Pattern Recognition 132 (2022) 108917

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

BSCA-Net: Bit Slicing Context Attention network for polyp segmentation

Yi Lin[a,1], Jichun Wu[a,1], Guobao Xiao[a,*], Junwen Guo[a], Geng Chen[b], Jiayi Ma[c]

[a] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China
[b] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, China
[c] Electronic Information School, Wuhan University, Wuhan, 430072, China

ARTICLE INFO

Article history:
Received 4 March 2022
Revised 18 July 2022
Accepted 20 July 2022
Available online 21 July 2022

Keywords:
Medical image segmentation
Polyp segmentation
Colonoscopy
Attention mechanism

ABSTRACT

In this paper, we propose a novel Bit-Slicing Context Attention Network (BSCA-Net), an end-to-end network, to improve the extraction ability of boundary information for polyp segmentation. The core of BSCA-Net is a new Bit Slice Context Attention (BSCA) module, which exploits the bit-plane slicing information to effectively extract the boundary information between polyps and the surrounding tissue. In addition, we design a novel Split-Squeeze-Bottleneck-Union (SSBU) module, to exploit the geometrical information from different aspects. Also, based on SSBU, we propose an multipath concat attention decoder (MCAD) and an multipath attention concat encoder (MACE), to further improve the network performance for polyp segmentation. Finally, by combining BSCA, SSBU, MCAD and MACE, the proposed BSCA-Net is able to effectively suppress noises in feature maps, and simultaneously improve the ability of feature expression in different levels, for polyp segmentation. Empirical experiments on five benchmark datasets (Kvasir, CVC-ClinicDB, ETIS, CVC-ColonDB and CVC-300) demonstrate the superior of the proposed BSCA-Net over existing cutting-edge methods.

1. Introduction

Image segmentation has aroused great concern from computer vision [1–3]. It focuses on classifying each pixel according to the context information in the image. In recent years, image segmentation methods are widely adopted to specific domains such as polyp segmentation which aims to identify and segment polyp from an image [4]. They have also been applied to medical image analysis, providing a valuable information for clinical diagnosis and pathology research, and pathology studies.

In clinical settings, polyp segmentation is essential to provide key information for prevention of the colorectal cancer. The colorectal cancer is the third most prevailing cancer and the fourth most fatal cancer on earth, imperiling human life and safety [5]. Early colonoscopy averts millions of deaths from colorectal cancer. According to the information of polyps from colonoscopy, doctors can remove colorectal polyps before they develop into colorectal cancer. With the help of automatic polyp segmentation, colonoscopy specialists can identify 20% more colon polyps to improve diagnostic accuracy [6].

Recently, convolutional neural networks (CNNs) are introduced for polyp segmentation and have achieved great success. CNN based polyp segmentation methods, such as, Psi-Net [7], Pra-Net [8] and SA-Net [9], are generally inspired by the salient object detection (SOD) since they pay more attention on object (polyp) region than the surrounding scene. Note that, the edge guidance (that plays an important role in SOD) requires additional edge data, and this often makes a polyp segmentation method suffer from computational inefficiency. To obtain the shape and boundary information, Psi-Net constructs three parallel decoders for polyp segmentation. Pra-Net introduces reverse attention [10] to reverse coarse segmentation map for obtaining boundary cues as the edge guidance. SA-Net proposes a shallow attention module based on the complementarity of different features, and multiplies shallow features with deep features to obtain clearer boundary information in deep features. However, the boundary information derived from Psi-Net, Pra-Net and SA-Net, is often ambiguous, which will lead to sub-optimal performance for polyp segmentation.

To address the problem, we develop a Bit-Slicing Context Attention (BSCA) mechanism to obtain the boundary information

* Corresponding author.
E-mail addresses: gbx@mju.edu.cn, x-gb@163.com (G. Xiao).
[1] Equal contribution.

https://doi.org/10.1016/j.patcog.2022.108917
0031-3203/

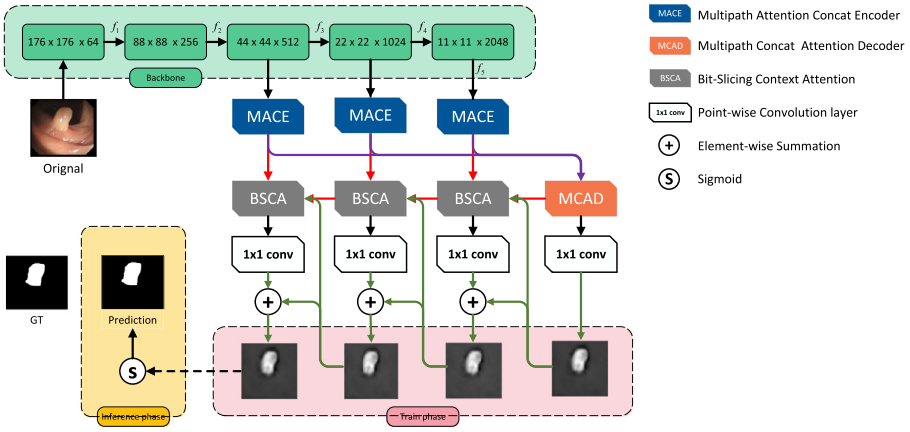

Fig. 1. Overall architecture of the proposed BSCA-Net..

by exploiting the bit-plane slicing information. Note that, high-bit planes involve the bulk of valid boundary information since there is little variability in high bit planes in a region. In contrast, low-bit planes include less boundary cues, and cost more computational resources due to their larger noisy information. Thus, we propose a novel BSCA module, which utilizes high bit planes to capture valid boundary information from a global feature map. For polyp segmentation, the proposed BSCA module is able to effectively extract boundary information when polyps are similar to their surroundings.

In addition, the existing U-Net structures, such as U-Net [11], U-Net++ [12], and ResU-Net++ [13], generally use simple addition or concatenation operations to fuse different level features progressively from the encoder and forward them to the decoder. For feature fusion, these two operations tend to generate lots of redundant information, which will weaken the really useful features and the characteristics of level-specific features simultaneously, leading to inaccurate details and rough boundaries of polyps. More specific, due to the redundant information generated by the fusion operation, the representation ability of feature maps may be decreased, which even results in the loss of really useful features that are essential to accurate polyp segmentation.

To address the above challenge, we propose a novel Split-Squeeze-Bottleneck-Union (SSBU) module, which consists of Split, Squeeze, Bottleneck and Union operations. The proposed SSBU is able to exploit the geometrical information from different aspects. Specifically, the Split operation generates multiple paths to exploit the geometrical information of feature maps from different aspects. The Squeeze operation compresses channel-wise information of feature maps to produce the channel descriptors. The Bottleneck operation extracts the channel dependence to learn for achieving the weight of each channel. The Union operator combines and aggregates the geometrical information from multiple paths. Then, we design an multipath concat attention decoder (MCAD) to enhance the representation ability of feature maps during fusing different levels of features, and an multipath attention concat encoder (MACE) to extract comprehensive information of level-specific feature maps.

Finally, with the proposed BSCA, SSBU, MCAD and MACE, we propose a novel simple and effective framework, termed as Bit-Slicing Context Attention Network (BSCA-Net), for polyp segmentation. We show the overall architecture of BSCA-Net in Fig. 1. The proposed BSCA-Net is able to improve the extraction ability of boundary information.

In summary, the contributions of this paper are threefold:

- We propose a novel attention mechanism by exploiting the bit-plane slicing information, for polyp segmentation. The proposed attention mechanism is able to capture the edge guidance from a global map and further enhance the representation ability of level-specific feature maps through high-bit plane learning.
- We design a novel SSBU module to extract rich contextual and geometric information. Based on the SSBU module, we propose two effective decoder and encoder to capture geometric information, for providing a comprehensive representation of the input features.
- Extensive experiments demonstrate that our approach advances state-of-the-art performance for polyp segmentation and outperforms most cutting-edge models. In particular, on CVC300 dataset, BSCA-Net achieves a 6.0% boost for mean IoU, and improves mean Dice from 88.8% to 92.7%.

The rest of the paper is organized as follows: In Section 2, we briefly introduce some related work in the area. Then, we provide the details of the proposed method in Section 3, and present the experimental results and discussions in Section 4. Finally, we draw a conclusion in Section 5.

2. Related work

In the section, we briefly review the learning-based polyp segmentation methods highly related to our paper. In addition, we also introduce some attention mechanisms, which are related to our work.

2.1. Polyp segmentation

Polyp segmentation is a pixel-level task, which exactly segments polyps from the colonoscopy image. However, exacting polyp segmentation is very challenging, since polyps are similar to its surrounding tissues in the colonoscopy image and polyps often include various size, color and texture. Thus, polyp segmentation networks generally focus on extracting semantic features with the detail information. To extract color, shape, texture and appearance features, early traditional methods utilize manually designed features [14,15]. Nevertheless, these methods often have high miss-detection rate due to the limited representation of manually designed features. Thus, polyp segmentation has gradual developments from traditional methods to deep learning methods [4,16,17]. These deep learning methods detect polyps by box-level prediction results, but they fail to locate accurate shape and contour of polyps. To improve the performance of polyp segmentation, a fully convolutional network (FCN) is used to identify and segment polyps from colonoscopy images [18,19]. However, FCN-based methods rely on low-resolution features to generate the fi-

nal prediction, resulting in rough segmentation results and fuzzy boundaries.

Recently, U-Net [11] is a typical structure network for polyp segmentation, due to its fusion ability for the semantic information and spatial details from different level features. For U-Net, the encoder block extracts feature maps from input images, and the decoder block optimizes features of encoder and designs the task for polyp segmentation. U-Net directly adopts simple skip connections to fuse feature maps from the encoder to the decoder. Simple skip connections, which mean that the feature maps of the encoder are directly received in the decoder, may depress the segmentation performance. To solve this problem, U-Net++ [12] reformulates the dense skip connections to combine feature maps between the encoder and the decoder and obtains promising performance. Compared with simple skip connections, the feature maps of the encoder undergo a dense convolution block, then they are received in the decoder for the dense skip connections. Later, ResU-Net++ [13] introduces four advanced techniques: residual computation [20], squeeze and excitation [21], atrous spatial pyramidal pooling [22], and attention mechanism, to further improve segmentation performance. Recently, EU-Net [23] develops a semantic feature Enhancement Module to enhance the semantic information by applying different sizes of patch-wise non-local attention block. In addition, Threshold-Net [24] proposes a two-branch network including a threshold branch and a segmentation branch. The threshold branch is a decoder to utilize semantic information for predicting the threshold map, while the segmentation branch decodes the same semantic information to predict the likelihood map. ACS-Net [25] designs an adaptive context selection based U-Net framework to enhance features according to the size of the polyp region. These methods concentrate on segmenting the polyp region, but they ignore the region-boundary relationship, which plays a critical role in the performance of polyp segmentation [8].

There are also some methods [7,8,26] based on U-Net, which restore the boundary for polyp segmentation by building the relationship between the boundary and region features. Psi-Net [7] constructs three parallel decoders to further obtain shape and boundary information. SFA-Net [26] proposes a selective feature aggregation structure, with a shared encoder and two mutually constrained decoders, to predict region and boundary of polyps. However, both two methods do not fully capture the relationship between region and boundary. Recently, Pra-Net [8] puts forward partial decoders to generate the global map, and utilizes the reverse global map to mine the dependence between the boundary and the polyp region. But this method lacks the expression of feature maps, which results inaccurate region location in the blurry boundary of polyps.

Our network is also based on U-Net, but it generates multiple paths to capture the geometrical information of feature maps from different aspects, which enhances a comprehensive understanding of the input features. Note that, our network is used for each scale to further exploit the information of feature maps.

2.2. Multi-scale representation

Multi-scale representation has been widely used in computer vision. Image pyramid is an effective multi-scale representation structure to interpret images with multi-resolution for processing multi-scale objects across visual tasks. In machine learning, some researchers [27,28] use different resolution classifiers to detect objects, which contribute to detecting small objects in relatively small window classifiers. In CNNS, some object detection tasks [29,30] extract features for each layer of the image pyramid. DeepLab [22] proposes the Atrous Spatial Pyramid Pooling method to fuse multi-scale context information through different scales of receptive fields in semantic segmentation. DeepLab continues pooling and downsampling results in reduced resolution, resulting in insufficient accuracy in semantic segmentation. In the segmentation task, the methods based on image pyramid extract different scales features from multi-scale input images for fusion, to improve the network performance. However, due to the multi-scale input of image pyramid, a large number of gradients are calculated and saved in memory, resulting in high requirements on hardware.

Bit-plane is introduced to obtain the significant and insignificant bits [31], which can cluster and generate more images for brain tumor segmentation. However, the information from two significant bits is not sufficient to capture complete information. 8 bit planes are used in the RGB dermoscopic images for each color component to segment skin lesion [32]. However, the above methods are designed with highly regular input data formats, i.e., RGB images. Note that, deep learning based polyp segmentation methods require the input data to be range $[0 - 1]$, which is beneficial to make the model training convergence stable.

In this paper, we try to introduce the multi-scale representation (the bit slicing method) and design a novel BSCA module to better capture boundary information at different scales for polyp segmentation.

2.3. Attention mechanism

The attention mechanism is a dynamic selection process, which adaptively weights features according to the importance of input. Attention mechanisms have been widely used in many visual tasks. For example, RMA [14] utilizes an attention mechanism to handle the image classification issue. PESA-Net [33] employs the permutation-equivariant split attention mechanism to correspondence learning. Reverse attention [10] proposes a novel attention to guide side-output residual learning for the salient object detection. Since the attention mechanism comes into polyp segmentation, the segmentation performance has made great advances [9,14,17,34,35]. Mnih et al. [14] adopted attention mechanisms to obtain promising performance for polyp segmentation. Sang et al. [34] applied attention gate to the skip connections between the encoder and the decoder for U-Net. Tomar et al. [35] introduced a squeeze and excitation layer to dynamically assign weights to each feature channel, which serves as a channel attention mechanism. PNS-Net [17] utilizes the normalized self-attention (NS) block to quicken the inference rate. For excluding the effects of color, SA-Net [9] designs the color exchange operation to capture the image contents and colors, and forces the model to focus on the polyp shape and structure. Although these methods realize some performance optimizations, they ignore the important value of the extraction ability of boundary information from feature maps.

In this work, we design a novel BSCA mechanism to obtain plenty of valid boundary information through the region-boundary relationship, for improving the polyp segmentation accuracy.

3. The proposed method

In this section, we demonstrate the architecture of the proposed Bit-Slicing Context Attention network (BSCA-Net). First we introduce the overall architecture. Then we introduce the details and specific functions of each module, including Bit Slicing Context Attention (BSCA) module, Multipath Attention Concat Encoder (MACE) and Multipath Concat Attention Decoder (MCAD).

3.1. Overall architecture

As shown in Fig. 1, we design the overall architecture of BSCA-Net, which consists of BSCA, MACE and MCAD. In BSCA-Net, we extract five levels of features (f_1, f_2, f_3, f_4 and f_5) from the Res2Net-

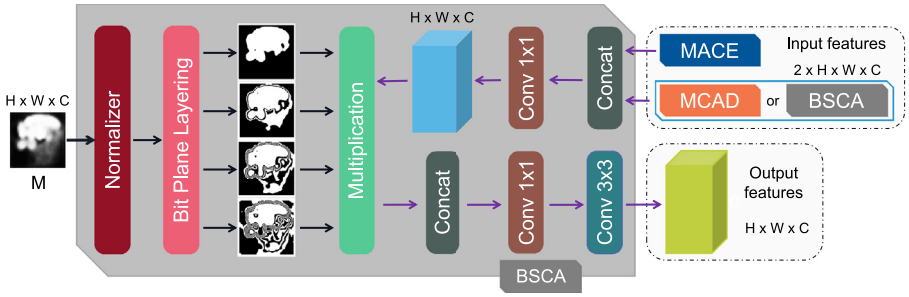

Fig. 2. The architecture of our Bit Slicing Context Attention module.

based backbone network [36], and take (f_1, f_2) and (f_3, f_4, f_5) as low-level features and high-level features, respectively. The reason behind this is that, compared with low-level features, high-level features use less computing resources since high-level features include less spatial resolution and more core features [10].

Based on the five levels of features from the backbone network, we add a novel BSCA module to extract boundary information. We also design a Split-Squeeze-Bottleneck-Union (SSBU) module to enhance the representation ability of feature maps during the fusion of different levels of features. For actively using SSBU, we implement an encoder (MACE) for BSCA-Net. This helps to extract global features from different levels of features. Also, we design a simple decoder (MCAD) corresponding to MACE, which contributes to aggregating the final features from different levels of encoder features.

Specifically, MCAD gathers features from three encoder modules to fusion features. Features from MCAD are used to predict an initial guidance map by an 1×1 Conv block. BSCA concatenates features from MACE and MCAD, and uses an initial guidance map to the context guidance. Note that, the output features of BSCA, not only are directly connected to the next BSCA, but also are used to generate a new global guidance map by an 1×1 Conv Block. Then, the global guidance map from BSCA is added with the initial guidance map from MCAD, which complements the global guidance map to obtain more accurate guidance for the next BSCA. After the first BSCA, we concatenate features of MACE and the output of previous BSCA in the next BSCA. Also, the guidance map from the previous BSCA is used as a context guidance in the next BSCA. After three recurrent BSCAs, the final output is computed with the sigmoid function to obtain a final prediction.

In summary, the overall architecture shows that the backbone features are encoded by MACE, and encoded features are forwarded to the decoder for the initial guidance map, which leads BSCA to learn a residual guidance map apart from the initial map. This helps the consequent BSCA focus more on the unclear area. To further attain boundary information, three BSCAs are used to build relationship between regions and boundaries. Recurrently using BSCA to extract boundary information, helps to calibrate some fault predictions.

3.2. Bit slicing context attention module

In this subsection, we propose a novel attention mechanism (BSCA), which exploits the bit plane slicing information to aggregate features and the context guidance for obtaining the boundary information, as shown in Fig. 2.

For the input guidance map M of BSCA, to ensure the stability of data from M, we implement the Min-Max Normalization to scale M to the range $[0 - 1]$. The normalized guidance map M_{norm} is computed as follows:

$$M_{norm} = \frac{M - M_{\min}}{M_{\max} - M_{\min}}, \tag{1}$$

where $M_{\min}$ and $M_{\max}$ represent the minimum and maximum values in M, respectively.

Given the normalized guidance map M_{norm}, we can obtain bit planes by Algorithm 1. We can see that, the first bit from our algorithm is about $1/2$ of M_{norm}, and the second bit is about $1/2$ of the remainder of the first bit, and so on to the eighth bit. Finally, we downsample the values of eight bit planes composed of a set of bits, resizing them to 0 or 1.

Algorithm 1 The proposed bit-plane slicing algorithm.

Input: Salience~map~M_{norm}.
1: $Bit_1 \longleftarrow \lfloor M_{norm}/2^{-1} \rfloor$
2: $M_{norm} \longleftarrow M_{norm} \bmod 2^{-1}$
3: $Bit_2 \longleftarrow \lfloor M_{norm}/2^{-2} \rfloor$
4: $M_{norm} \longleftarrow M_{norm} \bmod 2^{-2}$
5: $Bit_3 \longleftarrow \lfloor M_{norm}/2^{-3} \rfloor$
6: $M_{norm} \longleftarrow M_{norm} \bmod 2^{-3}$
7: $Bit_4 \longleftarrow \lfloor M_{norm}/2^{-4} \rfloor$
Output: Four bit planes (Bit_1, Bit_2, Bit_3 and Bit_4).

We show an example of the slices from first bit to eighth bit planes in Fig. 3. We realize that the first bit plane contains the most significant bit. However, the sequent bit planes have other information, which is highly related to boundary information. Instead of using a single plane, we argue that the combination of multiple bit planes is beneficial to extract comprehensive boundary information. In contrast to using all bit planes, actively selecting bit planes to compute can reduce the impact of noises and computing resources. In Section 4.3, we show that the first four bit planes are the most effective combination.

To obtain useful boundary information by exploiting the first four bit planes, we propose a novel attention mechanism, i.e., BSCA. In BSCA, we concatenate the input features from MACE and previous BSCA (or MCAD) as the input feature f_{in}:

$$f_{in} = \text{Cat}\left\{f_{mace}, f_{previous}\right\}, \tag{2}$$

where f_{mace} denotes features from MACE module and $f_{previous}$ denotes features from previous BSCA or MCAD. Then, we use a point-wise convolution on f_{in} to get the global features f_g, as follows:

$$f_g = \omega(f_{in}) \tag{3}$$

where $\omega(\cdot)$ represents a point-wise convolution.

We eliminate the last four bit planes to remove noises and only use the first four bit planes to extract useful boundary information. We use the first bit plane to keep the most significant data and employ the next bit planes to complement more boundary information. Then, we can obtain the first four bit planes (Bit_1, Bit_2, Bit_3 and Bit_4) by Algorithm 1.

Then, we exploit the first four bit planes $\{Bit_i, i = 1, 2, 3, 4\}$ from the guidance map M_{norm} as bit attention weights to implement

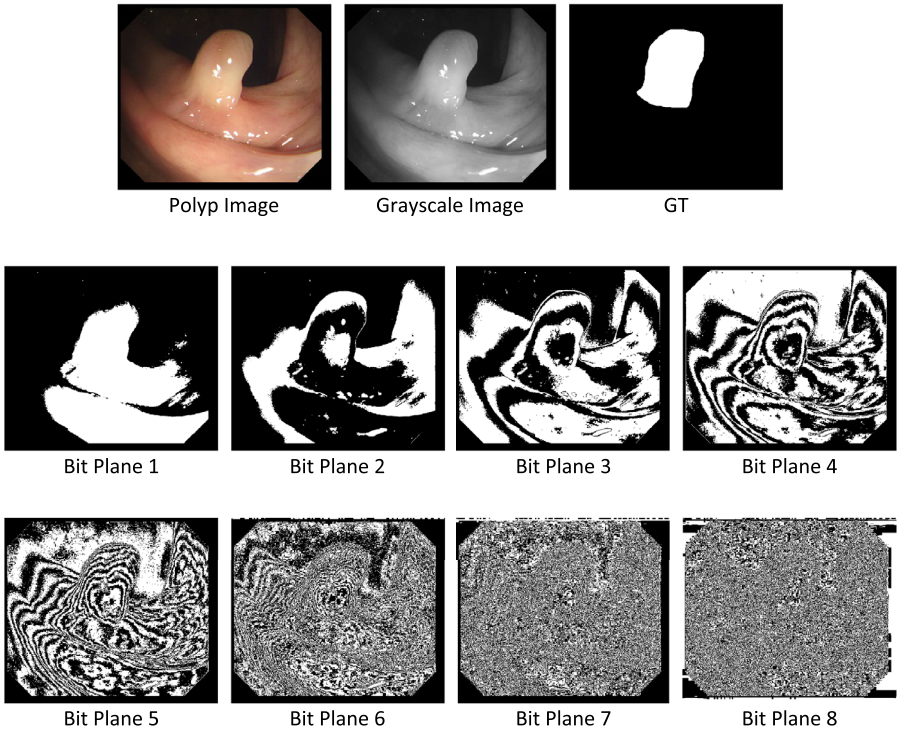

Fig. 3. The results of dividing the image into 8 bit planes.

context guidance with the global feature f_g. We write the output bit attention features B_i, as follows:

$$B_i = f_g \odot Bit_i, \tag{4}$$

where $\odot$ represents an element-wise multiplication between f_g and Bit_i. Then, we concatenate all four bit planes with respect to the channel axis, which contributes to the merge of features. After that, we adopt Conv blocks to generate output features f_{out} with the point-wise convolution, as follow:

$$f_{out} = \psi_3(\psi_1(\mathrm{Cat}\{B_1, B_2, B_3, B_4\})) \tag{5}$$

where $\psi_1(\cdot)$ and $\psi_3(\cdot)$ denote one 1×1 Conv block and one 3×3 Conv block, respectively.

3.3. Multipath attention concat encoder and multipath concat attention decoder

Most deep learning networks for segmentation task focus on foreground by erasing semantic features from background, since foreground and background are difficult to be directly separated. To this end, BAM [37] proposes the bottleneck attention mechanism, which combines channel attention and spatial attention to get an 3D attention. Nonetheless, this method cannot capture global contextual information.

In this subsection, we propose an SSBU module to capture global contextual information for separating background from a target. Specifically, we implement SSBU through four operations: *Split, Squeeze, Bottleneck and Union*, as illustrated in Fig. 4.

Split and Squeeze: We divide features $F_{in} \in \mathbb{R}^{H\times W\times C}$ into four groups of features $FG \in \mathbb{R}^{H\times W\times C/4}$, where H and W denote the size of the input features and C is the number of channels for the input features. Then, we employ the global average pooling to capture channel-wise information GC of feature maps $\{FG_s, s = 1, 2, 3, 4\}$:

$$GC = \frac{1}{4}\sum_{s=1}^{4} FG_s. \tag{6}$$

Bottleneck: To take advantage of channel-wise information GC obtained by global average pooling, we put channel-wise information GC into the bottleneck attention module. That is, we use the bottleneck attention BA to distinguish the target from the background:

$$BA = \mathcal{BAM}(GC), \tag{7}$$

where $\mathcal{BAM}$ is the bottleneck attention module [37]. We extend the channel number of BA to C, and divide BA into four groups $\{BA'_s, s = 1, 2, 3, 4\}$. Then we multiply BA'_s by the feature group FG_s to obtain the final feature group FG'_s. The specific process is described as follows:

$$FG'_s = \sum_{s=1}^{4} BA'_s \odot FG_s, \tag{8}$$

where $\odot$ is element-wise multiplication.

Union: To combine the geometric context information in multiple paths, we use an union manner to aggregate all the feature-map groups. Concretely, we axially connect final feature groups to obtain the output feature F_{out}:

$$F_{out} = \mathrm{Cat}\left\{FG'_1, FG'_2, FG'_3, FG'_4\right\}. \tag{9}$$

We propose to integrate SSBU into MACE and MCAD to represent the globally refined feature. First, we design MACE to aggregate the low-level features from backbone. Pra-Net directly flows features from the backbone into the attention mechanism without reducing the number of channels, which generates redundant information and parameters. Redundant information hinders the performance of the network and superfluous parameters cause unnecessary time overhead during training and testing. To solve these problems, we design MACE combined with SSBU and RFB [38] to enhance the representation ability of network features and extract global features through different sensing fields. As shown in Fig. 5, we flow features from the backbone into different-sized receptive field paths. We add SSBUs on every receptive field path to enable

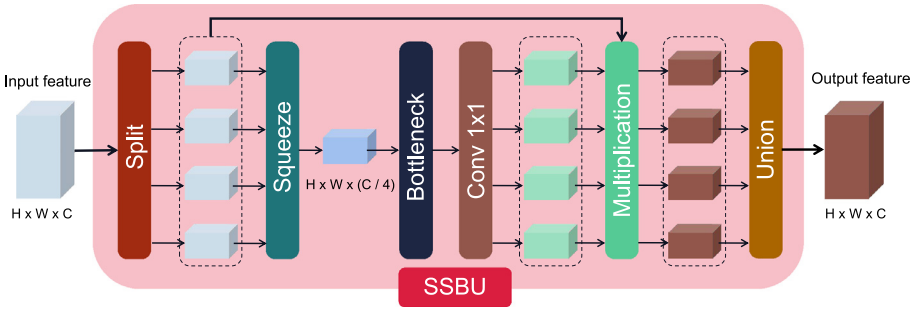

Fig. 4. The architecture of our SSBU module. .

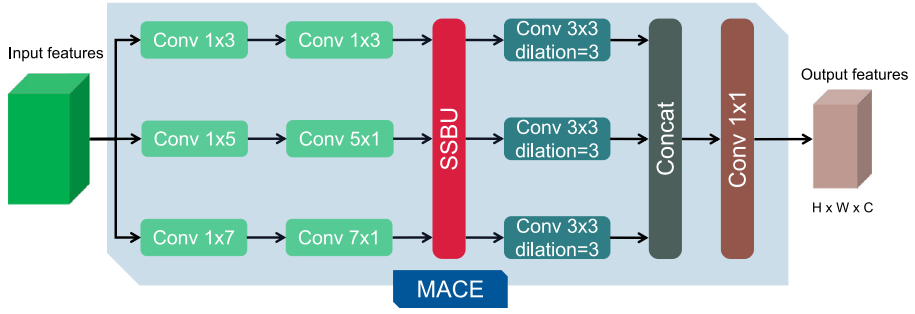

Fig. 5. The architecture of our MACE..

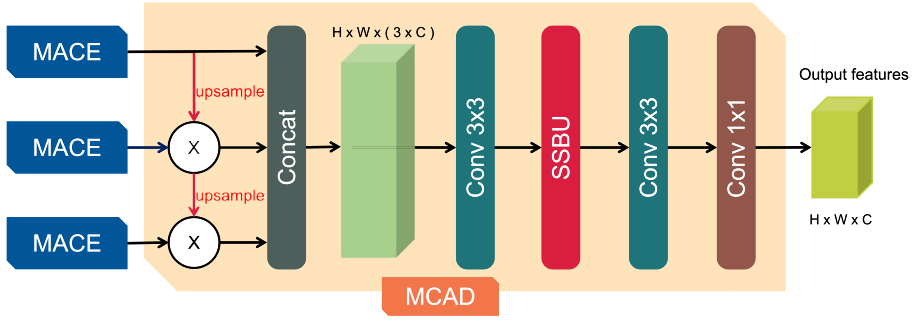

Fig. 6. The architecture of our MCAD..

each path to get global refined features. At the same time, we design a simple MCAD corresponding to MACE. As shown in Fig. 6, we adjust features from three MACEs to the same size, and use the dot product operation to reduce noises in the low-level features. In the end, we get the final feature by aggregating all features together through the concatenation operation.

4. Experiments

In this section, we describe the details of implementation, compare our model with the seven previous state-of-the-art models, and show our ablation experiments.

4.1. Implementation details

Baselines: We compare our BSCA-Net with fourteen medical image segmentation methods: U-Net [11], U-Net++ [12], ResUNet [39], ResUNet++ [13], SFA [40], ACS-Net [25], Pra-Net [8], Threshold-Net [24], HarDNet-MSEG [41], TransUnet [42],TransFuse-S/L [43], EU-Net [23] and SA-Net [9]. Among them, U-Net, U-Net++, ResUNet and ResUNet++ are classic image segmentation methods; Pra-Net, SFA, ACS-Net, Threshold-Net, HarDNet-MSEG, TransUnet, TransFuse-S/L, EU-Net and SA-Net are recently advanced polyp segmentation methods.

Datasets and Training Settings: Our experiment is conducted on five datasets: CVC300 [44], CVC-ClinicDB [45], CVC-ColonDB [46], ETIS [5] and Kvasir [47].

Kvasir includes three important anatomical landmarks, three clinically significant findings and two categories of images related to endoscopic polyp removal. The size of images varies from 332×487 to 1920×1072. Also, the polyps in the images vary in size and shape.

CVC-ClinicDB (also called CVC-612), contains 612 public polyp images from 25 colonoscopy videos. The size of images is 384×288.

CVC-ColonDB is a small training set, which contains 15 different colonoscopy sequences and 380 polyp images.

CVC300 is a part of EndoScene, which also contains some images from CVC-ClinicDB. Since the CVC-ClinicDB images are used to train, we only use images from CVC300 for generalization evaluation of the model.

ETIS is an early dataset containing 196 images from 34 colonoscopy videos. The size of images is 1225×966, and it is the largest size among all five datasets. Note that, the polyps images in this dataset are mostly small and hard to find, which makes this dataset more challenging.

To ensure the fairness of our experimental results, we fully follow the recommendations of [8,9]. That is, we use 80% of im-

Table 1
Quantitative results obtained by fifteen medical image segmentation methods for the learning ability evaluation on Kvasir and CVC-ClinicDB datasets. '-' denotes that the corresponding value is not reported.

Dataset	Kvasir		ClinicDB	
	Dice	IoU	Dice	IoU
U-Net (MICCAI'15)	0.818	0.746	0.823	0.750
U-Net+ (TMI'19)	0.821	0.743	0.794	0.729
ResUNet	0.791	-	0.779	-
ResUNet+	0.813	0.793	0.796	0.796
SFA (MICCAI'19)	0.723	0.611	0.700	0.607
ACS-Net (MICCAI'20)	0.898	0.838	0.882	0.826
Pra-Net (MICCAI'20)	0.898	0.840	0.899	0.849
Threshold-Net (TMI'20)	0.798	0.708	0.859	0.796
HarDNet-MSEG (arxiv)	0.912	0.857	0.932	0.882
TransUnet (arxiv)	0.913	0.857	0.935	0.887
TransFuse-S (MICCAI'21)	0.918	0.868	0.918	0.868
TransFuse-L (MICCAI'21)	0.918	0.868	0.934	0.886
EU-Net (CRV'21)	0.908	0.854	0.902	0.846
SA-Net (MICCAI'21)	0.904	0.847	0.916	0.859
BSCA-Net (Ours)	0.910	0.855	0.926	0.874

ages in two datasets (i.e., Kvasir and CVC-ClinicDB) for training, 10% for testing and 10% for verification. Then, we use all images of the other three datasets (i.e., CVC300, ETIS and CVC-ColonDB) for testing. To sum up, 1450 training images are selected entirely from Kvasir and CVC-ClinicDB, while 798 test images are from all five datasets. Before processing, we uniformly resize the images to 352×352.

Our BSCA-Net is built in PyTorch framework. We use Adam optimizer with learning rate of $1e^{-4}$. Each model is trained by 20 epochs and 16 batchsizes.

Metrics and Loss Function: We use two popular metrics (mean Dice and mean IoU) to evaluate the effect of our model. Our loss function $\mathcal{L}$ is defined as $\mathcal{L} = \mathcal{L}_{IoU} + \mathcal{L}_{BCE}$, where $\mathcal{L}_{BCE}$ and $\mathcal{L}_{IoU}$ are binary cross entropy loss and intersection over union loss. We conduct the deep supervision on four prediction maps, which are obtained by MCAD and BSCA, and we use the sum of four prediction maps as the final loss function. We also report $\mathcal{P} - value$, which is the probability of obtaining values at least as extreme as the observed values of a statistical hypothesis test. $\mathcal{P} - value$ is smaller and the evidence in favor of the alternative hypothesis is stronger.

4.2. Comparison with state-of-the-art methods

In this subsection, we compare our proposed BSCA with several State-of-the-Art methods on five popular datasets for polyp segmentation. We report quantitative results for the learning ability evaluation and generalization ability evaluation in Tables 1 and 2, respectively. Note that, we cite all values of competing methods from their literatures. We also provide all $\mathcal{P} - value$ in Table 3, and the information of Epoch, Learning Rate and Inference in Table 4. In order to feel the convergence more intuitively, we provide the curve of our loss in Fig. 7. In addition, we also show some qualitative results in Fig. 8.

For the learning ability on two training sets, i.e., Kvasir [47] and CVC-ClinicDB [45], as shown in Table 1, most of all competing methods are able to achieve good scores. Here, HarDNet-MSEG, TransUnet, TransFuse-S/L and BSCA achieve the best values in mean Dice and mean IoU, where mean Dice exceeds 91% in both datasets.

To evidence the applicability of all methods, we use three training sets: CVC300, CVC-ColonDB and ETIS. As shown in Table 2, the test result of our network is better than the current networks for three datasets. Compared to the most advanced models, all metrics of BSCA-Net are improved by 4% average. At the same time, the mean Dice for CVC300 dataset reaches 92.7%. This can show that

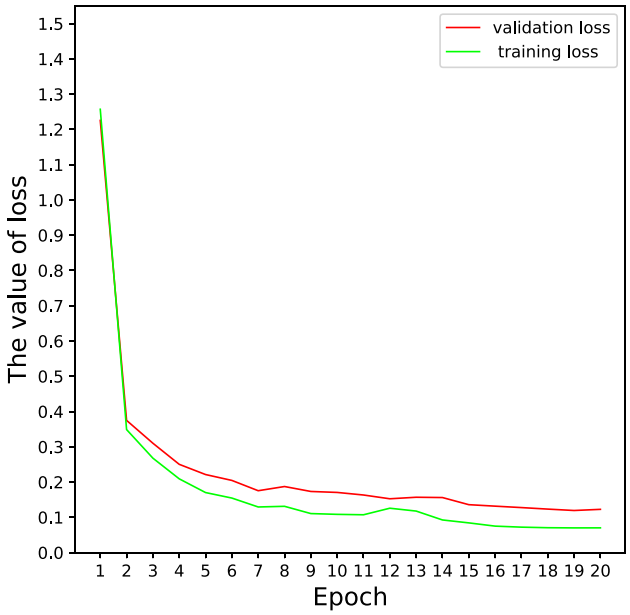

Fig. 7. The structure loss obtained by the proposed BSCA in training with different epoches.

our model has better generalization ability than other methods for polyp segmentation.

At the same time, we provide $\mathcal{P} - value$ in Table 3. When we set the confidence intervals to 95%, all $\mathcal{P} - value$ are less than 0.05. This can show that the proposed approach is an efficient strategy for polyp segmentation.

As shown in Table 4, we present the epoch, learning rate and inference time of BSCA-Net and current SOTA approaches. Note that, BSCA-Net is faster than TransFuse-L and slower than TransFuse-S, but BSCA-Net's score is higher than TransFuse-S and lower than TransFuse-L. Therefore, we believe that the three networks have their own advantages in inference speed and accuracy. Simultaneously, Our network converges when training closes to 20 epochs. We reduce the number of channels from 2048 to 64 when the backbone features are streamed into the MACE, which helps our network to be trained quickly. Comparing to prior CNN-based methods, our BSCA-Net using only 24.07M parameters, less then HarDNet-MSEG (33.3M) and TransFuse-S (26.3M). Moreover, the MACs (Multiply Accumulate Operations) of our method is $1.1e^{10}$, and it is also less then and TransFuse-S ($1.2e^{10}$). This can show the effectiveness of our proposed method.

From Fig. 7, we can see that our network converges only after several training epochs, since our network can backpropagate the loss to earlier layers (red flow in Fig. 1). We also can see that both the training loss (green line) and the validation loss (red line) converge after several epochs, which means our network does not have the problem of overfitting.

It is worth pointing out that, our BSCA-Net is able to achieve the final results within about $74FPS$ on $RTX\ 2080Ti\ GPU$, which guarantees that BSCA-Net can be used for colonoscopy video.

Qualitative Results: To show our model intuitively, we provide results of our model on different testing sets, and we compare testing results with three representative advanced models. As shown in Fig. 8, for the input image in the first line, other three networks produce blurred edges due to the loss of boundary information; for the input image in the second and third lines, the size of polyps is too small and the color of polyps is similar to the surrounding tissues, resulting in the ignorance of the boundary information from the deep features; for the input image in the last line, there are

Table 2
Quantitative results obtained by thirteen medical image segmentation methods for the generalization ability evaluation on CVC-ColonDB, CVC-300 and ETIS datasets.

Dataset	ColonDB		CVC300		ETIS	
	Dice	IoU	Dice	IoU	Dice	IoU
U-Net (MICCAI'15)	0.512	0.444	0.710	0.627	0.398	0.335
U-Net+ (TMI'19)	0.483	0.410	0.707	0.624	0.401	0.344
SFA (MICCAI'19)	0.469	0.347	0.467	0.329	0.297	0.217
ACS-Net (MICCAI'20)	0.716	0.649	0.863	0.787	0.578	0.509
Pra-Net (MICCAI'20)	0.709	0.640	0.871	0.797	0.628	0.567
Threshold-Net (TMI'20)	0.788	0.720	0.897	0.824	0.587	0.640
HarDNet-MSEG (arXiv)	0.731	0.660	0.887	0.821	0.677	0.613
TransUnet (arXiv)	0.781	0.699	0.893	0.824	0.731	0.660
TransFuse-S (MICCAI'21)	0.773	0.696	0.902	0.833	0.733	0.659
TransFuse-L (MICCAI'21)	0.744	0.676	0.904	0.838	0.737	0.661
EU-Net (CRV'21)	0.756	0.681	0.837	0.765	0.687	0.609
SA-Net (MICCAI'21)	0.753	0.670	0.888	0.815	0.750	0.654
BSCA-Net (Ours)	0.783	0.720	0.927	0.875	0.768	0.714

Table 3
All $\mathcal{P} - value$ for the scores in BSCA-Net.

Dataset	Kvasir		ClinicDB		ColonDB		CVC300		ETIS	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
scores	0.910	0.855	0.926	0.874	0.783	0.720	0.927	0.875	0.768	0.714
$\mathcal{P} - value$	0.004	0.015	0.011	0.021	0.014	0.022	0.031	0.011	0.002	0.004

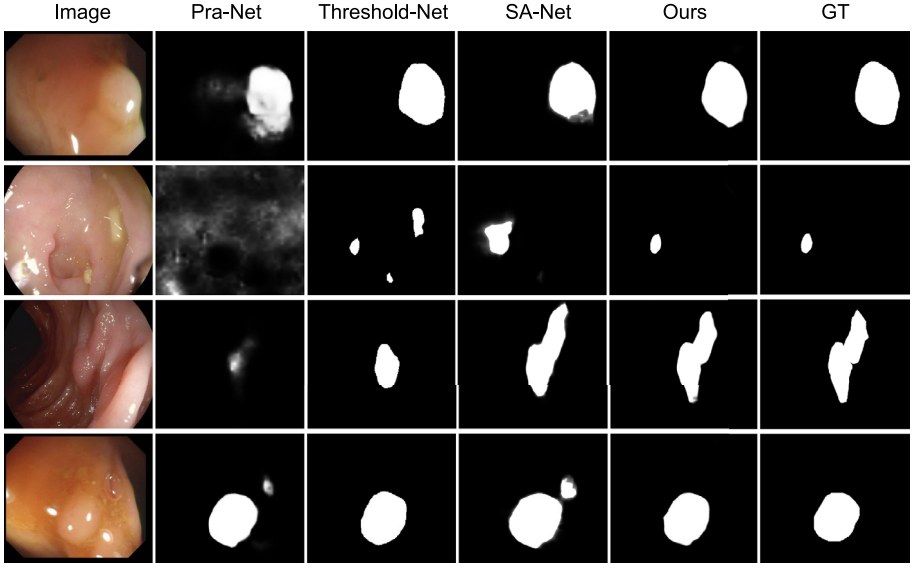

Fig. 8. Qualitative results obtained by Pra-Net, Threshold-Net, SA-Net and our BSCA-Net, on four benchmarks.

Table 4
The comparison of epoch, learning rate and inference in different networks. We provide each network's score on CVC-ClinicDB for comparison of each network.

Method	Epoch	Learning Rate	Inference	Mean Dice
U-Net (MICCAI'15)	30	$3e^{-4}$	8 fps	0.823
U-Net+ (TMI'19)	30	$3e^{-4}$	7 fps	0.794
SFA (MICCAI'19)	500	$1e^{-2}$	40 fps	0.700
ACS-Net (MICCAI'20)	150	$1e^{-3}$	-	0.882
Threshold-Net (TMI'20)	500	$1e^{-3}$	8 fps	0.859
HarDNet-MSEG (arXiv)	100	$1e^{-2}$	-	0.932
TransFuse-S (MICCAI'21)	30	$1e^{-4}$	98 fps	0.918
TransFuse-L (MICCAI'21)	30	$1e^{-4}$	68 fps	0.934
SA-Net (MICCAI'21)	128	$4e^{-2}$	72 fps	0.916
BSCA-Net (Ours)	20	$1e^{-4}$	74 fps	0.926

objects like bubble at the top right of polyps. Due to the lack of similar examples in the training process, other methods regard the object like bubble as the polyp. Other advanced methods have the same problem, and our BSCA-Net aims to address the problem. Our network is able to pay more attention to the boundary information between polyps and the surroundings by exploiting bit-plane slicing technology. We visualize four bit planes for BSCA-Net at the same time.

4.3. Ablation study

In this section, we show the necessity of each module in the network in the form of ablation experiment.

Proof of module effectiveness: To investigate the importance of our modules, we perform experiments for each module on the CVC300 and CVC-ClinicDB datasets. As shown in Table 5, each module has a positive impact on the final structure of the network. Combining each module together allows our network to achieve the state-of-the-art performance.

Table 5
The comparison of BSCA with different modules on CVC300 and CVC-ClinicDB datasets.

Backbone	BSCA	MACE	MCAD	CVC300		ClinicDB	
				Dice	IoU	Dice	IoU
√	-	-	-	0.726	0.631	0.747	0.668
√	√	-	-	0.864	0.788	0.913	0.861
√	√	√	-	0.893	0.828	0.916	0.865
√	√	√	√	0.927	0.875	0.922	0.872

Table 6
The impact of SSBU module on the network. (not SSBU) represents BSCA without the SSBU modules and $\mathcal{G}$ represents the number of split feature-map groups in SSBU.

Dataset	BSCA-Net	Dice	IoU
ClinicDB	(not SSBU)	0.910	0.858
	$\mathcal{G}=1$	0.907	0.849
	$\mathcal{G}=2$	0.919	0.867
	$\mathcal{G}=4$	0.922	0.872
ETIS	(not SSBU)	0.707	0.637
	$\mathcal{G}=1$	0.706	0.633
	$\mathcal{G}=2$	0.742	0.689
	$\mathcal{G}=4$	0.768	0.714

Table 7
Ablation experiments obtained by BSCA-Net with different numbers of bit planes in CVC300. X represents the number of bit planes.

$X=2$		$X=3$		$X=4$		$X=5$	
Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
0.882	0.810	0.900	0.843	0.927	0.875	0.912	0.847

Effectiveness of Bit Slicing Context Attention: We investigate the effect of the BSCA module. The results are in the first and second columns of Table 5. We observe that BSCA improves the performance of backbone: for CVC300 dataset, BSCA improves the mean IoU from 63.1% to 78.8%; for CVC-ClinicDB dataset, BSCA improves the mean IoU from 66.8% to 86.1%. These improvements indicate that the introduction of BSCA is able to help our network accurately distinguish polyps from the input images.

Effectiveness of MACE: We investigate the importance of MACE. As shown in the second and third columns of Table 5. Based on BSCA, MACE improves mean Dice and mean IoU on the CVC300 dataset by 2.9% and 4%, respectively. This result shows that MACE is effective for improving network performance. Note that, at the end of MACE, we reduce the number of channels, which also ensures that our network does not incur excessive time overhead.

Effectiveness of MCAD: We further investigate the role of the cascaded mechanism (MCAD), the results are shown in the third and fourth rows of Table 5. On the CVC300 dataset, mean Dice and mean IoU are improved to 92.7% and 87.5%, respectively. These two data are current state-of-the-art results for polyp segmentation tasks. At the same time, MCAD also brings an improvement on the metrics of the CVC-ClinicDB dataset. In short, the cascaded mechanism is essential for increasing performance.

Effectiveness of SSBU module: To demonstrate the validity of SSBU module, we design a new encoder and decoder whose specific architecture is identical to MACE and MCAD except for SSBU module. We experiment with the new encoder and decoder on CVC-ClinicDB and ETIS datasets. As shown in Table 6, the SSBU module has a stable improvement for each metric, since the SSBU module provides effective access to 3D attention features and rich contextual information through Split and Union operations. The results show that the SSBU module is a good module in encoder and decoder. We test the performance of BSCA-Net with different numbers of split feature-map groups, and report the results in Table 6. At $\mathcal{G}=4$, our network achieves the best results on four metrics. On ETIS datasets, BSCA-Net with $\mathcal{G}=4$ improves at least 6% on ETIS over the version with $\mathcal{G}=1$.

Effectiveness of Bit Planes: We demonstrate the relationship between the number of planes X obtained after bit-plane slicing and network performance. As shown in Table 7, we place the experiment on CVC300. We make five ablation experiments from 2 layers to 6 layers. When the number of layers is 4, BSCA achieves the best performance because the top four bit planes already contain the most of the key information in the image.

To illustrate the effectiveness of BSCA, we also visualize different bit planes in BSCA. As shown in Fig. 9, bit plane slicing technology is able to distinguish the foreground from background through high-level bit planes, and find some fuzzy areas that may be difficult to distinguish from mucosa and colon surface. For most polyp images, we can segment most of the foreground and background through the first two bit planes. For some polyp images with high similarity between background and polyps, plane 3 and 4 are able to notice more detailed pixel changes to obtain the boundary information of polyps. This also shows that selecting the highest four bit planes as the attention is able to improve the effect better compared with other bit planes.

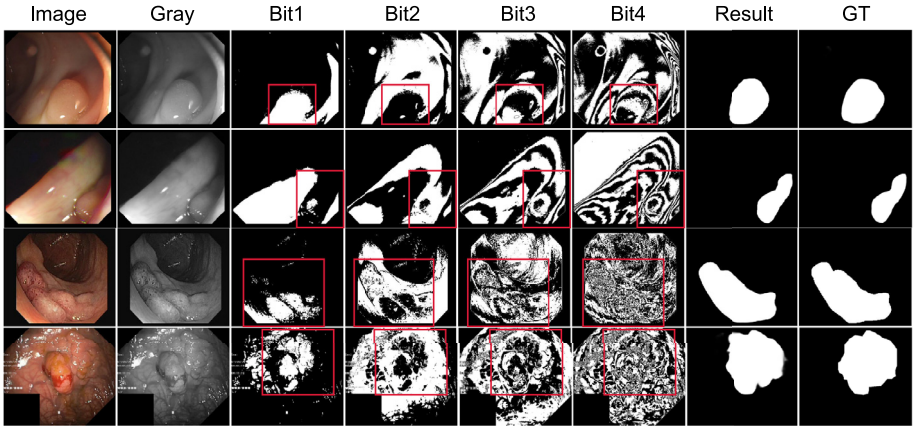

Fig. 9. Qualitative results of comparison with Bit Slicing.

5. Conclusion

In this paper, we propose the Bit-Slicing Context Attention Network (BSCA-Net), to augment the extraction ability of boundary information for accurate polyp segmentation. Specifically, we first design a novel Bit-Slicing Context Attention mechanism, which exploits the bit-plane slicing information to further capture the boundary information from global feature maps through high bit-plane learning, for addressing the difficulty of extracting the boundary between polyps and the surrounding tissues. Then, to enhance a comprehensive understanding of the input features, we propose the SSBU module to capture contextual and geometric information in features. After that, we propose an multipath concat attention decoder (MCAD) and an multipath attention concat encoder (MACE), to further improve the network performance for polyp segmentation. Both qualitative and quantitative results reveal that the proposed BSCA-Net is able to achieve the salient improvement over existing methods on five typical datasets (Kvasir, CVC-ClinicDB, ETIS, CVC-ColonDB and CVC-300). We hope that the development of polyp segmentation can be further promoted through this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants 62072223, and supported by the Natural Science Foundation of Fujian Province under Grant 2020J01131199.

References

[1] Y. Yang, R. Wang, X. Shu, C. Feng, R. Xie, W. Jia, C. Li, Level set framework with transcendental constraint for robust and fast image segmentation, Pattern Recognit 117 (2021) 107985.
[2] A. Oulefki, S. Agaian, T. Trongtirakul, A. Laouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, Pattern Recognit 114 (2021) 107747.
[3] Q. Yu, Y. Gao, Y. Zheng, J. Zhu, Y. Dai, Y. Shi, Crossover-net: leveraging vertical-horizontal crossover relation for robust medical image segmentation, Pattern Recognit 113 (2021) 107756.
[4] R. Zhang, Y. Zheng, C. Poon, D. Shen, J. Lau, Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker, Pattern Recognit 83 (2018) 209–219.
[5] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, Int J Comput Assist Radiol Surg 9 (2) (2013) 283–293.
[6] P. Tripathi, G. Urban, T. Alkayali, M. Mittal, F. Jalali, A. Patel, J. Kim, A. Ninh, G. Albers, K. Chang, J. Samarasena, P. Baldi, W. Karnes, Computer-assisted polyp detection identifies all polyps found by expert colonoscopists, Gastroenterology 154 (6) (2018) S–36.
[7] B. Murugesan, K. Sarveswaran, S. Shankaranarayana, K. Ram, J. Joseph, M. Sivaprakasam, Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2019, pp. 7223–7226.
[8] D. Fan, G. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Pranet: Parallel reverse attention network for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 263–273.
[9] J. Wei, Y. Hu, R. Zhang, Z. Li, S.K. Zhou, S. Cui, Shallow attention network for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 699–708.
[10] S. Chen, X. Tan, B. Wang, X. Hu, Reverse attention for salient object detection, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 234–250.
[11] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.
[12] Z. Zhou, M. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, pp. 3–11.
[13] D. Jha, P. Smedsrud, M. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H. Johansen, ResUNet++: An advanced architecture for medical image segmentation, in: 21st IEEE International Symposium on Multimedia, 2019, pp. 225–2255.
[14] A. Mamonov, I. Figueiredo, P. Figueiredo, Y. Tsai, Automated polyp detection in colon capsule endoscopy, IEEE Trans Med Imaging 33 (7) (2014) 1488–1502.
[15] N. Tajbakhsh, S. Gurudu, J. Liang, Automated polyp detection in colonoscopy videos using shape and context information, IEEE Trans Med Imaging 35 (2) (2015) 630–644.
[16] L. Yu, H. Chen, Q. Dou, J. Qin, P. Heng, Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos, IEEE J Biomed Health Inform 21 (1) (2016) 65–75.
[17] G. Ji, Y. Chou, D. Fan, G. Chen, H. Fu, D. Jha, L. Shao, Progressively normalized self-attention network for video polyp egmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention (2021) 142–152.
[18] P. Brandao, E. Mazomenos, G. Ciuti, R. Caliò, F. Bianchi, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo, D. Stoyanov, Fully convolutional neural networks for polyp segmentation in colonoscopy, in: Medical Imaging 2017: Computer-Aided Diagnosis, volume 10134, 2017, p. 101340F.
[19] M. Akbari, M. Mohrekesh, E. NasrEsfahani, S. Soroushmehr, N. Karimi, S. Samavi, K. Najarian, Polyp segmentation in colonoscopy images using fully convolutional network, in: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2018, pp. 69–72.
[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[21] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[22] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans Pattern Anal Mach Intell 40 (4) (2017) 834–848.
[23] K. Patel, A. Bur, G. Wang, Enhanced u-net: a feature enhancement network for polyp segmentation, Conference on Robots and Vision (2021) 1–8.
[24] X. Guo, C. Yang, Y. Liu, Y. Yuan, Learn to threshold: thresholdnet with confidence-guided manifold mixup for polyp segmentation, IEEE Trans Med Imaging 40 (4) (2021) 1134–1146.
[25] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, Y. Yu, Adaptive context selection for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 253–262.
[26] Y. Fang, C. Chen, Y. Yuan, K. Tong, Selective feature aggregation network with area-boundary constraints for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 302–310.
[27] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 1, Ieee, 2001. I–I
[28] R. Lienhart, J. Maydt, An extended set of haar-like features for rapid object detection, in: Proceedings. International Conference on Image Processing, volume 1, 2002. I–900
[29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A. Berg, SSD: Single shot multibox detector, in: European Conference on Computer Vision, 2016, pp. 21–37.
[30] B. Singh, L. Davis, An analysis of scale invariance in object detection snip, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3578–3587.
[31] T. Tuan, T. Tuan, P. Bao, Brain tumor segmentation using bit-plane and UNET, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 2019, pp. 466–475.
[32] M. Rizzi, C. Guaragnella, Skin lesion segmentation using image bit-plane multilayer approach, Applied Sciences 10 (9) (2020) 3045.
[33] Z.A. Zhen, A. Gx, B. Sw, C. Lw, D. Xz, Pesa-net: permutation-equivariant split attention network for correspondence learning, Information Fusion 77 (2022) 81–89.
[34] D. Sang, T. Chung, P. Lan, D. Hang, D. VanLong, N. Thuy, Ag-curesnest: a novel method for colon polyp segmentation, Image and Video Processing (2021) 1–28.
[35] N. Tomar, D. Jha, S. Ali, H. Johansen, D. Johansen, M. Riegler, P. Halvorsen, Ddanet: Dual decoder attention network for automatic polyp segmentation, in: International Conference on Pattern Recognition, 2021, pp. 307–314.
[36] S. Gao, M.M. Cheng, K. Zhao, X.Y. Zhang, P. Torr, Res2net: a new multi-scale backbone architecture, IEEE Trans Pattern Anal Mach Intell 43 (2) (2021) 652–662.
[37] J. Park, S. Woo, J. Lee, I. Kweon, Bam: Bottleneck attention module, in: British Machine Vision Conference, 2018, pp. 1–14.
[38] S. Liu, D. Huang, Y. Wang, Receptive field block net for accurate and fast object detection, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 385–400.
[39] X. Yang, X. Li, Y. Ye, R. Lau, X. Huang, Road detection and centerline extraction via deep recurrent convolutional neural network u-net, IEEE Trans. Geosci. Remote Sens. 57 (9) (2019) 7209–7220.
[40] J. Su, J. Li, Y. Zhang, C. Xia, Y. Tian, Selectivity or invariance: Boundary-aware salient object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3799–3808.

[41] C.H. Huang, H.Y. Wu, Y.L. Lin, Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps, arXiv, 2021.
[42] J. Chen, Y. Lu, Transunet: Transformers make strong encoders for medical image segmentation, arXiv, 2021.
[43] Y. Zhang, H. Liu, Q. Hu, Transfuse: Fusing transformers and CNNs for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention, 2021, pp. 14–24.
[44] V. David, B. Jorge, S.F. Javier, F.E. Gloria, A.M. López, R. Adriana, D. Michal, C. Aaron, A benchmark for endoluminal scene segmentation of colonoscopy images, J Healthc Eng 2017 (2017) 1–9.
[45] J. Bernal, F. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, Computerized Medical Imaging and Graphics 43 (2015) 99–111.
[46] N. Tajbakhsh, S.R. Gurudu, J. Liang, Automated polyp detection in colonoscopy videos using shape and context information, IEEE Trans Med Imaging 35 (2) (2016) 630–644.
[47] D. Jha, P. Smedsrud, M. Riegler, P. Halvorsen, T. deLange, D. Johansen, H. Johansen, Kvasir-SEG: A segmented polyp dataset, in: International Conference on Multimedia Modeling, 2020, pp. 451–462.

Yi Lin is currently pursuing the bachelor's degree from Minjiang University. His research interests include computer vision, machine learning, and pattern recognition.

Jichun Wu is currently pursuing the bachelor's degree from Minjiang University. His research interests include computer vision, machine learning, and pattern recognition.

Guobao Xiao received the B.S. degree in information and computing science from Fujian Normal University, China, in 2013 and the Ph.D. degree in Computer Science and Technology from Xiamen University, China, in 2016. From 2016–2018, he was a Postdoctoral Fellow in the School of Aerospace Engineering at Xiamen University, China. He is currently a Professor at Minjiang University, China. He has published over 50 papers in the international journals and conferences including IEEE TPAMI/TIP/TITS/TIE, IJCV, PR, ICCV, ECCV, AAAI, etc. His research interests include machine learning, computer vision and pattern recognition. He has been awarded the best PhD thesis in Fujian Province and the best PhD thesis award in China Society of Image and Graphics (a total of ten winners in China). He also served on the program committee (PC) of CVPR, ICCV, ECCV, AAAI, WACV, etc. He was the General Chair for IEEE BDCLOUD 2019.

Junwen Guo received the bachelor's degree in information and computing science from Huaqiao University, China, in 2019. He is currently pursuing the M.S. degree from Fuzhou University, where he is also attached to the Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University. His research interests include computer vision, machine learning, and pattern recognition.

Geng Chen is a Professor at Northwestern Polytechnical University, China. He received his Ph.D. from Northwestern Polytechnical University, China, in 2016. He was a research scientist at the Inception Institute of Artificial Intelligence, UAE, from 2019 to 2021, and a postdoctoral research associate at the University of North Carolina at Chapel Hill, USA, from 2016 to 2019. He has published over 50 papers in peer-reviewed international conference proceedings and journals. His research interests lie in medical image analysis and computer vision.

Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or co-authored more than 150 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, ECCV, etc. His research interests include computer vision, machine learning, and remote sensing. Dr. Ma has been identified in the 2020 and 2019 Highly Cited Researcher lists from the Web of Science Group. He is an Area Editor of Information Fusion, an Associate Editor of Neurocomputing, Sensors and Entropy, and a Guest Editor of Remote Sensing.