BREAST CANCER STAGE CLASSIFICATION IN HISTOPATHOLOGY IMAGES

Byungjae Lee and Kyunghyun Paeng

Lunit Inc., Seoul, Korea

ABSTRACT

Predicting TNM stage is the major determinant of breast cancer patient's prognosis and treatment. The essential part of TNM stage classification is whether the cancer has metastasized to the regional lymph nodes (N-stage). Pathologic N-stage (pN-stage) is commonly performed by pathologists detecting metastasis in histological slides. However, this diagnostic procedure is prone to misinterpretation and time-consuming. Automated detection of lymph node metastasis and pN-stage prediction has a great potential to reduce their workload and help the pathologist. In this paper, we present a framework to automatically predict pN-stage from whole slide histopathology images. pN-stage is predicted by combining convolutional neural network (CNN) based metastasis detector and slide-level lymph node classifier module. Our framework is evaluated on Camelyon17 which is recently introduced challenging benchmark dataset.

Index Terms— Camelyon17, Convolutional neural network, Deep learning, Metastasis detection

1. INTRODUCTION

The TNM stage system [1] is widely used to classify the magnitude of cancer spread which is a significant component of breast cancer control and surveillance. The essential part of TNM stage classification is whether the breast cancer has spread to the regional lymph nodes (N-stage) since lymph nodes are the first place breast cancer is likely to metastasize. N-stage is commonly determined by metastasis detection which is performed in lymph node histological slides. However, the diagnostic procedure examined by pathologists to detect metastases is prone to misinterpretation and time-consuming and tedious. Automated detection of lymph node metastasis and pN-stage prediction has a great potential to reduce their workload and help the pathologist.

In the last few years, considerable improvements have been emerged in the computer vision task using convolutional neural network (CNN) [2]. Followed by this paradigm, CNN based computer assisted metastasis detection has been proposed in recent years [3, 4]. In [3], the author proposed the unified framework for tumor proliferation score prediction in breast histopathology which was used to win the Tumor Proliferation Assessment Challenge at MICCAI 2016 [5]. [4] suggested CNN based lymph node breast tumor detection framework which obtained state-of-the-art results on the Camelyon16 [6] dataset.

In this paper, we propose an automatic framework to predict pathologic N-stage (pN-stage) from patient's whole slide histopathology images. The proposed framework is conducted by integrating three modules: a region of interest (ROI) extraction module, a CNN-based metastasis detection module, and a slide-level lymph node classification module. First, ROI extraction module proposes candidate tissue regions from whole slide images. Second, CNN-based metastasis detection module predicts cancer metastasis within extracted ROIs. Third, the predicted scores extracted from ROI are converted to a feature vector based on the morphological and geometrical information which is used to build a slide-level lymph node classifier. Finally, patient-level pN-stage is determined by aggregating slide-level predictions.

2. METHODOLOGY

We introduce an efficient framework for pN-stage prediction from the patient's histological lymph node whole slide images. Figure 1 shows the overall scheme of our framework which the details are illustrated in this section.

2.1. Regions of Interests Extraction

A whole slide image (WSI) is approximately 200000 x 100000 pixels on the highest resolution level. If we deal with all regions, enormous computation time is required because of the huge size of the slide. In order to extract tissue regions from the WSIs, Otsu threshold [7] or gray value threshold [8] is commonly used in recent studies. We observed that metastasis regions are commonly located at the edge of the tissue regions. Therefore, careful tissue region extraction method is needed. We determine to use gray value threshold method which is obtained a metastasis region's sensitivity 0.9752 on Camelyon16 train set. In detail, we convert RGB to gray from 32-times down-sampled WSI and then extract tissue regions by thresholding gray value > 0.8 [8].

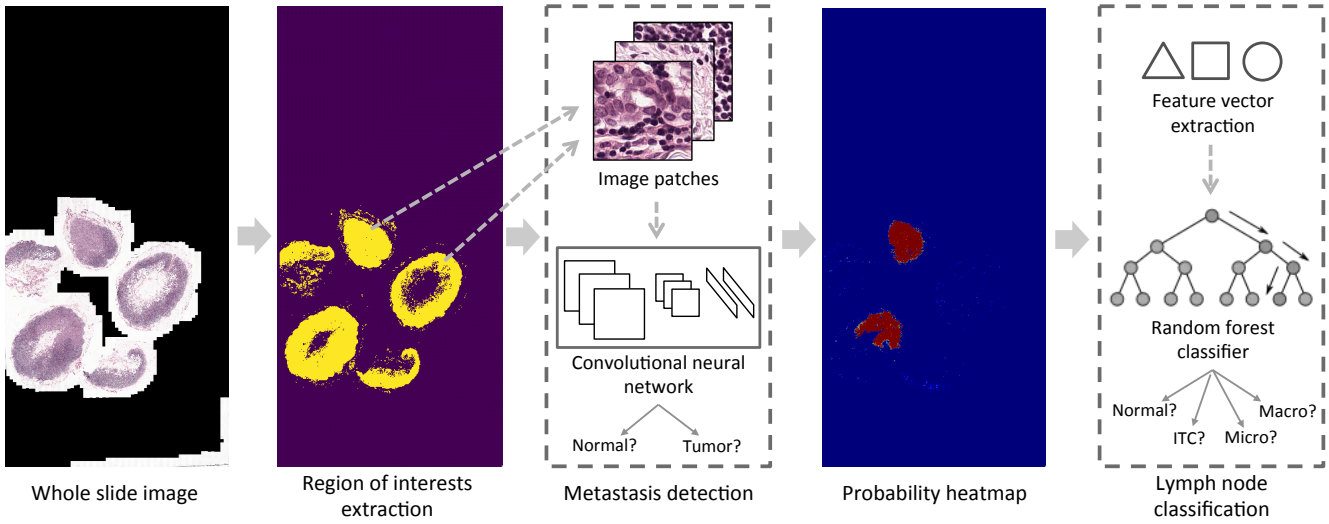

Fig. 1. Overall architecture of a pN-stage prediction framework.

2.2. CNN based Metastasis Detection

Some annotated metastasis regions include fat tissues and non-metastasis area since accurate pixel-level annotation is difficult in large size WSIs. Deep learning model becomes robust to noisy labels when a larger dataset is available [9]. Therefore, we build a large scale dataset by extracting small patches from WSIs to deal with those noisy labels. After the ROIs are founded from WSIs, we extract 256x256 patches within ROIs with stride 128. We label a patch as tumor if over 75% pixels in the patch are annotated as a tumor.

Our metastasis detection module is based on the well-known CNN architecture ResNet101 [2] for patch classification to discriminate between tumor and non-tumor patches. Training CNN model with extracted patches from WSIs is challenging because a number of extracted patches is various per WSI. To deal with this imbalance, we followed similar patch sampling approach used in [4]. In detail, we sample the same number of tumor/normal patches where patches are sampled from each slide with uniform distribution. To combat with the variety of hematoxylin and eosin (H&E) stained color because of chemical preparation difference per slide, extensive color augmentation is performed by applying random hue, saturation, brightness, and contrast. Since the classes of histopathology image exhibit rotational symmetry, we include data augmentation by randomly rotating over angles between 0 and 360, and random left-right flipping.

2.3. Lymph Node Classifier

To determine each patient's pN-stage, lymph node should be classified into four classes (Normal, Isolated tumor cells (ITC), Micro, Macro). For each lymph node WSI, we obtain the 128-times down-sampled tumor probability heatmap through the CNN based patch classifier described beforehand. Each heatmap is converted into a feature vector which is used to build a slide level lymph node classifier. We define 11 types of features based on the morphological and geometrical information. By using converted features, random forest classifier is trained to automatically classify the lymph node into four classes. Finally, each patient's pN-stage is determined by aggregating all lymph node predictions with the given rule.

3. EXPERIMENTS

In this section, we describe the details about the dataset, experiment setting, and demonstrate the performance of our framework.

3.1. Dataset

We evaluate our framework on Camelyon16 [6] and Camelyon17 [10] dataset. The Camelyon16 dataset contains 400 WSIs with region annotations for all its metastasis slides. The WSIs are collected from two different medical centers. The Camelyon17 dataset contains 1000 WSIs with 5 slides per patient: 500 slides for the train set, 500 slides for the test set. The WSIs are collected from five different medical centers.

Since the Camelyon17 dataset provides only 50 slides with region annotations, we split 100 patients (500 slides) into 43 patients for the train set, 57 patients for the validation set for hyperparameter tuning. In detail, if patient's any slide include region annotation, we allocate that patient as a train set. For training of the patch-level CNN based classifier, 400 WSIs from Camelyon16 dataset and 160 WSIs (50 region annotated WSIs and 110 negative WSIs) from Camelyon17 train set are used. For training of the slide-level lymph node classifier, we use 285 WSIs (57 patients) from Camelyon17

Table 1. Slide-level lymph node classification accuracy and patient-level quadratic weighted kappa score result on the Camelyon17 validation set

Model	Slide-level accuracy	Patient-level kappa
ResNet101 A	0.9351	0.9017
ResNet101 B	0.9284	0.9151
ResNet101 C	0.9251	0.9194
Ensemble	0.9390	0.9455

validation set. We use the Camelyon17 evaluation metric, five-class quadratic weighted kappa for the patient-level pN-stage classification. Slide-level lymph node classification accuracy is also evaluated for the validation purpose.

3.2. Experimental Setup

During training and inference, we extracted 256x256 patches from WSIs at the highest magnification level of $0.243\,\mu\mathrm{m}$/pixel resolution. We trained ResNet101 [2] with initial parameters from ImageNet pretrained model to speed up convergence. The deep learning framework Tensorflow [11] was used to train our model with a NVIDIA GTX TITAN X GPU. We used the Adam optimization method [12] with a learning rate 1e-4. The network was trained for 2 epoch using stochastic gradient descent with a batch size 32 per GPU.

For the post-processing, we thresholded the tumor probability heatmap with a threshold of $t = 0.9$. Given a heatmap, we extracted 11 features including the major axis length of the tumor region, maximum probability score, average probability score, and a total area of the tumor region. We built a random forest classifier to discriminate lymph node classes using extracted features. Finally, each patient's pN-stage was determined by the given rule with the 5 lymph node slide prediction result.

3.3. Results

We evaluated our framework on our Camelyon17 validation set with 5-fold cross validation setting. Our slide-level lymph node classification model and patient-level pN-stage prediction achieved 0.9351 accuracy and 0.9017 quadratic weighted kappa score, respectively, using a single CNN model (ResNet101 A). We trained additional models with different data augmentation hyperparameters (ResNet101 B, C). Finally, three model was ensembled by averaging probability heatmap and reached 0.9390 slide-level accuracy and 0.9455 patient-level quadratic weighted kappa score.

4. CONCLUSION

We have introduced a fully automatic framework to predict pN-stage from histopathology images, using CNN based metastasis detection and random forest based lymph node classification. We demonstrated competitive performance on the Camelyon17 dataset. In future work, we would like to build an end-to-end learning framework for pN-stage prediction from WSIs.

5. REFERENCES

[1] Leslie H Sobin, Mary K Gospodarowicz, and Christian Wittekind, *TNM classification of malignant tumours*, John Wiley & Sons, 2011.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[3] Kyunghyun Paeng, Sangheum Hwang, Sunggyun Park, Minsoo Kim, and Seokhwi Kim, "A unified framework for tumor proliferation score prediction in breast histopathology," *arXiv preprint arXiv:1612.07180*, 2016.

[4] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al., "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017.

[5] "Tumor proliferation assessment challenge 2016," `http://tupac.tue-image.nl/`, Accessed: 2017-07-24.

[6] "Camelyon 2016," `https://camelyon16.grand-challenge.org/`, Accessed: 2017-07-24.

[7] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[8] Andrew Janowczyk and Anant Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of pathology informatics*, vol. 7, 2016.

[9] David Rolnick, Andreas Veit, Serge Belongie, and Nir Shavit, "Deep learning is robust to massive label noise," *arXiv preprint arXiv:1705.10694*, 2017.

[10] "Camelyon 2017," `https://camelyon17.grand-challenge.org/`, Accessed: 2017-07-24.

[11] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[12] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.