Boundary Refinement Network for Colorectal Polyp Segmentation in Colonoscopy Images

Guanghui Yue, *Member, IEEE*, Yuanyan Li, Wenchao Jiang, Wei Zhou, and Tianwei Zhou

Abstract—Precise polyp segmentation is vitally essential for detection and diagnosis of early colorectal cancer. Recent advances in artificial intelligence have brought infinite possibilities for this task. However, polyps usually vary greatly in shape and size and contain ambiguous boundary, bringing tough challenges to precise segmentation. In this letter, we introduce a novel Boundary Refinement Network (BRNet) for polyp segmentation. To be specific, we first introduce a boundary generation module (BGM) to generate boundary map by fusing both low-level spatial details and high-level concepts. Then, we utilize the boundary-guided refinement module to refine the polyp-aware features at each layer with the help of boundary cues from the BGM and the prediction from the adjacent high layer. Through top-down deep supervision, our BRNet can localize the polyp regions accurately with clear boundary. Extensive experiments are carried out on five datasets, and the results indicate the effectiveness of our BRNet over seven recently reported methods.

Index Terms—Deep neural network, polyp segmentation, boundary refinement, colonoscopy image.

I. INTRODUCTION

COLORECTAL cancer (CRC) is a severe disease [1] that originates from abnormal tissue hyperplasia in the colon, gradually evolving from polyps. Early detection and removal of polyps helps to prevent the development of CRC [2]. Segmentation algorithms that are conducive to lesion detection and localization, is of great importance for clinics, attracting increasing attention in recent years [3]–[6].

Given that the polyp and non-polyp regions usually differ in appearances, traditional methods primarily detect polyps via analysis of the structural, textural, and color characteristics using hand-crafted features [7], [8]. Due to the limited representation abilities of hand-crafted features, such methods usually result in missed and false detection, especially when coping with challenging cases. Generally, precise polyp segmentation in colonoscopy images faces two main challenges: 1) Polyps with diverse shapes and sizes are easily overlooked. 2) It is difficult to distinguish polyps with ambiguous boundaries against normal tissue background.

G. Yue and Y. Li are with the School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen 518060, China (email: yueguanghui@szu.edu.cn; 2021220012@email.szu.edu.cn).

W. Jiang is with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China (e-mail: jiangwenchao@gdut.edu.cn).

W. Zhou is with the School of Computer Science and Informatics, Cardiff University, UK (e-mail: zhouw26@cardiff.ac.uk).

T. Zhou is with the College of Management, Shenzhen University, Shenzhen 518060, China (e-mail: tianwei@szu.edu.cn).

Recently, advances in convolutional neural networks (CNNs) have enabled many CNN-based methods for medical image segmentation. Representative works are U-Net [9] and its variants, such as U-Net++ [10], CENet [11], and HF-UNet [12]. Inspired by the success of U-Net families, many specific polyp segmentation methods based on U-Net have been proposed. Patel *et al.* [13] enhanced the polyp representation capability using a semantic feature enhancement module and introduced an adaptive global context module to help the network focus on hard fine-grained features. Likewise, Mahmud *et al.* [14] embedded several sequential depth dilated inception blocks in the encoder and updated the traditional skip connection with a newly designed deep fusion skip module. Wu *et al.* [15] applied a well-designed multi-scale transformer attention mechanism to the U-shape structure, obtaining an impressive performance on the polyp segmentation task with a proper balance between inference speed and accuracy.

Despite achieving great progress, most methods mainly focus on the polyp region while ignoring important boundary information. To address this limitation, some recent works have focused on improving boundary representation. One simple idea is adaptively mining boundary cues using a well-designed module, e.g., the reverse module [16], the boundary-aware attention module [17], and the uncertainty exploration module [18]. Another way is taking boundary prediction as an auxiliary task of polyp segmentation. To generate the boundary information, specific module is introduced by analyzing the features of the first layer [19], fusing the features of the first two layers [20], fusing the features of the last three layers [21], and fusing the features of all layers [22]. Additionally, boundary-aware data augmentation strategy is also used [23]. Although most methods have been validated effective, they usually produce poor results when handling polyps with complex shapes, diverse sizes, and blurred boundary. Some methods have unsatisfactory inference speeds due to the complex network structure. Moreover, how to effectively incorporate the extracted boundary cues with the network for further performance improvement remains an open question.

This letter proposes a novel Boundary Refinement Network (BRNet) for precise polyp segmentation. Specifically, a boundary generation module (BGM) is utilized to generate the boundary map by fusing multi-level features. Additionally, a boundary-guided refinement module (BRM) is introduced to refine the polyp-aware features at each layer under the guidance of the extracted boundary map from BGM and the segmentation result from the adjacent high layer. Experimental results indicate that our BRNet works well on the polyp segmentation task and obtains superior performance over seven

recent methods on five datasets. The contributions of this study are three-fold: 1) Considering the blurred boundary of polyps, we propose a simple yet effective network by leveraging boundary-aware information for precise segmentation; 2) Considering the diversity of polyp size and shape, we propose a BGM to fuse multi-level features for capturing the boundary cues; 3) We propose a new module for refining the polyp-aware features at each layer.

II. METHODS

A. Overview of Architecture

Fig. 1 illustrates the framework of BRNet, which is similar to the classical encoder-decoder based U-Net. BRNet takes PVTv2-B2 [24] as the backbone and incorporates it with two key components, i.e., BGM and BRM. To be specific, an image is first fed into the backbone to obtain multi-level features F_i ($i = 1, 2, 3, 4$). Then, F_2, F_3 and F_4 are fused to produce a coarse map S_5, which serves as one of the guidance maps of BRM. During feature fusion, all these three features are first processed by a 1×1 convolution operation and two 3×3 convolution operations to adjust their channel size to 32. After that, the resultant features are fed into the aggregation module introduced by PraNet [16] to output S_5. The reason why we do not consider F_1 is that the first layer usually consumes more computational resources due to the larger spatial resolutions than other layers, but contributes less to performance [25]. Additionally, F_2, F_3 and F_4 are also fed into BGM to generate a boundary map, which serves as another guidance map for BRM. Then, the two guidance maps help to refine polyp-aware information in BRM. Finally, we can obtain three more segmentation maps S_i ($i = 2, 3, 4$) with different resolutions from three BRMs, and take S_2 as the final segmentation result.
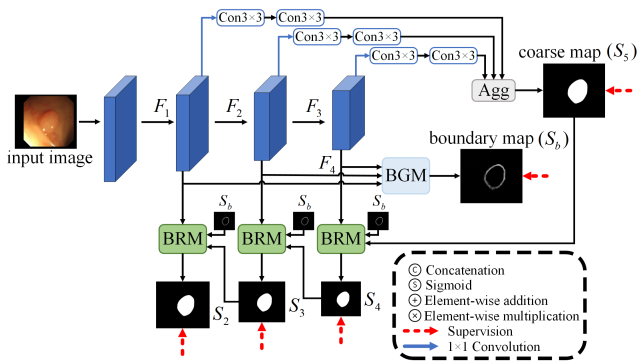

Fig. 1. Framework of the proposed boundary refinement network.

B. Boundary Generation Module (BGM)

Accurate boundary extraction is crucial for the segmentation task. Generally, low-level features contain rich local boundary details, but lack global contextual information. In contrast, high-level features characterize overall image semantics, but have less discrimination of local boundary details. In view of these, we propose a BGM that combines both low-level and high-level features to generate precise boundaries.

As illustrated in the left part of Fig. 2, F_2, F_3 and F_4 first perform feature adjustments using a 1×1 convolution operation $Con_1(\cdot)$. After that, the resultant features are processed by an up-sampling operation (if needed) to resize their spatial sizes to that of F_2 for convenient concatenation:

$$
\begin{cases}
F_2' = Con_1(F_2), \\
F_3' = (Con_1(F_3)) \uparrow_2, \\
F_4' = (Con_1(F_4)) \uparrow_4 .
\end{cases}
\tag{1}
$$

where $\uparrow_2$ and $\uparrow_4$ denote the up-sampling operation with the up-sampling rate of 2 and 4, respectively. Then, the resultant features F_2', F_3', and F_4' are concatenated along the channel direction and the concatenated feature is further processed by a 1×1 convolution operation, resulting in a fused feature F:

$$
F = Con_1(\copyright(F_2', F_3', F_4'))
\tag{2}
$$

where $\copyright(\cdot, \cdot, \cdot)$ is the concatenation operation. Finally, we feed F into a convolutional block (denoted as **CB**) to get the boundary map S_b. The **CB** includes a 1×1 convolution operation, two 3×3 convolution operations, a ReLU activation function, and a 1×1 convolution operation.
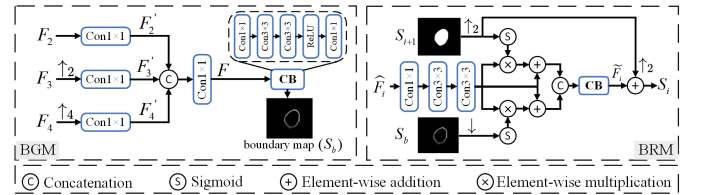

Fig. 2. Structure of the proposed boundary generation module (left) and boundary-guided refinement module (right). Here, $\uparrow$ and $\downarrow$ denote the up-sampling and down-sampling operations, respectively.

C. Boundary-guided Refinement Module (BRM)

Traditional methods usually extract and refine lesion-aware features at a single layer, which limits their abilities to use complementary information across different layers. In this study, we propose the BRM to obtain polyp-aware feature representations by introducing boundary guidance.

As shown in the right part of Fig. 2, BRM takes F_i of the current layer, the prediction S_{i+1} of the adjacent high layer, and the boundary map S_b as the inputs and outputs the prediction S_i of the current layer. Specifically, F_i is first processed by two 3×3 convolution operations, and the resultant feature $\widehat{F}_i$ is element-multiplied with S_{i+1} and S_b, respectively. With such operations, the feature is refined to focus more on the polyp-aware regions. Note that, since S_{i+1} and S_b have different sizes to $\widehat{F}_i$, they should be processed by an up-sampling or down-sampling operation before the multiplication operation. We use the bilinear interpolation for these re-sampling operations. Then, each result of the product is added to $\widehat{F}_i$, and the results of addition are concatenated. Finally, we process the concatenated feature with a convolutional block to obtain an intermediate prediction. We then add the intermediate prediction with S_{i+1} to obtain the prediction S_i of the current layer:

$$
S_i = \mathcal{CB}(\copyright(\widehat{F}_i \otimes \sigma(S_{i+1}) + \widehat{F}_i, \widehat{F}_i \otimes \sigma(S_b) + \widehat{F}_i)) + S_{i+1}, \tag{3}
$$

where $\mathcal{CB}(\cdot)$ denotes the operation of the convolutional block, $\otimes$ means the element-wise multiplication, $\sigma(\cdot)$ is the Sigmoid activation function.

D. Loss Function

The overall loss function of the proposed multi-task polyp segmentation method includes two types of supervision, each of which corresponds to a single task:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_B^{\omega}(S_b, G_b) + \sum_{i=2}^{5} \mathcal{L}_s(S_i, G), \quad (4)$$

where α is the weight to balance the two terms, and we set it to 1 based on the ablation studies in Section III-C. $\mathcal{L}_B^{\omega}$ is the weighted binary cross-entropy loss for polyp boundary segmentation. $\mathcal{L}_s$, formulated as the linear combination of $\mathcal{L}_B^{\omega}$ and weighted intersection over union (IoU) loss, is the loss for polyp region segmentation. We place $\mathcal{L}_s$ and $\mathcal{L}_B^{\omega}$ next to the dotted arrows to indicate supervision in Fig. 1. G and G_b are the ground truth masks of the polyp region and boundary. Since most datasets do not provide the boundary mask, we apply the Canny edge detector to obtain this information.

E. Implementation Details

We adopt PVTv2-B2 [24] as backbone of BRNet, which is pre-trained on ImageNet [26]. To optimize BRNet, we utilize the Adam algorithm with an initial learning rate of $5e - 4$. BRNet is trained for 50 epochs under the batch size of 32. We decay the learning rate by a rate of 0.1 every 20 epoch. To reduce over-fitting, several data augmentation strategies are used, i.e., random horizontal flipping/rotation. Moreover, following PraNet [16] we resize the image to 352×352, and utilize a multi-scale training strategy, which further adjusts the image sizes by randomly choosing a scale from $\{0.75, 1, 1.25\}$ to train the network. During the inference stage, we merely resize the images to 352×352 with no data augmentation.

III. EXPERIMENTS

A. Experimental Details

1) Datasets: For model evaluation, we choose five public datasets, including Kvasir [27], CVC-ClinicDB [28], CVC-ColonDB [32], ETIS [1], and CVC-300 [33]. The same data settings of PraNet [16] are used in this study. Specifically, 90% of images from CVC-ClinicDB and Kvasir are used to train the model, while the remaining 10% of images from these two datasets and partial images of other three datasets are used to test the model.

2) Evaluation Metrics: Six metrics are selected to quantitatively evaluate the segmentation ability of our proposed model, including mean dice score ($mDice$), mean IoU ($mIoU$), weighted F-measure (F_{β}^{ω}) [34], E-measure (E_{ϕ}^{max}) [35], S-measure (S_{α}) [36], and mean absolute error (MAE). $mDice$ and $mIoU$ measure the regional similarity between the prediction and ground truth, while S_{α} evaluates the structural similarity between them. F_{β}^{ω} evaluates the performance by considering both recall and precision. E_{ϕ}^{max} computes the results at the image and pixel levels. MAE calculates the mean of the absolute error between the prediction and ground truth. Higher values of the first five metrics yet smaller value of MAE shows a better segmentation performance.
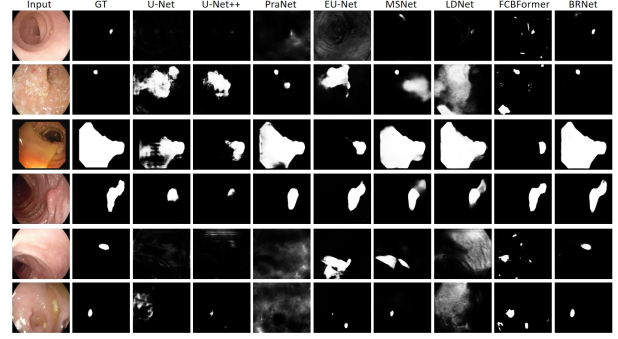

Fig. 3. Qualitative results of different methods. Zoom-in for more details.

B. Results

1) Quantitative Results: To validate the superiority of our BRNet, we compare it with 7 methods, including U-Net [9], U-Net++ [10], PraNet [16], EU-Net [13], MSNet [29], LDNet [30], and FCBFormer [31]. TABLE I and TABLE II show the results, in which the best results are marked in bold.

The results of in-domain tests on Kvasir and CVC-ClinicDB are presented in TABLE I. As shown, our BRNet outperforms competing methods in most of metrics. For instance, BRNet surpasses FCBFormer [31], the second best method, by approximately 1.4%, 2.5%, and 1.3% in terms of $mDice$, $mIoU$, and F_{β}^{ω} on Kvasir. Compared to PraNet, BRNet achieves 1.9%, 2.0%, 1.6%, and 1.3% improvements in terms of $mDice$, $mIoU$, F_{β}^{ω}, and S_{α} respectively on CVC-ClinicDB. Also, it is clear that our BRNet has relatively smaller standard deviation values compared to other methods. These results indicate that BRNet is more competent for accurate polyp segmentation than other methods under different scenarios. TABLE II presents the results of three out-of-domain tests. As seen, BRNet has the top performance across datasets. Compared to LDNet, BRNet achieves 1.5%, 1.2%, 2.1%, 1.6% and 0.7% gains in terms of $mDice$, $mIoU$, F_{β}^{ω}, E_{ϕ}^{max} and S_{α}, respectively, on CVC-ColonDB. Similarly, in terms of $mDice$, $mIoU$, F_{β}^{ω}, E_{ϕ}^{max}, S_{α} and MAE, BRNet achieves 4.2%, 3.0%, 3.9%, 3.1%, 1.7% and 0.3% improvements over MSNet on ETIS, which is one of the most challenging datasets. Additionally, BRNet is ahead of FCBFormer by 1.4%, 2.4%, 2.4%, and 1.9% in terms of $mDice$, $mIoU$, F_{β}^{ω}, and S_{α} on CVC-300. These findings confirm the superior generalization ability of our BRNet. A possible reason of above results is that our BRNet mines and leverages the boundary cues through two well-designed modules to improve the representation ability of polyps.

2) Qualitative Results: Fig. 3 provides a visual comparison of segmentation results among seven methods under different lesion scenarios. When it comes to small polyps, shown in the $1st$ and $2nd$ rows, BRNet achieves more precise segmentation while some competing methods fail to detect small polyps. The irregular shapes and large sizes of the polyps (see the $3rd$ and $4th$ rows) present tough challenges for accurate segmentation. Despite this, BRNet detects the polyp regions more accurately than other methods. As presented in the $5th$ and $6th$ rows, almost all competing methods fail to recognize the polyps with blurred boundaries, but our BRNet still shows excellent performance and produces continuous and clear boundaries. Overall,

TABLE I
QUANTITATIVE RESULTS (MEAN ± STANDARD DEVIATION ACROSS TEST IMAGES) ON KVASIR AND CVC-CLINICDB.

Methods	Kvasir [27]						CVC-ClinicDB [28]					
	$mDice$	$mIoU$	F_β^ω	E_ϕ^{max}	S_α	MAE	$mDice$	$mIoU$	F_β^ω	E_ϕ^{max}	S_α	MAE
U-Net [9]	0.820±0.152	0.755±0.151	0.793±0.155	0.898±0.091	0.857±0.078	0.055±0.020	0.824±0.177	0.767±0.174	0.811±0.185	0.917±0.118	0.889±0.092	0.019±0.007
U-Net++ [10]	0.823±0.152	0.752±0.155	0.808±0.154	0.906±0.780	0.861±0.081	0.048±0.023	0.797±0.166	0.741±0.172	0.785±0.167	0.898±0.092	0.872±0.087	0.022±0.008
PraNet [16]	0.901±0.091	0.848±0.093	0.885±0.094	0.943±0.054	0.915±0.049	0.030±0.014	0.902±0.111	0.858±0.111	0.896±0.115	0.957±0.069	0.935±0.060	0.009±0.010
EU-Net [13]	0.877±0.081	0.822±0.089	0.847±0.093	0.936±0.052	0.897±0.046	0.039±0.015	0.881±0.105	0.832±0.115	0.869±0.120	0.957±0.067	0.913±0.062	0.018±0.011
MSNet [29]	0.907±0.080	0.862±0.085	0.892±0.088	0.944±0.050	0.923±0.043	0.028±0.017	0.921±0.102	**0.880±0.105**	**0.913±0.113**	**0.971±0.068**	0.942±0.053	0.008±0.007
LDNet [30]	0.897±0.085	0.854±0.090	0.879±0.092	0.942±0.052	0.920±0.048	0.027±0.015	0.903±0.110	0.860±0.115	0.892±0.117	0.960±0.066	0.938±0.065	0.011±0.014
FCBFormer [31]	0.904±0.080	0.846±0.089	0.892±0.092	**0.958±0.048**	0.908±0.045	0.027±0.015	0.860±0.092	0.806±0.101	0.851±0.107	0.939±0.057	0.899±0.055	0.020±0.011
BRNet (ours)	**0.918±0.053**	**0.871±0.059**	**0.905±0.061**	0.957±0.037	**0.928±0.031**	0.026±0.015	**0.921±0.067**	0.878±0.073	0.912±0.078	0.967±0.046	**0.948±0.039**	**0.008±0.005**

TABLE II
QUANTITATIVE RESULTS (MEAN ± STANDARD DEVIATION ACROSS TEST IMAGES) ON CVC-COLONDB, ETIS AND CVC-300.

Methods	CVC-ColonDB [32]					
	$mDice$	$mIoU$	F_β^ω	E_ϕ^{max}	S_α	MAE
U-Net [9]	0.511±0.124	0.440±0.122	0.491±0.128	0.759±0.080	0.710±0.068	0.059±0.021
U-Net++ [10]	0.489±0.120	0.411±0.126	0.467±0.121	0.761±0.065	0.692±0.070	0.061±0.024
PraNet [16]	0.716±0.077	0.645±0.077	0.699±0.079	0.847±0.048	0.820±0.043	0.043±0.014
EU-Net [13]	0.744±0.074	0.661±0.080	0.710±0.086	0.870±0.049	0.831±0.044	0.042±0.016
MSNet [29]	0.755±0.072	0.678±0.076	0.736±0.080	0.883±0.049	0.836±0.040	0.041±0.017
LDNet [30]	0.780±0.078	0.711±0.081	0.755±0.084	0.876±0.047	0.859±0.046	0.032±0.015
FCBFormer [31]	0.754±0.065	0.668±0.071	0.724±0.076	0.875±0.040	0.828±0.039	0.037±0.015
BRNet (ours)	**0.795±0.048**	**0.723±0.053**	**0.776±0.056**	**0.892±0.030**	**0.866±0.03**	**0.031±0.014**

Methods	ETIS [1]					
	$mDice$	$mIoU$	F_β^ω	E_ϕ^{max}	S_α	MAE
U-Net [9]	0.406±0.144	0.343±0.137	0.366±0.147	0.645±0.104	0.682±0.076	0.036±0.006
U-Net++ [10]	0.413±0.133	0.342±0.136	0.390±0.130	0.704±0.085	0.681±0.070	0.035±0.008
PraNet [16]	0.630±0.096	0.576±0.089	0.600±0.094	0.792±0.061	0.791±0.052	0.031±0.009
EU-Net [13]	0.651±0.088	0.575±0.088	0.594±0.095	0.803±0.055	0.788±0.052	0.035±0.010
MSNet [29]	0.718±0.079	0.666±0.075	0.677±0.085	0.828±0.055	0.840±0.043	0.020±0.006
LDNet [30]	0.659±0.091	0.596±0.087	0.614±0.090	0.795±0.054	0.797±0.053	0.043±0.013
FCBFormer [31]	0.663±0.093	0.581±0.098	0.608±0.103	0.814±0.057	0.789±0.056	0.027±0.006
BRNet (ours)	**0.760±0.056**	**0.696±0.057**	**0.716±0.064**	**0.859±0.041**	**0.857±0.033**	**0.017±0.005**

Methods	CVC-300 [33]					
	$mDice$	$mIoU$	F_β^ω	E_ϕ^{max}	S_α	MAE
U-Net [9]	0.717±0.051	0.639±0.053	0.684±0.056	0.867±0.039	0.842±0.029	0.022±0.004
U-Net++ [10]	0.714±0.048	0.636±0.050	0.687±0.050	0.884±0.022	0.838±0.027	0.018±0.004
PraNet [16]	0.873±0.029	0.804±0.038	0.843±0.039	0.938±0.023	0.924±0.015	0.010±0.003
EU-Net [13]	0.856±0.023	0.782±0.029	0.817±0.030	0.932±0.018	0.909±0.012	0.011±0.003
MSNet [29]	0.869±0.028	0.808±0.030	0.848±0.030	0.942±0.019	0.925±0.012	0.010±0.003
LDNet [30]	0.862±0.020	0.795±0.022	0.828±0.023	0.923±0.017	0.920±0.009	0.010±0.003
FCBFormer [31]	0.884±0.024	0.812±0.032	0.854±0.031	0.953±0.015	0.922±0.013	0.008±0.003
BRNet (ours)	**0.898±0.020**	**0.836±0.026**	**0.878±0.026**	**0.956±0.016**	**0.941±0.009**	**0.007±0.002**

our BRNet performs better than these competing methods when facing diverse challenging factors and complex situations in polyp segmentation. Through observation, we have found that our BRNet may also fail in some special situations, e.g., polyps with extremely flat appearance (see the left part of Fig. 4) and polyp with camouflage characteristics (see the right part of Fig. 4). In the future, we will improve the ability to differentiate the polyps from background through uncertainty quantification. Also, more boundary extraction methods will be used to further enhance the boundary representation.
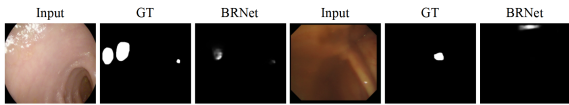

Fig. 4. Illustration of failure cases.

C. Ablation Study

Here, we quantify the influence of each module, i.e., BGM and BRM, through ablation experiments. All experiments are carried out under the same settings as described in Section II-E. The results are shown in TABLE III, in which only the values of $mDice$ and $mIoU$ are given due to space limitation.

1) Effectiveness of BGM: BGM aims to strengthen the model's ability to learn and extract boundary information. Here, we remove BGM while keeping other parts unchanged.

In this case, the input S_b of BRM is also removed. As illustrated by the first and third rows of TABLE III, removing BGM leads to obvious performance drop. Concretely, $mDice$ decreases by 3.7% and 5.7%, while $mIoU$ decreases by 3.6% and 5.9% on ETIS and CVC-300, respectively. These indicate the positive role of BGM in accurate segmentation.

2) Effectiveness of BRM: In our BRM, we utilize S_{i+1} and S_b to help to refine the polyp-aware features in an element-wise multiplication operation. Here, we explore the effectiveness of such an operation. For this purpose, we directly concatenate F_i, S_{i+1}, and S_b and use the convolutional block to process the concatenated feature for generating the prediction S_i. As shown in the second and last rows of TABLE III, the performance decreases obviously after changing BRM. For example, it results in decreases of 3.4%, 4.1%, and 4.7% for $mDice$, $mIoU$, and F_β^ω on CVC-300. These results indicates that the design concept of BRM contributes to achieving excellent segmentation performance.

TABLE III
RESULTS OF ABLATION STUDIES ON ETIS AND CVC-300.

Methods	ETIS [1]		CVC-300 [33]	
	$mDice$	$mIoU$	$mDice$	$mIoU$
w/o BGM	0.723	0.660	0.841	0.777
w/o BRM	0.743	0.683	0.864	0.795
BRNet	**0.760**	**0.696**	**0.898**	**0.836**

3) Effectiveness of the Weight α: In Eq. (4), α is used to balance the contribution of the boundary loss and the polyp loss. Here, we investigate its impact by setting it to $\{0.5, 1, 1.5, 2\}$. As shown in Fig. 5, we can obtain relatively better results when $\alpha = 1$. Of course, the performance may be further improved under other settings. We leave the optimal weight selection as a possible future work.
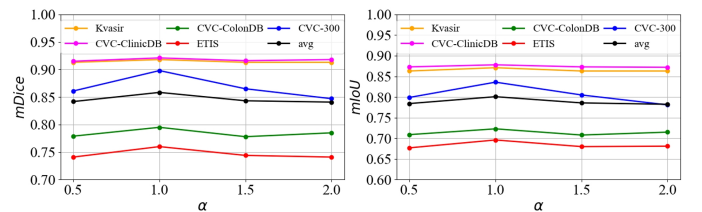

Fig. 5. Results of our method under different settings of α in Eq. (4).

IV. CONCLUSION

In this letter, we propose a novel BRNet for polyp segmentation. Specifically, we first utilize a BGM to generate boundary map by integrating multi-level features. Then, we utilize BRM at each layer to refine polyp-aware features. Experiments on five datasets support the effectiveness of our design concepts. BRNet can achieve over 0.918 $mDice$ and 0.871 $mIoU$ in two in-domain tests and over 0.760 $mDice$ and 0.696 $mIoU$ in three out-of-domain tests.

REFERENCES

[1] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado, "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer," *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, pp. 283–293, 2014.

[2] D. A. Corley, C. D. Jensen, A. R. Marks, W. K. Zhao, J. K. Lee, C. A. Doubeni, A. G. Zauber, J. de Boer, B. H. Fireman, J. E. Schottinger *et al.*, "Adenoma detection rate and risk of colorectal cancer and death," *New England Journal of Medicine*, vol. 370, no. 14, pp. 1298–1306, 2014.

[3] J.-D. Sun, C. Yao, J. Liu, W. Liu, and Z.-K. Yu, "Gnas-u 2 net: A new optic cup and optic disc segmentation architecture with genetic neural architecture search," *IEEE Signal Processing Letters*, vol. 29, pp. 697–701, 2022.

[4] B. Yang, X. Zhang, S. Li, R. Higashita, and J. Liu, "Ha-net: Hierarchical attention network based on multi-task learning for ciliary muscle segmentation in as-oct," *IEEE Signal Processing Letters*, vol. 30, pp. 1342–1346, 2023.

[5] Q. Ming and X. Xiao, "Towards accurate medical image segmentation with gradient-optimized dice loss," *IEEE Signal Processing Letters*, 2023.

[6] W. Li, X. Xiong, S. Li, and F. Fan, "Hybridvps: Hybrid-supervised video polyp segmentation under low-cost labels," *IEEE Signal Processing Letters*, accepted, in press, DOI: 10.1109/LSP.2023.3342613, 2023.

[7] C. Van Wijk, V. F. Van Ravesteijn, F. M. Vos, and L. J. Van Vliet, "Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 688–698, 2010.

[8] S. Y. Park, D. Sargent, I. Spofford, K. G. Vosburgh, A. Yousif *et al.*, "A colon video analysis framework for polyp detection," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1408–1418, 2012.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[10] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.

[11] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[12] K. He, C. Lian, B. Zhang, X. Zhang, X. Cao, D. Nie, Y. Gao, J. Zhang, and D. Shen, "Hf-unet: learning hierarchically inter-task relevance in multi-task u-net for accurate prostate segmentation in ct images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 8, pp. 2118–2128, 2021.

[13] K. Patel, A. M. Bur, and G. Wang, "Enhanced u-net: A feature enhancement network for polyp segmentation," in *2021 18th Conference on Robots and Vision (CRV)*. IEEE, 2021, pp. 181–188.

[14] T. Mahmud, B. Paul, and S. A. Fattah, "Polypsegnet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images," *Computers in Biology and Medicine*, vol. 128, p. 104119, 2021.

[15] H. Wu, Z. Zhao, and Z. Wang, "Meta-unet: Multi-scale efficient transformer attention unet for fast and high-accuracy polyp segmentation," *IEEE Transactions on Automation Science and Engineering*, accepted, in press, DOI: 10.1109/TASE.2023.3292373, 2023.

[16] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2020, pp. 263–273.

[17] N. Ta, H. Chen, Y. Lyu, and T. Wu, "Ble-net: boundary learning and enhancement network for polyp segmentation," *Multimedia Systems*, vol. 29, no. 5, pp. 3041–3054, 2023.

[18] Q. Guo, X. Fang, L. Wang, and E. Zhang, "Polyp segmentation of colonoscopy images by exploring the uncertain areas," *IEEE Access*, vol. 10, pp. 52 971–52 981, 2022.

[19] Y. Fang, C. Chen, Y. Yuan, and K.-y. Tong, "Selective feature aggregation network with area-boundary constraints for polyp segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*. Springer, 2019, pp. 302–310.

[20] T. Zhou, Y. Zhou, K. He, C. Gong, J. Yang, H. Fu, and D. Shen, "Cross-level feature aggregation network for polyp segmentation," *Pattern Recognition*, vol. 140, p. 109555, 2023.

[21] Z. Qiu, Z. Wang, M. Zhang, Z. Xu, J. Fan, and L. Xu, "Bdg-net: boundary distribution guided network for accurate polyp segmentation," in *Medical Imaging 2022: Image Processing*, vol. 12032. SPIE, 2022, pp. 792–799.

[22] Y. Fang, D. Zhu, J. Yao, Y. Yuan, and K.-Y. Tong, "Abc-net: Area-boundary constraint network with dynamical feature selection for colorectal polyp segmentation," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11 799–11 809, 2020.

[23] N.-Q. Nguyen, D. M. Vo, and S.-W. Lee, "Contour-aware polyp segmentation in colonoscopy images using detailed upsampling encoder-decoder networks," *IEEE Access*, vol. 8, pp. 99 495–99 508, 2020.

[24] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 568–578.

[25] Z. Wu, L. Su, and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3907–3916.

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

[27] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*. Springer, 2020, pp. 451–462.

[28] 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," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.

[29] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "Ms-net: multi-site network for improving prostate segmentation with heterogeneous mri data," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2713–2724, 2020.

[30] R. Zhang, P. Lai, X. Wan, D.-J. Fan, F. Gao, X.-J. Wu, and G. Li, "Lesion-aware dynamic kernel for polyp segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 99–109.

[31] E. Sanderson and B. J. Matuszewski, "Fcn-transformer feature fusion for polyp segmentation," in *Annual Conference on Medical Image Understanding and Analysis*. Springer, 2022, pp. 892–907.

[32] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, 2015.

[33] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdzal, A. Courville *et al.*, "A benchmark for endoluminal scene segmentation of colonoscopy images," *Journal of Healthcare Engineering*, vol. 2017, 2017.

[34] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1597–1604.

[35] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment measure for binary foreground map evaluation," *arXiv preprint arXiv:1805.10421*, 2018.

[36] D. Fan, M. Cheng, Y. Liu, T. Li, and A. Borji, "A new way to evaluate foreground maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 245484557, 2017.