AUTOMATIC DIAGNOSIS OF BREAST CAMCER ON PATIENT-LEVEL VIA DEEPLAB-V3+

Zhixing Chen, Shenghua Cheng, Shaoqun Zeng

Huazhong University of Science and Technology, Wuhan, China

ABSTRACT

Breast cancer's automatic diagnosis is an essential task that promises more accessible healthcare. In this paper, based on the machine learning algorithm, we describe the process of predicting patient-level cancer metastasis. Firstly, the region of interest is extracted from the whole slide image. Next, the foreground region is split into smaller patches which are classified in pixel-level with Deeplab-v3+ model. Then, based on pixel-level predicting probabilities, slide-level heatmap is generated and features are extracted. Moreover, with XGBoost classification, the slide is classified into negative, itc, micro and macro. Finally, patient-level pN stage are determined by 5 slide-level predictions.

Index Terms – Cancer metastases prediction, Deep neural networks, Deeplab v3+, Camelyon17

1. METHODS

The overall algorithm illustrated in Figure 1 is shown as follows:

1) Extracting region of interests.
2) Extracting features based on deep learning model:

   Train a sematic segmentation model based on Deeplab-v3+.

   Predict pixel-level heatmaps.

   Extract selected features.
3) Predicting slides based on XGBoost classification algorithm.

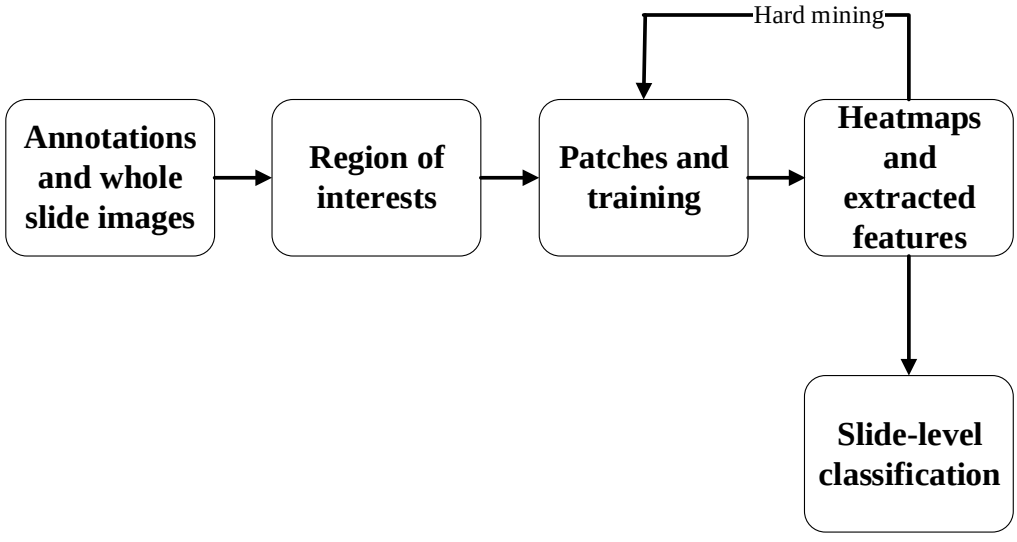

Fig. 1. The framework of cancer metastases detection

1.1. Data Set

The data set comes from Camelyon16 and 17 challenges [2]. The Camelyon16 challenge contains 160 labeled tumor WSIs and 240 normal WSIs. The Camelyon17 challenge's training set contains 100 patients

with pN-stage classification. 50 WSIs in 500 WSIs from Camelyon17 are annotated. In the process of training a feature extraction model, normal or annotated WSIs are used.

1.2. Extracting Region of Interests

Most regions of a whole slide image do not have tissue. To reduce computations and focus on region of interest, thresholds algorithm based on HSV value are adopted. It is tested that HSV threshold algorithm has better generalizing ability, compared with other algorithms such as Otsu or RGB threshold. Since the WSIs in the highest resolution have approximately 200000*100000 pixels, extracting tissue region can be time consuming. Thus, using 32-times down-sampled WSI (level 5) can be more sufficient. The maximum value for each channel in HSV space is 200 and the minimum value is 30. After the operation is done, holes and small points, which will influence feature extraction operation, still exist. So morphological algorithms are operated to fill holes and exterminate points.

1.3. Extracting features with pixel-level classifier

The main algorithm is based on Deeplabv3+. The model framework is depicted in Figure 2.

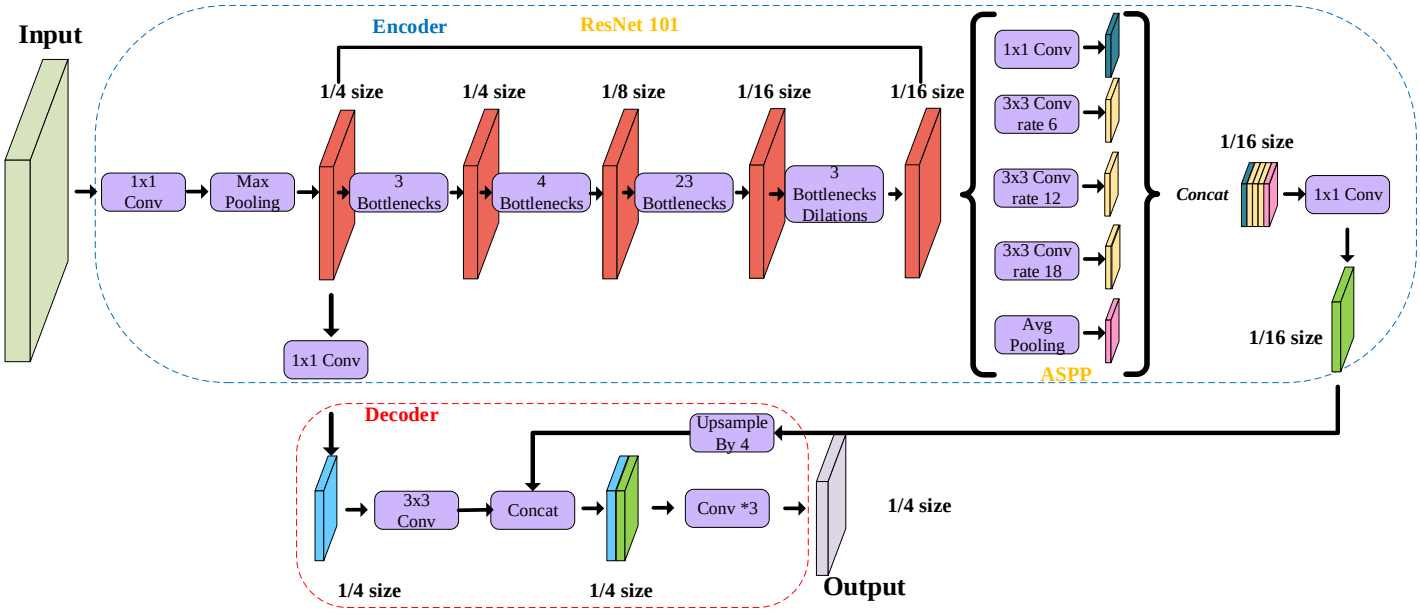

Fig. 2. Deeplabv3+

1.3.1. Preparation of training samples

For sematic segmentation network, image patches and pixel-wise ground truth patches are extracted. To minimize computing time and the number of weight parameters, the input resolution is chosen as level 1 and the output resolution is chosen as level 3. Moreover, level 1 patch with the size larger than 512 proves one patch has enough cells. To maintain the number of cells and to consider the limitation by computing resources, input image patch size is 768*768, and the output mask size is 192*192. For each annotated tumor WSI, tumor region mask is generated for making positive patches. Based on the mask, 1000 positive patches are randomly selected. In order to prepare for the scaling data augmentation, image patch is extracted with the size of 960*960. Every positive patch has more than 30000 positive pixels. For those small tumor regions whose total size is smaller than 30000 positive pixels, 1000 positive patches are extracted anyway. Besides, in each tumor WSI, 2000 negative patches are made on the foreground region which is extracted above. Equivalently, 2000 negative patches are made in each normal WSI.

Except for the training set, 3 annotated tumor WSIs and 5

negative WSIs in the Camelyon16 test set are reserved for validation. The total number of positive patches in validation set is 3000, and the number of negative patches is 16000.

After neural networks are trained, sliding window method is used on all the tissue region to generate a tumor probability heatmap for each slide. In the resolution of level 1, the sliding stride is 768 and the sliding window size is 1024. The overlap output uses the maximum probability to generate a whole slide heatmap.

1.3.2. Sematic segmentation Network training with data augmentation and hard-mining process

To predict pixel-level heatmap, deeplab-v3+ [1] is chosen as the segmentation model. Synchronized batch normalization is used in every batch-norm layer. And the feature extraction backbone is preferably to be ResNet-101 The network is trained with Adam [3] method. The loss function is preferred to be MSE-loss. The initiative learning rate is 0.001, but the learning rate in ASPP module is 10 times larger than other parts of the model. For each training block, model inferences each patch with criterions: IoU (intersection over union) and AP (accuracy of pixel).

Since there are much more normal regions than tumor regions, to address this problem, each training block contains 5000 positive patches and 5000 negative patches. The batch size of 4 is used on each GPU. There are six 2080ti GPUs used for training this model.

Training the segmentation network includes three stages. First of all, to initiate model weight, no augmentation algorithm is operated and learning rate is fixed in 0.001. The ResNet-101 module's weight is initialized with pre-trained model on Image-Net dataset.

Secondly, after the mean IoU in validation set reaches 0.84, the next stage begins. The second stage adds data augmentation and learning rate decay to the training process. Three sorts of augmentation methods are used to incorporate all possibilities:

(1) To handle with micron per meter problems, the initial image patch size 960 is randomly scaled or cropped to 768 pixels. The scale factor is 20%, which means the images is scaled 20% larger or smaller than the normal output size. Then, all the image patches are normed to the size of 768*768, and the mask patches are normed to 192*192.

(2) Random flipping and rotation are also applied to predict heatmaps robustly.

(3) To deal with staining colors' difference, color jitter augmentation is randomly applied with parameters containing PCA jittering, contrast, brightness, HSV perturbation, RGB perturbation, Gaussian blur.

The learning rate starts with 1e-3 and the weight decays 0.7 per 50 blocks.

Finally, slide-level heatmap prediction is run on all annotated or negative slides in Camelyon17 for hard example mining. For the negative patch, if the number of suspected pixel, whose probability is larger than 0.7, is not zero, this patch is saved as false positive example. For the positive patch, if the number of suspected pixel is smaller than 100 or the IoU is smaller than 0.94, the patch is also saved. Other 300 blocks are trained with initial learning rate of 1e-4, weight decay and data augmentation.

1.3.3. Feature extraction based on probability heatmap

Each heatmap is converted into a feature vector after being thresholded with 3 threshold value, i.e., 0.5,0.7,0.9. Then, 37 types of features (**Table 1**) are extracted according to the morphological and geometrical information.

.

Table 1. 37 types of features.

Feature Number	Feature Name	Feature Number	Feature Name
1	Largest probability	10	0.7, equivalent diameter of the largest component
2	0.5, total area of components	11	0.7, number of components
3	0.7, total area of components	12	0.7, area of components/ area of foreground region
4	0.9, total area of components	13-17	0.9, max, mean, variance, skewness, and kurtosis of 'area' of components
5	0.7, mean probability in the largest component	18-22	0.9, max, mean, variance, skewness, and kurtosis of 'perimeter' of components
6	0.7, max probability in the largest component	23-27	0.9, max, mean, variance, skewness, and kurtosis of 'eccentricity' of components
7	0.7, area of the largest component	28-32	0.9, max, mean, variance, skewness, and kurtosis of 'extent' of components
8	0.7, major axis length of the largest component	33-37	0.9, max, mean, variance, skewness, and kurtosis of 'solidity' of components
9	0.7, minor axis length of the largest component		

1.4. Predicting slides based on XGBoost classification

After feature vectors are extracted, XGBoost [4] classification is selected as the patient-level classification algorithm. Since the total number of WSIs, which are classified in four classes, are only 500, it is important to prevent overfitting. Thus, K-fold cross validation and other hyper-parameters, such as subsample or max depth, are adopted for preventing overfitting. Firstly, grid-search algorithm is used to search for the best hyper-parameters. Then, the XGBoost model is trained with ten-fold cross validation. Finally, the best model is selected and the 100 patients in the Camelyon17 test set are predicted.

2. RESULTS

Mean IoU of pixel-level classification is 0.901 with threshold 0.7 in validation set, when 1500 positive patches and 1500 negative patches are tested. The mean pixel accuracy is 0.962. For the XGBoost classification, the best slide-level accuracy in the ten-fold cross validation is 0.951, when 50 slides are tested.

3. REFERENCES

[1] George Papandreou Florian Schroff Hartwig Adam Liang-Chieh Chen, Yukun Zhu, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in ECCV 2018, 2018

[2] Geessink, O. et al. Camelyon17: Grand challenge on cancer metastasis detection and classification in lymph nodes, 2017.

[3] Kingma, D. et al. Adam: A method for stochastic optimization. *arXiv preprint arXiv*, 1412.6980 (2014).

[4] Chen T Guestrin C. XGBoost: A Scalable Tree Boosting System[J]. 2016.