Detection of Gastric Cancer from Histopathological Image using Deep Learning with Weak Label

Yeong Won Kim[1], **Dong-Il Kim**[2] **and Kyu-hwan Jung**[1]
[1]VUNO Inc., [2]Green Cross Laboratories
*ywon@vuno.co, +ilkim@gclabs.co.kr, †khwan.jung@vuno.co

Abstract

We propose an automatic method to detect gastric cancer on whole slide images. To do this, we collected cases with slide-level and region-level labels and trained deep neural network for tissue classification. To improve the tissue classification performance, we exploited slide-level weak label for training the model with patches without region-level label. For whole slide classification, we extracted features representative for whole slide characteristics. The proposed method archived 92.51% accuracy on 3 class classification. Macro-average and micro-average AUROC on the test set is 0.9802 and 0.9788.

1 Introduction

Gastric cancer is one of the most common causes of cancer death worldwide. For early detection of gastric cancer to reduce mortality, endoscopic biopsy of the stomach is commonly performed. Although most cases are not malignant, pathologists have to inspect carefully to not miss small cancer tissues. Because of this, pathologists are often overloaded and the quality of pathologic diagnosis is degraded. Thus, automatic screening systems could enable pathologists by removing the large proportion of normal cases among endoscopic biopsy specimens.

Due to growing availability of whole slide scanners, digital pathology has become popular and more whole slide images (WSIs) have become accessible. On the other hand, deep learning has shown outperforming performance in many fields and it is widely used. Especially, deep learning performs well on image analysis tasks such as detection and classification.[2] With these developments, numerous deep learning studies aim to detect and classify cancer.[1] However, unlike breast cancer, only a few studies have focused on gastric cancer with deep learning because of lack of the available datasets.[4, 5] Since labeling WSIs is laborious and time-consuming, it is hard to collect rich training datasets for deep learning.

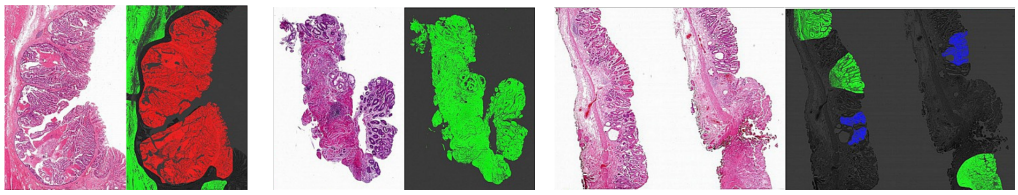
Therefore, we propose a two phase framework for the detection of gastric cancer using deep learning. Our main contribution is that we collected only typical regions as region-level labels to reduce labeling time and guarantee high inter-observability. Moreover, we propose a loss function with the slide-level label as weak labels to utilize the unlabeled regions. As a result, the proposed method showed higher accuracy than baseline methods.

2 Method

2.1 Data preparation

First, we collected WSIs for training and testing of the proposed method. Among several tissue phenotypes, this paper conducts tissue classification into three categories: *Normal, Tubular adenoma (TA)* and *Cancer*. The WSIs were captured by whole slide scanner Leica Versa.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.



(a) Cancer case (b) Normal case (c) Tubula adenoma case

Figure 1: The figure shows parts of the WSIs and its annotation. Red, green and blue indicate *Cancer, Normal* and *TA* categories, respectively.

For the development set, 90 cases of hematoxylin and eosin (H&E) stained resection and endoscopic biopsy specimens were collected at Korea University Guro Hospital and Green Cross Laboratories. Each class has 30 cases and at least two pathologists confirmed and labeled each case. For the test set, 548 cases of endoscopic biopsy specimens were collected under the same condition as the training set. Note that we used only endoscopic biopsy specimens excluding resection specimens for testing because the proposed method focused on biopsy screening.

Cancer and *TA* cases contain both normal and tumor tissues. Thus, the training sets were annotated with the region-level label (*Normal, Cancer, TA*) by one pathologist. To guarantee high inter-observability, only typical regions were annotated. Figure 1 shows samples of the collected cases.

2.2 Training and Prediction

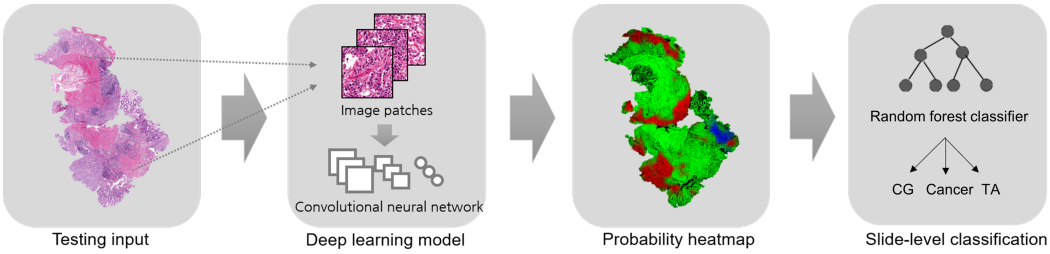


Figure 2: Illustration of the proposed method

We propose a method to predict slide-level labels of a WSI. Because the size of a WSI is too large to take an end-to-end approach, the proposed method is divided into two steps, the region-level inference and the slide-level inference. Figure 2 shows the processes of the proposed method. The region-level inference classifies small image patches into 3 classes; *Normal, Cancer, TA*. This is performed on whole slides and generates a probability map. After that, the slide-level inference predicts the class of the case based on the probability map.

For the region-level inference, we adopt the network architecture of [3], that showed the best results on the CAMELYON16 challenge. It is based on InceptionV3 with the last fully connected layer re-defined by the number of output classes. Input patch size is 299x299x3 with 100x scale. Output is the probability of each category with softmax normalization. Cross entropy between output and one-hot encoded label are used as the loss function. For optimization, RMSprop is used.

To utilize unlabeled regions shown as the dark region in Figure 1 to further improve the tissue classification performance, we exploit slide-level weak label as a prior information. As *TA* cases do not contain *Cancer* tissues, we can train our model with patches from unlabeled regions by minimizing the output of *Cancer* class. Similarly, as *Cancer* cases rarely contains *TA* tissues, we can train our model by minimizing *TA* class output for the patches generated from unlabeled region of *Cancer* slides.

From the probability map generated from region-level inference, the summation of classification output for each pixel are calculated as the features for WSI classification, for which we used random forest classifier. We excluded the *Normal* region because the size of the *Normal* region is not critical to make a diagnosis.

3 Experiments and Results

Table 1: Confusion matrix of test results of the baseline method with mini-batch size 16

		Prediction		
		Normal	TA	Cancer
Diagnosis	Normal	272	0	2
	TA	31	96	0
	Cancer	39	6	102

Table 2: Confusion matrix of test results of the proposed method

		Prediction		
		Normal	TA	Cancer
Diagnosis	Normal	268	0	6
	TA	16	107	4
	Cancer	9	6	132

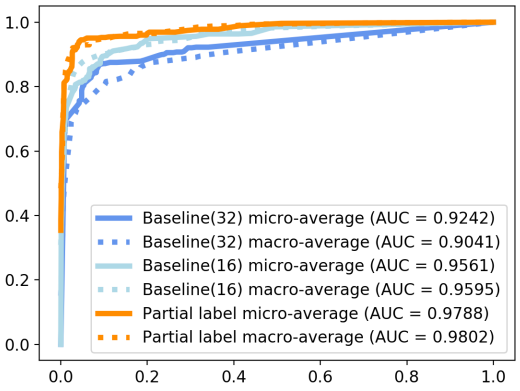


Figure 3: Macro-average and micro-average ROC curves on the test set.

We used 80% of the development set for training and the rest for validation. For the fast convergence of the training, we initialized the network with the parameters of pre-trained network for ImageNet. The mini-batch size was 16 and 32. Samples were uniformly sampled between 3 classes. For generalization, each sample was augmented by transition, rotation, flip, color. For the proposed method, each mini-batch contains 16 images with full labels and 16 images with weak labels. The initial learning rate is 0.002 with decay rate 0.5 on every 20000 iterations.

On the test set, the baseline method achieved 85.76% and 84.31% accuracy with mini-batch size 16 and 32, respectively. The proposed method achieved a 92.51% accuracy. Table 1 and 2 shows the confusion matrix of test results of the baseline method and the proposed method. The proposed method showed significantly better results than the baseline method. Especially, the proposed method performed well on the *TA* class. Both macro-average and micro-average ROC curves as shown in Figure 3, macro-average and micro-average AUROC of the proposed method achieved 0.9802 and 0.9788, respectively.

4 Conclusion

We proposed an automatic method for the detection of gastric cancer on whole slide images exploiting slide-level weak labels. In the experiments, our methods showed significantly better results compared to baseline methods. In the future, we will validate the method with more datasets from multiple laboratories.

References

[1] Zhongyi Han et al. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific reports*, 7(1):4172, 2017.

[2] Andrew. Janowczyk and Anant. Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7(1):29, 2016.

[3] Yun Liu et al. Detecting cancer metastases on gigapixel pathology images. *CoRR*, abs/1703.02442, 2017.

[4] Kosuke Oikawa et al. Pathological diagnosis of gastric cancers with a novel computerized analysis system. *Journal of pathology informatics*, 8, 2017.

[5] Harshita Sharma et al. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Computerized Medical Imaging and Graphics*, 61:2–13, 2017.