algorithms

MDPI

Article

HPANet: Hierarchical Path Aggregation Network with Pyramid Vision Transformers for Colorectal Polyp Segmentation

Yuhong Ying [1,2], Haoyuan Li [1,2], Yiwen Zhong [1,2] and Min Lin [1,2,*]

[1] College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 5221139004@fafu.edu.cn (Y.Y.); 12311093028@fafu.edu.cn (H.L.); yiwzhong@fafu.edu.cn (Y.Z.)
[2] Key Laboratory of Smart Agriculture and Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China
* Correspondence: linmin@fafu.edu.cn

Abstract: The automatic segmentation technique for colorectal polyps in colonoscopy is considered critical for aiding physicians in real-time lesion identification and minimizing diagnostic errors such as false positives and missed lesions. Despite significant progress in existing research, accurate segmentation of colorectal polyps remains technically challenging due to persistent issues such as low contrast between polyps and mucosa, significant morphological heterogeneity, and susceptibility to imaging artifacts caused by bubbles in the colorectal lumen and poor lighting conditions. To address these limitations, this study proposed a novel pyramid vision transformer-based hierarchical path aggregation network (HPANet) for polyp segmentation. Specifically, firstly, the backward multi-scale feature fusion module (BMFM) was developed to enhance the ability of processing polyps with different scales. Secondly, the forward noise reduction module (FNRM) was designed to learn the texture features of the upper and lower layers to reduce the influence of noise such as bubbles. Finally, in order to solve the problem of boundary ambiguity caused by repeated up and down sampling, the boundary feature refinement module (BFRM) was developed to further refine the boundary. The proposed network was compared with several representative networks on five public polyp datasets. Experimental results show that the proposed network achieves better segmentation performance, especially on the Kvasir SEG dataset, where the mDice and mIoU coefficients reach 0.9204 and 0.8655.

Keywords: medical image segmentation; polyp segmentation; pyramid vision transformers; computer vision

check for updates

Academic Editors: Laura Antonelli and Junzo Watada

Received: 2 April 2025
Revised: 3 May 2025
Accepted: 9 May 2025
Published: 11 May 2025

Citation: Ying, Y.; Li, H.; Zhong, Y.; Lin, M. HPANet: Hierarchical Path Aggregation Network with Pyramid Vision Transformers for Colorectal Polyp Segmentation. *Algorithms* **2025**, *18*, 281. https://doi.org/10.3390/a18050281

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Colorectal cancer (CRC) is considered as one of the three major malignant tumors in the world [1]. It is estimated that there are more than 1.85 million cases and 850,000 deaths every year in the world, and especially in low-income areas, the incidence rate and mortality rate have increased significantly [2]. Early detection and precise segmentation of colonic mucosal polyps are crucial for reducing CRC-related deaths. Optical colonoscopy (OC) is a gold standard screening technique that can locate and characterize colorectal polyps, promoting their removal before malignant transformation. However, the accurate identification and segmentation of polyps during colonoscopy remain clinically demanding procedures. Significant variations in polyp morphology (size, color, texture) combined with low contrast against surrounding mucosa substantially increase the risks of both missed detection and false-positive diagnoses.

To address these issues, numerous scholars have conducted research on colorectal polyp segmentation algorithms. Colorectal polyp segmentation methods can be broadly categorized into traditional methods relying on manual feature extraction and deep learning-based methods. Traditional methods involve extracting low-level features such as color, contour, edge, and texture, and then using classifiers to distinguish between polyps and normal colorectal mucosa. However, these methods heavily rely on the expertise of the designer, and due to the differences in size, shape, and texture of different polyps, they often perform poorly, leading to issues such as missed detection and false positives.

In recent years, with the development of deep learning technology, Convolutional Neural Networks (CNNs) have been widely applied in the field of medical image segmentation. For instance, a symmetric U-shaped network [3] capable of integrating multi-scale features was proposed to achieve precise pixel-level segmentation through its unique encoder–decoder architecture. The Attention U-Net [4] architecture introduced attention gates to enhance accuracy by focusing on regions of interest, effectively allocating computational resources to diagnostically significant areas. ABC-Net [5] introduced a dual-branch architecture with a shared encoder and mutually constrained decoders for simultaneous segmentation of colorectal polyp regions and their anatomical boundaries. SegT [6] improved the boundary between benign regions and polyps by designing a separate edge guidance module, including a separator and an edge guidance block. PraNet [7] utilized Parallel Partial Decoders (PPDs) to aggregate high-level features and established relationships between target regions and boundaries through an inverse attention module. As another representative of deep learning, the Transformer has been applied by numerous scholars in fields such as medical image segmentation due to its outstanding performance in the field of natural language processing. For example, TransUNet [8] combined the Transformer as an encoder with a CNN decoder to address the ambiguous boundary between pancreatic tissue and surrounding organs through global context modeling. PMTrans [9] utilized a pyramid-shaped Transformer branch to fuse with CNN features, effectively capturing multi-scale morphological features of retinal blood vessels. Med-Former [10] employed a hierarchical Transformer architecture to process 3D cardiac MRI data, modeling the spatial topological relationships between ventricles and atria through a self-attention mechanism.

Despite various advancements made by existing segmentation models in polyp segmentation, with some algorithms achieving relatively accurate segmentation results, the problem of polyp segmentation still poses many challenges:

- Noise Interference: when the polyp image is disturbed by noise such as bubbles, it is easy to cause discontinuity and blurred segmentation edges.
- Large Differences in Shape and Size: the polyp scale is large, and the shape characteristics are also different, making it easy to generate incorrect segmentation when dealing with small polyps.
- Semantic Information Loss: through continuous sampling operations during the segmentation process, the semantic information contained in the polyp context is easily gradually lost.

Inspired by the Pyramid Vision Transformer v2 (PVTv2) [11], this paper proposes a hierarchical path aggregation network with Pyramid Vision Transformer (HPANet) for colorectal polyp segmentation. This network effectively extracts multi-scale features of polyps and reduces noise interference, thereby improving segmentation accuracy. Three key technical innovations are introduced, which are outlined below:

(1) A novel network, HPANet, for colorectal polyp segmentation was proposed. This network utilizes PVTv2 to extract multi-level polyp image features and then designs a backward multi-scale feature fusion module (BMFM). This module sequentially fuses adja-

cent feature maps from the back to the front to generate spatial attention maps, effectively addressing the scale variation issue of colorectal polyps.

(2) A forward noise reduction module (FNRM) was developed to mitigate noise interference such as bubbles and illumination on the image by gradually integrating and learning the contextual texture features from front to back, thereby improving segmentation accuracy.

(3) A boundary feature refinement module (BFRM) was designed to address the boundary blurring issue caused by repeated up-down sampling operations. BFRM combines channel and spatial attention to preserve more high-frequency boundary details during feature alignment.

The remaining sections of the paper are arranged as follows: Section 2 introduces related works, including polyp segmentation networks and vision Transformer, Section 3 describes the structure of HPANet, Section 4 details the experimental settings and analysis of experimental results, and Section 5 concludes with a summary and discussion.

2. Related Works

2.1. Polyp Segmentation

The segmentation of colorectal polyps is a classic problem in medical image segmentation. Its accurate and effective segmentation not only promotes the solution of other medical image segmentation problems, but also promotes the clinical practice of early colorectal cancer screening. Recent advances in deep learning have significantly enhanced the performance of polyp segmentation. PraNet [7], a pioneering framework using Res2Net [12] as the backbone, was proposed to address multi-scale polyp variations. It enhanced segmentation accuracy on Kvasir-SEG through two key designs: (a) a Parallel Partial Decoder (PPD) for hierarchical feature aggregation, and (b) a reverse attention module for establishing correlation in boundary regions. However, its fixed pyramid levels limit its adaptability to extreme size variations. To alleviate the issues of color inconsistency and pixel imbalance in colonoscopy images, the authors of [13] designed a shallow attention gate to achieve precise localization of small polyps, and implemented probability calibration during the inference process to enhance the continuity of segmentation. However, manually designed color enhancement may degrade performance under low-light conditions. TGANet [14] demonstrated the feasibility of integrating polyp size and texture as text attention cues. It encoded ResNet50 features through text-guided spatial attention, effectively improving segmentation performance. This indicates that multimodal supervision can enhance the discriminability of features, but generating text in real-time remains technically challenging.

Compared with traditional CNN-based models, Transformer-based models for medical image segmentation offer several advantages. CNNs rely on local receptive fields and convolutional operations, which can capture local features well but may have limitations in modeling long-range dependencies. In contrast, transformers, with their self-attention mechanism, can capture global context information more effectively. This is particularly useful in medical image segmentation, where understanding the global structure of organs or lesions is crucial. For example, in polyp segmentation, a polyp's context within the colon environment is important for accurate segmentation. Transformers can better capture the relationships between different parts of the polyp and the surrounding tissues, reducing the risk of missegmentation. Additionally, transformers can handle variable-sized input more flexibly, which is beneficial when dealing with medical images that may have different resolutions or cropping sizes.

UViT-Seg [15] employed a visual transformer to extract long-range semantic information and utilized a convolutional neural network (CNN) module integrated with squeeze-and-excitation and dual attention mechanisms to capture low-level features in important

regions of the image. The limited receptive field of conventional U-Nets in capturing global polyp patterns was addressed by Wang et al. [16] through a hybrid architecture that synergistically combines multi-scale dense nested UNet with Transformer layers. HSNet [17] first employed a Transformer branch to capture long-range dependencies, then applied a CNN branch to capture local details of appearance, and finally bridged the gap between low-level and high-level features through an interaction mechanism, thereby enhancing the performance of polyp segmentation in this network. HIGF-Net [18] adopted a hierarchical guidance strategy to mine the deep global semantic information and shallow local spatial features of images. It also proposed an independent refinement module to refine the contour of polyps in uncertain areas, highlighting the differences between polyps and the background. The most advanced polyp segmentation algorithm, Polyp-PVT [19], abandoned the hierarchical sampling method based on CNN and introduced the Pyramid Vision Transformer to extract multi-scale features of polyps, effectively addressing the issue of poor edge segmentation accuracy. However, the design of Polyp-PVT led to the loss of important details and edge spatial information. In summary, despite various improved algorithms being proposed in the field of polyp segmentation, these algorithms all have certain limitations and shortcomings, which motivates us to conduct further research and experiments.

2.2. Vision Transformer

The latest advancements in Vision Transformers (ViTs) have reshaped the landscape of visual representation learning. The foundational work by Dosovitskiy et al. [20] pioneered the application of a pure Transformer architecture to images. Specifically, images are divided into fixed-sized patches, global feature extraction is performed through a Transformer encoder, and classification is carried out via Multi-Layer Perceptron (MLP). Building on this, Han et al. [21] developed Transformer-in-Transformer (T2T), introducing a recursive Tokenize-Transformer mechanism. Specifically, it reorganizes layer outputs into a spatial format and applies overlapping patch segmentation to capture multi-scale contextual relationships, thereby enhancing hierarchical feature learning. To alleviate the problem of high model training difficulty when the ViT becomes deeper, CaiT [22] and DeepViT [23] have improved the attention mechanism. To alleviate the limitations of ViT in terms of computational efficiency and pixel-level task adaptability, Liu et al. [24] proposed Swin Transformer, a hierarchical architecture with a sliding window operation. This design enables localized self-attention within windows while allowing cross-window communication through systematic window shifting, striking a balance between global modeling and computational practicality for dense prediction tasks. However, Swin Transformer has its limitations in polyp segmentation: polyps have the characteristics of complex morphology and large-scale differences, and Swin Transformer may not fully capture the long-range dependencies of polyp images, resulting in potential wrong segmentation. Chen et al.'s TransUNet [8] combined Transformer as an encoder with a CNN decoder to address the problem of blurred polyp boundaries through global context modeling. However, it extracts polyp features at a fixed resolution. This fixed-resolution feature extraction may not be able to adapt well to the diverse scales of polyps. Unlike traditional ViT for extracting columnar structural feature maps, Wang et al. [25] proposed Pyramid Vision Transformer (PVT) to generate pyramid feature maps for tasks such as semantic segmentation and object detection. Based on PVT, Wang et al. [11] further improved it and proposed PVTv2. PVTv2 reduces the computational complexity of PVT to linear and demonstrates superior performance compared to both PVT and Swin Transformer. Inspired by this, this paper adopted PVTv2 as the backbone of the image encoder to extract multi-scale image features,

and performs fusion, denoising, and other tasks on the generated feature maps to further enhance the accuracy of polyp segmentation.

3. Materials and Methods

3.1. Overall Architecture

As shown in Figure 1, the HPANet comprises an encoder, a backward multi-scale feature fusion module (BMFM), a forward noise reduction module (FNRM) and a boundary feature refinement module (BFRM).

The encoder adopts PVTv2 as the backbone to capture multi-scale features and details of polyps. Through the PVTv2, the encoder outputs four layers of feature maps with different scales, which are represented as $\{M_1, M_2, M_3, M_4\}$. The BMFM receives M_i $(i \in \{1, 2, 3, 4\})$ as input and uses three spatial attention map(SAM) to aggregate the adjacent low-level feature map M_i and high-level feature map M_{i+1} to generate $D_i(i \in \{1, 2, 3, 4\})$. SAM is based on a multi-scale fusion strategy, which can effectively capture the morphological information of different sizes of polyps in the image. To better learn the texture features of polyps and reduce noise interference, the FNRM uses three boundary enhance blocks (BEBs) to aggregate each pair of adjacent D_i and D_{i+1}, and progressively generates feature maps $N_i(i \in \{1, 2, 3, 4\})$. Based on N_i, the BFRM incorporates the Convolutional Block Attention Module (CBAM) [26], which makes it easier for the network to focus on important areas such as boundaries, and then adjusts the number of channels in the feature map through a convolutional head. Finally, the four feature maps are upsampled to the original image size and aggregated by addition to produce the final segmentation result.

3.2. Encoder

In recent years, Transformers have demonstrated better performance than CNNs in the field of image segmentation [27]. The encoder employs PVTv2 as the backbone to extract the features of polyp images. Unlike traditional transformers, such as ViT [20], which outputs fixed resolution images, PVTv2 generates feature maps with a 4-layer pyramid hierarchical structure and represents them as $\{M_1, M_2, M_3, M_4\}$. M_1, M_2, M_3, and M_4 represent feature maps of different scales, with sizes of $(H/4) \times (W/4) \times 64$, $(H/8) \times (W/8) \times 128$, $(H/16) \times (W/16) \times 320$, and $(H/32) \times (W/32) \times 512$, respectively. These feature maps contain rich contextual and semantic information, and are not suitable for generating segmentation results through simple fusion. Therefore, this study proposes three improved modules to enhance the performance of polyp segmentation from different perspectives.

3.3. Backward Multi-Scale Feature Fusion Module (BMFM)

Due to the huge differences in morphology and scale of colorectal polyps, some tiny polyps are difficult to segment, and relying solely on features of a single scale often cannot effectively pay attention to both global and local feature information. The high-level feature map output by the encoder contains the semantic and location information of the polyps, while the low-level feature map contains hidden information such as polyp edges and textures. In order to capture the spatial feature information of colorectal polyp images based on feature maps of different scales, the BMFM is proposed inspired by [28] to fuse high- and low-level feature maps to generate attention maps to enhance the image's ability to capture multi-scale features.

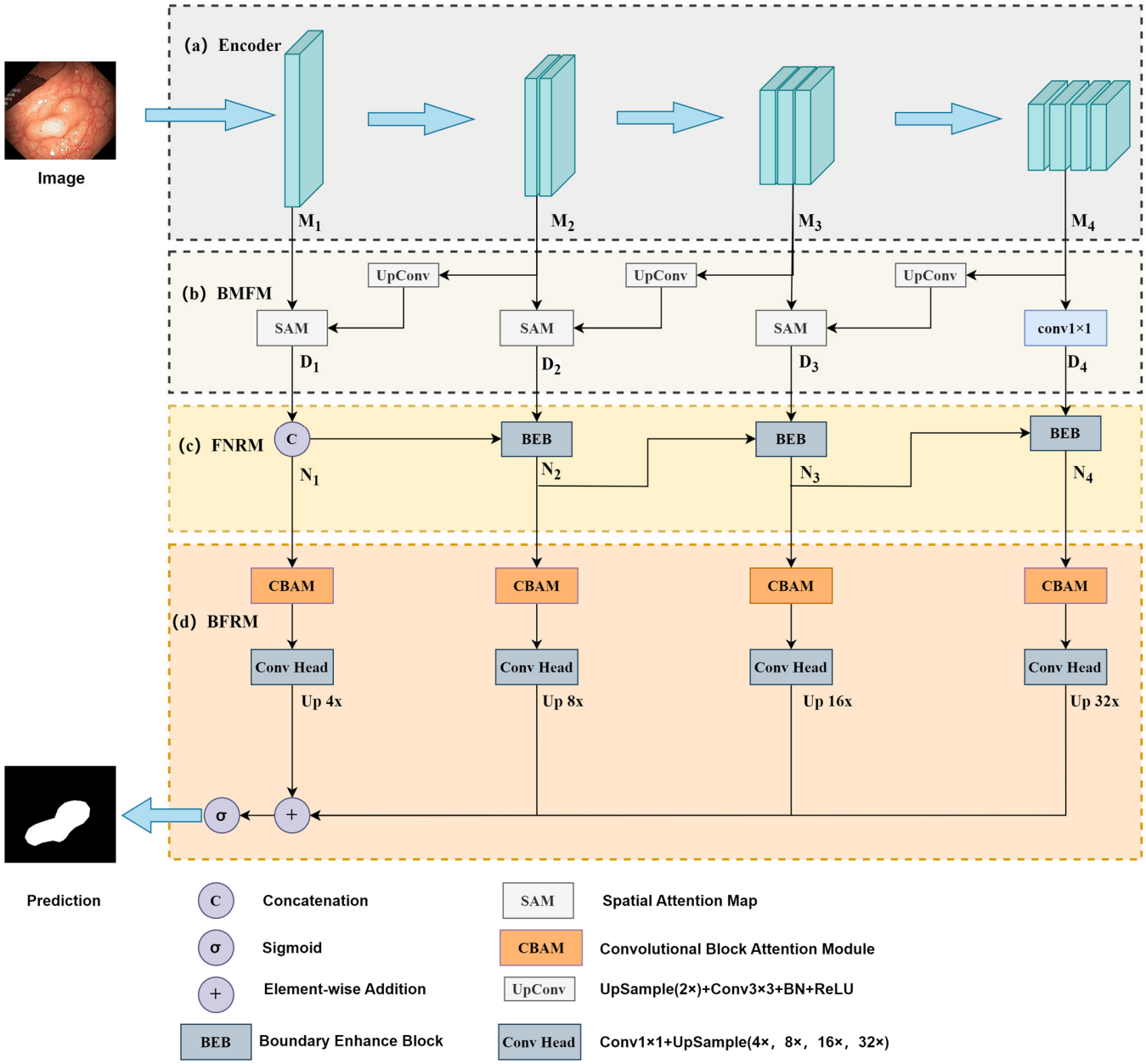


Figure 1. Structure of HPANet.

As shown in Figure 1, the BMFM is mainly composed of three spatial attention map (SAM) blocks, and the structure of the SAM is shown in Figure 2. Firstly, the high-level feature map M_{i+1} generated by the aforementioned encoder is upsampled to the same size as the low-level feature map M_i. Then, $M_i (i \in \{1, 2, 3\})$ and M_{i+1} are transmitted to the SAM block to generate a spatial attention map, which is multiplied with M_i to obtain D_i. In SAM, the channel dimensions of M_i and M_{i+1} are adjusted by a convolution block (ConvBlock), respectively, which includes a 1×1 convolution, *BN* and *ReLU* operations. Subsequently, M_i and M_{i+1} are added element by element and *ReLU* activation is performed to complete the fusion of feature information, and the fusion feature is expressed as α, as shown in Equation (1).

$$\alpha = ReLU(ReLU(BN(Conv_{1\times1}(M_i))) + ReLU(BN(Conv_{1\times1}(M_{i+1}))) \quad (1)$$

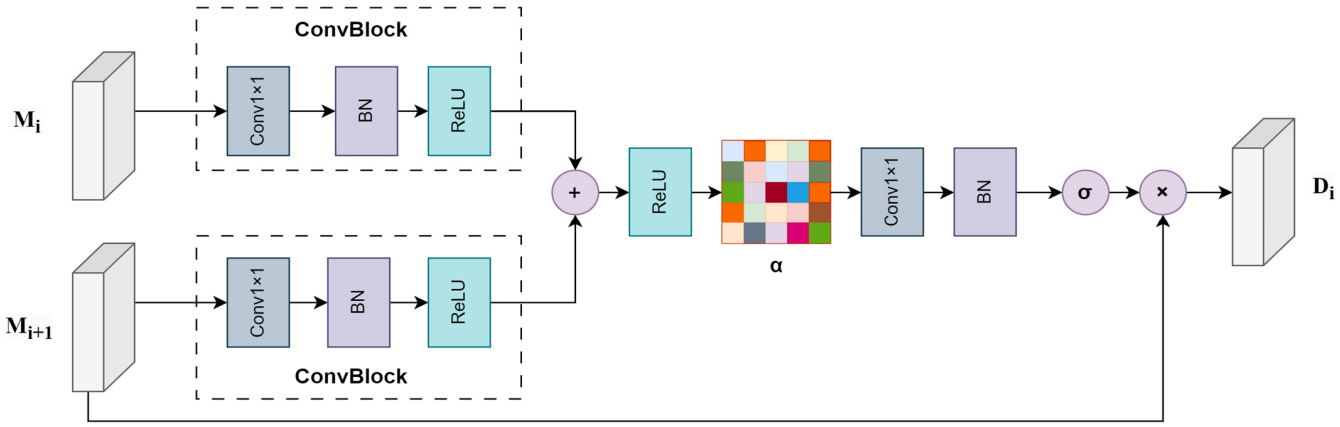


Figure 2. Structure of SAM.

Next, the following steps are taken: perform a 1×1 convolution, BN (batch normalization), and sigmoid activation operation on the fused feature α, map the feature values to the [0, 1] interval, generate a spatial attention weight map, and multiply the attention weight map element by element with the initial M_i, so that the multiplied feature map D_i can assign higher weights to key regions in low-level feature maps based on the semantic information contained in high-level feature maps, improve the ability to capture important features, and suppress the interference of irrelevant information. The calculation steps of D_i can be summarized using Equation (2) as follows:

$$D_i = SAM(M_i, M_{i+1}) = M_{i+1} * Sigmoid(BN(Conv_{1\times1}(\alpha))) \quad (2)$$

3.4. Forward Noise Reduction Module (FNRM)

Considering the low discrimination between polyps and background boundaries, and the edges of polyps are often disturbed by noise such as light, bubbles, and intestinal cleanliness, this may lead to inaccurate segmentation. In order to reduce these noise interferences, this section designed a FNRM that gradually integrates low-level implicit features such as texture into high-level features to enhance the boundary feature response of polyps.

The FNRM gradually enhances from the lowest level D_1 to D_4 through 3 BEBs (boundary enhance blocks), as shown in Figure 3. Each BEB first conducts global average pooling (GAP) on $D_i (i \in \{1,2,3\})$ to reduce its spatial dimension to 1×1, and then uses a 1×1 convolution to obtain the channel weight information r. Next, it peforms a 3×3 convolution operation on D_{i+1} to map it to the same number of channels and conducts point-wise multiplication with the weight information r to obtain D'_{i+1}. In this way, the noise information in the feature maps is cancelled out to reduce the model's sensitivity to noise. Finally, D_i is added to D'_{i+1} after a 3×3 convolution operation with a stride of 2 to finally output N_{i+1}. The composition of BEB is shown in Figure 4, and this process can be represented by Equations (3) and (4).

$$r = Conv_{1\times1}(GAP(D_i)) \quad (3)$$

$$N_{i+1} = Conv_{3\times3,stride=2}(D_i) + r \otimes Conv_{3\times3}(D_{i+1}) \quad (4)$$

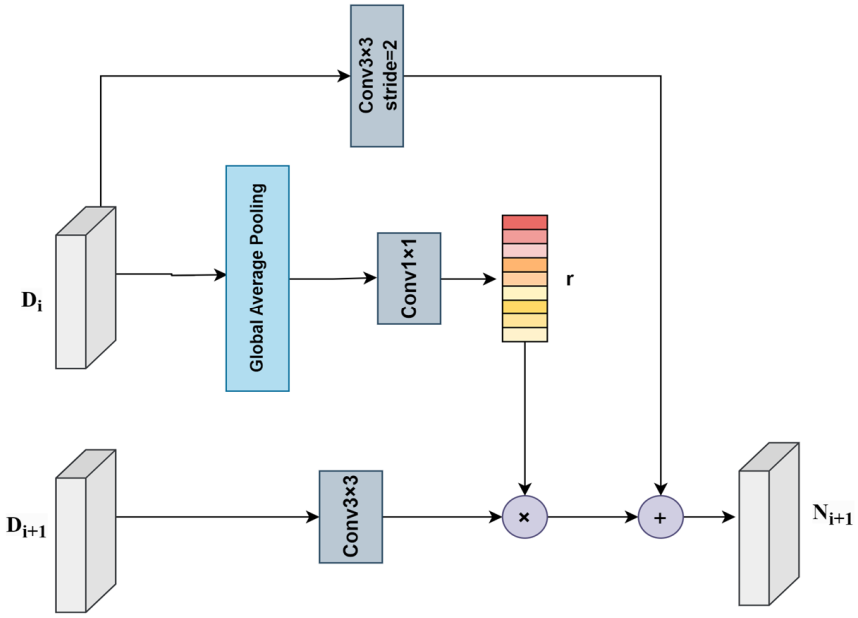


Figure 3. Structure of BEB.

3.5. Boundary Feature Refinement Module (BFRM)

Most of the classical colorectal polyp segmentation networks use a combination of up-sampling and down-sampling, which is prone to blurred boundaries when performing polyp segmentation. The Convolutional Block Attention Module (CBAM) [26] is an attention mechanism module designed for CNNs, which aims to improve the expressiveness of the model by enhancing its attention to important features such as boundaries. Inspired by this, the BFRM was proposed, which incorporates four CBAMs to process the feature maps generated by the previous layer. Subsequently, the channel dimension is adjusted using a 1×1 convolutional head, followed by bilinear interpolation upsampling to restore the four sharpened feature maps P_i to their original spatial resolution. Finally, these multi-scale maps are aggregated element-wise by summation and normalized using sigmoid to generate the final segmentation mask, as shown in Equation (5).

$$\text{BFRM_}out = sigmoid(\sum_{4}^{i=1} Up(Conv_{1\times1}(CBAM(N_i))) \tag{5}$$

In Equation (5), the *Up* function represents the upsampling of four feature maps with scaling factors of $4\times$, $8\times$, $16\times$, and $32\times$ through bilinear interpolation. The CBAM function represents the CBAM processing process, which consists of two stages: channel attention (CA) module and spatial attention (SA) module, as illustrated in Figure 4. For the input feature map N_i, the CA module performs both average pooling and max pooling simultaneously and inputs it into an MLP to generate two $1 \times 1 \times C$ channel attention maps. These two maps are then summed element-wise and activated by sigmoid function to obtain the refined channel attention map N_i', as formalized in Equation (6). The SA module performs max pooling and average pooling along the channel direction on N_i' to obtain two feature maps of size $1 \times H \times W$, concatenates these two feature maps, and then generates the final attention map through a 7×7 convolution and sigmoid operation, as formalized in Equation (7).

$$N_i' = sigmoid(MLP(AvgPool(N_i))) + sigmoid(MLP(MaxPool(N_i))) \tag{6}$$

$$CBAM(N_i) = sigmoid\left(Conv_{7\times7}\left(\left[AvgPool\left(N_i'\right); MaxPool\left(N_i'\right)\right]\right)\right) \tag{7}$$

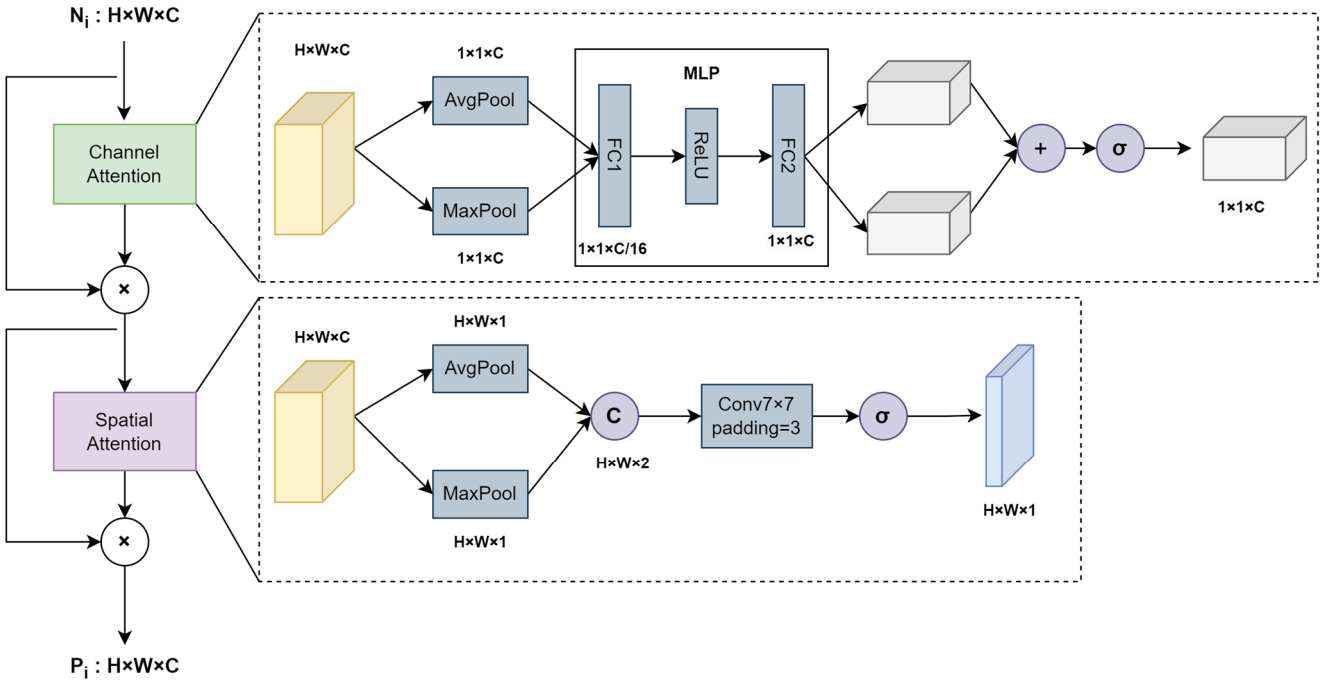


Figure 4. Structure of CBAM.

3.6. Loss Function

To address the specific characteristics of the object segmentation task, the loss function is selected as a weighted combination of Intersection over Union (IoU) and Binary Cross Entropy (BCE). The specific calculation steps are referred to in Equations (8)–(10).

$$\mathcal{L}_{total} = \mathcal{L}^w_{BCE} + \mathcal{L}^w_{IoU} \tag{8}$$

$$\mathcal{L}^w_{BCE} = -\sum_{i \in I} y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i) \tag{9}$$

$$\mathcal{L}^w_{IoU} = 1 - \frac{\sum_{i \in I} y_i \hat{y}_i}{\sum_{i \in I} y_i + \hat{y}_i - y_i \hat{y}_i} \tag{10}$$

4. Experiments and Results

4.1. Datasets and Training Settings

Five public colonoscopy image datasets are used for the study, including CVC-ClinicDB [29], Kvasir [30], CVC-300 [31], CVC-ColonDB [32], and ETIS-LaribPolypDB [33]. In this study, the Kvasir dataset was divided into training, validation, and testing sets in an 8:1:1 ratio, while the CVC ClinicDB, CVC-300, CVC-ColonDB, and ETIS-LaribPolypDB datasets were used as the testing sets. Five representative models, UNet [3], ResUnet [34], PraNet [7], CaraNet [35], and Polyp-PVT [19], were selected and compared with the model proposed in this paper.

The experimental framework utilized PyTorch 1.13.1 with GeForce GTX 3060 12 G GPU. The configuration standardized input images at 352 × 352-pixel resolution, maintained a batch size of 8 samples per iteration, and implemented 100-epoch maximum training cycles with a multi-scale training strategy similar to PraNet. Initialization employed ImageNet pretrained weights for the backbone network, coupled with gradient clipping at 0.5 threshold to stabilize training dynamics. The AdamW optimizer operated with uniform hyperparameters: 1×10^{-4} learning rate and 1×10^{-4} weight decay coefficient.

The image segmentation performance of models were evaluated through five complementary metrics: the Mean Dice Coefficient (mDice) quantifying spatial overlap accuracy

between predictions and ground truth; the Mean Intersection over Union (mIoU) assessing pixel-level alignment precision; recall (sensitivity) evaluating true positive identification completeness; precision determining the reliability of positive predictions; and the F2-score prioritizing recall while maintaining precision–recall equilibrium to adaptively balance performance under diverse clinical scenarios

4.2. Learning Capacity

To validate the learning capabilities of the proposed HPANet network, comparative experiments were conducted under the experimental configurations outlined in the previous section. The HPANet was rigorously evaluated on the Kvasir dataset against four traditional method-based models, including U-Net, as well as the PVT-based Poly-PVT approach. Quantitative results in Table 1 highlight HPANet's superior performance across key metrics, including mIoU, mDice, recall, precision, and F2-score.

Table 1. Comparison results on the Kvasir dataset. The best results have highlighted in bold.

Model	mIoU	mDice	Recall	Precision	F2
U-Net [3]	0.6497	0.742	0.6959	0.902	0.7094
ResUnet [34]	0.7637	0.8417	0.8372	0.8875	0.8344
PraNet [7]	0.8155	0.8802	0.8718	0.9239	0.8714
CaraNet [35]	0.8135	0.8758	0.8978	0.8952	0.8819
Polyp-PVT [19]	0.8595	0.9144	0.929	0.9201	0.9204
HPANet	**0.8655**	**0.9204**	**0.9323**	**0.927**	**0.9251**

As shown in Table 1, HPANet demonstrates enhanced learning capabilities for colonoscopy image segmentation through comprehensive benchmarking on the Kvasir dataset. Compared to the five selected models, this model achieved excellent performance in multiple metrics. Notably, when evaluated against the PVT-based Polyp-PVT model sharing the same backbone architecture (PVTv2), HPANet exhibits a 0.6% improvement in both mIoU and mDice metrics, which means that the segmentation results have higher spatial overlap and better pixel-level alignment accuracy with the real polyp region. It is helpful for doctors to clearly observe the boundary and shape of polyps and reduce the occurrence of misdiagnosis. The enhancements extend to recall (0.33% higher), precision (0.69% elevation), and F2-score (0.47% increase), indicating balanced precision–recall optimization. The improvement of metrics reduces the missed detection of small polyps or polyps with complex morphology. Furthermore, comparative analysis with PraNet reveals more substantial advancements: 5.00% mIoU, 4.02% mDice, and 6.05% recall improvements, alongside 0.31% precision and 5.37% F2-score gains. These metrics collectively validate the model's robustness in handling complex colonoscopic feature representations under clinical simulation conditions.

4.3. Generalization Capability

Through comprehensive testing on four benchmark datasets, the generalization performance of the HPANet has been rigorously validated, and detailed segmentation metrics comparisons are shown in Table 2. From the comparison results, the HPANet has statistically significant advantages over the benchmark model in five key evaluation indicators. On the CVC-ClinicDB dataset, compared to the Polyp-PVT, the proposed network achieves performance enhancements of 2.66% in mIoU, 2.19% in mDice, 3.31% in recall, 2.26% in precision, and 2.43% in F2-score. For the CVC-300 dataset, improvements extend to 4.71% in mIoU, 3.78% in mDice, 2.09% in recall, 3.31% in precision, and 3.24% in F2-score. In addition, while maintaining dominance in mIoU, mDice, recall, and F2-score across both CVC-ColonDB and ETIS-LaribPolypDB datasets, the network exhibits a marginal 0.18% pre-

cision deficit relative to Polyp-PVT. Specifically, on the CVC-ColonDB dataset, the method delivers gains of 1.31% in mIoU, 0.94% in mDice, 1.55% in recall, and 1.11% in F2-score. For ETIS-LaribPolypDB, corresponding improvements reach 2.08% in mIoU, 1.03% in mDice, 0.22% in recall, and 0.67% in F2-score. Furthermore, the Wilcoxon signed-rank test demonstrated statistically significant improvement achieved by HPANet ($p < 0.05$).

Table 2. Comparison results on CVC-ClinicDB, CVC-300, CVC-ColonDB and ETIS-LaribPolypDB. The best results have highlighted in bold.

Datasets	Model	mIoU	mDice	Recall	Precision	F2
CVC-ClinicDB	U-Net [3]	0.4051	0.4902	0.4534	0.7521	0.4619
	ResUnet [34]	0.5558	0.6314	0.6059	0.7899	0.6128
	PraNet [7]	0.7382	0.8029	0.8218	0.8228	0.8082
	CaraNet [35]	0.5578	0.6213	0.7612	0.6262	0.6658
	Polyp-PVT [19]	0.7273	0.8045	0.8301	0.836	0.8154
	HPANet	**0.7539**	**0.8264**	**0.8632**	**0.8586**	**0.8397**
CVC-300	U-Net [3]	0.2224	0.3025	0.2389	0.7127	0.2584
	ResUnet [34]	0.7278	0.8041	0.8207	0.822	0.8122
	PraNet [7]	0.7879	0.8621	0.9423	0.8182	0.9007
	CaraNet [35]	0.7269	0.7962	0.8622	0.8178	0.8217
	Polyp-PVT [19]	0.7799	0.8587	0.9404	0.8232	0.8978
	HPANet	**0.827**	**0.8965**	**0.9613**	**0.8563**	**0.9302**
CVC-ColonDB	U-Net [3]	0.1855	0.252	0.2301	0.5649	0.2273
	ResUnet [34]	0.4806	0.5507	0.5226	0.7056	0.5307
	PraNet [7]	0.556	0.6327	0.7078	0.6435	0.6486
	CaraNet [35]	0.4987	0.5635	0.6045	0.6488	0.5714
	Polyp-PVT [19]	0.6913	0.7773	0.7995	**0.817**	0.7842
	HPANet	**0.7044**	**0.7867**	**0.815**	0.8099	**0.7953**
ETIS-LaribPolypDB	U-Net [3]	0.257	0.3053	0.312	0.4734	0.2977
	ResUnet [34]	0.4656	0.5327	0.5519	0.6082	0.5386
	PraNet [7]	0.4626	0.5239	0.5881	0.5204	0.5461
	CaraNet [35]	0.4026	0.4782	0.4778	0.6013	0.4676
	Polyp-PVT [19]	0.606	0.6862	0.7436	**0.7361**	0.7131
	HPANet	**0.6268**	**0.6965**	**0.7458**	0.7061	**0.7198**

4.4. Comparative Analysis of Polyp Segmentation Performance Across Diverse Imaging Scenarios

To verify the segmentation performance of the HPANet for polyps in different scenarios, based on the experimental results in Sections 4.2 and 4.3, five representative imaging scenes were selected from five datasets to observe the segmentation results, with the segmentation outcomes visualized in Figure 5. The image (Row 1, Column 1) demonstrates large polyps with partial blurred margins. Conventional segmentation models (U-Net, ResUnet, PraNet) exhibit incomplete localization of polyp margins, while CaraNet and Polyp-PVT generate boundaries with reduced sharpness. In contrast, the proposed HPANet can accurately locate the polyp and produce clear and complete segmentation boundaries. The image (Row 2, Column 1) was selected from the CVC-ClinicDB dataset, where the polyps exhibit adhesions but have relatively clear boundaries. Compared to the other five models, the HPANet matches the manually annotated results very well in terms of overall shape and position, and the segmentation boundaries are also smoother. The image (Row 3, Column 1) selected from the CVC-300 dataset is characterized by the presence of artifacts and glare caused by the camera's movement in polyps. The U-Net and Polyp-PVT segmentation models may produce incorrect segmentation, while ResUnet, PraNet, and CaraNet can roughly locate the position of polyps, but artifacts may appear near the segmentation boundary. The image (Row 4, Column 1) is from the CVC-ColonDB dataset.

The contrast between the polyps in this image and the surrounding mucosa is low, leading to segmentation errors in U-Net, ResUnet, PraNet, and CaraNet, while Polyp-PVT and HPANet only show a certain degree of under segmentation. The image (Row 5, Column 1) is from the ETIS-LaribPolypDB dataset, where the proportion of polyps to the entire image is very small and there are interference factors such as yellow residues. The U-Net, ResNet, PraNet, and CaraNet models generate incorrect segmentation results, identifying background pixels as polyps, while Polyp-PVT and HPANet are also affected by interference factors, resulting in segmentation errors but with clear segmentation boundaries.

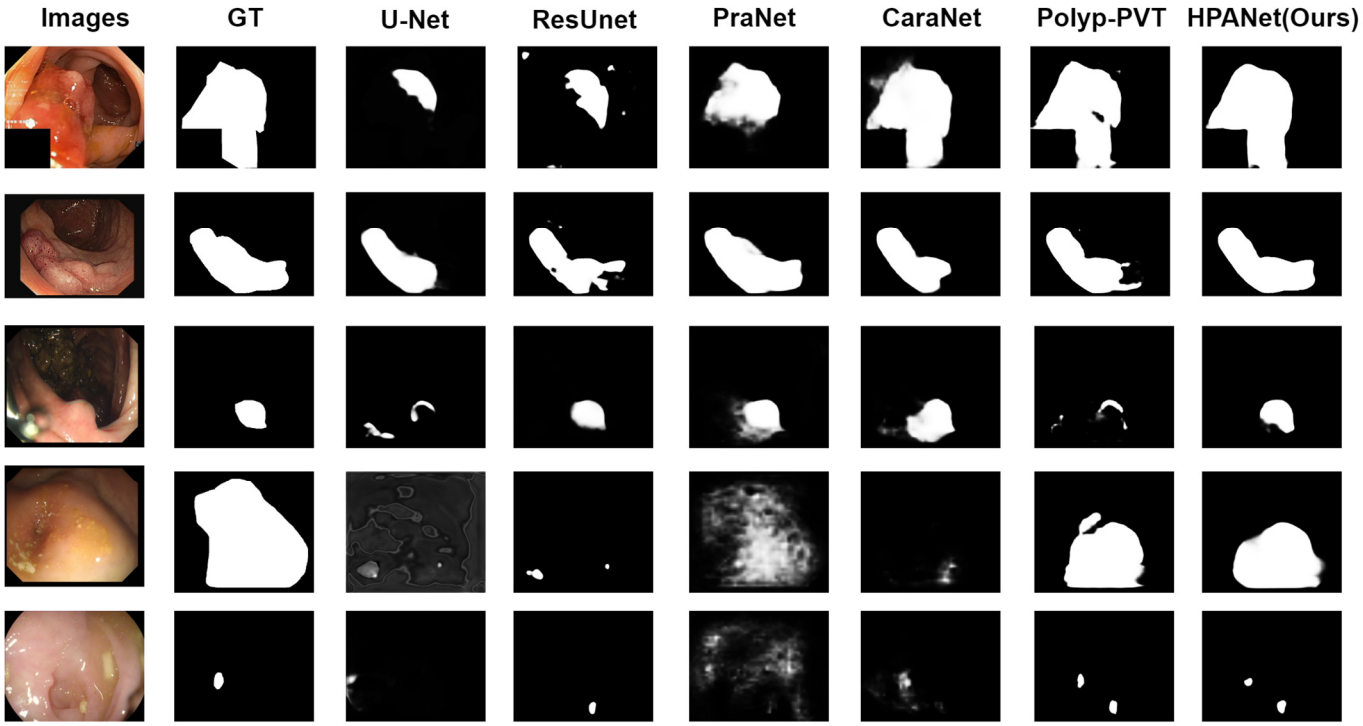


Figure 5. Qualitative results of different methods.

In summary, compared to other models, HPANet performs well in capturing fine details and segmenting boundaries, thereby improving segmentation accuracy.

4.5. Ablation Study

To systematically validate the contributions of individual modules in the proposed framework, an ablation study was performed on the Kvasir dataset. The quantitative results are shown in Table 3, where the symbol "×" indicates the cancellation of the module and "√" indicates the activation of the module. The visualization results of the ablation study are shown in Figure 6. From Table 3, it can be seen that the integration of the BMFM improved the baseline model's mIoU from 0.8392 to 0.8455 (+0.63%), mDice from 0.8955 to 0.9004 (+0.49%), and recall from 0.9061 to 0.9207 (+1.84%). These enhancements demonstrate the ability of BMFM to aggregate multi-scale spatial features, which is particularly important for maintaining fine-grained object boundaries in complex segmentation scenes. For example, the hierarchical feature fusion mechanism of this module improves the accuracy of boundary localization by adaptively weighting low-level texture details and high-level semantic context.

The BFRM further improved segmentation performance, with an mIoU of 0.8623 (an increase of 2.31% compared to baseline) and F2 score of 0.9127 (an increase of 2.93%). This improvement can be attributed to the dual channel attention mechanism of BFRM, which dynamically prioritizes channel dependence and spatial feature distribu-

tion. Specifically, the module amplifies discriminative features through learnable spatial channel recalibration, thereby reducing false positives in low contrast areas such as subtle lesion edges.

Table 3. Ablation study results of different modules. The best results have highlighted in bold.

BMFM	BFRM	FNRM	mIoU	mDice	Recall	Precision	F2
×	×	×	0.8392	0.8955	0.9061	0.9189	0.8934
√	×	×	0.8455	0.9004	0.9023	0.9276	0.8961
√	√	×	0.8623	0.9129	0.9207	**0.9292**	0.9127
√	√	√	**0.8655**	**0.9204**	**0.9323**	0.927	**0.9251**

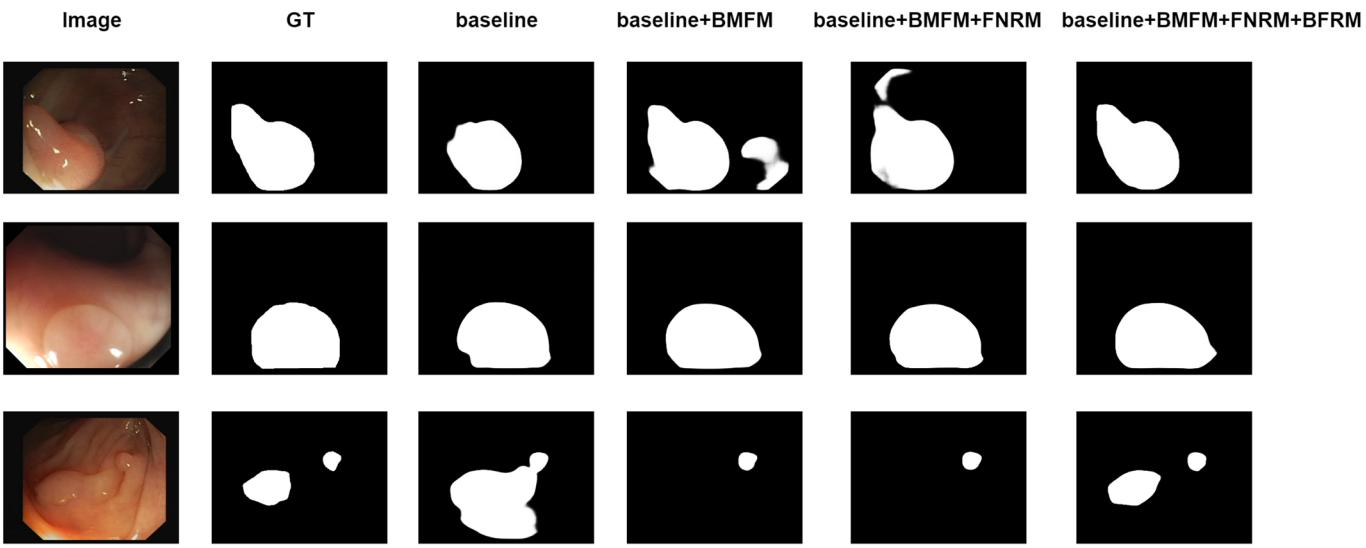


Figure 6. Segmentation result of the fusion of different modules.

The FNRM achieved the most significant performance improvement, with a peak mDice of 0.9204 (an increase of 2.49% compared to baseline) and a recall rate of 0.9323 (an increase of 2.62%). This indicates that the module effectively reduces interference such as bubble noise, enhances and adjusts various aspects of feature processing, expands the perception range, and thus improves the performance of the model in segmentation tasks.

The complete model integrating BMFM, BFRM, and FNRM achieves state-of-the-art performance, with an mIoU of 0.8655 and an F2 score of 0.9251, which are 3.2% and 3.5% higher than some configurations, respectively. The hierarchical integration of BMFM, BFRM, and FNRM establishes a balanced framework while optimizing local texture preservation and global semantic consistency.

4.6. Comparison of Inference Time and Model Complexity

The complexity of the model is a key consideration, which not only affects the training and inference efficiency of the model, but also is closely related to the computational resource requirements. In Table 4, the number of parameters (Params), floating point operations (Flops) and inference time comparison of the six models are presented. From the dimension of parameter number, the Params of the proposed HPANet is only 26.3 M, which is about 41% less than that of CaraNet and 6.4% less than that of Polyp_PVT. From the perspective of Flops, the floating-point number of HPANet is only 15.4 G, which is far lower than 227.53 G of Unet and 156.3 G of ResUnet. In terms of inference time, HPANet achieves 42 FPS. Although PraNet leads with 46 FPS, HPANet still provides good real-time performance and is faster than U-Net (32 FPS) and ResUnet (35 FPS).

Table 4. Comparison of inference time and model complexity of different models. "↓"indicates lower values are better. "↑" denotes higher values are better.

Model	Params ↓ (M)	Flops ↓ (G)	Inference Time ↑ (FPS)
U-Net [3]	17.26	227.53	32
ResUnet [34]	32.63	156.3	35
PraNet [7]	30.5	13.15	46
CaraNet [35]	44.59	21.75	38
Polyp-PVT [19]	28.11	17.02	40
HPANet	26.3	15.4	42

5. Conclusions

The proposed HPANet framework constructs a new intestinal polyp segmentation network by designing three key modules. The BMFM systematically solves the problem of polyp scale variation by reversely fusing the feature maps generated by the PVT-based encoder to establish hierarchical feature dependencies. The FNRM removes noise in the polyp image through forward progressive fusion of features, solving the problem of unclear and discontinuous segmentation caused by noise. The BFRM reduces segmentation errors by re-calibrating boundary details under the guidance of channel and spatial attention.

The proposed method was tested on five public image datasets and compared with representative U-Net, ResUNet, PraNet, CaraNet, and Polyp-PVT segmentation models. The experimental results show that the proposed architecture has more accurate segmentation results than traditional CNN-based methods, and has better performance than the same polyp-PVT model based on PVT. At the same time, in order to prove the effectiveness of the three modules proposed in this article, ablation experiments were conducted. The experimental data showed that these three modules improved the performance of the model in different aspects.

Author Contributions: Conceptualization, Y.Y.; methodology, Y.Y.; software, Y.Y.; validation, M.L., H.L. and Y.Z.; formal analysis, M.L.; investigation, Y.Y.; resources, Y.Y.; data curation, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Z. and M.L.; visualization, Y.Y. and H.L.; supervision, Y.Y. and M.L.; project administration, Y.Y.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Fujian Province, grant number 2023J01078; and the special fund for scientific and technological innovation of Fujian Agriculture and Forestry University, grant number No. KFB23151.

Data Availability Statement: This study used datasets that can be accessed through the following link: https://drive.google.com/file/d/1pFxb9NbM8mj_rlSawTlcXG1OdVGAbRQC/view?usp=sharing (accessed on 29 March 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CRC	Colorectal Cancer
OC	Optical Colonoscopy
CNN	Convolutional Neural Network
PVTv2	Pyramid Vision Transformer v2
ViT	Vision Transformer
MLP	Multi-Layer Perceptron
GAP	Global Average Pooling

CBAM	Convolutional Block Attention Module
SA	Spatial Attention
CA	Channel Attention
IoU	Intersection over Union
BCE	Binary Cross-Entropy
mDice	Mean Dice Coefficient
PPD	Parallel Partial Decoder
SAM	Spatial Attention Map
BEB	Boundary Enhance Block

References

1. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 233–254. [CrossRef] [PubMed]
2. Biller, L.H.; Schrag, D. Diagnosis and Treatment of Metastatic Colorectal Cancer: A Review. *JAMA* **2021**, *325*, 669–685. [CrossRef] [PubMed]
3. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
4. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.
5. Fang, Y.; Zhu, D.; Yao, J.; Yuan, Y.; Tong, K. ABC-Net: Area-Boundary Constraint Network with Dynamical Feature Selection for Colorectal Polyp Segmentation. *IEEE Sens. J.* **2020**, *21*, 11799–11809. [CrossRef]
6. Chen, F.; Ma, H.; Zhang, W. SegT: Separated Edge-Guidance Transformer Network for Polyp Segmentation. *Math. Biosci. Eng* **2023**, *20*, 17803–17821. [CrossRef]
7. Fan, D.P.; Ji, G.P.; Zhou, T.; Chen, G.; Fu, H.; Shen, J.; Shao, L. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*; Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12266, pp. 263–273, ISBN 978-3-030-59724-5.
8. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
9. Zhang, Z.; Zhang, W. Pyramid Medical Transformer for Medical Image Segmentation. *arXiv* **2022**, arXiv:2104.14702.
10. Chowdary, G.J.; Yin, Z. Med-Former: A Transformer Based Architecture for Medical Image Classification. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2024*; Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A., Eds.; Lecture Notes in Computer Science; Springer Nature Switzerland: Cham, Switzerland, 2024; Volume 15011, pp. 448–457, ISBN 978-3-031-72119-9.
11. Wang, W.; Xie, E.; Li, X.; Fan, D.-P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. PVT v2: Improved Baselines with Pyramid Vision Transformer. *Comput. Vis. Media* **2022**, *8*, 415–424. [CrossRef]
12. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 652–662. [CrossRef]
13. Wei, J.; Hu, Y.; Zhang, R.; Li, Z.; Zhou, S.K.; Cui, S. Shallow Attention Network for Polyp Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*; De Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 12901, pp. 699–708, ISBN 978-3-030-87192-5.
14. Tomar, N.K.; Jha, D.; Bagci, U.; Ali, S. TGANet: Text-Guided Attention for Improved Polyp Segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*; Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S., Eds.; Lecture Notes in Computer Science; Springer Nature Switzerland: Cham, Switzerland, 2022; Volume 13433, pp. 151–160, ISBN 978-3-031-16436-1.
15. Oukdach, Y.; Garbaz, A.; Kerkaou, Z.; El Ansari, M.; Koutti, L.; El Ouafdi, A.F.; Salihoun, M. UViT-Seg: An Efficient ViT and U-Net-Based Framework for Accurate Colorectal Polyp Segmentation in Colonoscopy and WCE Images. *J. Imaging Inform. Med.* **2024**, *37*, 2354–2374. [CrossRef]
16. Wang, Z.; Liu, Z.; Yu, J.; Gao, Y.; Liu, M. Multi-Scale Nested UNet with Transformer for Colorectal Polyp Segmentation. *J. Appl. Clin. Med. Phys.* **2024**, *25*, e14351. [CrossRef]
17. Zhang, W.; Fu, C.; Zheng, Y.; Zhang, F.; Zhao, Y.; Sham, C.-W. HSNet: A Hybrid Semantic Network for Polyp Segmentation. *Comput. Biol. Med.* **2022**, *150*, 106173. [CrossRef] [PubMed]

18. Wang, J.; Tian, S.; Yu, L.; Zhou, Z.; Wang, F.; Wang, Y. HIGF-Net: Hierarchical Information-Guided Fusion Network for Polyp Segmentation Based on Transformer and Convolution Feature Learning. *Comput. Biol. Med.* **2023**, *161*, 107038. [CrossRef] [PubMed]
19. Dong, B.; Wang, W.; Fan, D.P.; Li, J.; Fu, H.; Shao, L. Polyp-PVT: Polyp Segmentation with Pyramid Vision Transformers. *CAAI Artif. Intell. Res.* **2023**, *2*, 9150015. [CrossRef]
20. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
21. Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; Wang, Y. Transformer in Transformer. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 15908–15919.
22. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going Deeper with Image Transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 32–42.
23. Zhou, D.; Kang, B.; Jin, X.; Yang, L.; Lian, X.; Jiang, Z.; Hou, Q.; Feng, J. DeepViT: Towards Deeper Vision Transformer. *arXiv* **2021**, arXiv:2103.11886.
24. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021.
25. Wang, W.; Xie, E.; Li, X.; Fan, D.-P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021.
26. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
27. Bhojanapalli, S.; Chakrabarti, A.; Glasner, D.; Li, D.; Unterthiner, T.; Veit, A. Understanding Robustness of Transformers for Image Classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021.
28. Khanh, T.L.B.; Dao, D.-P.; Ho, N.-H.; Yang, H.-J.; Baek, E.-T.; Lee, G.; Kim, S.-H.; Yoo, S.B. Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging. *Appl. Sci.* **2020**, *10*, 5729. [CrossRef]
29. Bernal, J.; Sánchez, F.J.; Fernández-Esparrach, G.; Gil, D.; Rodríguez, C.; Vilariño, F. WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps from Physicians. *Comput. Med. Imaging Graph.* **2015**, *43*, 99–111. [CrossRef]
30. Jha, D.; Smedsrud, P.H.; Riegler, M.A.; Halvorsen, P.; de Lange, T.; Johansen, D.; Johansen, H.D. Kvasir-SEG: A Segmented Polyp Dataset. In Proceedings of the MultiMedia Modeling, Daejeon, Republic of Korea, 5–8 January 2020; Ro, Y.M., Cheng, W.-H., Kim, J., Chu, W.T., Cui, P., Choi, J.W., Hu, M.C., De Neve, W., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 451–462.
31. Vázquez, D.; Bernal, J.; Sánchez, F.J.; Fernández-Esparrach, G.; López, A.M.; Romero, A.; Drozdzal, M.; Courville, A. A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. *J. Healthc. Eng.* **2017**, *2017*, 4037190. [CrossRef]
32. Geetha, K.; Rajan, C. Automatic Colorectal Polyp Detection in Colonoscopy Video Frames. *Asian Pac. J. Cancer Prev.* **2016**, *17*, 4869–4873. [CrossRef]
33. Silva, J.; Histace, A.; Romain, O.; Dray, X.; Granado, B. Toward Embedded Detection of Polyps in WCE Images for Early Diagnosis of Colorectal Cancer. *Int. J. Comput. Assist. Radiol. Surg.* **2014**, *9*, 283–293. [CrossRef] [PubMed]
34. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
35. Lou, A.; Guan, S.; Loew, M. CaraNet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects. *J. Med. Imaging* **2023**, *10*, 014005. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.