
Polyp Segmentation Using Wavelet-Based Cross-Band Integration for Enhanced Boundary Representation

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Accurate polyp segmentation is essential for early colorectal cancer detection, yet
2 achieving reliable boundary localization remains challenging due to low mucosal
3 contrast, uneven illumination, and color similarity between polyps and surrounding
4 tissue. Conventional methods relying solely on RGB information often struggle
5 to delineate precise boundaries due to weak contrast and ambiguous structures
6 between polyps and surrounding mucosa. To establish a quantitative foundation
7 for this limitation, We analyzed polyp-background contrast in the wavelet domain,
8 revealing that grayscale representations consistently preserve higher boundary
9 contrast than RGB images across all frequency bands. This finding suggests that
10 boundary cues are more distinctly represented in the grayscale domain than in the
11 color domain. Motivated by this finding, we propose a segmentation framework that
12 integrates grayscale and RGB representations through complementary frequency-
13 consistent interaction, enhancing boundary precision while preserving structural
14 coherence. Extensive experiments on four benchmark datasets demonstrate that the
15 proposed approach achieves superior boundary precision and robustness compared
16 to conventional methods.

17 1 Introduction

18 Accurate polyp segmentation is vital for early colorectal cancer detection but remains difficult due
19 to low mucosal contrast, uneven illumination, and strong visual similarity between polyps and
20 surrounding mucosa [1]. These conditions blur polyp boundaries, particularly in small or flat cases,
21 making reliable delineation challenging. Although recent boundary-aware methods improve edge
22 perception, their dependence on RGB inputs limits robustness under contrast variations. To investigate
23 this limitation, we performed a wavelet-based contrast analysis between RGB and grayscale images.
24 The contrast index (CI), defined as $CI = |\mu_{\text{polyp}} - \mu_{\text{background}}| / (\mu_{\text{polyp}} + \mu_{\text{background}} + \epsilon)$, measures the
25 distinction between polyp and background regions using the mean of absolute wavelet coefficients.
26 As shown in Fig. 1, grayscale consistently achieves higher CI across all sub-bands, indicating that
27 boundary cues are more distinct in the intensity domain. Building on this evidence, we propose a
28 segmentation approach that integrates grayscale and RGB representations through frequency-band
29 interaction. By encouraging interaction between corresponding wavelet sub-bands of both modalities,
30 contrast-rich grayscale features refine RGB-derived spatial structures, improving boundary precision
31 while preserving overall coherence.

32 2 Related Work

33 Polyp segmentation has become a key task in computer-aided colorectal cancer screening, where
34 precise boundary delineation is essential for accurate diagnosis and treatment planning. A variety of

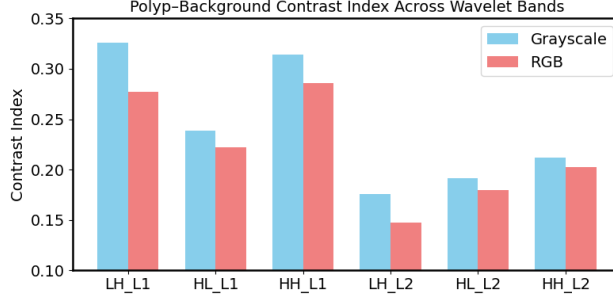


Figure 1: Structural contrast comparison between RGB and grayscale images in the wavelet domain, showing consistently higher contrast for grayscale across all detail sub-bands.

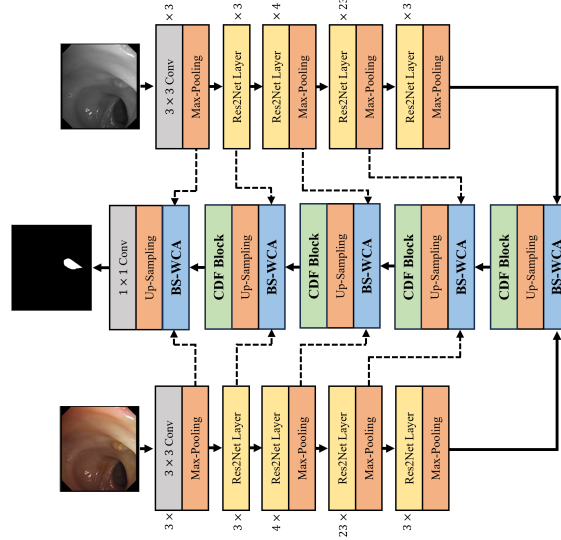


Figure 2: Proposed wavelet-based cross-band integration framework that fuses frequency-consistent information from RGB and grayscale features for enhanced boundary representation.

35 deep learning-based models have been developed for polyp segmentation [11, 15, 14, 8, 5]. Among
 36 these, boundary-aware models such as PraNet [6], CaraNet [10], MEGANet [4], and Polyper [12] aim
 37 to improve boundary localization through attention- and edge-guided mechanisms for more accurate
 38 polyp delineation. Nevertheless, most of these approaches primarily rely on RGB representations,
 39 which capture chromatic appearance but insufficiently describe structural contrast.

40 3 Method

41 The proposed model adopts a dual-encoder structure designed to leverage the complementary char-
 42 acteristics of RGB and grayscale modalities. Each encoder is based on Res2Net [7] and extracts
 43 hierarchical feature representations: the RGB encoder captures chromatic and textural cues, while
 44 the grayscale encoder focuses on contrast-driven structural patterns that are effective for boundary
 45 discrimination. Features from corresponding encoder stages are processed within the decoder through
 46 two key components: the Band-Specific Window Cross-Attention (BS-WCA) module and the Cascade
 47 Dilated Fusion (CDF) block. The BS-WCA module performs frequency-aligned interaction
 48 between the RGB and grayscale features by selectively exchanging information across identical
 49 wavelet sub-bands, allowing high-frequency grayscale details to refine RGB-derived structural fea-
 50 tures. The CDF block then integrates the refined multi-scale features through dilated convolutions,
 51 preserving both fine-grained boundary precision and global contextual consistency. As illustrated

Table 1: Quantitative comparison of polyp segmentation methods across four benchmark datasets, evaluated by mean Dice and IoU scores (mean $\pm$ standard deviation) averaged over 10 random runs.

Methods	Kvasir		ClinicDB		ColonDB		ETIS	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
Ours	0.885 $\pm$ 0.021	0.822 $\pm$ 0.019	0.926 $\pm$ 0.014	0.862 $\pm$ 0.023	0.913 $\pm$ 0.021	0.840 $\pm$ 0.042	0.922 $\pm$ 0.029	0.821 $\pm$ 0.029
Polyper	0.867 $\pm$ 0.014	0.796 $\pm$ 0.020	0.914 $\pm$ 0.019	0.841 $\pm$ 0.021	0.868 $\pm$ 0.042	0.796 $\pm$ 0.035	0.888 $\pm$ 0.047	0.760 $\pm$ 0.047
MEGANet	0.863 $\pm$ 0.011	0.802 $\pm$ 0.018	0.909 $\pm$ 0.011	0.801 $\pm$ 0.077	0.802 $\pm$ 0.083	0.704 $\pm$ 0.059	0.747 $\pm$ 0.097	0.598 $\pm$ 0.077
CRCNet	0.879 $\pm$ 0.015	0.815 $\pm$ 0.018	0.910 $\pm$ 0.052	0.854 $\pm$ 0.018	0.866 $\pm$ 0.064	0.800 $\pm$ 0.046	0.665 $\pm$ 0.027	0.582 $\pm$ 0.129
CaraNet	0.727 $\pm$ 0.023	0.628 $\pm$ 0.035	0.836 $\pm$ 0.006	0.702 $\pm$ 0.014	0.760 $\pm$ 0.056	0.651 $\pm$ 0.038	0.784 $\pm$ 0.078	0.633 $\pm$ 0.105
ConvSegNet	0.856 $\pm$ 0.008	0.765 $\pm$ 0.022	0.902 $\pm$ 0.020	0.810 $\pm$ 0.025	0.884 $\pm$ 0.033	0.756 $\pm$ 0.044	0.859 $\pm$ 0.055	0.667 $\pm$ 0.077
DUCKNet	0.818 $\pm$ 0.016	0.751 $\pm$ 0.019	0.878 $\pm$ 0.026	0.791 $\pm$ 0.033	0.683 $\pm$ 0.161	0.570 $\pm$ 0.073	0.383 $\pm$ 0.205	0.321 $\pm$ 0.099
PraNet	0.650 $\pm$ 0.021	0.524 $\pm$ 0.032	0.793 $\pm$ 0.032	0.648 $\pm$ 0.030	0.784 $\pm$ 0.063	0.620 $\pm$ 0.037	0.666 $\pm$ 0.090	0.423 $\pm$ 0.092
UNet	0.775 $\pm$ 0.013	0.668 $\pm$ 0.025	0.855 $\pm$ 0.019	0.762 $\pm$ 0.024	0.802 $\pm$ 0.083	0.704 $\pm$ 0.059	0.549 $\pm$ 0.186	0.382 $\pm$ 0.099

in Fig. 2, the integration of BS-WCA and CDF within a dual-encoder–decoder framework enables model to exploit contrast-rich intensity information while maintaining the contextual advantages of RGB representations, resulting in more stable and accurate polyp boundary segmentation.

4 Experimental Results

Datasets and Evaluation The experiments utilized four widely used polyp segmentation datasets: Kvasir-SEG (1,000 images) [9], CVC-ClinicDB (612 images) [3], CVC-ColonDB (380 images) [2], and ETIS (196 images) [13]. All datasets were partitioned into training, validation, and testing sets to ensure consistent evaluation across varying data distributions. Performance was measured using the Dice coefficient and Intersection over Union (IoU), following standard practices in polyp segmentation.

Implementation Details All experiments were implemented using PyTorch and Torchvision and conducted on a single NVIDIA GPU (CUDA 11.7, Ubuntu 20.04, Python 3.9, PyTorch 1.13, Torchvision 0.14). The proposed model and all baseline models were trained and evaluated with a batch size of 8.

Results. As shown in Table 1, the proposed method consistently outperforms existing approaches across all four benchmark datasets in terms of both Dice and IoU scores. The improvement is particularly evident in the overall average performance, demonstrating the benefit of jointly leveraging grayscale and RGB representations. Compared to recent conventional models, the proposed model achieves higher mean Dice and IoU values while maintaining stable performance across datasets of varying scales and imaging conditions. These results indicate that the integration of intensity-based and color-based representations provides a more comprehensive feature space for polyp segmentation.

5 Discussion

This study primarily evaluates region-level segmentation performance using Dice and IoU. To better understand the behavior of the proposed model, further analyses should include boundary-sensitive metrics for contour accuracy, qualitative comparisons across imaging conditions, and frequency-domain ablations to assess modality contributions. Such evaluations would provide a more comprehensive view of how grayscale–RGB integration influences segmentation performance.

6 Conclusion

This work introduced a dual-encoder segmentation framework that integrates grayscale and RGB representations through frequency-band interaction. Although the proposed model does not explicitly model boundaries, its design naturally facilitates boundary-aware representation through high-frequency information exchange. The design effectively combines contrast-rich intensity cues with color-based features, achieving consistent performance improvements across four benchmark datasets. Future work will extend this framework with boundary-sensitive evaluations and frequency-domain analyses to further clarify its contribution to robust polyp boundary segmentation.

87 7 Potential Negative Societal Impact

88 While the proposed method aims to improve segmentation reliability in medical imaging, several
89 potential risks should be noted. Generalization across imaging domains may be limited, as variations
90 in endoscopic devices or acquisition settings can influence performance consistency. In addition,
91 limited diversity in publicly available training data could result in biased or uneven segmentation
92 outcomes across populations. To mitigate these risks, careful cross-domain validation, transparent
93 dataset reporting, and responsible integration into medical workflows are essential to ensure equitable
94 and trustworthy application.

95 References

- 96 [1] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, et al. Comparative validation of polyp
97 detection methods in video colonoscopy: Results from the miccai 2015 endoscopic vision
98 challenge. *IEEE Transactions on Medical Imaging*, 36(6):1231–1249, 2017.
- 99 [2] Jorge Bernal, F. Javier Sánchez, and Fernando Vilarino. Towards automatic polyp detection
100 with a polyp appearance model. *Pattern Recognition*, 45(9):3166–3182, 2012.
- 101 [3] Jorge Bernal, F. Joan Sánchez, Gloria Fernández-Esparrach, Debora Gil, Carme Rodríguez, and
102 Fernando Vilarino. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation
103 vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111,
104 2015.
- 105 [4] Nguyen-Tien Bui, Son Hua, Hieu Le, Minh-Triet Nguyen, Nhat Ho, and Thanh-Dat Do.
106 Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In
107 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*,
108 pages 6622–6631, Waikoloa, HI, USA, 2024. IEEE.
- 109 [5] Razvan-Gabriel Dumitru, Darius Peteleaza, and Catalin Craciun. Using duck-net for polyp
110 image segmentation. *Scientific reports*, 13(1):9803, 2023.
- 111 [6] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling
112 Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *Proceedings of the*
113 *International Conference on Medical Image Computing and Computer-Assisted Intervention*
114 *(MICCAI)*, pages 263–273, Lima, Peru, 2020.
- 115 [7] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip
116 Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis*
117 *and machine intelligence*, 43(2):652–662, 2019.
- 118 [8] Ayokunle Olalekan Ige, Nikhil Kumar Tomar, Felix Ola Aranuwa, Oluwafemi Oriola, Alaba O
119 Akingbesote, Mohd Halim Mohd Noor, Manuel Mazzara, and Benjamin Segun Aribisala.
120 Convsegnet: automated polyp segmentation from colonoscopy using context feature refinement
121 with multiple convolutional kernel sizes. *IEEE Access*, 11:16142–16155, 2023.
- 122 [9] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag
123 Johansen, and Håvard D. Johansen. Kvasir-seg: A segmented polyp dataset. In *Proceedings of*
124 *the International Conference on Multimedia Modeling (MMM)*, pages 451–462, Daejeon, South
125 Korea, 2020. Springer.
- 126 [10] Ange Lou, Shuyue Guan, and Murray Loew. Caranet: context axial reverse attention network
127 for segmentation of small medical objects. *Journal of Medical Imaging*, 10(1):014005–014005,
128 2023.
- 129 [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for
130 biomedical image segmentation. In *Proceedings of the International Conference on Medical*
131 *Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, Munich,
132 Germany, 2015.
- 133 [12] Hao Shao, Yang Zhang, and Qibin Hou. Polyper: Boundary sensitive polyp segmentation.
134 In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 4731–4739,
135 Vancouver, Canada, 2024.

- 136 [13] Jorge Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward em-
137 bedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International*
138 *Journal of Computer Assisted Radiology and Surgery*, 9(2):283–293, 2014.
- 139 [14] Lei Yang, Chenxu Zhai, Yanhong Liu, and Hongnian Yu. Cfha-net: A polyp segmentation
140 method with cross-scale fusion strategy and hybrid attention. *Computers in Biology and*
141 *Medicine*, 164:107301, 2023.
- 142 [15] Jianbo Zhu, Mingfeng Ge, Zhimin Chang, and Wenfei Dong. Crcnet: Global-local context
143 and multi-modality cross attention for polyp segmentation. *Biomedical Signal Processing and*
144 *Control*, 83:104593, 2023.