Lesion-Aware Dynamic Kernel for Polyp Segmentation

Ruifei Zhang[1], Peiwen Lai[1], Xiang Wan[2,4], De-Jun Fan[3], Feng Gao[3], Xiao-Jian Wu[3], and Guanbin Li[1,2(✉)]

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
liguanbin@mail.sysu.edu.cn
[2] Shenzhen Research Institute of Big Data, Shenzhen, China
[3] The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[4] Pazhou Lab, Guangzhou, China

Abstract. Automatic and accurate polyp segmentation plays an essential role in early colorectal cancer diagnosis. However, it has always been a challenging task due to 1) the diverse shape, size, brightness and other appearance characteristics of polyps, 2) the tiny contrast between concealed polyps and their surrounding regions. To address these problems, we propose a lesion-aware dynamic network (LDNet) for polyp segmentation, which is a traditional u-shape encoder-decoder structure incorporated with a dynamic kernel generation and updating scheme. Specifically, the designed segmentation head is conditioned on the global context features of the input image and iteratively updated by the extracted lesion features according to polyp segmentation predictions. This simple but effective scheme endows our model with powerful segmentation performance and generalization capability. Besides, we utilize the extracted lesion representation to enhance the feature contrast between the polyp and background regions by a tailored lesion-aware cross-attention module (LCA), and design an efficient self-attention module (ESA) to capture long-range context relations, further improving the segmentation accuracy. Extensive experiments on four public polyp benchmarks and our collected large-scale polyp dataset demonstrate the superior performance of our method compared with other state-of-the-art approaches. The source code is available at https://github.com/ReaFly/LDNet.

1 Introduction

Colorectal Cancer (CRC) is one of the most common cancer diseases around the world [17]. However, actually, most CRC starts from a benign polyp and gets progressively worse over several years. Thus, early polyp detection and removal make essential roles to reduce the incidence of CRC. In clinical practice, colonoscopy is a common examination tool for early polyp screening. An accurate and automatic polyp segmentation algorithm based on colonoscopy images can greatly support clinicians and alleviate the reliance on expensive labor, which is of great clinical significance.

L. Wang et al. (Eds.): MICCAI 2022, LNCS 13433, pp. 99–109, 2022.
https://doi.org/10.1007/978-3-031-16437-8_10

However, accurate polyp segmentation still remains challenge due to the diverse but concealed characteristics of polyps. Early traditional approaches [12, 19] utilize hand-craft features to detect polyps, failing to cope with complex scenario and suffering from high misdiagnosis rate. With the advance of deep learning technology, plenty of CNN-based methods are developed and applied for polyp segmentation. Fully convolution network [11] is first proposed for semantic segmentation, and then its variants [1,3] also make a great breakthrough in the polyp segmentation task. UNet [16] adopts an encoder-decoder structure and introduces skip-connections to bridge each stage between encoder and decoder, supplying multi-level information to obtain a high-resolution segmentation map through successive up-sampling operations. UNet++ [27] introduces more dense and nested connections, aiming to alleviate the semantic difference of features maps between encoder and decoder. Recently, to better overcome the above mentioned challenges, some networks specially designed for polyp segmentation task have been proposed. For example, PraNet [6] adopts a reverse attention mechanism to mine finer boundary cues based on the initial segmentation map. ACSNet [23] adaptively selects and integrates both global contexts and local information, achieving more robust polyp segmentation performance. CCBANet [13] proposes the cascading context and the attention balance modules to aggregate better feature representation. SANet [21] designs the color exchange operation to alleviate the color diversity of polyps, and proposes a shallow attention module to select more useful shallow features, obtaining comparable segmentation results. However, existing methods mainly focus on enhancing the network's lesion representation from the view of feature selection [6,13,23] or data augmentation [21], and no attempts have been made to consider the structural design of the network from the perspective of improving the flexibility and adaptability of model feature learning, which limits their generalization.

To this end, we design a Lesion-aware Dynamic Network (LDNet) for the polyp segmentation task. Inspired by [9,24], we believe that a dynamic kernel can adaptively adjust parameters according to the input image, and thus achieving stronger feature exploration capabilities in exchange for better segmentation performance. Specifically, our unique kernel (also known as segmentation head) is dynamically generated basing on the global features of the input image, and generates one polyp segmentation prediction in each decoder stage. Accordingly, these segmentation results serve as clues to extract refined polyp features, which in turn update our kernel parameters with better lesion perception. For some complex polyp regions, the dynamic kernel generation and update mechanism we designed can step-wisely learn and mine discriminative regional features and gradually improve the segmentation results, enhancing the generalization of the model. Besides, we design two attention modules, *i.e.,* Efficient Self-Attention (ESA) and Lesion-aware Cross-Attention (LCA). The former is used to capture global feature relations, while the latter is designed to enhance feature contrast between lesions and other background regions, further improving the segmentation performance. In summary, the contributions of this paper mainly include three folds: (1) We design a lesion-aware dynamic network for polyp segmenta-

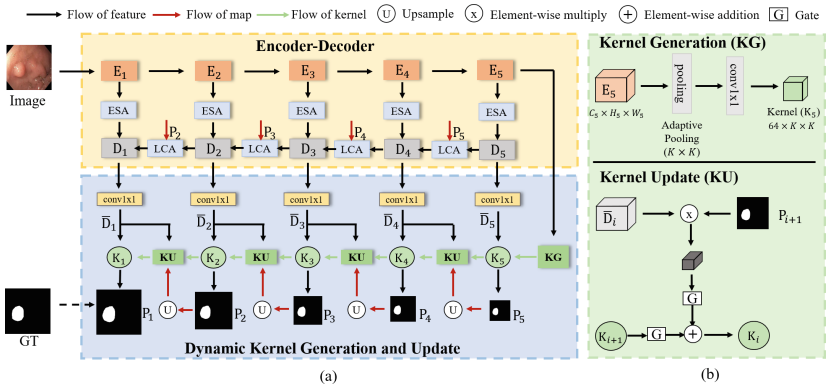

Fig. 1. (a) Overview of our LDNet. (b) Illustration of kernel generation and update.

tion. The introduction of a dynamic kernel generation and update mechanism endows the model with generalizability to discriminate polyp regions with diverse shapes, sizes, and appearances. (2) Our tailored ESA and LCA modules enhance the polyp feature representation, which helps to mine concealed polyps with low visual contrast. (3) Extensive experiments on four public polyp benchmarks and our collected large-scale polyp dataset demonstrate the effectiveness of our proposed method.

2 Methodology

The overview of our LDNet is shown in Fig. 1, which is a general encoder-decoder structure, incorporated with our designed dynamic kernel scheme and attention modules. The Res2Net [8] is utilized as our encoder, consisting of five blocks. The generated feature map of each block is denoted as $\{\mathbf{E}_i\}_{i=1}^{5}$. Accordingly, five-layer decoder blocks are adopted and their respective generated features are defined as $\{\mathbf{D}_i\}_{i=1}^{5}$. 1×1 convolution is utilized to unify the dimension of $\mathbf{D}_i$ to 64, denoted as $\bar{\mathbf{D}}_i$, which are adaptive to subsequent kernel update operations. In contrast to previous methods [6,7,21,23] with a static segmentation head, which is agnostic to the input images and remains fixed during the inference stage, we design a dynamic kernel as our segmentation head. The dynamic kernel is essentially a convolution operator used to produce segmentation result, but its parameters are initially generated by the global feature $\mathbf{E}_5$ of the input, and iteratively updated in the multi-stage decoder process based on the current decoder features $\bar{\mathbf{D}}_i$ and its previous segmentation result $\mathbf{P}_{i+1}$, which is employed to make a new prediction $\mathbf{P}_i$. For the convenience of expression, we denote the sequential updated kernels as $\{\mathbf{K}_i\}_{i=1}^{5}$. Each segmentation prediction is supervised by the corresponding down-sampled Ground Truth, and the prediction $\mathbf{P}_1$ of the last decoder stage is the final result of our model. We detail the dynamic kernel scheme and attention modules in the following sections.

2.1 Lesion-Aware Dynamic Kernel

Kernel Generation. Dynamic kernels can be generated in a variety of ways and have been successfully applied in many fields [9,14,22,24]. In this paper, We adopt a simple but effective method to generate our initial kernel. As shown in Fig. 1, given the global context feature $\mathbf{E}_5$, we first utilize an adaptive average pooling operation to aggregate features into a size of $K \times K$, and then perform one 1×1 convolution to produce the initial segmentation kernel with a reduced dimension of 64. To be consistent with the sequence of decoder, we denote our initial kernel as $\mathbf{K}_5 \in \mathbb{R}^{1\times 64\times K\times K}$. $\mathbf{K}_5$ is acted on the unified decoder features $\bar{\mathbf{D}}_5$ to generate the initial polyp prediction $\mathbf{P}_5$.

Kernel Update. Inspired by [24], we design an iterative update scheme based on the encoder-decoder architecture to improve our dynamic kernel. Given the i-th unified decoder features $\bar{\mathbf{D}}_i \in \mathbb{R}^{64\times H_i\times W_i}$ and previous polyp segmentation result $\mathbf{P}_{i+1} \in \mathbb{R}^{1\times H_{i+1}\times W_{i+1}}$, we first extract lesion features as:

$$\mathbf{F}_i = \sum^{H_i}\sum^{W_i} up_2(\mathbf{P}_{i+1}) \circ \bar{\mathbf{D}}_i, \tag{1}$$

where up_2 denotes up-sampling the prediction map by a factor of 2 to keep a same size with feature map. '$\circ$' represents the element-wise multiplication with broadcasting mechanism.

The essential operation of the kernel update is to integrate the lesion representations extracted by the current decoder features into previous kernel parameters. In this way, the kernel can not only perceive the lesion characteristics to be segmented in advance, but gradually incorporate multi-scale lesion information, thus enhancing its discrimination ability for polyps. Since the previous polyp prediction may be inaccurate, as in [24], we further utilize a gate mechanism to filter the noise in lesion features and achieve an adaptive kernel update. The formulation is:

$$\mathbf{K}_i = \mathbf{G}_i^F \circ \phi_1(\mathbf{F}_i) + \mathbf{G}_i^K \circ \phi_2(\mathbf{K}_{i+1}), \tag{2}$$

where ϕ_1 and ϕ_2 denote linear transformations. $\mathbf{G}_i^F$ and $\mathbf{G}_i^K$ are two gates, which are obtained by the element-wise multiplication between the variants of $\mathbf{F}_i$ and $\mathbf{K}_{i+1}$ followed by different linear transformation and Sigmoid function (σ), respectively:

$$\mathbf{G}_i = \phi_3(\mathbf{F}_i) \circ \phi_4(\mathbf{K}_{i+1}) \tag{3}$$

$$\mathbf{G}_i^K = \sigma(\phi_5(\mathbf{G}_i)), \mathbf{G}_i^F = \sigma(\phi_6(\mathbf{G}_i)) \tag{4}$$

The updated kernel $\mathbf{K}_i$ is acted on the specific decoder feature to make a new prediction $\mathbf{P}_i$. Both of them are sent to the $(i-1)$-th decoder stage to iteratively perform the above update scheme.

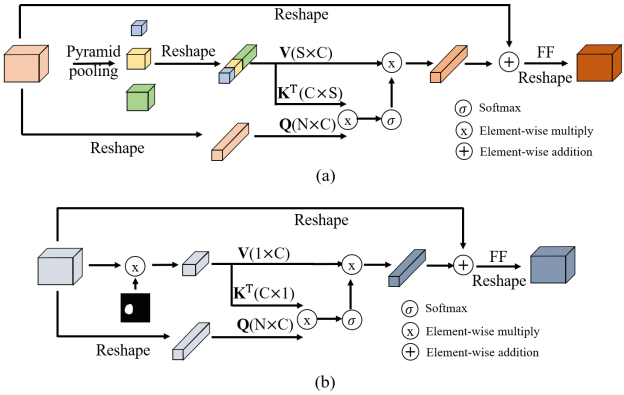

Fig. 2. (a) Illustration of ESA. (b) Illustration of LCA. FF denotes the feed-forward layer. We omit the residual addition between the input and output of FF for simplicity.

2.2 Attention Modules

Efficient Self-attention. Self-attention mechanism is first proposed in Transformer [20], and recently has played a significant role in many tasks [4,5] due to its strong long-range modeling capability, however is criticized for prohibitive computation and memory cost. To overcome these challenges, we borrow the idea from [26,28] and design our ESA module. As shown in Fig. 2, we follow the component of Transformer but replace the original self-attention with our ESA layer, followed by a feed-forward layer and a reshaping operation. We also perform a multi-head parallel scheme to further improve the performance. Specifically, given one encoder feature map $\mathbf{E}_i \in \mathbb{R}^{C_i \times H_i \times W_i}$, details of our ESA layer are formulated as follows:

$$\mathrm{ESA}(\mathbf{E}_i) = \phi_o(\mathrm{concat}(\mathrm{head}^0, ..., \mathrm{head}^n)), \tag{5}$$

$$\mathrm{head}^j = \mathrm{Attention}(\phi_q^j(\mathbf{Q}), \phi_k^j(\mathbf{K}), \phi_v^j(\mathbf{V})), \tag{6}$$

where ϕ_o, ϕ_q^j, ϕ_k^j, ϕ_v^j denote the linear projections, and n is the number of heads. $\mathbf{Q} \in \mathbb{R}^{N_i \times C_i} (N_i = H_i \times W_i)$ is reshaped from the $\mathbf{E}_i$. $\mathbf{K}$, $\mathbf{V} \in \mathbb{R}^{S \times C_i}$ are obtained by the pyramid pooling operation [26], which includes 1×1, 3×3, 5×5 adaptive average pooling to down-sample the feature map, followed by reshaping and concatenating operations. Thanks to such a sampling process, we utilize fewer representative global features to perform the standard attention [20], not only introducing global relations to original features, but significantly saving the computation overhead ($S = 1 \times 1 + 3 \times 3 + 5 \times 5 \ll N_i$). Attention$(\cdot)$ is formulated as:

$$\mathrm{Attention}(\mathbf{q}, \mathbf{k}, \mathbf{v}) = \mathrm{softmax}(\frac{\mathbf{q}\mathbf{k}^{\mathrm{T}}}{\sqrt{d_k}})\mathbf{v}, \tag{7}$$

where d_k is the dimension of each head, equal to $\frac{C_i}{n}$.

Lesion-Aware Cross-Attention. Besides our lesion-aware dynamic kernel, the predicted polyp result is also utilized to enhance the features. Specifically, given the decoder feature $\mathbf{D}_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ and the prediction $\mathbf{P}_i \in \mathbb{R}^{1 \times H_i \times W_i}$, the extracted lesion representations by Eq. 1 (w/o up_2) serve as the $\mathbf{K}$ and $\mathbf{V} \in \mathbb{R}^{1 \times C_i}$ to perform the cross-attention, which is similar to the above mentioned self-attention. Through such an operation, the more similar the region to the lesion, the further enhancement of lesion characteristics, which significantly improves the feature contrast and benefits to detect concealed polyps.

3 Experiments

3.1 Datasets

Public Polyp Benchmarks. We evaluate our proposed LDNet on four public polyp datasets, including Kvasir-SEG [10], CVC-ClinicDB [2], CVC-ColonDB [19] and ETIS [18]. Following the same setting in [6,21], we randomly select 80% images respectively from Kvasir-SEG and CVC-ClinicDB and fuse them together as our training set, 10% as validation set. The remaining data of Kvasir-SEG and CVC-ClinicDB, and other two unseen datasets are used for testing.

Our Collected Large-Scale Polyp Dataset. We also evaluate LDNet on our collected polyp dataset, which has 5175 images in total. This dataset is randomly split into 60% for training, 20% for validation, and the remaining for testing.

3.2 Implementation Details and Evaluation Metrics

Our method is implemented based on PyTorch framework [15] and runs on an NVIDIA GeForce RTX 2080 Ti GPU. We simply set $K = 1$ in the kernel generation and $n = 8$ in the multi-head attention mechanism. The SGD optimizer is utilized to train the model, with batch size of 8, momentum of 0.9 and weight decay of 10^{-5}. The initial learning rate is set to 0.001, and adjusted by a poly learning rate policy, which is $lr = lr_{init} \times (1 - \frac{epoch}{nEpoch})^{power}$, where $power = 0.9$, $nEpoch = 80$. All images are uniformly resized to 256×256. To avoid overfitting, data augmentations including random horizontal and vertical flips, rotation, random cropping are used in the training stage. A combination of Binary Cross-Entropy loss and Dice loss is used to supervise the training process.

As in [7,23], eight common metrics are adopted to evaluate polyp segmentation performance, including *Recall*, *Specificity*, *Precision*, *Dice Score*, *IoU for Polyp (IoUp)*, *IoU for Background (IoUb)*, *Mean IoU (mIoU)* and *Accuracy*.

3.3 Experiments on the Public Polyp Benchmarks

We compare our LDNet with several state-of-the-art methods, including UNet [16], ResUNet [25], UNet++ [27], ACSNet [23], PraNet [6], SANet [21], CCBANet [13], on the public polyp benchmarks. As shown in Table 1, our LDNet

Table 1. Comparison with other state-of-the-art methods on four benchmark datasets. The best three results are highlighted in red, green and blue, respectively.

	Methods	*Rec*	*Spec*	*Prec*	*Dice*	*IoUp*	*IoUb*	*mIoU*	*Acc*
Kvasir	UNet [16]	87.04	97.25	84.28	82.60	73.39	93.89	83.64	95.05
	ResUNet [25]	84.70	97.17	83.00	80.50	70.60	93.19	81.89	94.43
	UNet++ [27]	89.23	97.20	85.57	84.77	76.42	94.23	85.32	95.44
	ACSNet [23]	91.35	98.39	91.46	89.54	83.72	96.42	90.07	97.16
	PraNet [6]	93.90	97.33	89.87	90.32	84.55	95.98	90.26	96.75
	CCBANet [13]	90.71	98.04	91.02	89.04	82.82	96.21	89.52	97.02
	SANet [21]	92.06	98.20	91.14	89.92	83.97	96.54	90.26	97.18
	Ours	92.72	98.05	92.04	90.70	85.30	96.71	91.01	97.35
CVC- ClinicDB	UNet [16]	88.61	98.70	85.10	85.12	77.78	97.70	87.74	97.95
	ResUNet [25]	90.89	99.25	90.22	89.98	82.77	98.18	90.47	98.37
	UNet++ [27]	87.78	99.21	90.02	87.99	80.69	97.92	89.30	98.12
	ACSNet [23]	93.46	99.54	94.63	93.80	88.57	98.95	93.76	99.08
	PraNet [6]	95.22	99.34	92.25	93.49	88.08	98.92	93.50	99.05
	CCBANet [13]	94.89	99.22	91.39	92.83	86.96	98.79	92.87	98.93
	SANet [21]	94.74	99.41	92.88	93.61	88.26	98.94	93.60	99.07
	Ours	94.49	99.51	94.53	94.31	89.48	98.95	94.21	99.08
CVC-ColonDB	UNet [16]	63.05	98.00	68.01	56.40	47.32	94.51	70.92	94.84
	ResUNet [25]	59.91	98.06	65.29	54.87	44.31	93.77	69.04	94.06
	UNet++ [27]	63.49	98.59	77.79	60.77	52.64	95.19	73.92	95.48
	ACSNet [23]	77.38	99.26	81.72	75.51	67.38	96.16	81.77	96.32
	PraNet [6]	81.85	98.54	78.43	76.24	68.29	96.06	82.17	96.26
	CCBANet [13]	82.34	98.39	77.79	75.36	66.57	95.89	81.23	96.14
	SANet [21]	75.21	99.09	81.43	73.50	65.47	96.19	80.83	96.40
	Ours	83.46	98.49	81.15	78.43	70.58	96.21	83.39	96.48
ETIS	UNet [16]	47.33	96.36	48.05	34.81	28.38	94.72	61.55	94.90
	ResUNet [25]	49.12	97.21	56.85	38.65	30.54	95.27	62.90	95.43
	UNet++ [27]	55.52	95.40	59.14	40.91	33.86	93.87	63.87	94.07
	ACSNet [23]	78.31	98.44	68.81	69.44	60.96	97.78	79.37	97.89
	PraNet [6]	81.20	98.73	72.23	72.38	64.07	98.29	81.18	98.38
	CCBANet [13]	78.70	97.19	61.12	62.63	53.81	96.52	75.17	96.66
	SANet [21]	77.08	99.04	72.73	72.26	63.33	98.47	80.90	98.54
	Ours	82.83	98.44	72.07	74.37	66.50	98.01	82.26	98.10

achieves superior performance over other methods across four datasets on most metrics. In particular, on the two seen datasets, *i.e.,* Kvasir and CVC-ClinicDB, the proposed LDNet obtains the best *Dice* and *mIoU* scores, outperforming other methods. On the other two unseen datasets, the LDNet also shows strong generalization ability and achieves 78.43% and 74.37% *Dice* scores, 2.19% and

Table 2. Comparison with other state-of-the-art methods and ablation study on our collected dataset.

Methods	*Rec*	*Spec*	*Prec*	*Dice*	*IoUp*	*IoUb*	*mIoU*	*Acc*
UNet [16]	87.89	97.27	87.23	85.00	77.48	93.95	85.71	95.64
UNet++ [27]	89.88	97.43	88.18	86.92	79.88	94.56	87.26	96.21
ACSNet [23]	92.43	97.79	90.94	90.54	84.64	95.75	90.19	97.11
PraNet [6]	92.86	97.87	90.52	90.64	84.60	95.91	90.25	97.28
CCBANet [13]	91.91	97.79	91.32	90.39	84.36	95.73	90.04	97.10
SANet [21]	92.18	98.22	91.67	90.75	84.98	96.02	90.50	97.27
Ours	93.22	98.15	92.16	91.66	86.28	96.39	91.34	97.55
Baseline	92.02	97.03	87.75	88.30	81.26	94.95	88.11	96.54
Baseline+DK	92.22	97.58	90.41	89.88	83.86	95.53	89.70	96.92
Baseline+DK+ESAs	91.76	98.25	92.14	90.74	84.91	95.85	90.38	97.14

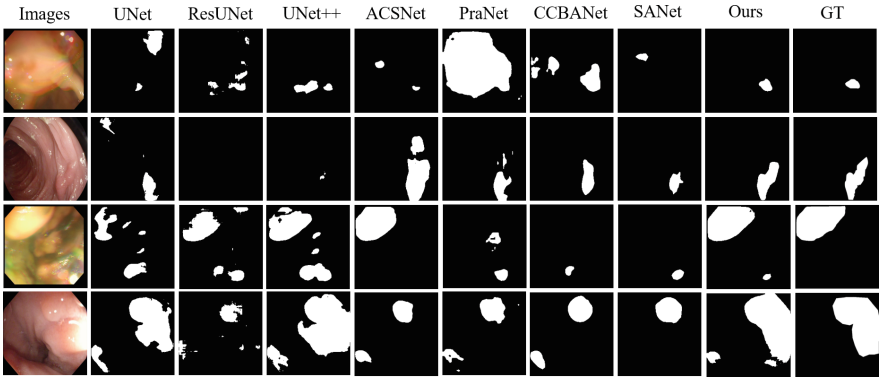

Fig. 3. Visual comparison of polyp segmentation results.

1.99% improvements over the second best approaches, further demonstrating the effectiveness of our approach. Some visualization examples are shown in Fig. 3.

3.4 Experiments on the Collected Large-Scale Polyp Dataset

On our collected large-scale polyp dataset, we compare the LDNet with UNet [16], UNet++ [27], ACSNet [23], PraNet [6], SANet [21] and CCBANet [13]. As shown in Table 2, our method again achieves the best performance, with a *Dice* of 91.66% and a *mIoU* of 91.34%, respectively.

3.5 Ablation Study

We conduct a series of ablation studies on our collected polyp dataset to verify the effectiveness of our designed dynamic kernel scheme and attention modules. Specifically, we utilize the traditional u-shape structure with a static segmentation head as our baseline, and gradually replace the static head with our designed dynamic kernels, then further add ESA and LCA modules, denoting as Baseline, Baseline+DK, Baseline+DK+ESAs and Ours respectively. As shown in Table 2, the introduction of the dynamic kernel significantly enhances the performance of the baseline, with a 1.58% improvement of *Dice* score. With the addition of our ESA and LCA modules, the scores of *Dice* and *mIoU* are further boosted by 0.86% and 0.68%, 0.92% and 0.96%, respectively.

4 Conclusion

In this paper, we propose the lesion-aware dynamic kernel (LDNet) for polyp segmentation, which is generated conditioned on the global information and updated by the multi-level lesion features. We believe that such a dynamic kernel can endow our model with more flexibility to attend diverse polyps regions. Besides, we also improve the feature representation and enhance the context contrast by two tailored attention modules, *i.e.,* ESA and LCA, which is beneficial for detecting concealed polyps. Extensive experiments and ablation studies demonstrate the effectiveness of our proposed method.

Acknowledgements. This work is supported in part by the Chinese Key-Area Research and Development Program of Guangdong Province (2020B0101350001), in part by the Guangdong Basic and Applied Basic Research Foundation (2020B 1515020048), in part by the National Natural Science Foundation of China (61976250), in part by the Guangzhou Science and technology project (20210202 0633), and is also supported by the Guangdong Provincial Key Laboratory of Big Data Computing, The Chinese University of Hong Kong, Shenzhen.

References

1. Akbari, M., et al.: Polyp segmentation in colonoscopy images using fully convolutional network. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 69–72 (2018)
2. 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. **43**, 99–111 (2015)
3. Brandao, P., et al.: Fully convolutional neural networks for polyp segmentation in colonoscopy. In: Medical Imaging 2017: Computer-Aided Diagnosis, vol. 10134, p. 101340F. International Society for Optics and Photonics (2017)
4. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 213–229. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58452-8_13

5. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
6. Fan, D.-P., et al.: PraNet: parallel reverse attention network for polyp segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 263–273. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_26
7. Fang, Y., Chen, C., Yuan, Y., Tong, K.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 302–310. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_34
8. 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. **43**(2), 652–662 (2019)
9. He, J., Deng, Z., Qiao, Y.: Dynamic multi-scale filters for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3562–3572 (2019)
10. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. In: Ro, Y.M., et al. (eds.) MMM 2020. LNCS, vol. 11962, pp. 451–462. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37734-2_37
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
12. Mamonov, A.V., Figueiredo, I.N., Figueiredo, P.N., Tsai, Y.H.R.: Automated polyp detection in colon capsule endoscopy. IEEE Trans. Med. Imaging **33**(7), 1488–1502 (2014)
13. Nguyen, T.-C., Nguyen, T.-P., Diep, G.-H., Tran-Dinh, A.-H., Nguyen, T.V., Tran, M.-T.: CCBANet: cascading context and balancing attention for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 633–643. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_60
14. Pang, Y., Zhang, L., Zhao, X., Lu, H.: Hierarchical dynamic filtering network for RGB-D salient object detection. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12370, pp. 235–252. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58595-2_15
15. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, pp. 8026–8037 (2019)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. CA Cancer J. Clin. **72**(1), 7–33 (2022)
18. 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. **9**(2), 283–293 (2014)
19. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans. Med. Imaging **35**(2), 630–644 (2015)
20. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems 30 (2017)

21. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S.K., Cui, S.: Shallow attention network for polyp segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 699–708. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_66
22. Zhang, J., Xie, Y., Xia, Y., Shen, C.: DoDNet: learning to segment multi-organ and tumors from multiple partially labeled datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1195–1204 (2021)
23. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 253–262. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_25
24. Zhang, W., Pang, J., Chen, K., Loy, C.C.: K-net: towards unified image segmentation. In: Advances in Neural Information Processing Systems 34 (2021)
25. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual U-Net. IEEE Geosci. Remote Sens. Lett. **15**(5), 749–753 (2018)
26. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890 (2017)
27. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
28. Zhu, Z., Xu, M., Bai, S., Huang, T., Bai, X.: Asymmetric non-local neural networks for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 593–602 (2019)