DEEP NEURAL NETWORK FOR NUCLEI SEGMENTATION IN HISTOPATHOLOGY IMAGES

Peter Naylor, Thomas Walter, J.P. Vert and E. Barillot

Insitute Curie/ Mines paristech Inserm

ABSTRACT

The abstract should appear at the top of the left-hand column of text, about 0.5 inch (12 mm) below the title area and no more than 3.125 inches (80 mm) in length. Leave a 0.5 inch (12 mm) space between the end of the abstract and the beginning of the main text. The abstract should contain about 100 to 150 words, and should be identical to the abstract text submitted electronically along with the paper cover sheet. All manuscripts must be in English, printed in black ink.

***Index Terms*—** Deep Networks, Nuclei Segmentation, Histpathology. Maybe more?

1. INTRODUCTION

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts.

2. RELATED WORK

Image segmentation and object recognition in medical imagining have been tackled for many years and many approaches 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 [1]. Segmentation methods based on a strong priori information are also possible and yield comparable results in terms of accuracy, in [2] they segment the picture using a ellispe fit which is considerably faster than other state of the art methods. Graph- based tools for segmentation are also possible in histopathology [3]. [4] focus on gland segmentation, they achieve good results by predicting image patches as a whole, via a classification scheme. The output confidence vector is then preprocessed to obtain the final segmentation. Alternative methods have arisen from the achieving Convolutional neural networks [5], in particular strong incentive rose with the very impressive results that they achieved in the pascal voc challenge with [6]. 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 very impressiv results, we can think of semantic segmentation problems, where [7] 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 Talk about related work.

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 (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 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 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 breast cells (epithelial and lobule cells), cancerous breast cells, fibroblasts, endothelial cells, fat

Thanks to XYZ agency for funding.

cells, macrophage cells and inflammatory cells (lymphocytes and plasmocytes). For the moment, cell types have not been seperated, which means that in this dataset every annotation corresponds to a global type of cell.

- Number of images: 33
- Number of cells: 2754
- Maximum number of cells in one sample: 293
- Minimum number of cells in one sample: 5
- Mean number of cells: 83
- Standard deviation of number of cells: 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 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 34, 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 (Acc), the F1 score and a performance (Perf) score which is the mean between the true positive rates and the true negative rates.

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 per patient. 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 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.

5. DIFFERENT ARCHITECTURES/RESULTS

We experience with 4 know arhitectures in semantic segmentation, one will be named BaochuanNet, Fully Convolutionnal Net (FCN), DeconvNet, UNet. The most basic architecture, BaochuanNet, consists of 4 convolutionnal layer where each convolutionnal layer has 8 feature map. 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. 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. The final architecture that we tried is the UNet. This network combines both contribution of the above networks, the UNet has a similar architecture to the DeconvNet with skip layers.

6. RESULT

	UNet	Baochuan	DeconvNet	FCN	Ensemble
Accuracy	0.89	0.94	0.97		
IU	0.70	0.76	0.86		
Recall	0.75	0.74	0.86		
Precision	0.64	0.74	0.86		
F1	0.69	0.74	0.86		
TP	0.75	0.74	0.86		
TN	0.92	0.96	0.98		
Performance	0.84	0.85	0.92		
Pixel error	0.11	0.06	0.03		

7. DISCUSSION

What's next?

8. FORMATTING YOUR PAPER

All printed material, including text, illustrations, and charts, must be kept within a print area of 7 inches (178 mm) wide by 9 inches (229 mm) high. Do not write or print anything outside the print area. The top margin must be 1 inch (25 mm), except for the title page, and the left margin must be 0.75 inch (19 mm). All *text* must be in a two-column format. Columns are to be 3.39 inches (86 mm) wide, with a 0.24 inch (6 mm) space between them. Text must be fully justified.

9. PAGE TITLE SECTION

The paper title (on the first page) should begin 1.38 inches (35 mm) from the top edge of the page, centered, completely capitalized, and in Times 14-point, boldface type. The authors' name(s) and affiliation(s) appear below the title in capital and lower case letters. Papers with multiple authors and affiliations may require two or more lines for this information.

10. TYPE-STYLE AND FONTS

To achieve the best rendering both in the proceedings and from the CD-ROM, we strongly encourage you to use Times-Roman font. In addition, this will give the proceedings a more uniform look. Use a font that is no smaller than nine point type throughout the paper, including figure captions.

In nine point type font, capital letters are 2 mm high. If you use the smallest point size, there should be no more than 3.2 lines/cm (8 lines/inch) vertically. This is a minimum spacing; 2.75 lines/cm (7 lines/inch) will make the paper much more readable. Larger type sizes require correspondingly larger vertical spacing. Please do not double-space your paper. True-Type 1 fonts are preferred.

The first paragraph in each section should not be indented, but all the following paragraphs within the section should be indented as these paragraphs demonstrate.

11. MAJOR HEADINGS

Major headings, for example, "1. Introduction", should appear in all capital letters, bold face if possible, centered in the column, with one blank line before, and one blank line after. Use a period (".") after the heading number, not a colon.

11.1. Subheadings

Subheadings should appear in lower case (initial word capitalized) in boldface. They should start at the left margin on a separate line.

11.1.1. Sub-subheadings

Sub-subheadings, as in this paragraph, are discouraged. However, if you must use them, they should appear in lower case (initial word capitalized) and start at the left margin on a separate line, with paragraph text beginning on the following line. They should be in italics.

12. PRINTING YOUR PAPER

Print your properly formatted text on high-quality, 8.5 x 11-inch white printer paper. A4 paper is also acceptable, but please leave the extra 0.5 inch (12 mm) empty at the BOTTOM of the page and follow the top and left margins as specified. If the last page of your paper is only partially filled, arrange the columns so that they are evenly balanced if possible, rather than having one long column.

In LaTeX, to start a new column (but not a new page) and help balance the last-page column lengths, you can use the command "\pagebreak" as demonstrated on this page (see the LaTeX source below).

13. PAGE NUMBERING

Please do **not** paginate your paper. Page numbers, session numbers, and conference identification will be inserted when the paper is included in the proceedings.

14. ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS

Illustrations must appear within the designated margins. They may span the two columns. If possible, position illustrations at the top of columns, rather than in the middle or at the bottom. Caption and number every illustration. All halftone illustrations must be clear black and white prints. If you use color, make sure that the color figures are clear when printed on a black-only printer.

Since there are many ways, often incompatible, of including images (e.g., with experimental results) in a LaTeX document, below is an example of how to do this [8].

15. FOOTNOTES

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced. Use Times 9-point type, single-spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

16. COPYRIGHT FORMS

You must include your fully completed, signed IEEE copyright release form when you submit your paper. We **must** have this form before your paper can be published in the proceedings. The copyright form is available as a Word file, a PDF file, and an HTML file. You can also use the form sent with your author kit.

17. REFERENCES

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [**?**].

18. REFERENCES

[1] Humayun Irshad, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A reviewcurrent status and future potential," *Biomedical Engineering, IEEE Reviews in*, vol. 7, pp. 97–114, 2014.

[2] 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.

[3] Vinh-Thong Ta, Olivier Lézoray, Abderrahim Elmoataz, and Sophie Schüpp, "Graph-based tools for microscopic cellular image segmentation," *Pattern Recognition*, vol. 42, no. 6, pp. 1113–1125, 2009.

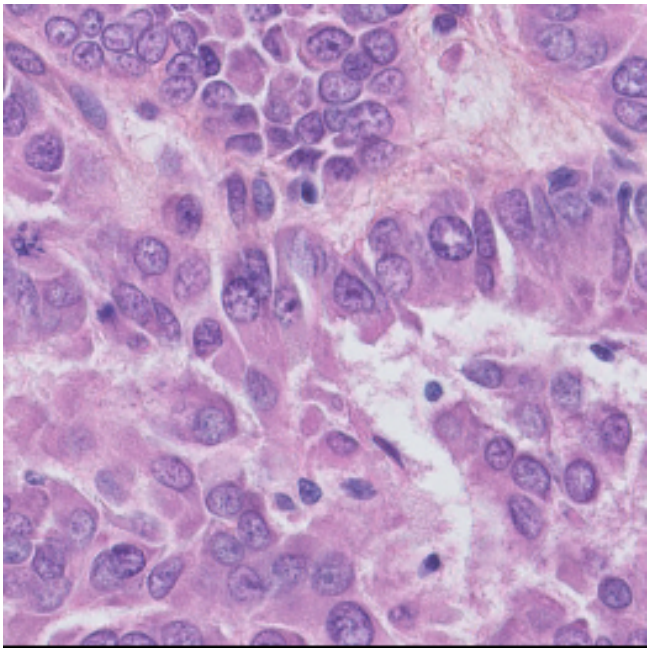
[4] Siyamalan Manivannan, Wenqi Li, Shazia Akbar, Jianguo Zhang, Emanuele Trucco, and Stephen J McKenna, "Local structure prediction for gland segmentation," in *International Symposium on Biomedical Imaging (ISBI'16), Prague, Czech Republic, April 13-16, 2016*, 2016.

[5] 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.

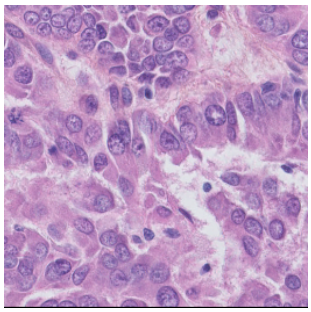
[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] 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.

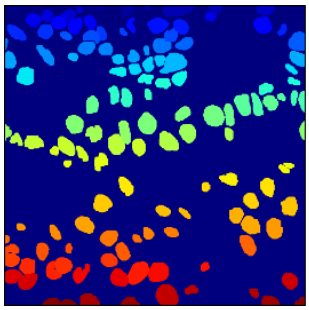
[8] A.B. Smith, C.D. Jones, and E.F. Roberts, "Article title," *Journal*, vol. 62, pp. 291–294, January 1920.



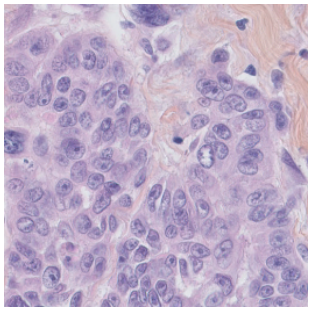
Patient 498959



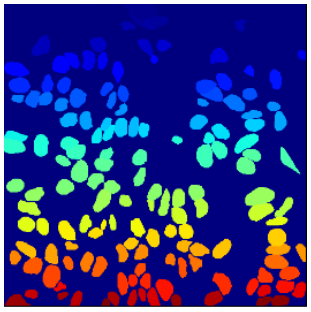
Sample from patient 498959



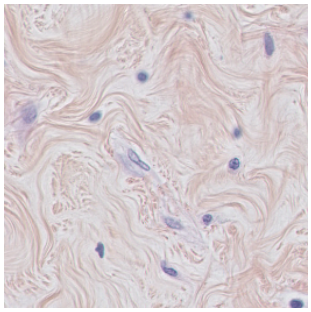
Associated ground truth



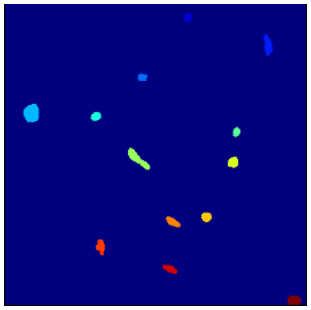
Sample from patient 581910



Associated ground truth



Sample from patient 581910



Associated ground truth

Fig. 1. Random annotated samples from the dataset