PATIENT-LEVEL CLASSIFICATION ON BREAST CANCER METASTASES IN LYMPH NODES

Kyuhyoung Choi and Sun Woo Kim

DeepBio

ABSTRACT

In this paper, we describe the process of estimating pathologic N-stage (pN-stage) of a patient from his/her HE stained slides of lymph nodes, which is the task of Camelyon17 challenge. For tumor region detection, we developed patch-based classifier, that is, a ResNet-like network. The number of edge pixels in a patch is counted to distinguish effective tissue patch from the background. Based on the patch-level evaluation, a confidence map (a.k.a heatmap) is generated after smoothing. To classify each slide, the heatmap should be thresheld properly to give tumor blobs from which features as input to SVM (support vector machine) are extracted. Since the goal is to achieve maximum Cohen's Kappa score (κ), we directly optimize the threshold and feature list by minimizing the cost as $-\kappa$ using PSO (particle swarm optimization).

***Index Terms*—** Breast cancer, deep neural net, SVM, PSO

1. INTRODUCTION

The task of this year's Camelyon challenge is to estimated the pN-stages of 100 patients for each of which five slides of lymph node are given. To make the challenge feasible, the simplified form of pN-state is used and it is categorized into five classes, namely, pN0, pN0(i+), pN1mi, pN1, pN2. [1] For five sides of lymph nodes, the classes are defined as following.

- pN0: No micro-metastases or macro-metastases or ITCs found.
- pN0(i+): Only ITCs found.
- pN1mi: Micro-metastases found, but no macro-metastases found.
- pN1: Metastases found in 13 lymph nodes, of which at least one is a macro-metastasis.
- pN2: Metastases found in 45 lymph nodes, of which at least one is a macro-metastasis.

Here, macro and micro metastases, and ITC(isolated tumor cells) are all defined by the size(length) of the tumor region as following.

- Macro-metastases: metastases greater than 2.0 mm.
- Micro-metastases: metastases greater than 0.2 mm or more than 200 cells, but smaller than 2.0 mm.
- ITC: strictly not a metastasis, but is rather defined as: single tumour cells or a cluster of tumour cells smaller than 0.2 mm or less than 200 cells.

So if we can classify each slide of lymph node correctly, best estimation of pN-stage is sure to follow. Since the size of tumor region matters, it is important to compute exact segmentation of tumor blobs and the threshold if it should be done from a heatmap. Again best heatmaps are generated from best patch-classifier if it a patch-based is used. In the following sections, we will describe our patient-level classification process in sequence.

2. POI EXTRACTION

Ideally PoI(patch of interest) should be that of tissue. However, it is hard to find one measure to distinguish tissue patches from the others since the slides are prepared from five different medical centers and it seems that . (In total, it is seven since we also used the dataset of Camelyon16 [2] which is from two different medical centers.) The criteria for PoI extraction is following

- When it is applied to ground truth tumor slides, any of the tumor region should not be missed
- Apparently non-tissue region should be excluded as much as possible

After trying a few methods to extract effective tissue patches, we decide to just take patches of enough edge pixels as those of tissue which is the last item of the following list of methods we tried

- Otsu thresholding on Hematoxylin channel : First convert RGB channels to HE channels. Then apply Otsu threshold with a proper threshold.

- Union of Otsu thresheld masks of hue and saturation channels. : This seems to be the method of [3]. [1]
- Intersection of Otsu thresheld masks of hue and saturation channels.
- Thresholding on RED channel : A patch is considered to be of tissue if its average RED value is above a threshold.
- Edge pixel count on a patch : A patch is taken as POI if there is enough edge pixels in it.

At first, the third of the above methods seemed to be very promising. However, after examining the resulted masks, we found that much of tissue region is missing in some of total 1400 slides (400 from Camelyon16 + 1000 from Camelyon17). This happens especially for the cases of Otsu thresholding when the major part of slide is not tissue. For example, when the fat is so dominant in a slide, Otsu thresholding gives very unstable results. Figure 1 shows the PoIs extracted by the above mentioned methods on the slide of $tumor\_077$. Note that the pink inner regions of tissue are missing.

3. PREPARATION OF PATCH SET AS INPUT TO DEEP NEURAL NET

The size of patch to classify is 240 $\times$ 240. Even though, the tumor region annotation is not provided for all the tumor slides and even some of them are not annotated exhaustively, it will be too large image set if we sample the patches from all the 900 slides (400 from C16 + 500 from C17). Before the release of C17 training dataset, we trained our deep neural net with almost 200,000 image patches sampled from 170 slides (C16 training set). Then we added more patches by hard-example mining on the 130 slides of C16 testset and 500 slides of C17 training set. In other words, we collected the false positives and false negatives from those slides and the final number of patches were 442,544. Among them, 274,815 patches belonged to $Normal$ and 336,976 patches went into training set. Figure 2 shows the example of how the $Normal$ and $Tumor$ patches are sampled from slides and annotation masks. Note that blue contour is annotation mask and yellow, blue, red and green squares stand for originally sampled normal, originally sampled tumor, hard-example mined normal, hard-example mined tumor patches respectively.

4. DEEP NEURAL NET AS PATCH CLASSIFIER

Our neural net is modified from original ResNet [4]. As shown in Figure 3, the biggest difference from the vanilla one is that ours have two main branches, one for small region (240 $\times$ 240) and the other for bigger region (720 $\times$ 720). [2] We trained two neural nets, one by retraining the already made one and the other from the scratch. We chose the retrained one since it reached to the convergence earlier. Figure 4 shows the plots of loss, precision and recall versus epoch for the retrained. Both of final precision and recall was around 0.9

5. HEATMAP AND TUMOR BLOBS

With the trained net, a slide is scanned in a sliding-window way to assign each tile the probability of being tumor. This results in a heatmap (confidence map) on which slide level classification is performed. Actually the heatmap is smoothed and scaled to give gray scale image. Figure 5 shows such an example of the scaled heatmap.

To measure the size of tumor region, the heatmap should be thresheld to give the binary image of tumor blobs. So it is important to compute the optimal threshold. Actually, just estimating the class of a slide by measuring the biggest tumor blob is very ideal case. Since there are many false positive and false negative tiles due to various reasons including noises in PoI extraction and neural net training, measuring the biggest blob is not enough.

6. OPTIMIZATION ON THRESHOLD AND FEATURES

Since just measuring the biggest tumor blob is risky, we need a classifier. To train a classifier, we need a set features. Since it is all about the size of blobs it is certain that blob size/area should be included in the feature. We decided to have a feature composed of confidence histogram of N bins and length/area of B biggest blob. As mentioned above, we need to optimize the confidence threshold T for the heatmap as well as N and B. We take the Cohen's Kappa score as the cost of minimization. We used PSO (particle swarm optimization) [5], a well-known population based minimization method. The lower and upper bounds for T, N, B was $[0.9, 1], [0, 10]$ and $[1, 10]$ respectively. Integer constraints on N and B were posed. Figure 6 shows the how particles moved through the generations by projecting each 2D dimension. The computed optimal values for T, N, B was 0.995, 8 and 3 and the best Kappa score was 0.62.

7. REFERENCES

[1] Oscar Geessink, Péter Bándi, Geert Litjens, and Jeroen van der Laak, "Camelyon17: Grand challenge on cancer metastasis detection and classification in lymph nodes," 2017.

[1] It was not clear when they said "the final mask images are generated by combining the masks from H and S channels". So we tried both union and intersection

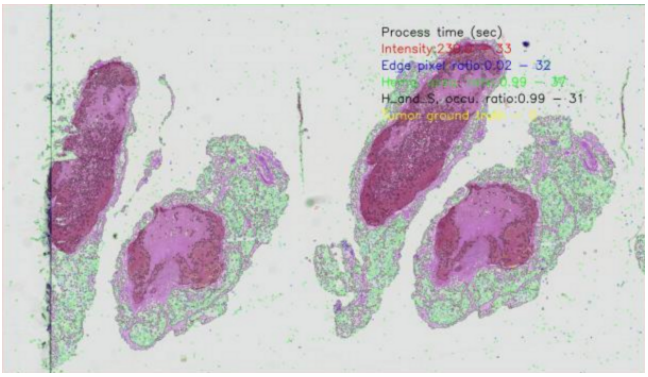
[2] Therefore, the actual image size of sampled patches is 720 $\times$ 720

[2] Babak Ehteshami Bejnordi and Jeroen van der Laak, "Camelyon16: Grand challenge on cancer metastasis detection in lymph nodes," 2016.

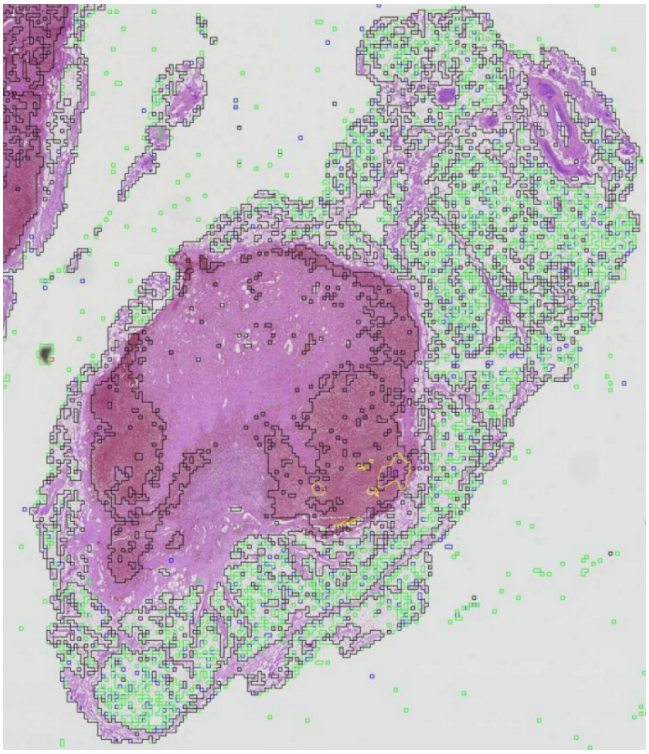
[3] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck, "Deep learning for identifying metastatic breast cancer.," *CoRR*, vol. abs/1606.05718, 2016.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[5] James Kennedy and Russell Eberhart, "Particle swarm optimization," 1995.



(a) Various POI extractions



(b) Zoomed subimage of (a)

Fig. 1. PoI extraction example

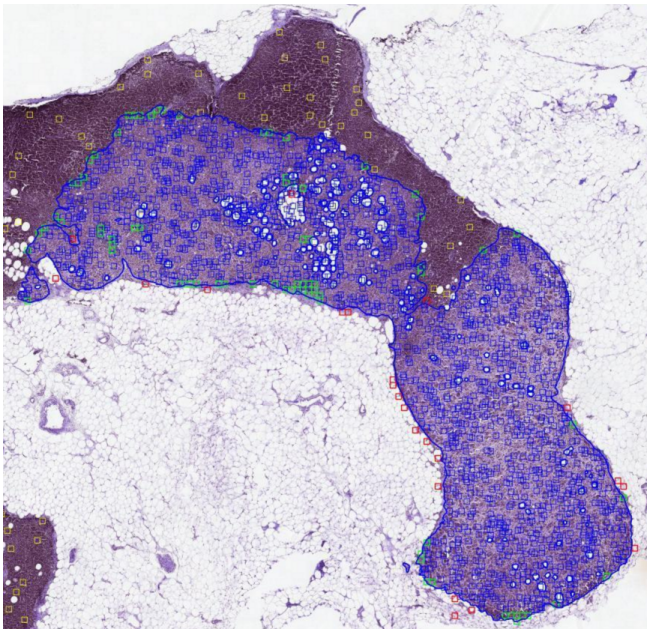


Fig. 2. Patch sampling

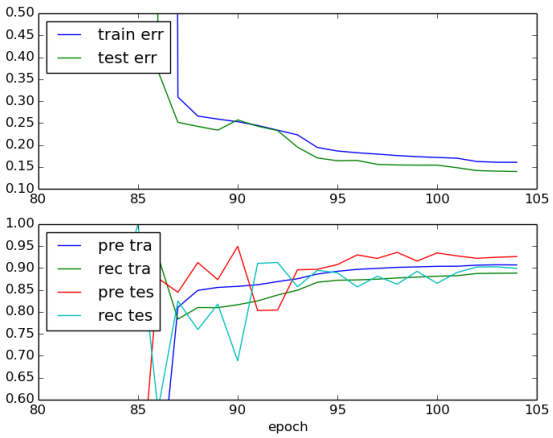


Fig. 4. Loss, precision and recall vs. epoch

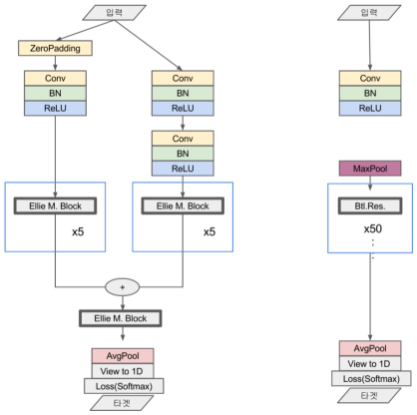


Fig. 3. Our net (left) vs. ResNet (right)

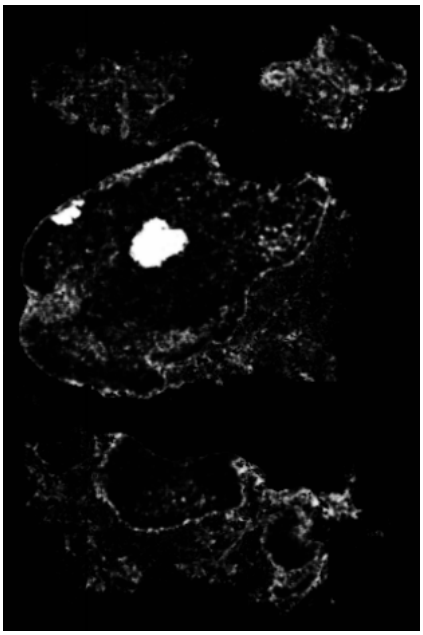
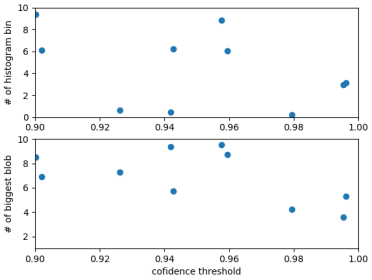
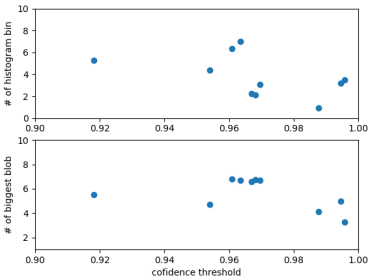


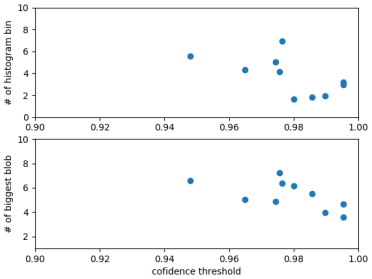
Fig. 5. Heatmap



1st generation



8th genearation



16th generation

Fig. 6. population change along the generation

This second submission differs from the first one in the following point.

We use optimization on the confidence threshold, the number of bins for confidence histogram, the number of biggest blobs.

For the 1st submission, the cost of optmization is -Kappa score.

For the 2st submission, the cost of optmization is average F-score.