

BREAST CANCER NODAL STAGING USING DEEP LEARNING

Adrian Levine¹, Jasleen Grewal², Colin Schlosser², Steve Jones², Stephen Yip¹

1. Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

2. Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada.

ABSTRACT

Several recent papers have demonstrated the potential for deep learning to improve the accuracy and efficiency of pathologic diagnosis, particularly the identification of cancer metastases in lymph nodes. In this paper we present our framework for the computer-aided staging of breast cancer axillary nodal dissections. Our method uses a fully convolutional neural network for pixel level segmentation of whole slide images. This generates tumor probability heatmaps, from which a vector of quantitative image features is classified using an extremely randomized trees classifier.

Index terms – deep learning, convolutional neural network, breast cancer, lymph node staging, whole slide imaging

1. INTRODUCTION

Accurate nodal staging is highly important for guiding management of breast cancer, however there is evidence of significant variability in the identification of metastatic deposits by pathologists [1]. Building on the recent advances in image analysis by convolutional neural networks, the CAMELYON 2016 challenge was a landmark work in medical applications of machine learning, demonstrating that computer algorithms can equal the performance of human experts in identifying metastatic cancer in lymph nodes [2]. Despite this success, extensive validation must still be undertaken prior to deploying these algorithms clinically. Subsequent work using models trained on the CAMELYON 2016 dataset have been validated on

external datasets [3] and demonstrated the ability to improve pathologist accuracy and efficiency in a simulated clinical workflow [4]. The follow-up CAMELYON 2017 competition further extends this work to the pN staging of artificial patients made up of sets of five whole slide images [5]. While the results from the initial submissions to this competition have recently been published, the results leaderboard has been kept open to incentivize further advances in the field. In this paper we present our submission to the ongoing CAMELYON 2017 leaderboard.

2. METHOD

2.1 Data

The CAMELYON 2017 training set contains 100 simulated patients with five slides each, of which 43 patients have one or more slides with annotations. We used all 50 tumor slides with annotations, as well as all 113 normal slides from these patients for training the classifier. Tumor slides from these patients without annotations were excluded from classifier training. The CAMELYON 2016 dataset contains 399 slides, with 270 in the training set for the 2016 challenge (110 tumor, 160 normal), and 129 in the test set, for which the annotations have now been released. We used all slides in the tumor and test sets. A total of 882,011 image patches were exported, which were divided into train and validation splits as outlined in Table 1.

For training and evaluating our slide level classifier, only the 500 slides from the CAMELYON 2017 training set were used. All 215 slides from the 43

patients with annotations were used for training the random forest classifier, and the remaining 285 slides from 57 patients were used for validation.

Table 1: Image patches exported for classifier training

| Slide set | Training | | Validation | |
|-----------|----------|---------|------------|---------|
| | Tumor | Normal | Tumor | Normal |
| 2016 | 70,002 | 148,369 | 0 | 0 |
| 2017 | 6284 | 421,206 | 3353 | 232,807 |

2.2 Pre-processing

Tissue regions were identified by converting the whole slide image on magnification level five to grayscale and applying Otsu thresholding. Tissue patches of size 768 x 768 were then exported, along with the accompanying annotation mask region, down-sampled by eight-fold to a size of 96 x 96. For patches that only contained normal tissue, the annotation mask would contain all zeros.

For image augmentation, only random horizontal and vertical flips were used. We experimented extensively with other image augmentation methods, including color jitter in the RGB and HSV spaces, other affine transformations, and elastic deformations, and found that these did not improve performance.

2.3 Patch based classifier

Our patch-based classifier was a modified fully convolutional version of the Inception V3 convolutional neural network (CNN) [6], pre-trained on Image net. We experimented with numerous other base models, including ResNets, Xception, Densenets, and Deeplab, and found that the Inception V3 model had the best performance. The final fully connected layer of the network was removed, and a top block was added, which contained upsample layers and several layers of depth-wise separable convolutions [7]. The output of the model was a two-class segmentation mask, downsampled by 8 compared to the input dimensions.

The CNN was trained using stochastic gradient descent with momentum 0.9 and an initial learning rate of 0.01, which was decreased by half for every epoch. Training was run on a Nvidia Tesla V100 GPU using Keras [8] with Tensorflow [9] backend. The models were trained for 4-5 epochs, which took several days.

2.4 Slide level classification

A slide level heatmap of tumor probabilities was generated with the trained classifier using a “sliding window” approach, in which each 768 x 768 image patch that was part of the tissue region was run through the classifier and made up a 96 x 96 patch in the resulting heatmap. The heatmap was binarized by thresholding at 0.5 and 0.9. At each of these thresholds, connected components were extracted and properties were measured using the Scikit-image regionprops function. A feature vector of length 13 was generated at each threshold using the following features:

- Slide level: maximum pixel value, total tumor area, tumor area/tissue area, total tumor probabilities/tissue area
- Largest connected component: area, major axis length, eccentricity, extent, solidity, perimeter, bounding box area, maximum pixel intensity, and mean pixel intensity

The resulting feature vectors of length 26 were classified using an extremely randomized trees classifier in Scikit-learn. In order to improve reproducibility between runs, ten classifiers were trained and their predictions were averaged. pN stages were then determined using the decision rules. From the CAMELYON 2017 dataset, 215 slides were used to train the classifier and the remaining 285 slides for validation.

3. RESULTS

On our patch level validation set during classifier training, our best classifier achieved a pixel level accuracy of 0.996, Dice score of 0.963, intersection over union of 0.958, and mean per class accuracy of 0.984. Given the significant class imbalance, with many more image patches containing normal tissue than tumor, we found that pixel level accuracy was a

poor metric, and the Dice and IOU correlated better with overall slide classification.

For slide level classification, our best model achieved 0.915 accuracy, and for patient level nodal staging the best model had a kappa of 0.922.

4. DISCUSSION

In this paper we have described our approach to cancer identification in lymph nodes, using a modified fully convolutional Inception V3 network, with a top block composed of depth-wise separable convolutions. Compared with typical image classification models with fully connected top layers, we found that the use of a fully convolutional model allowed us to input larger image patches while maintaining fine resolution of the resulting heatmap, and was less prone to overfitting. As has been previously reported in this challenge, our approach performed well with macro- and micro-metastases, but had particular difficulty in identifying isolated tumor cells.

REFERENCES

1. Vestjens, J.H. et al. (2012) Relevant impact of central pathology review on nodal classification in individual breast cancer patients. *Ann Oncol* 23 (10), 2561-6.
2. Ehteshami Bejnordi, B. et al. (2017) Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* 318 (22), 2199-2210.
3. Liu, Y. et al. (2018) Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection. *Arch Pathol Lab Med*.
4. Steiner, D.F. et al. (2018) Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol* 42 (12), 1636-1646.
5. Bandi, P. et al. (2018) From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. *IEEE Transactions on Medical Imaging*, 1-1.
6. Szegedy, C. et al. (2015) Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9.
7. Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv:1610.02357*.
8. Chollet, F. (2015) Keras. <https://keras.io>.

9. Abadi, M. et al. (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from www.tensorflow.org.