Boundary-compensated Lesion Aware Transformer for Polyp Segmentation

Yuan Ju, Bingxin Wei, Guoqi Liu*, Haewoon Nam
tangerine@hanyang.ac.kr, wei0911@hanyang.ac.kr, liuguoqi080408@163.com, hnam@hanyang.ac.kr
Hanyang University, Henan Normal University*

Abstract

In recent years, the mortality rate due to colorectal cancer has been increased significantly. Moreover, colorectal cancer is mainly caused by polyps. Although deep learning has made significant progress in the field of polyp segmentation in recent years, polyp segmentation is still a challenging task because the polyp is very similar to the background resulting in very blurred boundaries. In this paper, we add the boundary guidance module to make the network more focused on the lesion region, which can reduce the problem of attention divergence. We conducted tests on five public datasets, CVCClinicDB, Kvasir, CVC-300, CVC-ColonDB and ETIS. The results show that our network outperforms the current mainstream networks on the five benchmark datasets.

Ⅰ. Introduction

Colorectal cancer is one of the most deadly cancers in the world. Colorectal adenomatous polyps are the initial manifestation of its disease. The earlier polyps are detected and removed, the better the chance of reducing the incidence and mortality of colorectal cancer.

Current deep learning based polyp segmentation work relies mainly on CNN or Transformer. Olaf Ronneberger et al. propose UNet [1] for medical image segmentation, which makes creative use of an encoder-decoder structure and uses skip connections to compensate for missing contextual information. PraNet [2] first roughly predicts the lesion area and then uses a reverse attention mechanism to mimic the clinician's labeling process for colon polyps to obtain fine boundary cues for segmenting polyp. Transformer has the ability to capture global information better than CNN, but it has the problem of attention scattering. SSFormer proposed [3] by Qiming Huang et al. design the LE module to enhance local features and the SFA module for feature stepwise fusion. Wenchao Zhang propose HSNet [4] and designed the CSA module to filter the noise information.

In this paper, we propose Boundary-compensated Lesion Aware Transformer (BLAFormer) to first determine the general area of the lesion, and then use the boundary attention map to determine the details of the lesion area.

Ⅱ. Method

An overview of our proposed model is shown in Fig. 1. We use PVT v2 [5] as encoder to capture different levels of features (i.e., E_1, E_2 ,E_3, E_4). We use a dual decoder structure, one for generating boundary

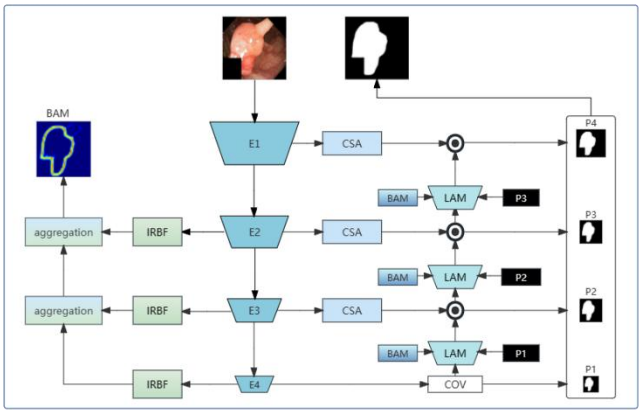

Fig. 1. An overview of BLAFormer

attention map (BAM) and use it to help the second decoder to generate the final polyp prediction map.

Inspired by BDGNet [6], first we put the last three layers of the encoder output into the Receptive Field Block (RBF) and aggregation to get boundary attention map is. In the lesion aware module (LAM), we first use cross-attention with the prediction map from the previous decoder layer, followed by multiplication with the boundary attention map (BAM). Finally, two more convolution operations are performed. This is in line with the physician's habit in clinical diagnosis, where the approximate location of the lesion area is first determined, and then its fine boundary is determined.

The output of each decoder layer and BAM are under supervision by the corresponding down-sampled ground truth. The loss function is shown below :

$$L_{total} = L_{bam} + L_{segi}, \mathrm{i} \in \{1,2,3,4\} \quad (1)$$

$$L_{bam} = \sum_{(i,j)} (b_{ij} - \hat{b}_{ij})^2 \bullet [(b_{ij} - \hat{b}_{ij})^2 > \lambda] \quad (2)$$

$$L_{seg} = L_{wbce} + L_{wiou} \quad (3)$$

Methods	Kvasir	CVC-ClinicDB	CVC-ColonDB	CVC-300	ETIS-Lairb
PraNet[2]	0.894	0.899	0.712	0.871	0.628
CaraNet[7]	0.880	0.871	0.750	0.886	0.647
LDNet[8]	0.912	0.921	0.780	0.861	0.701
SSForme[3]	0.917	0.916	0.772	0.887	0.767
BLAFormer	0.923	0.931	0.820	0.893	0.809

Table 1: Comparison of method proposed in this paper with other mainstream methods for mDice on five public datasets

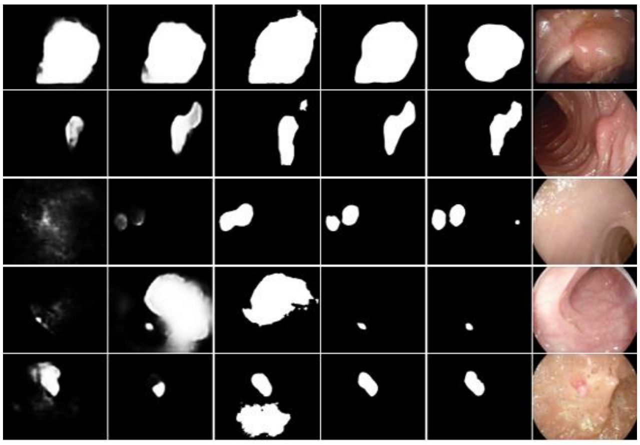

Fig. 2. Visual comparison of our method with 3 other segmentation methods with the dataset ETIS-LaribPolypDB

We validate the feature modeling capability of the proposed framework using the Kvasir-SEG and ClinicDB datasets. Kvasir-SEG is divided into 900 training images and 100 test images, and ClinicDB is divided into 548 training images and 64 test images. Similarly, we validated the generalization performance of our method on unseen ColonDB, 300, and ETIS datasets. The results are shown in Table 1 and Fig. 2.

Ⅲ. Conclusion

In this work, the boundary-compensated lesion aware transformer imitates the clinical diagnosis method of doctors, which determines the general area of the lesion part before determining the detailed boundary to achieve a better segmentation effect and can effectively reduce the distraction. Experimental results show that our method also performs well on different data sets. In future work, we will further investigate how to improve the model's ability to integrate different information.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (2022R1A2C1011862).

REFERENCES

[1] Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp.234-241). Springer, Cham.

[2] Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J. and Shao, L., 2020, October. Pranet: Parallel reverse attention network for polyp segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 263-273). Springer, Cham.

[3] Wang, J., Huang, Q., Tang, F., Meng, J., Su, J., & Song, S. (2022). Stepwise Feature Fusion: Local Guides Global. International Conference on Medical Image Computing and Computer-Assisted Intervention.

[4] Zhang, W., Fu, C., Zheng, Y., Zhang, F., Zhao, Y., & Sham, C. (2022). HSNet: A hybrid semantic network for polyp segmentation. Computers in biology and medicine, 150, 106173 .

[5] Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P. and Shao, L., 2022. Pvt v2: Improved baselines with pyramid vision transformer. *Computational Visual Media*, *8* (3), pp.415-424.

[6] Qiu, Z., Wang, Z., Zhang, M., Xu, Z., Fan, J., & Xu, L. (2022). BDG-Net: boundary distribution guided network for accurate polyp segmentation. Medical Imaging.

[7] Lou, A., Guan, S., & Loew, M.H. (2021). CaraNet: context axial reverse attention network for segmentation of small medical objects. Medical Imaging.

[8] Zhang, R., Lai, P., Wan, X., Fan, D., Gao, F., Wu, X., & Li, G. (2022). Lesion-Aware Dynamic Kernel for Polyp Segmentation. MICCAI.