

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

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 lymph nodes whole-slide images (WSIs). The detection is performed in slide-level and patient-level processing. The pN-stage for every patient is determined by the number of positive lymph nodes that consists of 5 categories. We used convolutional neural networks for slide level detection of tumor cells. We found that by using test-time color augmentation and false positive samples bootstrapping, the prediction improves significantly. The pN-stage evaluation has been done as post processing stage for detected positive regions by using blob analysis and *DBSCAN* clustering. We have evaluated our approach on our validation set and unrevealed test data of Camelyon17 dataset.

Index Terms— Pathology, convolutional neural networks, 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 prone to unquantified error. Computer-aided detection (CADe) can yield a great deal by empowering the pathologists with comprehensive evaluation of WSI in short time. Processing of WSI can be consider as low-level processing of histopathology slides by tumor cells detection and determining the metastases or as high level processing by examining multiple samples per patient for categorization of the stage of cancer. For finding the tumor cells, we used convolutional neural networks, a powerful machine learning approach as (super)pixel classifier. For determining the pN-stage of patients, we used blob analysis and clustering of detected positive regions.

2. METHOD

The schematic of our method has been shown in figure 1. We are going to explain its different modules in this section.

2.1. Preprocessing and data augmentation

For detecting the non-relevant empty regions of a slide which do not contain examined ductal carcinoma in situ (DCIS), first we need to segment the image into two regions. Although, recently some fancy methods has been proposed for this task [1], we used local adaptive thresholding of image for finding a close to optimum threshold for each patch in image followed by using a relaxation to smooth the threshold map. After computing the threshold map, we deliberately add a small bias values to it in favor of false positives. This helps to decrease the uncertainty for missing foreground because missing any regions at this stage will be left for the preceding processing modules. For the sake of computational time, all the computations have been done on down-sampled images.

Data augmentation has been used for tackling with the over-fitting problem. The patches have been flipped in two directions and have been rotated by 90, 180 and 270 degrees on fly in learning phase. This rotation angles have been chosen for their fast matrix calculations. Also we found that the color augmentation has a great impact on training the models for different WSIs which provided with different color characteristics. The color augmentation has been used not only for the over-fitting problem but also for learning the variability of contrasting dyes in histological staining in different medical centers. This approach can be consider as a different perspective to the methods that try to normalize and standardize the WSI's color space [2].

Color augmentation in training phase has been done by rotating the hue channel in HSV cylindrical-coordinate by adding a random offset. Saturation and brightness (value channel) have been changed with a random offset. Furthermore, the brightness has been scaled randomly. Figure 2 shows an example of color augmentation for input image patch.

2.2. ConvNet patch classification

As a well-know approach for segmenting the images, we used patch-based Convolutional Neural Networks (ConvNets) classifier [3]. While the WSIs which we are working with, have been stored in 512 tiles, the size of RGB patches sets to 256x256. We randomly extracted about 600K patches for both classes for training the network. The labels of marginal patches which contain partially both classes and more than %75 of one class have been assigned

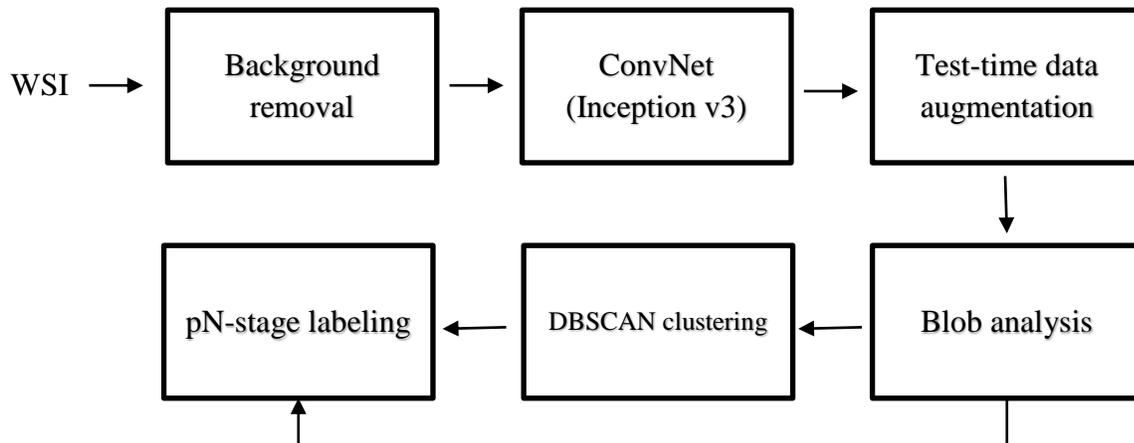


Fig. 1. Schematic of model in prediction mode

to the dominant class and the rest of marginal patches have been ignored in trainset. Recently, Wang et al. [4] have used GoogleNet [5] for detecting tumor cells in WSIs. In similar approach, we used GoogleNet but the inception v3 model [6] as our patch image classifier.

2.2. Test-time color augmentation

We computed predictions across different color augmented versions of the input RGB test images and then we select the most likely prediction (based on network output probability) as the final prediction outcome. We found that this techniques that we called test-time color augmentation increases the prediction accuracy. Based on superiority of HSD to HSI color transformation for analyzing the stained tissue [7], we first transformed the test RGB image patches to HSD space similar to [2], then we modified the chromatic distribution of the input samples to be similar to multiple medical centers which the data have been collected. For each center, we estimate the color distribution on both positive and negative regions individually. For example, if the dataset contains sample of five different medical centers, we generate ten different color distributions (two per center) and the input test samples transform (augments) to have a similar chromatic distribution to each of those. Our approach has two main differences to [2]. First we don't intend to normalize the color characteristic of input slides for training the classifier but we train the classifier to learn 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, we do not consider the color pattern of hematoxylin, eosin and background, individually but the positive and negative regions separately. So we don't need to detect nuclei in WSIs.

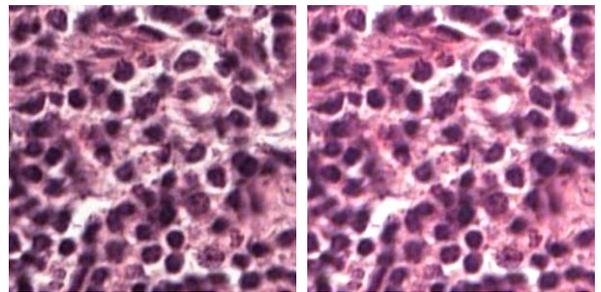


Fig. 2. Example of training patch and it color augmented version

2.3. False positive bootstrapping

Because of computational time issues, balancing the classes or reducing the information redundancy in trainset, people usually use a random small subset of possible extracted patches for training the network. This naive random selection of patches causes the whole input space of training data is not explored by the network, hence the generalization performance of learning will decrease. One common solution can be finding candidate regions by examining some features which computed from image patches and then selecting the patches more wisely but this needs elaborating some features that represents effectively the complexity of data. In another approach that is more straightforward, we used false positive bootstrapping [8]. We first trained the models by randomly extracted patches from training set and then we classify all the patches in training set. After finding the false predicted samples, we add the false positive patches to the previous training patches and by using the previous optimized parameters of the network, we fine-tuned them on modified train set. We observed that this

bootstrapping approach significantly increase the model prediction precision by a bit drop in recall. In bootstrapping, we just used false positives because the population of negative samples are much higher than positive samples and this causes pruning the negative samples in the process of balancing the two classes.

2.4. pN-stage labeling

After detecting the metastases regions and isolated tumor cells (itcs) in slide, we need to classify the WSI into four classes. The negative class does not have any itc or metastases. The slides in itc class have small tumor cell blobs with less than 0.2 mm axis length or less than 200 cells. The metastases with greater size or cell number than what is defined for itc, account for micro metastases which have less than 2mm major axis length. Consequently, the metastases with major axis length greater than 2 mm account for macro class. For estimating the major axis of the tumor blobs, we fit an ellipse to each individual blobs and we measure the major axis of the ellipse and the diameter of a circle with the same area as the region. According these two scaler, we initially label the tumor blobs. Because apart from the above definition the distribution of the tumor blobs in WSI is important for considering them as separate or a unified big blob, we need to consider some clustering concept for classifying between itc and micro or between itc and macro classes. For this purpose, we used density-based algorithm for discovering clusters (DBSCAN) [9]. Of course, we utilized this clustering method just for some special cases which have several itc blobs, close to each others. Furthermore, after clustering, we evaluate the distance between the center of blobs within a cluster and their dimension for merging them and validating the clustering outcome.

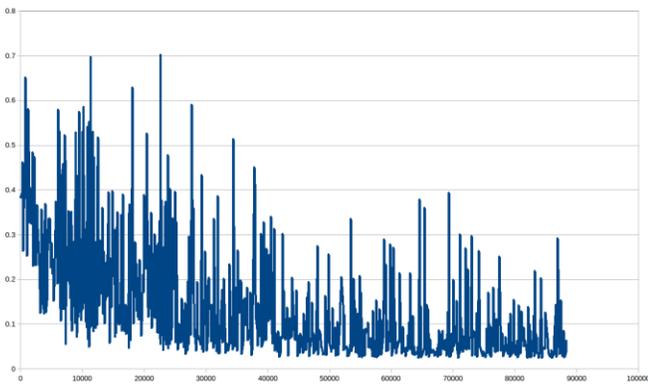


Fig 3. Training loss

3. EVALUATION

The evaluation of our method has been done on unrevealed Camelyon17 test data. Because the ground truth is not

released at the time that we are writing this paper, we just tested the method on a small validation set.

3.1. Histopathology image dataset

We used Camelyon16 and Camelyon17 dataset consist of a total 400 and 1000 WSIs, respectively. Camelyon16 provides and reveals region annotations for all its positive samples. The sample are collected from two institutions. Camelyon17 provides and reveals 50 slides with region annotations and the rest of training set have been labeled in lymph node class and pN-stage level per patient. The Camelyon17 samples collected from five different medical centers. The labels of the test set has unrevealed for Camelyon17 competition.

For training of the classifier, we used the whole Camelyon16 and training set of Camelyon17.

3.2. Experiments and results

While the ground truth of test set of Camelyon17 are unrevealed at time that we report this paper, we evaluate our models on our small validation set. The evaluation of result in Camelyon17 challenge are the five class quadratic weighted kappa where the classes are the pN-stages.

We retrained inception v3 with initial parameters trained on 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 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 to %98.7 accuracy for patch classification on our validation set and it increased to %99.5 after false positive bootstrapping of training set. The training loss values is shown in figure 3.

Our initial evaluation on validation set shows kappa score of about 0.8 on slide level. We did not compute the kappa score on patient level for our validation set because it has less useful feedback for our algorithm development. Although, it will be reported by organizer after submission our result on test set.

4. DISCUSSION AND CONCLUSION

In this paper we explained the models that we used for Camelyon17 challenge for determining pN-stage of breast cancer in WSIs. We used ConvNet for detecting the tumor cells and for evaluating the expansion of cancer regions in WSI, we used DBSCAN clustering when needed. We have found that color augmentation and false negative bootstrapping can increase the prediction performance significantly. In different method to color normalization but

with similar concept, we found that test-time color augmentation can improve the performance.

5. REFERENCES

- [1] 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.
- [2] Bejnordi B. E. et al. "Stain specific standardization of whole-slide histopathological images." *IEEE transactions on medical imaging* 35.2 (2016): 404-415.
- [3] Ciresan, Dan, et al. "Deep neural networks segment neuronal membranes in electron microscopy images." *Advances in neural information processing systems*. 2012.
- [4] Wang, Dayong, et al. "Deep learning for identifying metastatic breast cancer." *arXiv preprint arXiv:1606.05718* (2016).
- [5] Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
- [6] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
- [7] 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.
- [8] 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)
- [9] Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." *Kdd*. Vol. 96. No. 34. 1996.