High resolution whole prostate biopsy classification using streaming stochastic gradient descent

Hans Pinckaers*, Wouter Bulten, Geert Litjens

Computational Pathology Group, Radboud University Medical Center, The Netherlands
*hans.pinckaers@radboudumc.nl

ABSTRACT

Prostate cancer is the most common cancer for men in Western countries, counting 1.1 million new diagnoses every year. The incidence is expected to increase further, due to the growing elderly population. This is leading to a significantly increased workload for pathologists. The burden of this time-consuming and repetitive workload has the potential to be decreased by computational pathology, e.g., by automatically screening prostate biopsies. The current state-of-the-art in many computational pathology tasks use patch-based convolutional neural networks. Developing such algorithms require detailed annotations of the task-specific classes on whole-slide images, which are challenging to create due to low availability of the pathologists. Therefore, it would be beneficial to be able to train using labels the pathologist already provides for regular clinical practice in the form of a report. However, these reports correspond to whole-slide images which are of such a high resolution that current accelerator cards cannot process them at once due to memory constraints. We developed a method, streaming stochastic gradient descent, to train a convolutional neural network end-to-end with entire high resolution images and slide-level labels extracted from pathology reports. Here we trained a neural network on 2812 whole prostate biopsies, at a input size of 8000x8000 pixels, equivalent to 50x total magnification, for a binary classification, cancerous or benign. We achieved an accuracy of 84%. These results show that we may not need expensive annotations to train classification networks in this domain.

1. INTRODUCTION

Prostate cancer (PCa) is the most prevalent cancer for men in Western countries.[1] Like most carcinomas, PCa develops from genetically damaged epithelium, resulting in uncontrolled cellular proliferation. Once the epithelial cells break through the basal layer they are considered invasive, i.e. carcinomatous. In low-grade tumors these epithelial cells still form glandular structures, however in the case of high-grade tumors, the glandular structures are eventually lost.[2] These morphological changes can be appreciated on histopathological slides of prostate biopsies taken during the diagnostic process. The percentage of cancerous glands in a prostate biopsy can differ substantially. For a pathologist, assessing all epithelial regions can be a time-consuming task, especially when considering the gigapixel-sized whole-slide images. This is exacerbated by the increasing incidence of PCa due to the aging of the population. An automated method to identify biopsies containing PCa can help pathologists become more efficient and deal with this increasing workload.

A straightforward approach to develop an automated method would be to train a deep convolutional neural network (CNN) to classify biopsies automatically. However, due to sheer size of whole-slide histopathology images, feeding in entire biopsies is not feasible because of memory limitations. Typically this is solved by using a patch-based system, which is trained on small samples from a whole-slide image. The labels of these samples are generally given by detailed outlines of the classes (e.g. tumor regions) by a experienced pathologist. However, this outlining is a tedious and time-consuming task which limits the dataset size for training deep networks. Several approaches exist to use slide-level labels to train a patch-based system, which circumvents the tedious outlining process. One strategy is to use multiple instance learning (MIL) approaches.[3,4] Additionally, recently, a reinforcement learning strategy has been proposed.[5] However, in both cases still

patch-based CNNs are trained, which has several disadvantages, such as the need to identify an appropriate patch size and developing a method to combine patch predictions into a slide-level classification.

In this paper we propose a novel method, using streaming stochastic gradient descent,[6] enabling the use of entire whole-slide biopsy images at high resolution directly. We show it is possible to train a convolutional neural network with the biopsy level labels this way. This technique opens the possibility to quickly gather enough data with labels from pathology reports to straightforwardly train convolutional neural networks without using MIL or reinforcement learning.

2. METHODS

We collected 442 prostate biopsy section glass slides, containing 2182 biopsies sections, from 442 patients at the Radboud University Medical Center (IRB approval 2016-2275). The slides were stained with standard hematoxylin and eosin. All glass slides were scanned using a *3DHistech Pannoramic Flash II 250 scanner* with a pixel resolution of $0.24\mu m$. As a glass slide can contain multiple biopsies, two trained non-experts outlined the individual biopsies and assigned each biopsy a label (the Gleason score or 'negative') using the original pathologist's report.

Individual biopsies were extracted at a pixel resolution of $1.92\mu m$, visually equivalent to 50x total magnification (i.e. 5x microscope objective with a standard 10x ocular lens), resulting in a total of 2182 whole prostate biopsies. The biopsies were divided into a train (1621 biopsies), validation (130 biopsies) and test (431 biopsies) set, stratifying for patient and the presence of cancer. The test set was not touched during the development of the method. The pixel size distribution of the individual biopsies is shown in Figure 1. We decided to zero-pad the images to 8000x8000 and center crop the few biopsies bigger than this size to obtain equivalent input size to the network for all biopsies.

Distribution of the morphological prostate cancer grade, the Gleason groups,[7] are shown in Figure 2. To improve clinically-relevant performance we oversampled worse prognostic Gleason scores (Gleason primary score 5) as these are typically underrepresented.

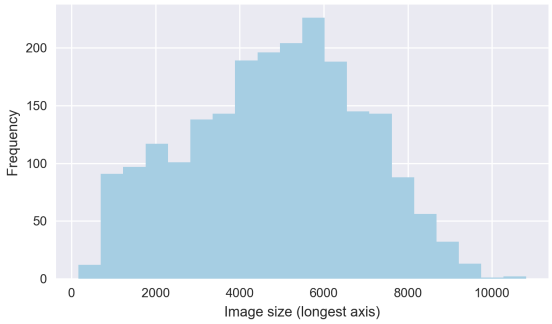

Figure 1: Histogram of image sizes (longest axis).

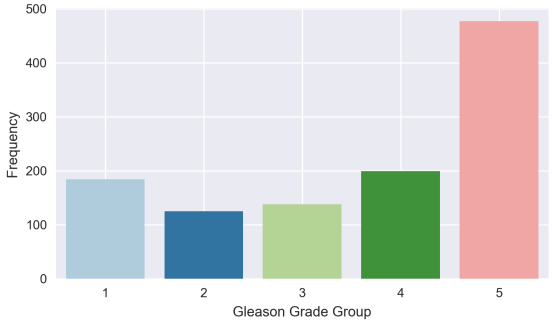

Figure 2: Histogram of Gleason Grade Groups.

Network architecture. We trained a 26-layer convolutional neural networks using StreamingSGD.[6] The network contained six blocks of three 3x3 convolutions ending with a 2x2 max-pooling layer with a stride of 2. In the first four blocks the amount of filters were doubled each block, with 16 filters in the first block. Before the final fully-connected layer the network contained a global max-pooling layer to obtain the highest feature response. Self-normalizing activation functions (SELU)[8] were used as non-linearity. The network was trained on a NVIDIA Titan V graphics card.

Table 1: Network architecture.

Block	Layers	Feature map shape
Input	Conv2d (3x3x16)	8000 x 8000 x 3
1	Conv2d (3x3x16)	7998 x 7998 x 16
	Conv2d (3x3x16)	7996 x 7996 x 16
	Conv2d (3x3x16)	7994 x 7994 x 16
	Maxpool2d (2x2)	7992 x 7992 x 16
2	Conv2d (3x3x16)	3996 x 3996 x 16
	Conv2d (3x3x32)	3994 x 3994 x 32
	Conv2d (3x3x32)	3992 x 3992 x 32
	Maxpool2d (2x2)	3990 x 3990 x 32
3	Conv2d (3x3x32)	1995 x 1995 x 32
	Conv2d (3x3x64)	1993 x 1993 x 64
	Conv2d (3x3x64)	1991 x 1991 x 64
	Maxpool2d (2x2)	1989 x 1989 x 64
4	Conv2d (3x3x64)	994 x 994 x 64
	Conv2d (3x3x128)	992 x 992 x 128
	Conv2d (3x3x128)	990 x 990 x 128
	Maxpool2d (2x2)	988 x 988 x 128
5	Conv2d (3x3x128)	494 x 494 x 128
	Conv2d (3x3x128)	492 x 492 x 128
	Conv2d (3x3x128)	490 x 490 x 128
	Maxpool2d (2x2)	488 x 488 x 128
6	Conv2d (3x3x128)	244 x 244 x 128
	Conv2d (3x3x128)	242 x 242 x 128
	Conv2d (3x3x128)	240 x 240 x 128
	Global max pool	238 x 238 x 128
Final	Fully connected layer	128
	Output	1

The network was optimized with stochastic gradient descent without momentum. The learning rate was 0.00005 with a batch-size of 16. During training, the data was augmented with random rotations, reflections, translations, and hue, brightness and saturation perturbations. The model was trained for a total of 100 epochs.

Streaming stochastic gradient descent. To obtain a result for a high resolution image an intermediate activation map of a convolutional neural network is reconstructed by doing partial forward passes with smaller parts, tiles, of the whole image up until the activation maps at a layer of choice (Figure 3). This reconstructed activation map can then be fed as a whole through the rest of the neural network resulting in a final output. This output subsequently be backpropagated through the individual tiles using partial forward passes to recover the intermediate activations. An implementation of this algorithm is available at `https://github.com/DIAGNijmegen/StreamingSGD`.

Saliency maps. We created saliency maps of the predictions by calculating the gradient of the input image with respect to the output of the network. The gradients were smoothed with a Gaussian window of with a size of 10 pixels and thresholded at 0.0001 times the maximum level. These saliency maps indicate which regions are important for the classification of the biopsy and should coincide with regions containing cancerous epithelium.

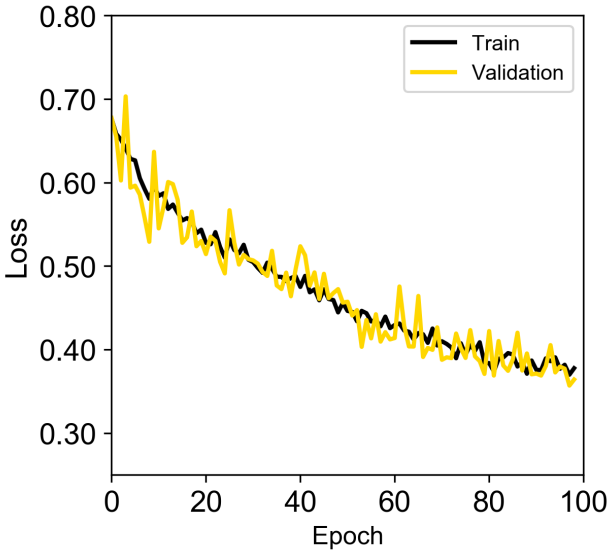

Figure 3: Train and validation loss curve.

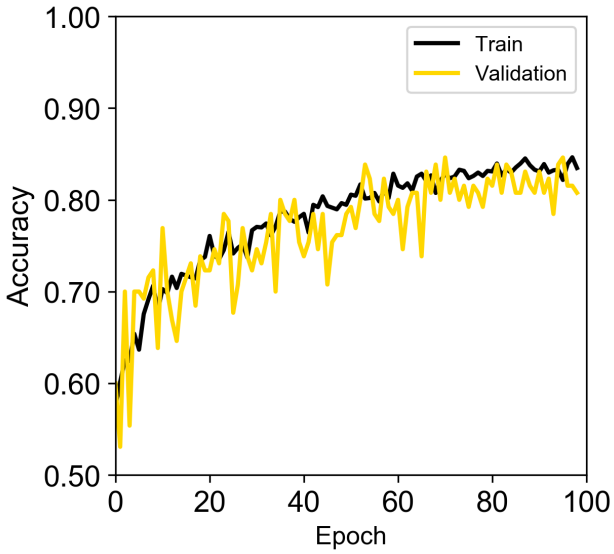

Figure 4: Train and validation accuracy curve.

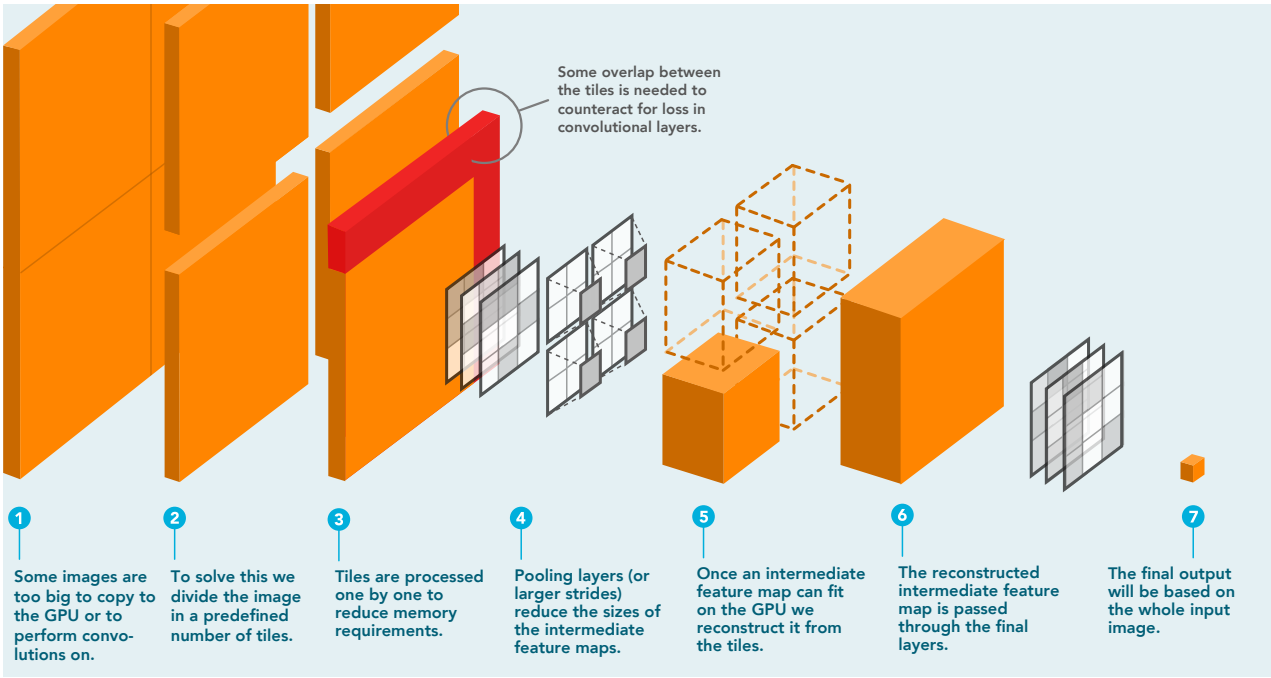

Figure 5: Schematic overview of the StreamingSGD method.

3. RESULTS

As shown in Table 1, the convolutional neural network was trained for 100 epochs and achieved accuracy of 84% on the test set (with 84% on train set) and a F1 score of 0.862. Saliency maps of true positive (Figure 6) shows highlighting of cancerous glands with adjacent stroma and no signal at normal glands. A saliency map of a false positive classified biopsy shows (Figure 6) highlights in a region with high cell density where the epithelium is forming a small glands.

Table 2: Results of the test set.

	Accuracy	**F1-score**	**Recall**	**Precision**
Test set	84.0%	0.862	0.884	0.840

Figure 6: (above) true positive, highlighting small area with cancer. (below) false positive, small gland is wrongly classified as tumorous.

4. DISCUSSION

We trained a convolutional neural network to detect prostate cancer on 8000x8000 pixel images of whole prostate biopsies at 50x total magnification. Our results show that a neural network can learn to extract relevant features from megapixel images using only a single label per biopsy.

Pathologist provide written reports to clinicians per examination using clinical diagnostic jargon. From these reports slide level labels can be extracted to quickly build large, diverse datasets. We show that these labels and slides can directly be used with StreamingSGD, as it provides the memory savings needed to make this end-to-end training possible. Training the proposed network architecture without StreamingSGD would require 377 gigabytes of memory.

For future work, we noticed that our test set accuracy (84%) was close to our train accuracy (84%), which suggests underfitting (see Figure 3 & 4). The logical next step would be training wider and deeper networks. Also, we expect the the results can be improved by using techniques such as skip connections (e.g. used in ResNet), which increase gradient flow to the lower layers of the network. Since StreamingSGD trades time for memory, a multiple GPU implementation should make it feasible to train at even higher resolutions without taking too much time.

Furthermore, saliency maps created with these networks have the potential to show regions of interest in new examinations. Recent improvements to these visualization methods[9] provide more precise maps, which might be used as surrogate segmentations.

To conclude, we use a novel method to train a convolutional neural network on 64 megapixel images of prostate biopsies. Our preliminary results show that convolutional neural networks can learn to classify cancer from a single binary label on these large images. As expected, saliency maps show the network is looking at the invasive epithelial cells. This work suggest it is worthwhile to train networks without expensive patch-level annotations.

Final details. This work has not been submitted for publication or presentation elsewhere.

REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-tieulent, and A. Jemal, "Global Cancer Statistics, 2012," *CA: a cancer journal of clinicians.*, vol. 65, no. 2, pp. 87–108, 2015.

[2] S. W. Fine, M. B. Amin, D. M. Berney, A. Bjartell, L. Egevad, J. I. Epstein, P. A. Humphrey, C. Magi-Galluzzi, R. Montironi, and C. Stief, "A contemporary update on pathology reporting for prostate cancer: Biopsy and radical prostatectomy specimens," *European Urology*, vol. 62, no. 1, pp. 20–39, 2012.

[3] P. Courtiol, E. W. Tramel, M. Sanselme, and G. Wainrib, "Classification and Disease Localization in Histopathology Using Only Global Labels: A Weakly-Supervised Approach," *arXiv preprint*, 2018.

[4] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based Deep Multiple Instance Learning," *arXiv preprint*, 2018.

[5] N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. P. Xing, "Reinforced Auto-Zoom Net: Towards Accurate and Fast Breast Cancer Segmentation in Whole-slide Images," *arXiv preprint*, 2018.

[6] H. Pinckaers and G. Litjens, "Training convolutional neural networks with megapixel images," *arXiv preprint*, 2018.

[7] J. I. Epstein, M. J. Zelefsky, D. D. Sjoberg, J. B. Nelson, L. Egevad, C. Magi-Galluzzi, A. J. Vickers, A. V. Parwani, V. E. Reuter, S. W. Fine, J. A. Eastham, P. Wiklund, M. Han, C. A. Reddy, J. P. Ciezki, T. Nyberg, and E. A. Klein, "A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score," *European Urology*, vol. 69, pp. 428–435, 3 2016.

[8] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks," *arXiv preprint*, 2017.

[9] P.-J. Kindermans, K. T. Schütt, M. Alber, K.-R. Müller, D. Erhan, B. Kim, and S. Dähne, "Learning how to explain neural networks: PatternNet and PatternAttribution," *arXiv preprint*, 2017.