CENTERNET-BASED DETECTION MODEL AND U-NET-BASED MULTI-CLASS SEGMENTATION MODEL FOR GASTROINTESTINAL DISEASES

Yoon Ho Choi[1], *Yeong Chan Lee*[2], *Sanghoon Hong*[2], *Junyoung Kim*[3], *Hong-Hee Won*[2†], *Taejun Kim*[2,3†]

[1] Dept. of Health Sciences & Tech., Samsung Advanced Institute for Health Sciences & Tech. (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea
[2] Dept. of Digital Health, Samsung Advanced Institute for Health Sciences and Tech. (SAIHST), & Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea
[3] Dept. of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

ABSTRACT

From the perspective of the computer-aided diagnosis system, it is important to build automated techniques that detect and diagnose lesions to reduce the missing rate of clinicians. Recently, various diagnosis techniques using computer vision and artificial intelligence have been developed.However, they need to diagnose various lesions more accurately to be used in actual clinical practice. Accordingly, we developed CenterNet-based object detection model and U-Net-based class-wise binary segmentation model. These models were trained with random augmentation methods including color and morphological changes. For the 43 test set images, our model shows 0.1932 ± 0.0622 of mean average precision with standard deviation in detection, and 0.2544 ± 0.2080 of semantic score in segmentation.

1. INTRODUCTION

Endoscopists can recognize diverse lesions related with digestive disorders in gastrointestinal organs through endoscopic examinations. The detected lesion is clinically managed or resected in compliance with medical guidelines. However, it is not typically diagnosed until the results of pathological examination are known. Some endoscopic examinations are effective for the early diagnosis and prevention of gastrointestinal disease, but detecting lesions is highly dependent on the skill and experience of the endoscopists. For example, some studies have reported that the missing rate of polyps during colonoscopy ranges from 17% to 28% [1].

Recently, computer-aided system has remarkably improved with medical imaging. Especially, recent studies have shown that artificial intelligence can meet the endoscopists' needs. A prospective randomized controlled trial showed that adenoma detection rates of colorectal polyps significantly increased when endoscopists co-worked with real-time automatic detection system [2]. Another randomized controlled trial showed that deep convolutional neural network using deep reinforcement learning achieved real-time monitoring blind spots with a high accuracy during esophagogastroduodenoscopy [3].

We participated in sub-challenge II: Endoscopic Disease Detection and Segmentation (EDD2020) of Endoscopy Computer Vision Challenges on Segmentation and Detection (EndoCV2020). Deep learning models were developed for detecting or segmenting lesions from 4 different organs for this challenge.

For this challenge, A CenterNet-based model was designed to detect lesions and a class-wise U-Net-based model was developed to segment lesions.

2. DATASETS

In total, 386 endoscopic images of the training set were obtained from 5 multi-centers [4]. Every image was assigned to at least 1 class from 5 disease classes with Barretts esophagus (BE), high grade dysplasia (HGD), cancer, polyp and suspicious region from 4 different organs. These images had corresponding bounding boxes and pixel-level labels of each lesion and were annotated by medical experts. The number of images in the entire training set was imbalanced across disease classes. (BE : 160, HGD : 74, cancer : 53, polyp : 127, and suspicious : 88).

3. METHODS

3.1. Image preprocessing

Class imbalance can lead to biased results towards a particular class during the training of the model. Thus, prior to image pre-processing, we randomly duplicated images

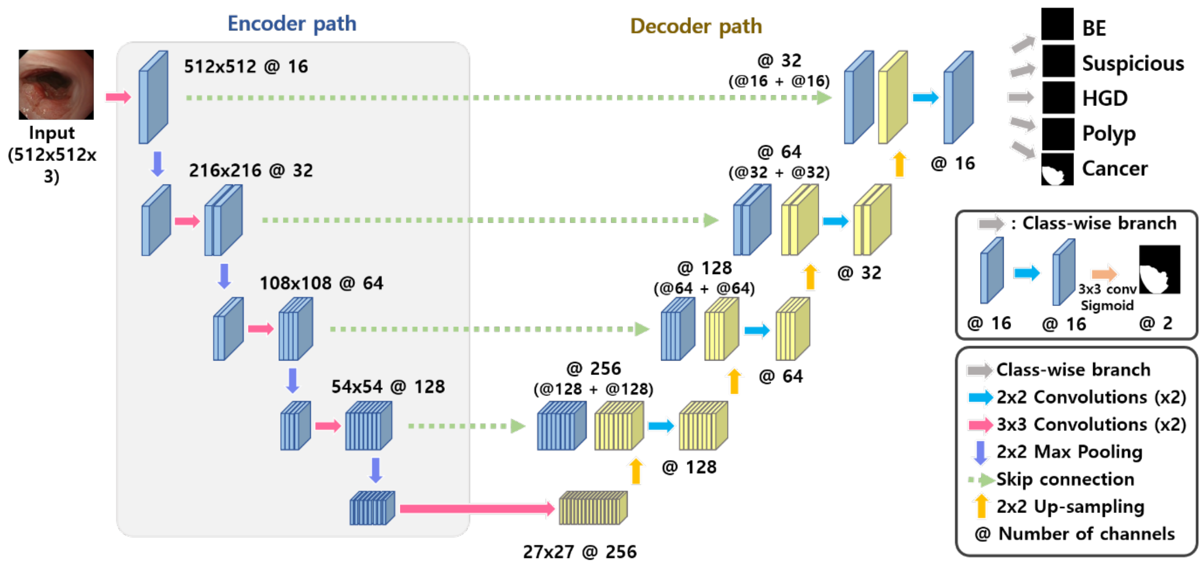

Fig. 1. The architecture of U-Net-based class-wise binary segmentation model

in insufficient classes to balance the number of images in all classes. At this point, it was important to minimize the number of duplicated images, since indiscriminately duplicated images may cause substantial bias in the trained model. Therefore, every round we identified the class with the highest number of images and the class with the lowest number. Then we randomly duplicated images of the lowest class. To ensure that, the images containing objects of the highest class were excluded from the random duplication.

After balancing the number of images belonging to each class, we preprocessed the training data to reduce overfitting of our models to it and generalize the models to the test data. Firstly, all images in the training data were standardized for each channel and randomly augmented 86 times using rotation, flipping, contrast enhancement, and brightness adjustment. Next, to train the model with invariant properties for the scale, we randomly changed the resolution of the original image from 320 to 602 every 10 epochs and then converted it to a size of 512 $\times$ 512 pixels.

3.2. Model development for detection

For disease detection, we focused on single-stage object detection model with fast execution speed that is appropriate for real-time object detection and can possibly be used in clinical practice because the endoscopic image consist of video frame images rather than still images.

CenterNet was shown to work more simply and efficiently by predicting both key points and bounding boxes of objects in images at the same time instead of sliding anchors that compute image features by identifying possible bounding boxes [5]. Because it has recently demonstrated excellent performance in real-time target detection, we applied CenterNet to endoscopic disease detection. Our CenterNet-based EDD detection model predicts the center points of the lesions, offsets to the x and y axes, and the width and height of bounding boxes.

The backbone architecture of our detection model is a ResNet50 [6] model pre-trained on the PASCAL VOC 2012 and EDD2020 datasets [4] for multiclass classification. We fine-tuned this detection model with the following training options. The batch size and epoch were 8 and 150 times, respectively, and the initial learning rate was 5e-4 and divided by 10 after every 80 epochs. The input image size was 512 and the test image was restored to its original size by applying an affine transformation. The threshold of the confidence score was set to 0.2.

3.3. Model development for segmentation

For disease segmentation, we modified the decoder part of Vanilla U-Net [7] to build a multi- class segmentation model that can infer independent result for each class. Because some classes overlap with other disease classes in the EDD2020 data, it would be inappropriate to implement general multi class segmentation that constitutes the final layer as softmax operation. Therefore, we replaced the final layer of vanilla U-Net with class-wise binary segmentation branches for multi class segmentation. As shown in Fig 1, we designed a branch structure in which the last up-convolution layer of U-Net performed segmentations for each class independently. Through these branches, the class-wise binary segmentation model was trained by dice similarity coefficient loss. The same backbone architecture used for the detection model was used for of our segmentation model. Training of our segmentation model was carried out with batch size of 4 and 150 epochs, and the initial learning rate was 5e-4 and divided by 10 after every 80 epochs.

4. RESULTS

For the 43 test set images, our model showed mean average precision of 0.1932 ± 0.0622 in detection, and semantic score of 0.2544 ± 0.2080 in segmentation.

5. DISCUSSION & CONCLUSION

EndoCV2020 is an annual global competition for detecting and segmenting lesions of endoscopic images from gastrointestinal organs. We developed deep learning models for each task. The detection model achieved mean average precision of 0.1932 ± 0.0622 and the segmentation model achieved semantic score of 0.2544 ± 0.2080 in the test dataset.

The challenging problem was extremely small data size. Only 386 images were given as a training set to classify and localize 5 imbalanced classes. Even suspicious class literally comprised unclear regions that endoscopists could not define. To overcome this problem, the images of minority classes from the training set were oversampled to balance with other classes, and all images were augmented through various image preprocessing techniques.

Further research is required to develop an artificial intelligence model that can fulfill the standard for practical endoscopic examination.

6. REFERENCES

[1] Nam Hee Kim, Yoon Suk Jung, Woo Shin Jeong, Hyo-Joon Yang, Soo-Kyung Park, Kyuyong Choi, and Dong Il Park. Miss rate of colorectal neoplastic polyps and risk factors for missed polyps in consecutive colonoscopies. *Intestinal research*, 15(3):411, 2017.

[2] Pu Wang, Tyler M Berzin, Jeremy Romek Glissen Brown, Shishira Bharadwaj, Aymeric Becq, Xun Xiao, Peixi Liu, Liangping Li, Yan Song, Di Zhang, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*, 68(10):1813–1819, 2019.

[3] Lianlian Wu, Jun Zhang, Wei Zhou, Ping An, Lei Shen, Jun Liu, Xiaoda Jiang, Xu Huang, Ganggang Mu, Xinyue Wan, et al. Randomised controlled trial of wisense, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut*, 68(12):2161–2169, 2019.

[4] Sharib Ali, Noha Ghatwary, Barbara Braden, Dominique Lamarque, Adam Bailey, Stefano Realdon, Renato Cannizzaro, Jens Rittscher, Christian Daul, and James East. Endoscopy disease detection challenge 2020. *arXiv preprint arXiv:2003.03376*, 2020.

[5] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet: Keypoint triplets for object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6569–6578, 2019.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. arxiv 2015. *arXiv preprint arXiv:1512.03385*, 2015.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.