Triple-Branch Deep Network for Polyp Image Segmentation

Muwei Jian[a], Yanjie Zhong [a], Kin-man Lam[b]
[a]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China. E-mail: jianmuweihk@163.com
[b]Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

Medical image segmentation is essential for accurately extracting tissue structures or pathological regions from medical images. However, medical image segmentation methods are often influenced by factors such as image noise and irregular shapes, making precise segmentation challenging. To tackle these challenges, this paper proposes a triple-branch medical image segmentation network (TB-Net) that incorporates implicit boundary priors. The boundary map, acquired through a boundary detection algorithm, is used to restrict the results of the boundary branch. Extensive experiments indicate that TB-Net achieves state-of-the-art performance on publicly available polyp datasets.
Keywords: medical image segmentation, boundary prior, polyp segmentation

1. INTRODUCTION

Medical image segmentation is a critical step in medical imaging analysis, enabling the accurate extraction of tissue structures and pathological regions, which is essential for diagnosis and treatment planning. As a vital component of computer-aided detection, medical image segmentation plays an indispensable role in various clinical practices, including tumor detection and organ localization [13,14,15]. It is widely applied in radiological imaging, such as Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI), providing clinicians with deeper insights into pathological areas and forming the basis for more precise treatment plans. With the rapid development of computer-assisted medical image segmentation techniques, this technology is revolutionizing healthcare by offering high-performance automated tools that enhance diagnostic accuracy and efficiency. However, segmentation is often hindered by challenges such as image noise, blurred boundaries, and the inherent complexity of anatomical structures. These issues, often caused by noisy labels and annotation errors, result in unclear boundaries, making precise segmentation difficult [1]. Boundary information is crucial in image segmentation, primarily tasked with detecting and localizing object boundaries by exploiting the discontinuity in intensity or color between different regions in an image [16,17,18]. By capturing significant changes between different regions in the image, boundary information assists segmentation algorithms in more accurately delineating foreground and background areas. This precise differentiation allows segmentation results to more accurately reflect the edge contours of various objects, thereby improving the overall quality of the segmentation outcome.

Effective utilization of boundary information also helps preserve boundary details in the segmentation results, enabling segmented images to maintain the important features of target boundaries without distortion, while ensuring accuracy. However, challenges remain in dealing with fuzzy boundaries, especially when there are noticeable changes in boundary sizes.

To address these limitations, we propose a novel triple-branch network (TB-Net) that optimizes medical image segmentation by incorporating implicit boundary priors. The boundary map, generated through a boundary detection algorithm, guides the boundary branch to focus on refining edges and improving segmentation accuracy, particularly in complex or irregular regions. By capturing subtle boundary features, the network preserves fine structures crucial for medical image analysis. In summary, the main contributions of this study are as follows:

1) We propose a novel medical image segmentation network named TB-Net. The network is composed of a semantic branch, a detail branch, and a boundary branch, fully integrating the advantages of detailed, semantic, and boundary information.

2) We introduce precise boundary features as a boundary prior, which imposes shape constraints on the boundary generated by the boundary branch, enabling the capture of image boundary features.

Name: Yanjie Zhong Email: 222115038@mail.sdufe.edu.cn

International Workshop on Advanced Imaging Technology (IWAIT) 2025, edited by Masayuki Nakajima, Chuan-Yu Chang, Chia-Hung Yeh, Jae-Gon Kim, Kemao Qian, Phooi Yee Lau, Proc. of SPIE Vol. 13510, 135100A
 · 0277-786X · doi: 10.1117/12.3057381

3) Experiments on publicly available polyp datasets show that TB-Net achieves excellent performance.

2. THE PROPOSED TB-NET

Overall Architecture

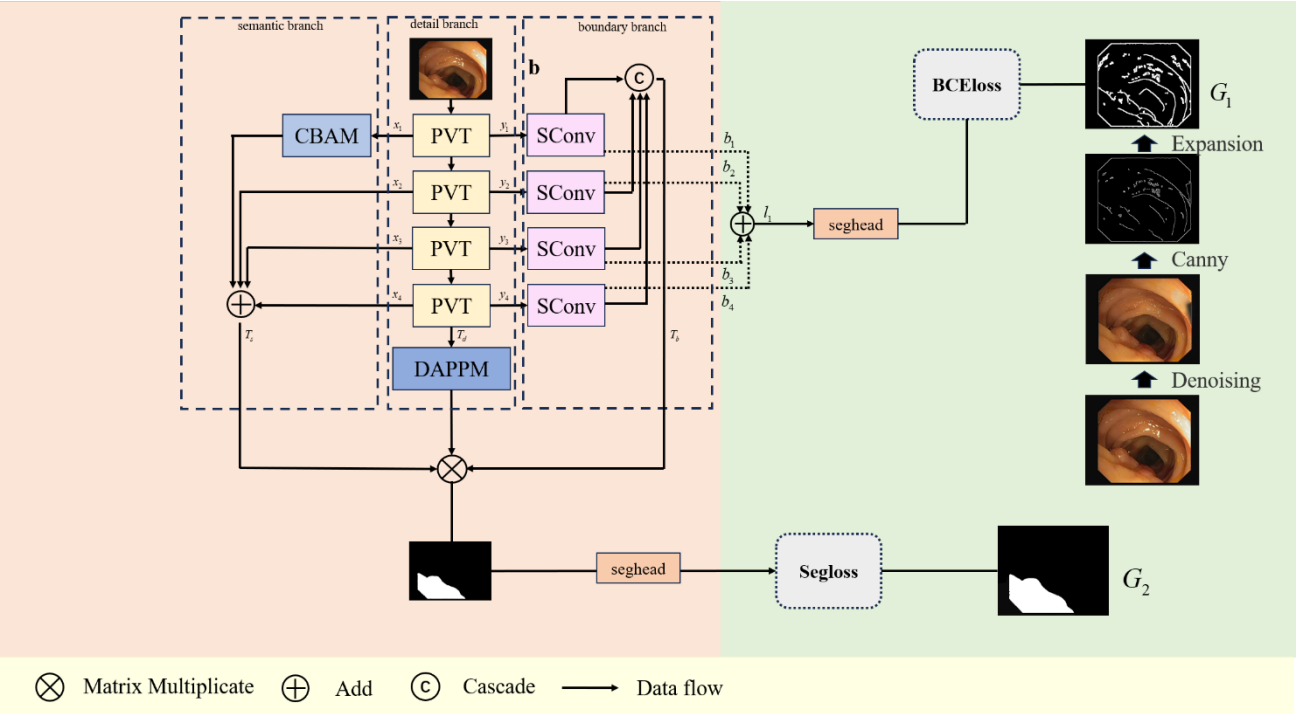

Fig.1. Framework of our proposed TB-Net.

Fig. 1 shows the overall structure of our proposed deep model for polyp image segmentation. Our network consists of three parts: a semantic branch, a detail branch, and a boundary branch. We use PVTv2 [2] as the encoder, which produces feature maps at different resolutions. These feature maps are extracted in the three branches to obtain detail features (T_d), semantic features (T_s), and boundary features (T_b), respectively. These features are then fused to produce the final output. We train our deep model using the segmentation loss between the detail branch output T_d and the final output T_f with the segmentation label G_d. High-resolution images are renowned for their rich textures, semantic information, and clear boundaries. We use the CBAM [19] attention module to extract crucial information from multiple dimensions of high-resolution feature maps. The module can focus on different scales and resolutions, capturing more details and semantic information. We also use the DAPPM [20] module to enrich feature representation to process low-resolution images. Low-resolution images typically contain more detailed information, and the DAPPM module helps extract deep features from the feature maps, allowing them to obtain richer low-resolution multi-scale information. By comprehensively applying different modules, the network can handle high-resolution and low-resolution images more effectively and efficiently, improving image analysis and processing.

Boundary prior

To reduce boundary noises, we first remove bright noise from the original images. Then, we use the Canny operator [3], a boundary detection algorithm, to extract the image boundaries. Observing that the extracted boundary textures are thin, we apply a dilation algorithm to thicken the boundaries, obtaining more prominent boundary maps to serve as our boundary labels.

We introduce the boundary map of training images as an implicit prior to train the network. In our network, we employ the snake convolution (SConv) [4] in the encoder network to generate feature maps at different resolutions, denoted as y_1, y_2, y_3, y_4. SConv is particularly effective in extracting line and boundary structures. The four feature maps are scaled to have the same resolution and are then concatenated. We combine the convolutional feature maps by concatenating them after resizing them to the same dimensions, producing the final boundary map, denoted as T_3. The specific implementation is as follows:

$$T_3 = Concat(y_1, y_2, y_3, y_4), \tag{1}$$

where $Concat(\cdot)$ denotes feature concatenation along the channel dimension. The four boundary maps, i.e., b_1, b_2, b_3, and b_4, at different resolutions are added to form l_1, which is then used to calculate the binary cross entropy loss, as follows:

$$l_1 = b_1 + b_2 + b_3 + b_4, \tag{2}$$

$$\mathcal{L}_1 = \mathcal{L}^{\omega}_{BCE}(l_1, G_1). \tag{3}$$

where L^{ω}_{BCE} denotes the BCE loss.

Overall loss function

We use the common binary cross entropy (BCE) loss [5] to calculate our boundary loss, and use the sum of intersection over union loss and binary cross entropy loss to calculate segment loss. The loss function is as follows:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_1 + \mathcal{L}_2, \\ \mathcal{L}_1 &= \mathcal{L}^{\omega}_{BCE}(l_1, G_1), \\ \mathcal{L}_2 &= \mathcal{L}^{\omega}_{IoU}(T_f, G_2) + \mathcal{L}^{\omega}_{BCE}(T_f, G_2), \end{aligned} \tag{4}$$

where $\mathcal{L}_1$ and $\mathcal{L}_2$ represent the boundary loss and main loss, respectively. $\mathcal{L}^{\omega}_{IoU}(\cdot)$ and $\mathcal{L}^{\omega}_{BCE}(\cdot)$ [38] denote the weighted intersection over union (IoU) loss and weighted binary cross entropy (BCE) loss, respectively.

3. EXPERIMENTAL RESULTS

Experiment Settings

We adopt the industry authoritative datasets for colorectal polyp segmentation training, consisting of 900 samples from Kvasir and 550 samples from CVC-Clinic. For quantitative evaluation, we applied several commonly used metrics: mean Dice and mean IoU. Our model is implemented using PyTorch framework and trained on a single 3090 GPU with a batch size of 16 for 100 epochs. The images are resized to 352×352 for both training and testing. We evaluated the performance of our network on various colorectal image datasets, including CVC-ClinicDB [6], Kvasir [7], and ETIS-LaribPolyDB [8]. The segmentation results on the polyp image datasets are shown in Table 1 and Fig. 2. We compared our TB-Net with U-Net [9], SANet [10], PraNet [11], and MSNet [12].

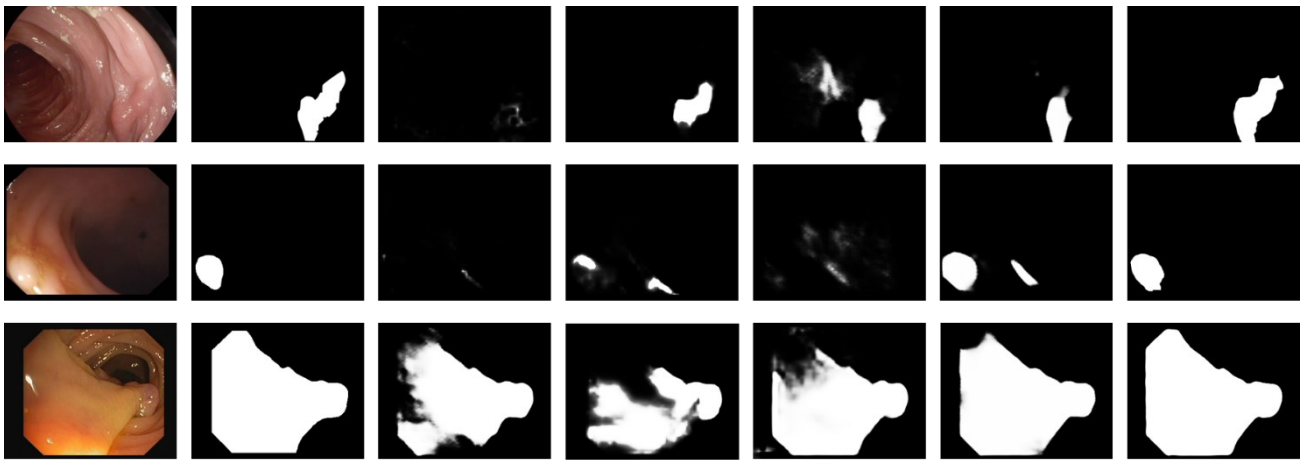

Fig.2. Visual comparison of existing methods on polyp datasets.

Table 1. Quantitative comparison with existing methods on CVC-ClinicDB, Kvasir and ETIS-LaribPolyDB datasets.

	Clinic-DB		ETIS		Kvasir	
	Dice	IoU	Dice	IoU	Dice	IoU
U-Net [9]	0.824	0.761	0.398	0.336	0.821	0.727
SANet [10]	0.909	0.859	0.739	0.661	0.901	0.843
PraNet [11]	0.918	0.838	0.679	0.617	0.897	0.843
MSNet [12]	0.930	0.879	0.734	0.664	0.915	0.816
Ours	**0.938**	**0.895**	**0.797**	**0.714**	**0.922**	**0873**

As shown in Table 1, the experimental results demonstrate that our method achieves the best performance across the Clinic-DB and Kvasir medical image segmentation datasets, showing exceptional segmentation capabilities. Specifically, on the Clinic-DB dataset, our model surpasses other methods with a Dice score of 0.938 and an IoU score of 0.895, significantly outperforming models like U-Net and PraNet. On the more challenging ETIS dataset, our model also stands out with a Dice score of 0.797 and an IoU score of 0.714, showing a substantial improvement over U-Net with a Dice score of 0.398 and an IoU score of 0.336, highlighting its advantage in handling complex images. On the Kvasir dataset, our method once again achieves the best performance, with Dice and IoU scores of 0.922 and 0.873, respectively, further validating its robustness and generalization across different datasets.

Ablation experiments

Table 2. Ablation study on the CVC-ClinicDB, Kvasir, and ETIS-LaribPolyDB datasets.

Components	CVC-ClinicDB		Kvasir		ETIS-LaribPolyDB	
Boundary	Dice↑	IoU↑	Dice↑	IoU↑	Dice↑	IoU↑
×	0.911	0.858	0.905	0.856	0.765	0.682
√	0.938	0.868	0.922	0.873	0.797	0.714

In the CVC-ClinicDB, Kvasir, and ETIS-LaribPolyDB datasets, including boundary information leads to improvements in both Dice and IoU scores to varying degrees. For instance, in the CVC-ClinicDB dataset, the Dice score increases from 0.911 to 0.938, while IoU improves from 0.858 to 0.868. Similarly, in the Kvasir dataset, the Dice score rises from 0.905 to 0.922, and IoU increases from 0.856 to 0.873. Notably, in the more complex ETIS-LaribPolyDB dataset, the Dice score improves from 0.765 to 0.797, and IoU increases from 0.682 to 0.714. These results demonstrate that the incorporation of boundary information effectively enhances the model's ability to capture the edges of target regions, resulting in more accurate and consistent segmentation, particularly in scenes with complex details and boundaries. Overall, boundary information plays a crucial role in improving the accuracy and robustness of the model's segmentation performance.

4. CONCLUSION

In this paper, we have introduced a triple-branch network with implicit boundary priors, which consists of three parts: a semantic branch, a detail branch, and a boundary branch. They effectively extract semantic, detailed, and boundary information from images. Our proposed model provides an innovative solution to medical image segmentation, overcoming the trade-off problem between boundary preservation and segmentation accuracy faced by existing methods.

ACKNOWLEDGEMENTS

This work was supported by Taishan Young Scholars Program of Shandong Province.

REFERENCES

[1] D. Karimi, H. Dou, S. K. Warfield, A. Gholipour, Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis, Medical Image Analysis, 65 (2020) 101759.
[2] W. Wang, E. Xie, X. Li, et al., PVT V2: Improved Baselines with Pyramid Vision Transformer, Computational Visual Media 8 (3) (2022) 415-424.
[3] J. Canny, A Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (6) (1986) 679-698.
[4] Y. Qi, Y. He, X. Qi, G. Yang, Dynamic Snake Convolution Based on Topological Geometric Constraints for Tubular Structure Segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 6070-6079.
[5] J. Wei, S. Wang, Q. Huang, F³Net: fusion, feedback and focus for salient object detection, **i**n: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 12321-12328.
[6] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, et al., WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, IEEE Transactions on Medical Imaging and Graphics 43 (2015) 99-111.
[7] D. Jha, P. H. Smedsrud, M. A. Riegler, et al., Kvasir-seg: A segmented polyp dataset, MultiMedia Modeling 26 (2020) 451-462.
[8] J. Silv31a, A. Histace, O. Romain, X. Dray, B. Granado, Toward Embedded Detection of Polyps in WCE Images for Early Diagnosis of Colorectal Cancer, International Journal of Computer Assisted Radiology and Surgery. 9 (2014) 283-293.
[9] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2015, pp. 234-241.
[10] J. Wei, Y. Hu, R. Zhang, et al., Shallow attention network for polyp segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted (MICCAI), 2021, pp. 699-708.
[11] D. P. Fan, G. P. Ji, T. Zhou, et al., "PraNet: Parallel reverse attention network for polyp segmentation," in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2020, pp. 263-273.
[12] X. Zhao, L. Zhang, H. Lu, Automatic polyp segmentation via multi-scale subtraction network, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2021, pp. 120-130.

[13] A. Norouzi, M. S. M. Rahim, A. Altameem, et al., Medical Image Segmentation Methods, Algorithms, and Applications, IETE Technical Review 31 (3) (2014) 199-213.
[14] G. C. Ates, P. Mohan, E. Celik, Dual Cross-Attention for medical image segmentation, Engineering Applications of Artificial Intelligence, 126 (2023) 107139.
[15] P. Song, Z. Yang, J. Li, H. Fan, DPCTN: Dual path context-aware transformer network for medical image segmentation, Engineering Applications of Artificial Intelligence, 124 (2023) 106634.
[16] Y. Yuan, J. Xie, X. Chen, J. Wang, Segfix: Model-agnostic boundary refinement for segmentation, in: Proceedings of the 16th European Conference on Computer Vision (ECCV), Glasgow, UK, Aug. 23–28, 2020, pp. 489-506.
[17] H. Ma, H. Yang, D. Huang, Boundary Guided Context Aggregation for Semantic Segmentation, British Machine Vision Conference (BMCV), 2021.
[18] J. Wang, F. Chen, Y. Ma, et al., XBound-Former: Toward Cross-Scale Boundary Modeling in Transformers, IEEE Transactions on Medical Imaging (2023).
[19] S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3-19.
[20] Y. Hong, H. Pan, W. Sun, Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Road Scenes, IEEE Transactions on Intelligent Transportation Systems 24 (3) (2023) 3448-3460.