AUTOMATED DETECTION AND CLASSIFICATION OF CANCER METASTASES IN WHOLE-SLIDE HISTOPATHOLOGY IMAGES USING DEEP LEARNING

F. Ghazvinian Zanjani, S. Zinger, P. H. N. de With

Electrical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

This paper presents and evaluates automatic breast cancer metastases detection in whole-slide images of lymph nodes. The classification is performed on patient level by inspecting several WSIs per patient. Every patient is assigned to one out of five pN-stages. We use convolutional neural networks for slide-level tumor detection. We found that the prediction performance improves by using test-time color augmentation and false positive bootstrapping. Also we use conditional random fields as post processing stage for improving the label assignment. The pN-stage is determined by measuring the major axes of detected blobs and the approximated number of cells inside by using blob analysis and DBSCAN clustering when needed. Our proposal shows 0.8759 quadratic weighted kappa score on Camelyon17 test set.

Index Terms— Histopathology, convolutional neural networks, conditional random field, color augmentation, Camelyon17, breast cancer detection

1. INTRODUCTION

Detecting tumor cells in microscopic examination of stained histological slides can be a tedious work for pathologists which is error-prone and depends on qualification and skills. Emerging fast digital scanners producing Gigapixel pathology images from examined tissue, are a major driving force for research in Computer-Aided Diagnosis (CAD) within automation of cancer detection. Designing such a CAD system close to human performance can be a highly challenging task because of the large variability in morphology, color patterns of stained slides and inter-patient visual contents. The CAD system can assist the pathologists by comprehensive evaluation of Whole-Slide Images (WSIs) in short time. Processing of WSIs can be considered as first-stage processing in case of detection of tumor cells, or as second-stage processing in case of examining multiple WSIs per patient for categorizing the stage of cancer. The main challenge for designing such a CAD system lies in precisely detecting tumor cells. Given the complexity of the data, a supervised Convolutional Neural Network (CNN) can be the best choice, since it showed outstanding results on image classification and segmentation tasks [1],[2] and especially on histopathology cancer detection [3]. In our study, we also use CNN as pixel classifier. We found that training the network on different color profiles of stained slides, which are collected from different medical centers, can be a good alternative for the color normalization approach [4]. In our method, we train the CNN with original images and their color-augmented versions. For decreasing the uncertainty of network prediction, we use test-time color augmentation, so that the test samples resemble color patterns of different medical centers and we accept the less uncertain output of the CNN. For assigning a binary label to the obtained probability output of the network, and for enhancing the label assignment by considering not only the local visual content of the input image, we use conditional random fields (CRF). They are defined as a fully connected graph on whole image pixels [5]. For determining the pN-stage of patients, we perform blob analysis and clustering of detected positive regions, if needed.

2. METHODOLOGY

The block scheme of our method is shown in Figure 1, of which the modules are explained in this section.

2.1. Pre-processing and data augmentation

For detecting non-relevant empty regions of a slide that do not contain examined tissues, the image is segmented into two regions. For doing so, there are some straightforward methods, e.g. by finding a global Otsu's threshold [8] and some more complex methods, like finding super-pixels of the image and training a classifier on features extracted from the super-pixels [6]. In this paper, we apply Otsu's adaptive image thresholding, but computed locally inside a sliding window. This allows us to find a threshold value for every pixel in the image by considering its local neighbors. The obtained threshold map is then smoothed and clipped by considering the global Otsu's threshold. At the end, for avoiding missing foreground pixels, we add a small bias value to the computed threshold map that benefits the

detection of false positives. This preserves all regions needed for further processing. For the reducing the computation time, all the calculations are performed on down-sampled WSIs.

Data augmentation is widely applied in training a deep CNN for addressing the problem of having a limited number of training samples. For training the CNN, we use a combination of flipping the input samples in two directions and rotating them by 90, 180 and 270 degrees. These rotation angles are chosen for fast matrix calculations. For tackling the problem of high color variation of stained images, a solution typically consists of normalizing the color of all slides to have the same color distribution as a reference slide [4]. Consequently, after training the network on the color-normalized images, the inference should be done on normalized test images. An alternative approach can be training a CNN to learn different color variations of training slides and boosting this learning by adding color-augmented samples to the original training set. We have adapted the second approach. Color augmentation in the training phase is performed by adding a random offset to hue, saturation and brightness signals, often referred to as channels. Furthermore, the brightness is scaled randomly.

2.2. CNN classifier

We used the well-known patch-based CNN classifier [7] for image segmentation. The patches are randomly extracted for both classes to train the network. The labels of marginal patches which contain partially two classes and more than 75% of one class, have been assigned to the dominant class and the rest of marginal patches have been ignored in the training set. Similarly to Wang *et al.* [8], we apply the GoogleNet [9] model as pixel classifier for detecting tumor cells in WSIs. We use the Inception v3 model [10], which is an extension to the preceding version of GoogleNet.

2.3. False positive bootstrapping

Because of computation time and balancing the classes or reducing the information redundancy in the training set, it is common to use a random small subset of possible extracted patches for training the network. This naive random selection of patches leads to the situation that the whole input space of training data is not explored by the network, hence the generalization performance of learning decreases. A possible solution can be finding candidate regions by examining some features which are computed from image patches and then selecting the patches based on it, but this needs elaborating a set of features that represent effectively the data complexity. We choose a straightforward approach: false positive bootstrapping [11].

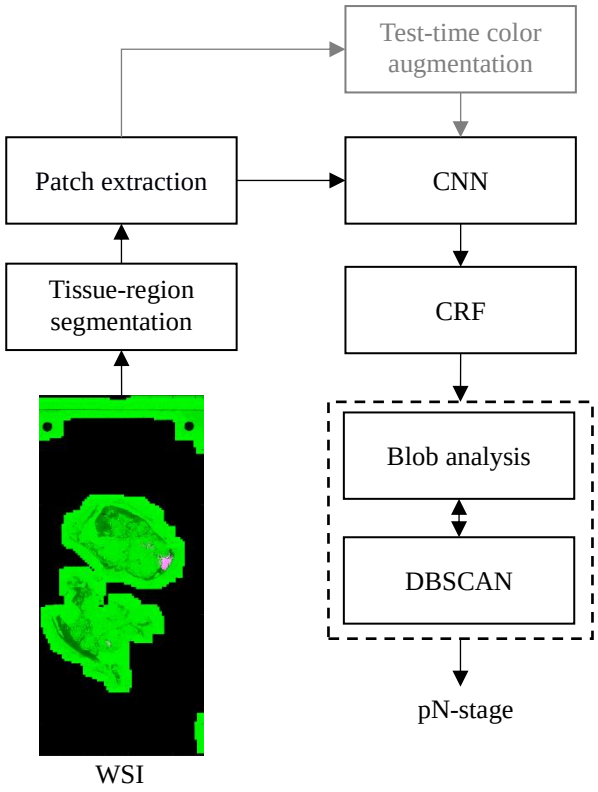

Fig 1. Schematic of our proposed model

We first train the models on randomly extracted patches from the training set and then we classify all the patches in the training set. After finding the falsely predicted samples, we add the false positive patches to the previously trained patches and by using the previously optimized parameters of the network, we fine-tune them on the modified training set. We have observed that this bootstrapping approach significantly increases the model prediction precision, leading to a small drop in recall only. In bootstrapping, we only use false positives because the population of negative samples is much higher than positive samples and this causes pruning the negative samples in the process of balancing the two classes.

2.4. Test-time color augmentation

After training the CNN on color-augmented training samples, we expect that the network learns a wide range of color varieties in data space. For getting a lower inference uncertainty of the network, we transform the color distribution of test samples to resemble the color distributions of different medical centers and accept a less uncertain prediction (based on network output probability). We have found that this technique, which we call test-time color augmentation, increases the prediction accuracy for

the post-processing stage. Based on advantages of HSD above the HSI color space for analyzing the stained tissue [12], we first transformed the test RGB image patches to HSD space similar to [4], then we modify the chromatic distribution of the input samples to be similar to multiple medical centers from which the data is collected (Figure 3). Our approach has two main differences compared to [4]. First, we do not intend to normalize the color characteristics of input slides for training the classifier, but we train the classifier to learn the diversity of color patterns in input space. Consequently, we augment the test patches with different color profiles and we select the less uncertain prediction outcome. Second, in color-conversion steps, we do not detect the nuclei individually in the image, as was done in [4] by using Hough transform. Instead, we proceed with using the color-deconvolution method [13] by converting the RGB-color space to hematoxylin-eosin-DAB space. Then by imposing an adaptive threshold on these channels, we obtain the ROI for computing hematoxylin, eosin and background color distribution from the HSD-color space similar to [4] (see also Figure 2).

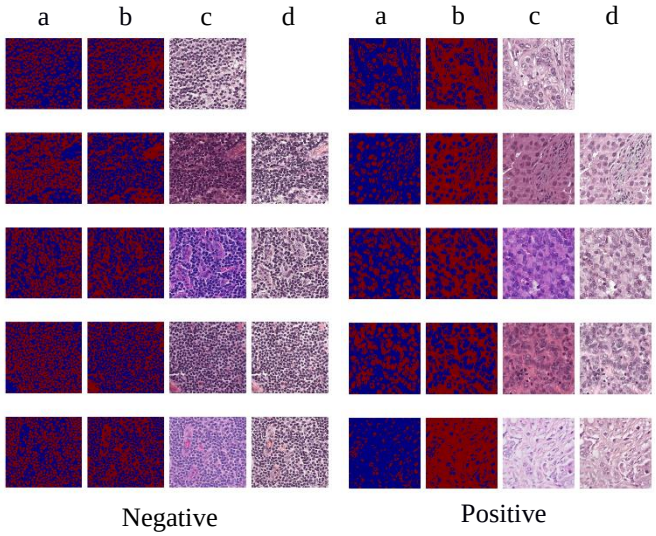

Fig. 2. Example of a color conversion for samples from five medical centers (in rows) to resemble Center 1, with (a) hematoxylin mask, (b) eosin mask, (c) original patches, and (d) color-converted patches

2.5. Conditional Random Field

Applying CRF for improving image segmentation is a well-known method [5],[14],[15]. Performing inference on WSIs with lower stride can produce a more precise probability map and consequently a better prediction. However, a lower inference stride needs more computational cost. Using CRF as a post processing stage can be applied to the up-sampled probability map, which is obtained by the pixel classifier and then doing inference on the high-resolution image. Our experiments show that using CRF can improve label assignment and enhance the prediction performance.

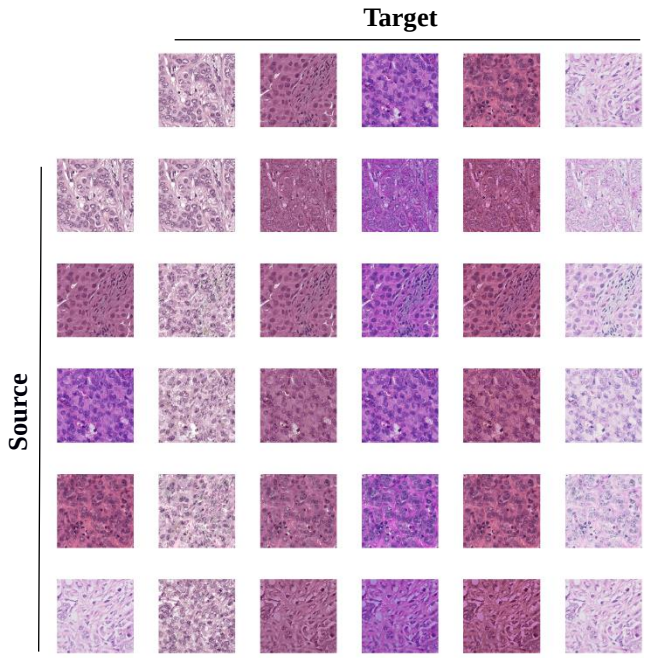

Fig 3. Example of color conversion between five medical centers

2.6. pN-stage labeling

After detecting the metastases regions and isolated tumor cells (itc) in a slide, we need to classify the WSI into four classes (normal/itc/micro/macro). The normal WSI does not contain any metastases or itc. The difference between the three other classes comes from the size of the tumor region or the number of cells that lie inside the detected blob. We classify the itc and micro classes by measuring the length of the major axis of a fitted ellipse to the tumor blob. Also, we need to count the number of cells inside the region. According to definition of the micro metastases, the number of cells inside the tumor blob should be larger than 200 cells, in spite of having a larger length than 0.2 mm. For doing so, instead of exact cell counting, we use the foreground area of the hematoxylin mask as a rough estimation of the cell count. The differentiation between the micro and macro classes is more straightforward and only based on measuring the length of major axis and considering a 2-mm threshold between them.

According to the problem definition, some isolated blobs that are not connected but densely distributed close to each other, should be clustered and considered as a one unified

larger blob. Since there are no explicit definition and criteria for clustering in this definition, it can be somewhat subjective. According to observing our data, we just need to evaluate the presence of such a cluster when several itc blobs in the vicinity of each other have been detected. We use the density-based algorithm for discovering clusters (DBSCAN) [16]. Furthermore, after clustering, we evaluate the distance between the center of blobs within a cluster and their dimensions for validating the clustering result.

3. EVALUATION

The evaluation of our method is performed on Camelyon17 test data.

3.1. Histopathology image dataset

We use the Camelyon16 and Camelyon17 datasets, consisting of a total of 400 and 1000 WSIs, respectively. Camelyon16 provides and reveals region annotations for all its positive samples. The samples are collected from two institutions. The Camelyon17 dataset provides and reveals 50 slides with region annotations and the rest of the training set is labeled by the lymph-node class and pN-stage level per patient. The Camelyon17 samples are collected from five different medical centers. The labels of the test set are unrevealed for the Camelyon17 competition. For training of the classifier, we used the whole Camelyon16 and training set of Camelyon17.

3.2. Experiments and results

The evaluation of result in Camelyon17 challenge is the five-class quadratic weighted kappa, where the classes are the pN-stages. We retrained Inception v3 with initial parameters trained on the ImageNet 2012 Challenge dataset, by changing the softmax layer for two output classes with one-hot encoding of target vectors. We changed the learning rate by monitoring the loss function value on the validation set. The batch size is equal to 32 and weight decay has been used for penalizing large values for the parameters and prohibiting the overfitting problem.
We achieved 98.7% accuracy for patch classification on our validation set and it increased to 99.5% after false positive bootstrapping of the training set. The reported quadratic kappa for our proposal is equal to 0.8759 on the test set.

4. DISCUSSION AND CONCLUSIONS

In this paper, we have explained the features of our method that we applied for the Camelyon17 challenge, to determine the pN-stage of breast cancer in WSIs. We have used a CNN for detecting the tumor cells as pixel classifier. For better training of the network, we used false positive bootstrapping and color augmentation. We found that test-time color augmentation can help for reducing the prediction uncertainty. Also, we have applied CRF as post-processing stage for enhancing label assignment. Since decision on type of metastases depends on the tumor major axis and the number of cells inside the blob, we used the estimated hematoxylin mask, blob analysis and DBSCAN clustering.

5. REFERENCES

[1] Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. Adv. in Neural Inf. Process. Syst. pp. 1097–1105 (2012)

[2] Long, J., et al.: Fully convolutional networks for semantic segmentation (2015)

[3] Wang, Dayong, et al. "Deep learning for identifying metastatic breast cancer." arXiv preprint arXiv:1606.05718 (2016).

[4] Bejnordi B. E. et al. "Stain specific standardization of whole-slide histopathological images." IEEE transactions on medical imaging 35.2 (2016): 404-415.

[5] Krähenbühl, Philipp, and Vladlen Koltun. "Efficient inference in fully connected crfs with gaussian edge potentials." Advances in neural information processing systems. 2011.

[6] Bejnordi B. E. et al. "Automated detection of DCIS in whole-slide H&E stained breast histopathology images." IEEE transactions on medical imaging 35.9 (2016): 2141-2150.

[7] Ciresan, Dan, et al. "Deep neural networks segment neuronal membranes in electron microscopy images." Advances in neural information processing systems. 2012.

[8] Wang, Dayong, et al. "Deep learning for identifying metastatic breast cancer." arXiv preprint arXiv:1606.05718 (2016).

[9] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[10] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[11] Rowley, H.a., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI) 20(1), 23–38 (1998)

[12] van der Laak, Jeroen AWM, et al. "Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy." Cytometry 39.4 (2000): 275-284.

[13] Ruifrok, Arnout C., and Dennis A. Johnston. "Quantification of histochemical staining by color deconvolution." *Analytical and quantitative cytology and histology* 23.4 (2001): 291-299.

[14] Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." arXiv preprint arXiv:1606.00915 (2016).

[15] Rajchl, Martin, et al. "Deepcut: Object segmentation from bounding box annotations using convolutional neural networks." IEEE transactions on medical imaging 36.2 (2017): 674-683.

[16] Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." Kdd. Vol. 96. No. 34. 1996.