

PATIENT-LEVEL CLASSIFICATION FOR BREAST CANCER METASTASES

Zixu Zhao[†], Huangjing Lin[†], Hao Chen^{†,§}, Pheng Ann Heng[†]

[†]Department of Computer Science and Engineering, The Chinese University of Hong Kong

[§]Insight Medical Technology, China

ABSTRACT

Patient-level classification is an essential task for breast cancer patient’s diagnosis and treatment. In this paper, we propose a method to automatically predict pN-stage for whole slide images by leveraging multi-scale features and path aggregation. We first train a slide-level classifier with hard case mining to generate heatmaps of each WSI. Next, features extracted from those heatmaps are used to train a random forest classifier by rule-based criteria. pN-stages of patients are finally determined by our prediction results.

Index Terms— Camelyon17, Deep learning, pN-stage Classification, Breast Cancer, multi-scale features

1. METHODS

The overall algorithmic steps of our method are: preprocessing, slide-level detection, feature extraction, and patient-level classification.

- (1) Preprocessing: non-informative regions on the WSIs are removed for efficiency.
- (2) Slide-level detection: we train a deep learning model based on ScanNet [1] with hard negative mining. We use 400 WSIs from Camelyon16 dataset [2] (including 208 slides with annotations) and 215 WSIs of 43 patients from Camelyon17 dataset [3] (including 50 slides with annotations) for training our model.
- (3) Feature extraction: we threshold the heatmaps we obtained and extract several features from the predicted cancerous area.
- (4) Patient-level classification: we train a random forest classifier based on features from each heatmaps.

1.1. Data Augmentation

To increase the variation in the training set, we perform extensive data augmentation including random flipping, scaling, rotation, and cropping. In addition, we perform color jittering and HSV augmentation to make our model robust to color variation caused by differences between scanners and staining protocols.

1.2. Slide-level Classifier Training

We propose a modified ScanNet as the slide-level classifier. Our model contains a classification branch and a segmentation branch. To handle with itc cases and control false positives, the segmentation branch is used to refine the classification result. We train the two branches at the same time with patch size being 692×692 . During the whole training process, we do hard case mining three times to find out the hard cases, e.g., false positive and boundaries, and add them to the training pool.

In the prediction phase, our model can process a WSI hundreds of times faster than patch-based classification methods with the same stride by leveraging the merit of ScanNet, which allows us to feed patches with a size up to 2228×2228 into our model for inference.

1.3. Feature Extraction

Each heatmap is converted into a feature vector after being thresholded with 5 threshold value, i.e., 0.4, 0.5, 0.6, 0.7, 0.8. Next, 10 types of features (e.g., major axis and tumor region) are extracted from each thresholded heatmap according to the morphological and geometrical information.

2. RESULTS

We test our model on 100 slides (20 patients) each time. The average slide-level accuracy of our model is 0.937, and average kappa score is 0.924.

3. REFERENCES

- [1] Huangjing Lin, Hao Chen, Qi Dou, Liansheng Wang, Jing Qin, and Pheng-Ann Heng, “Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide images,” 03 2018.
- [2] “Camelyon 2016,” <http://camelyon16.grand-challenge.org/>, 2017, [Online; accessed 24-July-2017].

[3] “Camelyon 2017,” <http://camelyon17.grand-challenge.org/>, 2017, [Online; accessed 24-July-2017].