A HIERARCHICAL AND ENSEMBLE FRAMEWORK FOR PATIENCE'S PN-STAGE PREDICTION

Wei Liu, Yaohua Wang and Yu Chen

JD.com Inc, Beijing, China

ABSTRACT

The pathologic N-stage (pN-stage) prediction can tell whether cancer has spread to the regional lymph nodes. However, this prediction need analysis on whole-slide image which is very huge and time consuming to detect the tumor by human. In this paper, we introduce the deep learning progress recent years in computer vision into tumor recognition and put forward an automated hierarchical and ensemble framework to predict patience's pN-stage. Our method is evaluated on Camelyon17 and achieved 0.913 Kappa Score on test set.

Index Terms—GoogLeNet, Random Forest, hard example mining, Metastasis Detection

1. INTRODUCTION

The TNM system [1] is widely accepted to classify the extent of cancer spread in patient. The pN-stage in this system shows the presence of metastases in lymph nodes which has important implications for breast cancer patients. However, detection of tumor by human is time consuming and subjective. Automated detection and classification procedure can reduce the workload and gain objective results.

Camelyon17 [2] is a challenge that using whole-slide image (WSI) with pixels annotated. The main task of this challenge is to determine the pathologic N-stage for every patient in the test dataset which means whether the cancer has spread to the regional lymph nodes.

Deep learning is a popular method in computer vision. It can learn a representation of image from large data with which classifier can distinguish the category easily.

Our work is using deep learning method on whole-slide images and classify each slide into four categories: Normal, Isolated tumor cells (ITC), Micro-metastases or Macro-metastases, and then predict the patience's pN-stage according to the given rules.

Our main contributions on this task are (1) we make morphological operation, dilation [3], on the extracted tissue region to raise the recall rate for the color on the edge is light (2) we flip and rotate each input patch and average the output probability when patch level model inference (3) we not only train one stage classifier for patches but also make hard example mining [4] to generate training dataset for the second stage classifier and at last combining the two stage results together for the final result.

2. METHODOLOGY

In this section, we introduce a hierarchical and ensemble framework to deal with the whole slide image of patient's histological lymph node for pN-stage prediction. Hierarchy makes the system efficient and ensemble makes it robust. We use two level models to predict the patient's pN-stage hierarchically, and integrate two training period models with combination 8 kinds of image augmentation results to gain robustness. Figure 1 shows the complete process of our framework and we will discuss it in detail in this section.

2.1. Tissue Region Extraction

A whole slide image is so huge while the tissue regions are only parts of it. In order to focus on the tissue regions and accelerate processing, we use Otsu [5] threshold method to extract the tissue regions. Otsu method can give us the threshold which we use do image segmentation according to the pixel statistics between foreground and background. However, segmentation result from Otsu method is not as good as we expected. As is shown on the first image in Figure 1, we found that hematoxylin and eosin (H&E) stained color on the edge of tissue region is light so these regions tends be missing while segmentation. In order to recall these areas, we firstly find the contours of the tissue regions and seek the connected domain [6], and then dilate the connected domain. We use the dilated result as segmented tissue region.

The second image in Figure 1 is an example of extracted tissue region result.

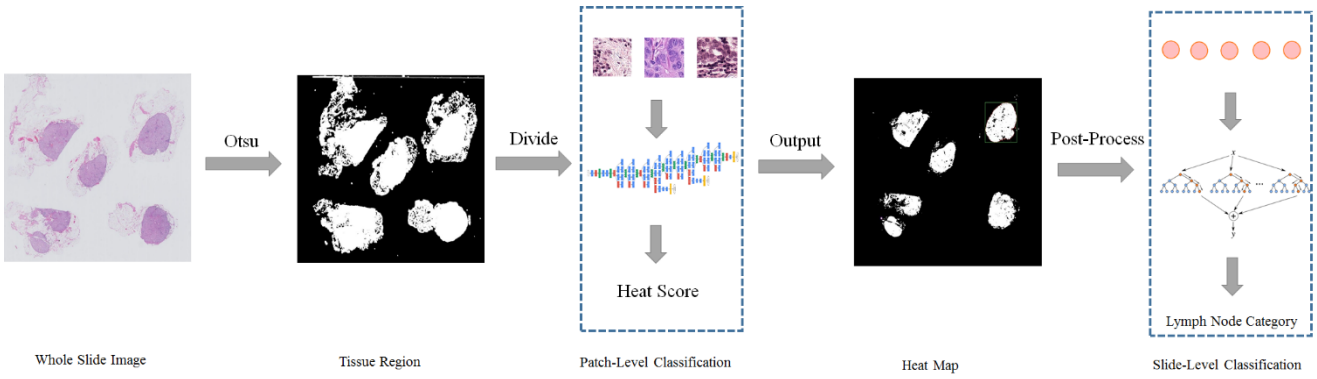

Fig. 1 Complete workflow of our framework on pN-stage prediction

2.2. Patch-Level Model

2.2.1. Patch division

In order to gain fine-grained features, we divide the tissue regions on WSI into a lot of 256*256 patches with stride of 128. Since the tumor region is rare, we label one patch as tumor if the pixels annotated as a tumor in this patch exceed 25%. With these patches, we gain enough samples to train a robust convolution neural network [7] (CNN) for tumor pattern recognition.

2.2.2. Data selection and augmentation

In the consideration of the balance of performance and speed, we choose GoogLeNet [8] as the patch classifier. During training period, we sample same tumor and non-tumor patches from one slide no matter how many tumor this slide has for that tumor cells may vary with each individual and the H&E stained color may also be slightly different because these slides are from different medical centers. For data augmentation, we randomly rotate the patch 0, 90, 180 and 270 degrees and randomly flip the patch left, right, up and down.

When inference, we also do the same operation on each patch so we get 8 probability for one patch. We average the 8 result for the final result of this stage CNN.

2.2.3. Model ensemble

However, GoogLeNet can only achieve 97.2% accuracy on these training data. In order to enhance the model's classification ability, we do hard example mining on the training data using the first stage classifier. We combine the original model and enhanced model together by averaging the final result of each CNN model as the final probability of this patch to be tumor. With this method, our model gain 1.9% improvement.

The third image in Figure 1 shows the procedure of patch-level prediction.

2.3. Slide-level Model

2.3.1. Feature extraction

With the output of patch-level model, we can get each slide's probability heat map. In detail, we use a threshold of t = 0.5 to segment the heat map and get the Region of Interest (ROI). We seek the connected domain of these ROIs, and then extract the geometric features from this connected domain, such as axis, area and statistical features such average and max value in this ROI.

2.3.3. Slide classification

Slide-level classification can help to determine the patient' pN-stage. We use 10 features in total to represent the slide. We choose an ensemble classifier, random forest [9] [10], to train slide-level model for its good performance in classification [11]. With these features from heat maps and the random forest classifier, we classify the slides into four classes (Normal, ITC, Micro, Macro) and then predict each patient's pN-stage by the all his lymph node slide categories according to the given rules.

The fourth image in Figure 1 shows an example of heat map and the fifth image shows the procedure of slide-level prediction.

3. EXPERIMENTS

In this section, we will discuss our dataset arrangement for two levels training and three levels testing. We will describe our experiment setup in detail and the scores we achieved in each step.

3.1. Dataset arrangement

For patch-level classifier training, we use all WSIs with region annotations from Camelyon16 [12] and Camelyon17. In detail, we use 400 WSIs from Camelyon16 and 50 WSIs from Camelyon17 and divide them into patches. The patches used on training and testing is 10:1 respectively.

For slide-level classifier training, we use the remaining 450 slides with 5-folder cross validation.

For patient-level prediction, we use the left 500 WSIs as test set.

3.2. Experiment Setup

To do this experiment, we use Caffe [13] framework and GoogLeNet pre-train model from Caffe Model Zoo [14] for fast convergence. We train our model on four NVIDIA Tesla P40 GPUs using SGD [15] optimizer with step decay [16] for about 2 days. We use a big batch size of 256 when training to gain stability of loss curve.

3.3. Results

Table 1. Three-level scores

Stage	Score
Patch-Level Accuracy	99.1%
Slide-Level Accuracy	91.0%
Patient-Level Kappa Score	0.913

With the ensemble of two stages CNN, we achieved 99.1% accuracy on our test patches. With the 10 features from the slide heat maps and the random forest classifier, we achieved 91.0% accuracy on slide-level classification with 5-folder cross validation. On the test set in Camelyon17, we achieved 0.913 Kappa Score [17] on the 500 lymph node metastasis slides.

4. CONCLUSION AND PROSPECT

In the paper, we introduce a hierarchical and ensemble framework to predict patient's pN-stage from metastasis slides. Our method is efficient and effective and perform well on Camelyon17 dataset. For further work, we will try deeper network, such as ResNet101 [18] for patch-level model training, try more geometrical and statistical features for slide-level model training.

5. REFERENCES

[1] "TNM", http://www.uicc.org/resources/tnm, Accessed: 2017-09-25.

[2] "Camelyon 2016", https://camelyon17.grand-challenge.org/, Accessed: 2017-09-25.

[3] "Dilation", https://en.wikipedia.org/wiki/Dilation_(morphology), Accessed: 2017-09-25.

[4] Shrivastava, Abhinav, Abhinav Gupta, and Ross Girshick. "Training region-based object detectors with online hard example mining." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[5] 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

[6] "Connected Domain", https://proofwiki.org/wiki/Definition:Connected_Domain_(Complex_Analysis), Accessed: 2017-09-25.

[7] LeCun, Yann. "LeNet-5, convolutional neural networks." URL: http://yann.lecun.com/exdb/lenet, 2015.

[8] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[9] Ho, Tin Kam. "Random decision forests." Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on. Vol. 1. IEEE, 1995.

[10] Ho, Tin Kam. "The random subspace method for constructing decision forests." IEEE transactions on pattern analysis and machine intelligence 20.8, pp.832-844, 1998.

[11] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. New York: Springer series in statistics, 2001.

[12] "Camelyon 2016", https://camelyon16.grand-challenge.org/, Accessed: 2017-09-25.

[13] "Caffe", http://caffe.berkeleyvision.org/, Accessed: 2017-09-25.

[14] "Caffe Model Zoo", https://github.com/BVLC/caffe/wiki/Model-Zoo, Accessed: 2017-09-25.

[15] "SGD", https://en.wikipedia.org/wiki/Stochastic_gradient_descent/, Accessed: 2017-09-25.

[16] "step decay", http://caffe.berkeleyvision.org/tutorial/solver.html, Accessed: 2017-09-25.

[17] "Kappa Score", https://en.wikipedia.org/wiki/Cohen%27s_kappa, Accessed: 2017-09-25.

[18] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.