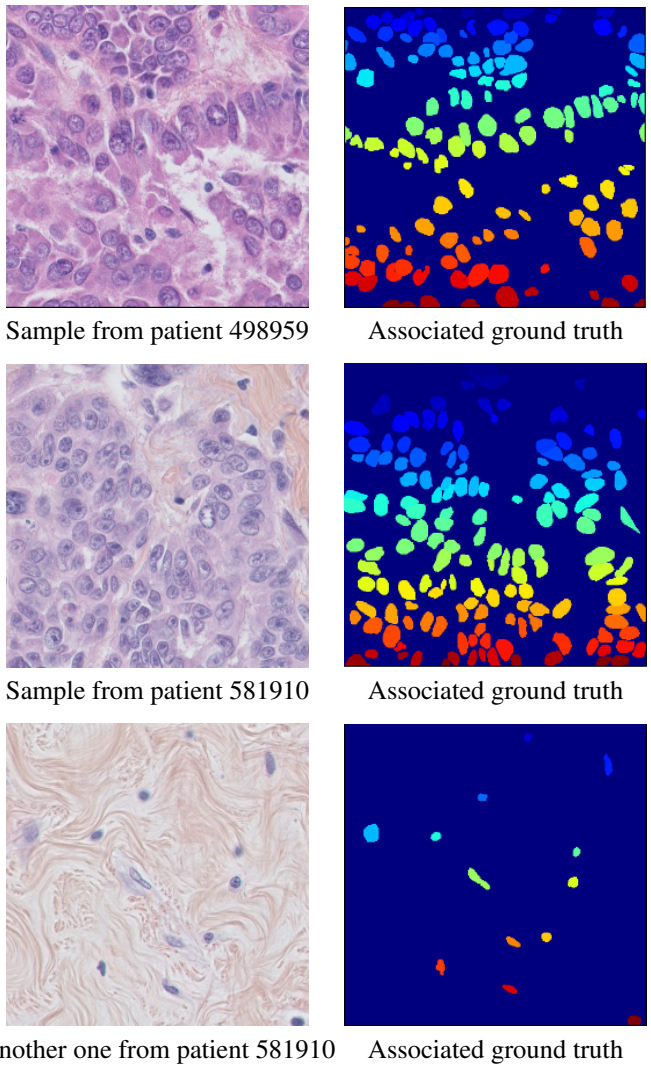

Fig. 1. Random annotated samples from the dataset

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[11] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig, "User-guided 3D active contour segmentation of anatomical structures : Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[12] Michel Grimaud, "New measure of contrast : the dyna-

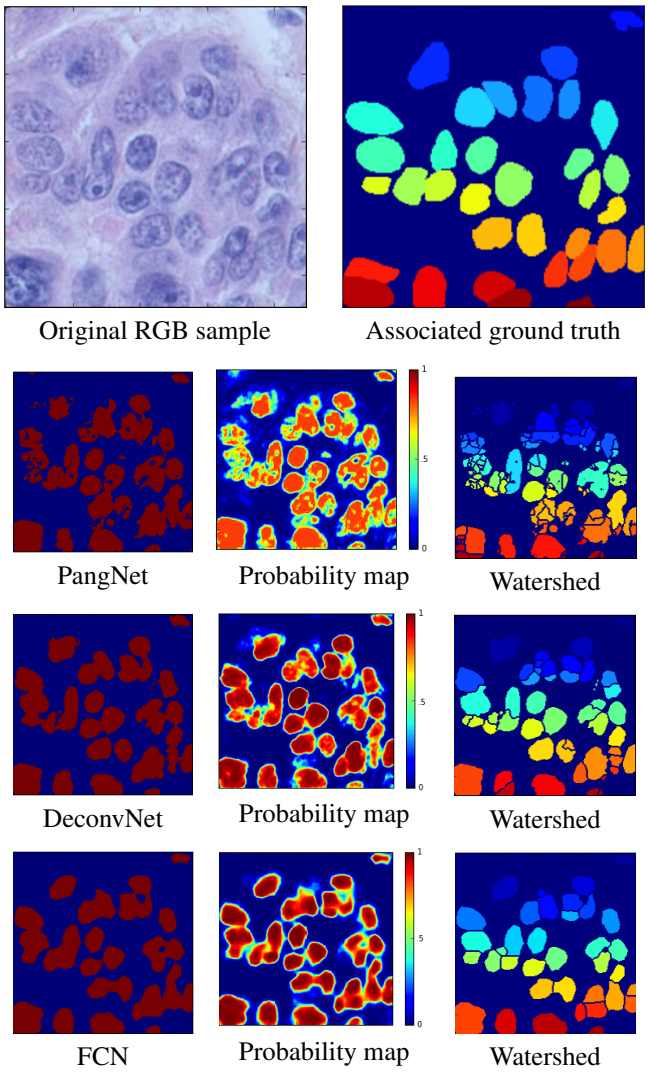

Fig. 2. Prediction via different classifiers of a random sample on the left out patient : 581910

mics," in *Proc. SPIE*, 1992, vol. 1769, pp. 292–305.

[13] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors, "scikit-image : image processing in Python," *PeerJ*, vol. 2, pp. e453, 6 2014.

[14] Baochuan Pang, Yi Zhang, Qianqing Chen, Zhifan Gao, Qinmu Peng, and Xinge You, "Cell nucleus segmentation in color histopathological imagery using convolutional networks," in *Pattern Recognition (CCPR), 2010 Chinese Conference on*. IEEE, 2010, pp. 1–5.

[15] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han, "Learning deconvolution network for semantic segmentation," *arXiv preprint arXiv :1505.04366*, 2015.

[16] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe : Convolutional architecture for fast feature embedding," *arXiv preprint arXiv :1408.5093*, 2014.