2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Weakly Supervised Attention Map Training for Histological Localization of Colonoscopy Images

Jangho Kwon and Kihwan Choi

Abstract— **We consider the problem of training a convolutional neural network for histological localization of colorectal lesions from imperfectly annotated datasets. Given that we have a colonoscopic image dataset for 4-class histology classification and another dataset originally dedicated to polyp segmentation, we propose a weakly supervised learning approach to histological localization by training with the two different types of datasets. With the classification dataset, we first train a convolutional neural network to classify colonoscopic images into 4 different histology categories. By interpreting the trained classifier, we can extract an attention map corresponding to the predicted class for each colonoscopy image. We further improve the localization accuracy of attention maps by training the model to focus on lesions under the guidance of the polyp segmentation dataset. The experimental results show that the proposed approach simultaneously improves histology classification and lesion localization accuracy.**

I. INTRODUCTION

Colon cancer is the third most common and fourth most fatal cancer in the world [1]. In order to prevent colon cancer, screening, which consists of colonoscopy and histologic analysis, is considered as the first option [2]. Colonoscopy is an endoscopic examination that observes and removes the abnormal colon tissues [3]. Typical histologic analysis is a microscopic biopsy of the tissues which were removed during the colonoscopy [4]. Over a decade, many researchers have tried to reduce the burden of histologic analysis, because it requires additional time and expense due to the microscopic biopsy [5]. In addition, the required cost increases as the number of tissues, which are removed during the colonoscopy, increases [6].

Recently, many studies has developed optical biopsy as a method for reducing unnecessary histologic analysis [7]. Optical biopsy is a technique that analyzes the abnormal tissues using optical devices during the colonoscopy. Using optical biopsy, endoscopists can predict histologic categories of the lesions without surgically removing the lesions. As a result, optical biopsy potentially reduces time and expense by replacing microscopic biopsy. However, existing optical biopsy methods need additional endoscopic devices to highlight the cancer-related optical features such as endocytoscopy [8] and laser-induced fluorescence spectroscopy [9]. These additional devices could entail changes in the existing endoscopic system, which further require additional costs.

This work was supported by Korea Institute of Science and Technology (KIST) Institutional Program (Project No. 2E31122). (Corresponding author: Kihwan Choi)

Jangho Kwon and Kihwan Choi are with the Center for Bionics, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea, (e-mail: g15007@kist.re.kr; kihwanc@kist.re.kr).

In the previous study [10], [11], we have shown that a computer-aided diagnosis (CAD) system is able to conduct optical biopsy without additional devices. Based on convolutional neural network (CNN), the CAD system takes white light endoscopic images and predicts the corresponding histological categories. In addition, the CAD system employs a model interpretation technique [12] and shows probability heatmaps with respect to the histological categories. From the predicted heatmaps, endoscopists can localize and determine which abnormal tissues to remove. However, the predicted heatmaps occasionally show high probabilities irrelevant to the colon cancer such as light reflection, wrinkles, and contrast difference. These inaccurate localizations might be due to the classifer's decision based on the irrelevant correlation in the training data [13].

For accurate lesion localization, we propose a multi-task learning framework, which combines the lesion localization task into the original histology classification task. With a dataset, which is originally dedicated for polyp segmentation, our approach learns to correct the predicted heatmaps derived from the histology classifier in a weakly supervised learning manner. Through a retrospective clinical study, we show that the weakly supervised learning approach can simultaneously improve lesion localization performance as well as histology classification accuracy.

II. MATERIALS AND METHODS

In order to enhance the performance of lesion localization, we employ weakly supervised multi-task learning. The proposed framework simultaneously learns two different tasks: histology classification and lesion localization. We first train the classifier with a histology report dataset. Then, we further train the classifier to predict lesion locations with another dataset, which is originally dedicated for polyp segmentation.

A. Histology Classification

As a histology classifier, we used ResNet-101 [14], which was pre-trained with the ImageNet dataset [15]. We adapted the last fully-connected layer to predict the four histology classes of colon cancer.

For the supervised learning of the classification task, we prepared a dataset from Korea University Medical Center (KUMC), Seoul, Korea. The KUMC dataset includes endoscopic images acquired during the colonoscopy and the corresponding histology reports after the microscopic biopsy. The images and the histology reports were carefully collected from the hospital's Picture Archiving and Communication System (PACS). After collecting data, we categorized the

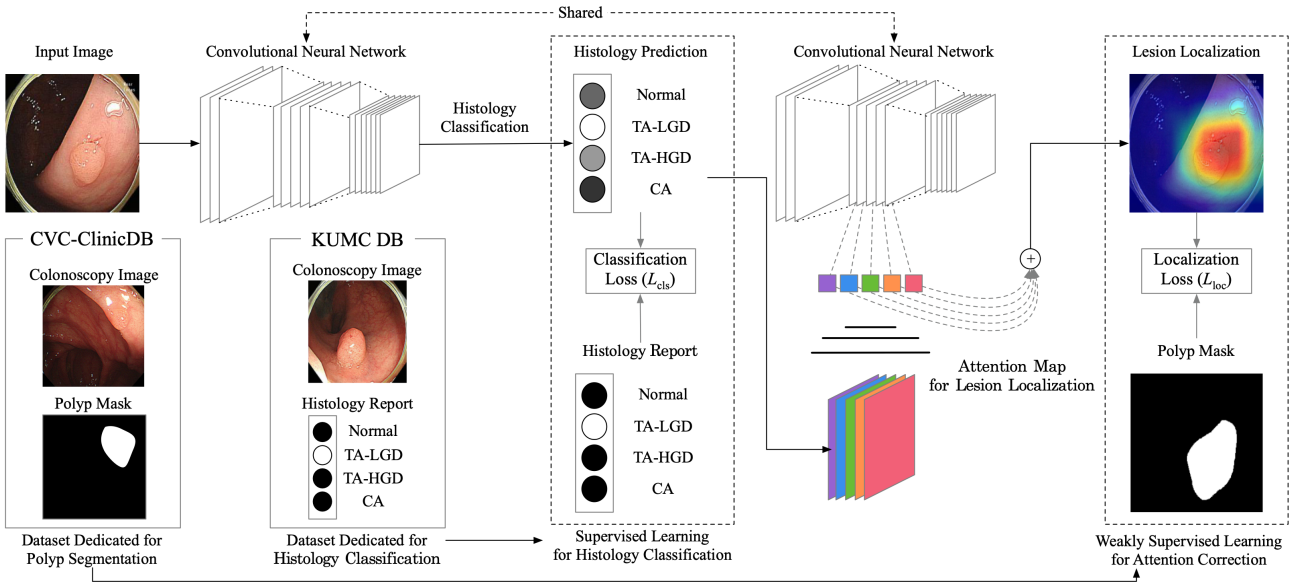

Fig. 1. Overview of histological localization with weakly supervised multi-task learning. The histological localization task consists of two different processes: histology classification and lesion localization. The histology classification process uses a dataset with 4 different histology categories for training. For lesion localization, we further train the classifier to produce attention maps consistent to masks from a polyp segmentation dataset.

colonoscopic images into four classes based on the corresponding histology reports. The histology classes include normal, tubular adenoma with low-grade dysplasia (TA-LGD), tubular adenoma with high-grade dysplasia (TA-HGD), and adenocarcinoma (CA). The acquired data include 1000 normal images, 1000 TA-LGD images, 500 TA-HGD images, and 500 CA images.

B. Attention Map for Lesion Localization

For lesion localization, the model predicts a probability heatmap, which is related with the predicted histology class, using a model interpretation technique. We used grad-CAM [12] to interpret the histology classifier. The grad-CAM synthesized an attention map from the classifier according to the gradients, which are computed with respect to the maximal output of the histology classification.

For training our model, we used the CVC-ClinicDB [16] dataset, which is dedicated for polyp segmentation. The CVC-ClinicDB consists of 612 colonoscopic images and polyp masks, which were collected from 29 colonoscopy videos containing colon polyps. Most of the colonoscopic images (478 images) contain adenoma lesions such as TA-LGD or TA-HGD, and part of the colonoscopic images (124 images) contain hyperplastic polyps. The polyp masks are binary masks indicating polyps within the image, and there is no detailed label for the histology information.

C. Weakly Supervised Learning for Histological Localization

In order to train the classifier with the two different datasets, we propose a multi-task learning framework consisting of supervised learning for histology classification and weakly supervised learning for attention correction. For histology classification, we trained the classifier with the KUMC dataset using classification loss. The classification loss L_{cls} is defined by categorical cross-entropy between the histology report and histology classification:

$$L_{\text{cls}} = \sum_{c} y^c \log \hat{y}^c, \tag{1}$$

where y^c indicates class based on histology reports c (normal, TA-LGD, TA-HGD, and CA), and $\hat{y}^c$ indicates predicted histology reports.

In order to improve the accuracy of attention maps, we further trained the classifier with the CVC-ClinicDB dataset using localization loss. The lesion localization loss L_{loc} is defined by weighted binary cross-entropy between the polyp mask and the predicted region.

$$L_{\text{loc}} = \sum_{i,j} \big(wP(i,j) \log \hat{P}(i,j) + (1 - P(i,j)) \log (1 - \hat{P}(i,j)) \big), \tag{2}$$

where $P(i,j) \in \{0,1\}$ indicates the lesion existence at pixel (i,j), $\hat{P}(i,j) \in [0,1]$ indicates the predicted heatmap probability at (i,j), and w is the lesion weight to minimize missed lesion localization. We set the w as 5 in this study.

For multi-task learning for histological localization, we trained the classifier by switching between the two different learning losses. After 5 training iterations for histology classification with the KUMC dataset, we trained the classifier using the lesion localization loss with the CVC-ClinicDB dataset. During the lesion localization training process, the classifier learns to produce attention maps that correspond to the segmentation masks. We trained the classifier using Adam optimizer [17] for 72 epochs of classification training. The learning rate was set to 10^{-5} with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$.

Histological Category	LGD	HGD	CA
Single-Task Model [11]	48.10%	74.17%	87.28%
Our Method	55.00%	74.38%	87.37%

TABLE I

DICE SCORES OF SUPERVISED SINGLE-TASK LEARNING METHOD AND WEAKLY SUPERVISED MULTI-TASK LEARNING METHOD FROM CROSS-VALIDATION WITH KUMC DATASET. WE EMPLOYED GRAD-CAM TO EXTRACT PROBABILITY HEATMAPS FROM THE TWO METHODS. THE REGIONS OF INTERESTS (ROIS) WERE PREDICTED BY THRESHOLDING THE PROBABILITY HEATMAPS WITH 0.5, 0.4, AND 0.4 FOR LGD, HGD, AND CA, RESPECTIVELY. WE MEASURED DICE SCORES BETWEEN THE PREDICTED ROIS AND TRUE LABELS FOR EACH HISTOLOGICAL CATEGORY.

III. RESULTS AND DISCUSSION

We evaluated the weakly supervised multi-task learning framework by comparing it with a typical supervised learning approach. For performance evaluation, we conducted five-fold cross-validation with the KUMC dataset. We divided the dataset into 5 splits where each split has 200 normal images, 200 TA-LGD images, 100 TA-HGD images, 100 CA images. In each validation, we trained the model with 4 splits and tested the trained model with the remaining split. By repeating the validation 5 times, we cross-validated the model with the entire dataset.

From the cross-validation with the KUMC dataset, we summarized the results using a confusion matrix of the predicted histology classes in terms of recall, precision, and classification accuracy.

For performance evaluation of lesion localization, we prepared a test set by annotating the colon lesions with bounding boxes. From model prediction to produce heatmaps of the colonoscopic images in the test set, we calculated the regions of interests (ROIs) by thresholding the predicted heatmaps. Then, we summarized the results using the Dice coefficient between the annotated bounding boxes and predicted ROIs. The evaluated threshold values were 0.4, 0.5, and 0.6.

A. Performance Evaluation of Lesion Localization

Table I compares the lesion localization performance between the single-task learning approach [11] and our multi-task learning approach. The lesion localization was improved for all histology classes in the weakly supervised multi-task learning approach. Especially, the lesion localization performance was lower in the classes with smaller sized colon lesion. The Dice Score of TA-LGD (48.10%) was lower than TA-HGD (74.17%), and the Dice score of TA-HGD was also lower than CA (87.28%). We also found that the performances were further improved for small-sized colon lesions by the weakly supervised multi-task learning method with the polyp segmentation dataset. The Dice score of the TA-HGD was higher (+6.90%) than the HGD (+0.21%) and the CA (+0.09%).

Figure 2 shows examples of the corrected attention maps. The supervised single-task learning model showed attention maps affected by light reflection spots, contrast difference, and bleeding areas. These irrelevant heatmaps might adversely affect endoscopists when understanding the predicted histology and determining the resection regions. Compared to the single-task learning approach, the weakly supervised multi-task learning approach resulted in attention maps focused on the colon lesions.

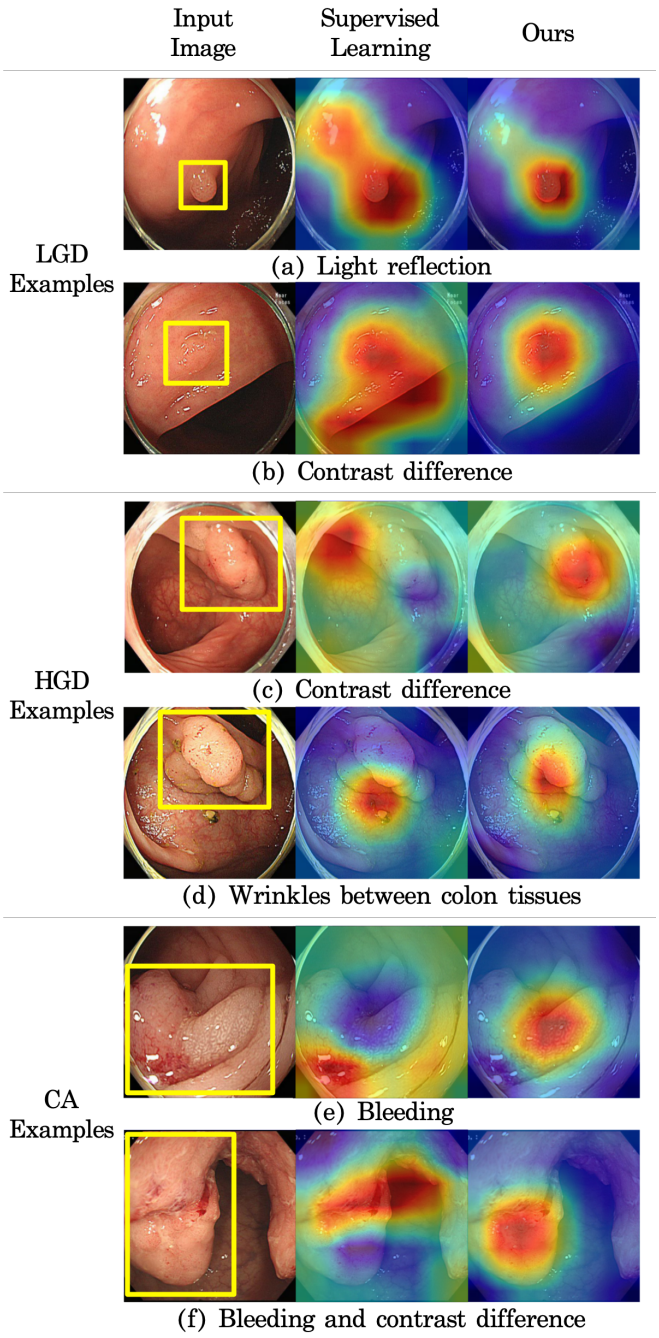

Fig. 2. Comparison of lesion localization results between supervised single-task learning and our weakly supervised multi-task learning. The heatmaps are acquired from the cross-validation using the KUMC dataset. Yellow boxes show the labeled bounding boxes for lesion locations. The irrelevant features highlighted by the supervised single-task learning model are listed.

	Normal	LGD	HGD	CA	Precision
Normal	949 (-12)	33 (+9)	7 (-4)	11 (+7)	94.9% (-1.2%)
LGD	40 (-3)	829 (+24)	100 (-16)	31 (-5)	82.9% (+2.4%)
HGD	6 (-3)	97 (+9)	292 (+13)	105 (-19)	58.4% (+2.6%)
CA	5 (-1)	26 (-7)	107 (+8)	362 (+0)	72.4% (+0.0%)
Recall	94.9% (+0.6%)	84.2% (-0.6%)	57.7% (+2.5%)	71.1% (+2.3%)	**ACC: 81.1% (+0.8%)**

TABLE II

CONFUSION MATRIX OF WEAKLY SUPERVISED LEARNING MULTI-TASK MODEL FOR HISTOLOGICAL LOCALIZATION. THE NUMBERS IN PARENTHESES INDICATE DIFFERENCES COMPARED TO THE SINGLE-TASK MODEL [11]. THE ROWS REPRESENT PREDICTED CLASSES, AND THE COLUMNS REPRESENT GROUND TRUTH.

B. Evaluation of Histology Classification Performance

Table II shows the confusion matrix of histology classification performance. The correctly classified images increased from 805 images (80.5%) to 829 images (82.9%) for the TA-LGD, and the correctly classified images also increased from 279 images (55.8%) to 292 images (58.4%) for the TA-HGD. As a result, the overall accuracy of histological localization was slightly improved in weakly supervised learning (81.1%) compared to supervised single-task learning (80.3%). Interestingly, the performance of histology classification improved despite training with a dataset only dedicated for polyp segmentation without histological class information.

Similar to real endoscopists, who tend to miss smaller sized colon lesions than larger colon lesions [18], the deep learning models show a lower localization performance with small-sized colon lesions in Table I. However, Tables I and II show that our weakly supervised multi-task learning approach improved histology classification performance for small-sized colon lesions. Along with the enhanced heatmaps, our multi-task model can potentially support endoscopists in localizing small colon lesions and predicting histological categories. In a future study, we will evaluate the lesion localization results with endoscopists in order to investigate whether the proposed method is clinically acceptable.

IV. CONCLUSION

In this study, we investigated weakly supervised learning for histological localization in order to improve lesion localization accuracy for CNN-based optical biopsy. We extracted attention maps from our classifier using a model interpretation technique. We further trained the classifier to correct the attention maps with a dataset originally dedicated for polyp segmentation. The result shows that weakly supervised multi-task learning improved the lesion localization performance without performance degradation in histology classification.

REFERENCES

[1] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global patterns and trends in colorectal cancer incidence and mortality," *Gut*, vol. 66, no. 4, pp. 683–691, Apr. 2017.

[2] K. Simon, "Colorectal cancer development and advances in screening," *Clin. Interv. Aging*, vol. 11, pp. 967–976, Jul. 2016.

[3] US Preventive Services Task Force *et al.*, "Screening for colorectal cancer: US preventive services task force recommendation statement," *JAMA*, vol. 315, no. 23, pp. 2564–2575, Jun. 2016.

[4] M. Fleming, S. Ravula, S. F. Tatishchev, and H. L. Wang, "Colorectal carcinoma: Pathologic aspects," *J. Gastrointest. Oncol.*, vol. 3, no. 3, pp. 153–173, Sep. 2012.

[5] W. R. Kessler, T. F. Imperiale, R. W. Klein, R. C. Wielage, and D. K. Rex, "A quantitative assessment of the risks and cost savings of forgoing histologic examination of diminutive polyps," *Endoscopy*, vol. 43, no. 8, pp. 683–691, Aug. 2011.

[6] D. K. Rex, "Risks and potential cost savings of not sending diminutive polyps for histologic examination," *Gastroenterol. Hepatol.*, vol. 8, no. 2, pp. 128–130, Feb. 2012.

[7] M. F. Byrne, N. Shahidi, and D. K. Rex, "Will Computer-Aided detection and diagnosis revolutionize colonoscopy?" *Gastroenterology*, vol. 153, no. 6, pp. 1460–1464.e1, Dec. 2017.

[8] Y. Maeda, S.-E. Kudo, Y. Mori, M. Misawa, N. Ogata, S. Sasanuma, K. Wakamura, M. Oda, K. Mori, and K. Ohtsuka, "Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video)," pp. 408–415, 2019.

[9] T. Rath, G. E. Tontini, M. Vieth, A. Nägel, M. F. Neurath, and H. Neumann, "In vivo real-time assessment of colorectal polyp histology using an optical biopsy forceps system based on laser-induced fluorescence spectroscopy," *Endoscopy*, vol. 48, no. 6, pp. 557–562, Jun. 2016.

[10] K. Choi, S. J. Choi, and E. S. Kim, "Computer-Aided diagonosis for colorectal cancer using deep learning with visual explanations," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2020, pp. 1156–1159, Jul. 2020.

[11] S. J. Choi, E. S. Kim, and K. Choi, "Prediction of the histology of colorectal neoplasm in white light colonoscopic images using deep learning algorithms," *Sci. Rep.*, vol. 11, no. 1, p. 5311, Mar. 2021.

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[13] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek, and K.-R. Müller, "Unmasking clever hans predictors and assessing what machines really learn," *Nature communications*, vol. 10, no. 1, pp. 1–8, 2019.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255.

[16] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imaging Graph.*, vol. 43, pp. 99–111, Jul. 2015.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014.

[18] J. C. van Rijn, J. B. Reitsma, J. Stoker, P. M. Bossuyt, S. J. van Deventer, and E. Dekker, "Polyp miss rate determined by tandem colonoscopy: a systematic review," *Am. J. Gastroenterol.*, vol. 101, no. 2, pp. 343–350, Feb. 2006.