This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ICME55011.2023.00389, 2023 IEEE International Conference on Multimedia and Expo (ICME)

ATENet: Adaptive Tiny-Object Enhanced Network for Polyp Segmentation

1st Xiaogang Du
Shaanxi Joint Laboratory of Artificial Intelligence, The School of Electronic Information and Artificial Intelligence Shaanxi University of Science and Technology
Xi'an, China
duxiaogang@sust.edu.cn

2nd Yinghao Wu
Shaanxi Joint Laboratory of Artificial Intelligence, The School of Electronic Information and Artificial Intelligence Shaanxi University of Science and Technology
Xi'an, China
211612059@sust.edu.cn

3rd Tao Lei
Shaanxi Joint Laboratory of Artificial Intelligence, The School of Electronic Information and Artificial Intelligence Shaanxi University of Science and Technology
Xi'an, China
leitao@sust.edu.cn

4th Dongxin Gu
Shaanxi Joint Laboratory of Artificial Intelligence, The School of Electronic Information and Artificial Intelligence Shaanxi University of Science and Technology
Xi'an, China
201611018@sust.edu.cn

5th Yinyin Nie
Shaanxi Joint Laboratory of Artificial Intelligence, The School of Electronic Information and Artificial Intelligence Shaanxi University of Science and Technology
Xi'an, China
201612049@sust.edu.cn

6th Asoke K. Nandi
Department of Electronic and Computer Engineering Brunel University London
London, United Kingdom
asoke.nandi@brunel.ac.uk

***Abstract*—Polyp segmentation is of great importance for the diagnosis and treatment of colorectal cancer. However, it is difficult to segment polyps accurately due to a large number of tiny polyps and the low contrast between polyps and the surrounding mucosa. To address this issue, we design an Adaptive Tiny-object Enhanced Network (ATENet) for tiny polyp segmentation. The proposed ATENet has two advantages: First, we design an adaptive tiny-object encoder containing three parallel branches, which can effectively extract the shape and position features of tiny polyps and thus improve the segmentation accuracy of tiny polyps. Second, we design a simple enhanced feature decoder, which can not only suppress the background noise of feature maps, but also supplement the detail information to improve further the polyp segmentation accuracy. Extensive experiments on three benchmark datasets demonstrate that the proposed ATENet can achieve the state-of-the-art performance while maintaining low computational complexity.**

***Index Terms*—Deep learning, Convolutional neural network, Multi-scale features, Polyp segmenation, Colonoscopy**

I. Introduction

Intestinal polyps usually cause colorectal cancer, which has a high fatality rate. Thus, it is of great significance to find and resect polyps for preventing colorectal cancer [1]. However, it is difficult for clinical doctors to extract polyps due to the largely varied sizes and blurred boundaries of polyps. Automatic and effective polyp segmentation is still a very challenging task in the field of medical image analysis [2] [3].

Corresponding author: Tao Lei.

With the continuous development of deep learning techniques, more and more segmentation methods have been proposed in recent years. For example, UNet [4] and its variants (UNet++ [5], U^{2}Net [6], nnUnet [7], and DefED-Net [8] etc.) employ symmetric encoder-decoder structure and reduce information loss by skip connections for improving the segmentation accuracy. These methods show high performance in cell and organ segmentation tasks. However, they have poor performance in polyp segmentation tasks. To improve polyp segmentation accuracy, PraNet [9] first uses a global map to detect the polyps location roughly, then designs the reverse attention modules to extract detail features to improve the segmentation effect on the polyp boundary areas. SFA [10] designs a two-branches network to extract features of the polyp boundaries and areas, respectively, and designs a boundary information-sensitive loss function to improve the segmentation accuracy. To improve the segmentation accuracy of small polyps, AcsNet [11] designs the global context module and the local context attention to extract global and local information. Although the above methods have made some progress in the polyp segmentation tasks, the segmentation effect of tiny polyps is still poor.

To address this drawback, we propose an Adaptive Tiny-object Enhanced Network, called ATENet for short. The main contributions of this paper are summarized as follows:

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/ICME55011.2023.00389, 2023 IEEE International Conference on Multimedia and Expo (ICME)

(1) We design an Adaptive Tiny-object Encoder (ATE) module. The ATE module can extract and fuse the fine-grained features, which improves the segmentation accuracy of tiny polyps effectively.

(2) We design an Enhanced Feature Decoder (EFD) module with a simple structure. The EFD module can enhance the feature maps passed by the shortcut connections and integrate the tiny-object prediction maps generated by ATE to reduce the information loss of the network effectively.

(3) Extensive experiments are carried out on three public available datasets and the experimental results demonstrate that ATENet achieves the better segmentation accuracy and the lower computational cost than other popular networks.

II. RELATED WORK

Scholars have proposed a variety of polyp segmentation methods in recent years. These methods can be divided into three categories roughly according to the implementation ideas.

(1) *Improving the backbone network.* There are three popular ways to improve the backbone network: optimizing CNN networks [12], introducing Transformers [13] [14], and combining CNNs with Transformers [15]. To improve the polyp segmentation accuracy on the basis of CNN further, Huang et al. [12] reduced the number of shortcut connections and increased feature map channels to reduce information loss of DenseNet, then designed the HardNet-MSEG using the improved DenseNet. Ding et al. [13] and Wang et al. [14] respectively proposed Polyp-PVT and SSFormer to improve the segmentation accuracy and generalization ability by introducing pyramid vision transformer as backbone. Zhang et al. [15] proposed Transfuse by meriting both Transformer and CNN, which can extract low-level spatial features and high-level semantic features effectively and improve the accuracy of polyp segmentation. Although introducing Transformers can enhance the generalization and feature extraction capability of the network, it also significantly increases the number of parameters and computational complexity.

(2) *Introducing feature enhancement strategies.* By introducing feature enhancement strategies, scholars have proposed many polyp segmentation networks, such as SANet [16], HRENet [17], LDNet [18], and LODNet [19]. SANet [16] designs a shallow attention to strengthen the feature maps, which improves the segmentation accuracy. HRENet [17] introduces the informative context enhancement module to extract fine-grained feature maps and improves the segmentation performance of the hard regions. LDNet [18] designs lesion-aware cross-attention to enhance the contrast of the lesion and the background regions, so as to improve the segmentation accuracy. However, LDNet has a high computational complexity due to the employment of many self-attentions. According to the theory that the oriented derivatives of pixels in boundary regions are larger, LODNet [19] captures and fuses the polyp boundary features with the high-level semantic features by designing the oriented-derivative network, which improves the segmentation accuracy on the boundary regions. However, LODNet can not effectively segment tiny polyps due to the lack of multi-scale features.

(3) *Utilizing multi-scale information.* MSNet [20] uses multi-stage cascaded subtraction to extract complementary information from low-level to high-level feature maps and improves the polyp segmentation accuracy comprehensively. BDGNet [21] introduces the receptive filed block to extract and fuse deep multi-scale feature maps in the encoder to generate a boundary distribution map, which guides the decoding operations and improves the segmentation accuracy of polyp boundaries. TGANet [22] employs convolutions with different dilated rates to extract multi-scale feature maps, and further strengthens the channel connections of feature maps by using the channel attention module, which improves the final segmentation accuracy. However, information loss may arise due to the fusion of features generated by textual attention and the semantic features of images through simple multiplication operations.

III. METHOD

A. Overall structure

We design ATENet based on feature enhancement and multi-scale information for tiny polyps segmentation. The overall structure of ATENet is illustrated in Figure 1. ATENet includes three key modules: backbone, the ATE module and the EFD module. ATENet employs the EfficientNet-B5 [23] as the backbone, which can extract effectively features while maintaining low computational complexity. Moreover, since shallow features cause high computational costs with limited performance gains [24], the ATE and EFD module do not use the shallow features extracted from encoder. In the ATE module, we first use Adaptive Tiny-object Detector(ATD) to extract the polyp boundary features and the position and shape features of tiny polyps. Then, we employ the dense aggregation to fuse feature maps of different scales to generate a Tiny-object Prediction Map(TPM), and pass TPM to the EFD module to further supplement detail information. TPM with rich details can not only improve the segmentation accuracy of tiny polyps, but also effectively improve the segmentation performance of polyp boundaries. In the EFD module, Enhancement module(EM) firstly enhances and fuses the feature maps passed by the skip connection. Then, the Fusion Decoder(FD) fuses the enhanced feature maps with the TPM. Finally, the segmentation result is obtained by the segmentation head.

ATENet has three advantages: First, the proposed ATE module can effectively extract the polyp boundary information and detail information, which improves the segmentation accuracy of polyp boundaries and tiny polyps. Second, the proposed EFD module is simple in structure, yet effective for improving the segmentation performance of ATENet by supplementing rich detail features. Third, ATENet adopts lightweight backbone, ATE, and EFD module, which can maintain low computational complexity while outperforming state-of-the-art models remarkably.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ICME55011.2023.00389, 2023 IEEE International Conference on Multimedia and Expo (ICME)

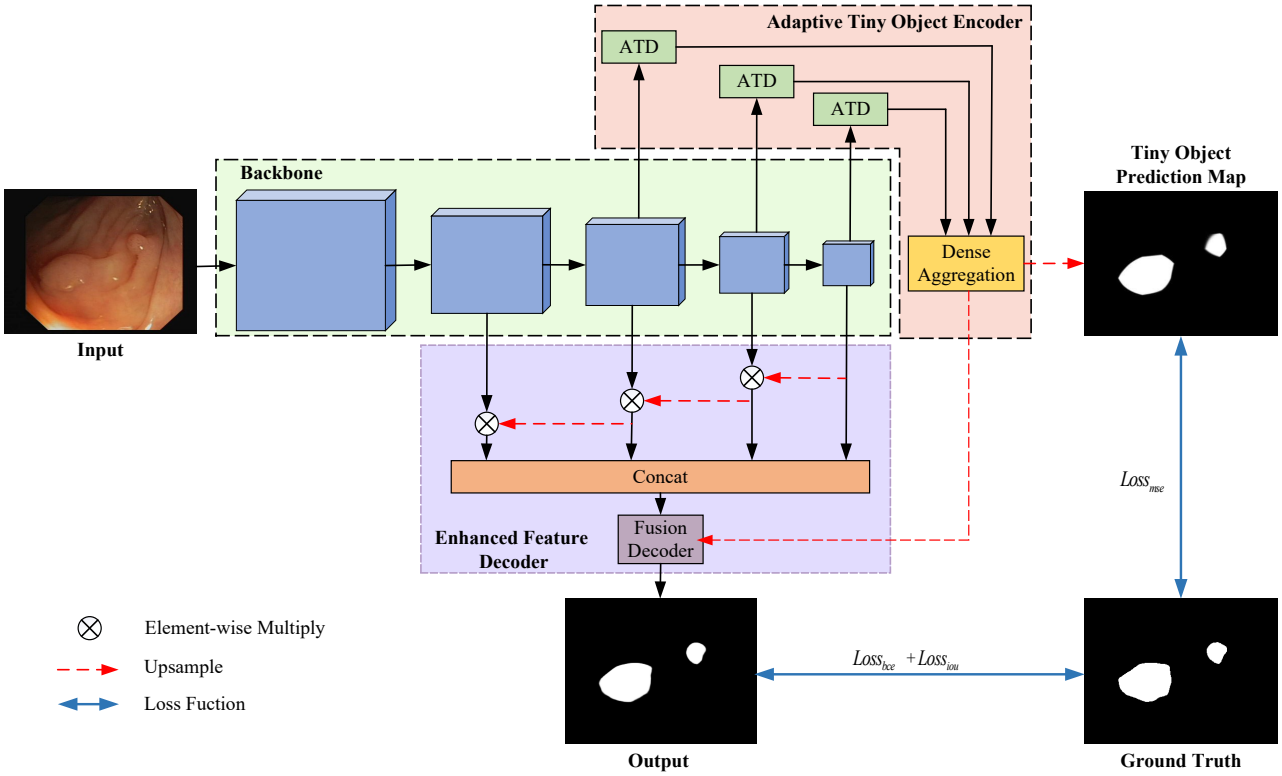

Fig. 1. The overall structure of the proposed ATENet.

B. Adaptive Tiny-object Encoder

We design the ATE module to extract and fuse features of polyp boundaries and tiny polyps. In Figure 1, the ATE module is composed of the ATD and dense aggregation.

1) Adaptive Tiny-object Detector: we design the ATD module to extract effectively features of polyp boundaries and tiny polyps. The structure of the ATD module is illustrated in Figure 2. Firstly, the ATD module reduces the input channels by a 1×1 convolution and passes the feature maps after dimension reduction into four parallel branches. Secondly, three different-sized convolutional kernels are used in three parallel branches, which can not only effectively extract local information such as polyp boundaries and tiny polyps, but also effectively adapt to polyps with different shapes. Then, three dilated convolutions with different dilation rates are used in the last branch to extract global information. Finally, three feature maps containing detail information, a feature map containing global information, and the feature map passed through the residual connection are fused. The ATD module is implementated in (1)-(6):

$$f = Conv_{1\times1}(input) \tag{1}$$

$$f_1 = DConv_{3,3}(DConv_{3,2}(DConv_{3,1}(f))) \tag{2}$$

$$f_2 = DConv_{3,3}(Conv_{3\times1}(Conv_{1\times3}(f))) \tag{3}$$

$$f_3 = DConv_{3,5}(Conv_{5\times1}(Conv_{1\times5}(f))) \tag{4}$$

$$f_4 = DConv_{3,7}(Conv_{7\times1}(Conv_{1\times7}(f))) \tag{5}$$

$$MS_i = Concat(f, f_1, f_2, f_3, f_4) \tag{6}$$

Where $Conv_{n\times n}$ represents the convolution with the size $n\times n$. $DConv_{n,r}$ denotes the dilated convolution with the kernel size $n\times n$ and the dilation rate r. f indicates the feature map after dimension reduction. f_1, f_2, f_3, and f_4 are the feature maps extracted by the four parallel branches, respectively. MS_i represents the output of ATE module, i is the number of encoder layers.

2) Dense aggregation: The ATE module employs the dense aggregation [9] to effectively fuse the multi-scale feature maps containing polyp boundaries and tiny polyps information. The dense aggregation is illustrated in Figure 3. In Figure 3, the original image resolution is $h \times w$, then the input size of the dense aggregation is $\frac{h}{2^{i-1}} \times \frac{w}{2^{i-1}}$, where i is the number of encoder layers, and $i = 3, 4, 5$. The dense aggregation fuses the feature maps with different sizes by multiplication and concatenation to generate TPM. The dense aggregation can be defined as (7):

$$TPM = Agg(MS_3, MS_4, MS_5) \tag{7}$$

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/ICME55011.2023.00389, 2023 IEEE International Conference on Multimedia and Expo (ICME)

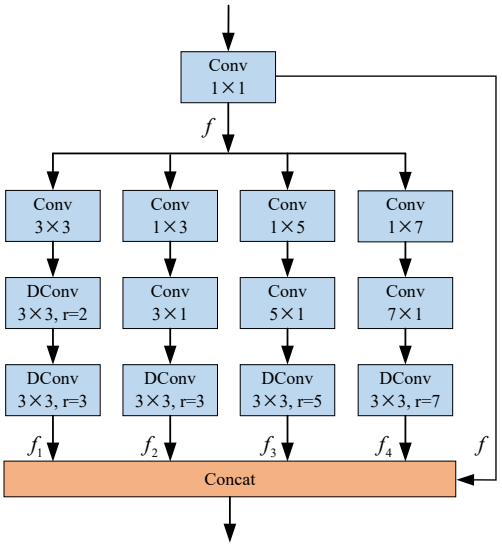

Fig. 2. Adaptive tiny-object detector.

Where MS_i represents the multi-scale information extracted on the i-th layer feature of the encoder.

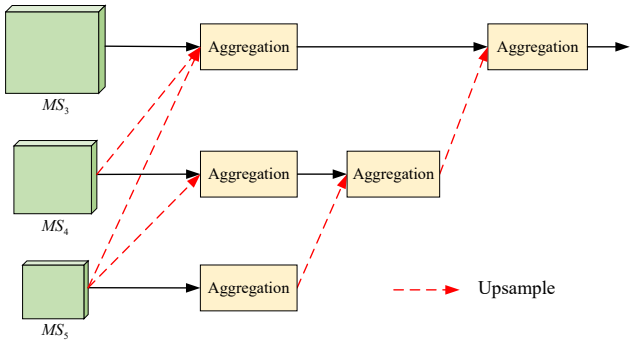

Fig. 3. Dense aggregation.

C. Enhanced Feature Decoder

We design the EFD module to enhance the feature maps and supplement the detail information. In Figure 1, the EFD consists of the EM and FD module.

Since the shallow feature maps usually contain strong background noise which decreases the segmentation accuracy, while the deep feature maps contain weak background noise. Therefore, in EM module, the shallow feature maps are multiplied with the deep feature maps to suppress the background noise. The EM can be denoted as (8)-(9):

$$f_{Ei} = E_i \times UP_2(E_{i+1}) \tag{8}$$

$$f_E = Concat(f_{E2}, UP_2(f_{E3}), UP_4(f_{E4}), UP_8(E_5)) \tag{9}$$

Where f_{Ei} represents the enhanced feature map of the i-th layer, E_i indicates the feature map of the i-th layer, $i = 4, 3, 2$, UP_2, UP_4, and UP_8 denote $2\times$, $4\times$, and $8\times$ upsampling. f_E indicates the enhanced feature map.

The FD is used to fuse the enhanced feature map f_E with the TPM to supplement the detail information. The structure of the FD is shown in Figure 4. Firstly, the TPM is upsampled to the size of f_E. Secondly, f_E and TPM are fused using element-wise multiplication. Then, f_E and the feature map obtained by multiplication are fused to generate the final feature map containing rich detail information. Finally, the final feature map is processed by using 1×1 convolution to obtain the segmentation result. The EFD is implementated as (10):

$$output = UP_4(Conv_{1\times1}(f_E \times out_{ATE} + f_E)) \tag{10}$$

Where out_{ATE} represents the TPM.

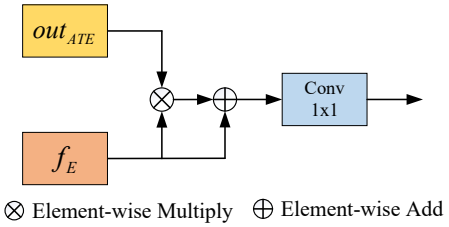

Fig. 4. Fusion decoder.

IV. EXPERIMENTS

A. Datasets

In our experiments, we refer to the training settings in [9] [21]. The training dataset contains 1450 images. 900 and 550 images are randomly selected from Kvasir [25] and CVC-ClinicDB [26], respectively. To evaluate the performance of ATENet for tiny polyp segmentation, we conducted experiments on three public available datasets ETIS [27], CVC-ColonDB [28] and CVC-T [29]. These datasets have many tiny polyps and low contrast with surrounding mucosa, which are of great segmentation challenge. ETIS, CVC-ColonDB, and CVC-T include 198, 380, and 60 images, respectively.

B. Experimental setup

Experiments are performed on a workstation with Intel(R) Xeon(R) Gold 6326 CPU @ 2.90GHz, 50GB RAM, single NVIDIA GeForce RTX 3090 GPU with graphic memory 24GB, Ubuntu 16.04.3, and PyTorch. We set the values of hyper-parameters to train the proposed ATENet. The batch size is set to 20. The initial learning rate is 0.0001. ATENet employs the AdamW optimizer for network training. The maximum epoch is set to 100.

C. Evaluation criteria

Our experiments employs commonly used evaluation criteria to evaluate quantitatively the segmentation effect of the model, including mean Dice($mDice$), mean IoU($mIoU$), mean F-measure ($avgF$), the weighted F-measure (F_β^ω), the recently released S-measure (S_α), and E-measure (E_φ^{max}). The higher the value of the six evaluation criteria, the better the segmentation effect.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/ICME55011.2023.00389, 2023 IEEE International Conference on Multimedia and Expo (ICME)

D. Accuracy

We compare the proposed ATENet with current popular networks, such as PraNet [9], HardNet-MSEG [12], MSNet [20], SANet [16], BDGNet [21], CaraNet [30], LDNet [18], and TGANet [22]. The experimental results on the CVC-ColonDB dataset are shown in Table 1.

TABLE I
QUANTITATIVE EVALUATIONS ON THE CVC-COLONDB DATASET. ↑ INDICATES THAT THE LARGER SCORES ARE BETTER.

Models	*mDice*↑	*mIoU*↑	F_β^ω ↑	S_α ↑	E_φ^{max} ↑	*avgF*↑
PraNet [9]	0.7045	0.6305	0.6840	0.8147	0.8465	0.7072
HardNet-MSEG [12]	0.7395	0.6672	0.7210	0.8314	0.8702	0.7486
MSNet [20]	0.7549	0.6782	0.7360	0.8364	0.8825	0.7360
SANet [16]	0.7712	0.6922	0.7544	0.8437	0.8852	0.7360
BDGNet [21]	0.7827	0.7121	0.7643	0.8601	0.8899	0.7759
CaraNet [30]	0.6988	0.6317	0.6864	0.8059	0.8562	0.7382
LDNet [18]	0.7787	0.7005	0.7582	0.8463	0.9003	0.7744
TGANet [22]	0.7068	0.6334	0.6912	0.8149	0.8303	0.7254
ATENet	**0.8111**	**0.7392**	**0.7942**	**0.8712**	**0.9087**	**0.8041**

In Table 1, the mean Dice and IoU score of ATENet are 0.8111 and 0.7392, respectively. Besides, the F_β^ω, S_α, E_φ^{max} and *avgF* of ATENet on the CVC-ColonDB dataset are 0.7942, 0.8712, 0.9087 and 0.08041, respectively. Therefore, in terms of the segmenation accuracy, ATENet is superior to other popular models on the CVC-ColonDB dataset.

To verify the effectiveness of ATENet, we also conducted comparative experiments on the ETIS and CVC-T datasets, and the experimental results are shown in Tables 2 and 3, respectively. It can be seen from the Tables 2 and 3 that ATENet also outperformed other methods in term of segmentation accuracy on the ETIS and CVC-T datasets.

TABLE II
QUANTITATIVE EVALUATIONS ON THE ETIS DATASET. ↑ INDICATES THAT THE LARGER SCORES ARE BETTER.

Models	*mDice*↑	*mIoU*↑	F_β^ω ↑	S_α ↑	E_φ^{max} ↑	*avgF*↑
PraNet [9]	0.6517	0.5880	0.6213	0.7961	0.8208	0.6301
HardNet-MSEG [12]	0.7098	0.6420	0.6713	0.8226	0.8440	0.6709
MSNet [20]	0.7185	0.6655	0.6767	0.8404	0.8276	0.6532
SANet [16]	0.7595	0.6807	0.7086	0.8448	0.8561	0.6954
BDGNet [21]	0.7614	0.6835	0.7139	0.8539	0.8829	0.7032
CaraNet [30]	0.6352	0.5742	0.6093	0.7537	0.8197	0.6093
LDNet [18]	0.6979	0.6245	0.6615	0.8142	0.8695	0.6702
TGANet [22]	0.6528	0.5779	0.6286	0.6113	0.8455	0.6286
ATENet	**0.7723**	**0.6989**	**0.7316**	**0.8613**	**0.8868**	**0.7256**

TABLE III
QUANTITATIVE EVALUATIONS ON THE CVC-T DATASET. ↑ INDICATES THAT THE LARGER SCORES ARE BETTER.

Models	*mDice*↑	*mIoU*↑	F_β^ω ↑	S_α ↑	E_φ^{max} ↑	*avgF*↑
PraNet [9]	0.8875	0.8189	0.8612	0.9306	0.9488	0.8385
HardNet-MSEG [12]	0.8875	0.8211	0.8617	0.9291	0.9450	0.8445
MSNet [20]	0.8690	0.8078	0.8480	0.9255	0.9416	0.8291
SANet [16]	0.8976	0.8344	0.8715	0.9264	0.9587	0.8485
BDGNet [21]	0.8952	0.8278	0.8707	0.9335	0.9567	0.8643
CaraNet [30]	**0.9113**	0.8457	0.8892	0.9390	**0.9747**	0.8625
LDNet [18]	0.8743	0.8056	0.8471	0.9234	0.9471	0.8387
TGANet [22]	0.8860	0.8289	0.8678	0.9292	0.9558	0.8685
ATENet	0.9105	**0.8486**	**0.8927**	**0.9418**	0.9006	**0.8792**

E. Ablation study

To prove the effectiveness of the proposed ATE and EFD, we conducted ablation study on the ETIS dataset. Firstly, EfficientNet-b5 is chosen as the baseline. Secondly, the ATE and EFD are added to baseline, respectively. Finally, we add the two modules to baseline for performance testing. The experimental results are shown in Table 4. As can be seen from Table 4, both ATE and EFD can effectively improve the segmentation accuracy.

TABLE IV
ABLATION STUDY ON THE ETIS DATASET. ↑ INDICATES THAT THE LARGER SCORES ARE BETTER.

Settings	*mDice*↑	*mIoU*↑	F_β^ω ↑	S_α ↑	E_φ^{max} ↑	*avgF*↑
Baseline	0.6771	0.5888	0.6226	0.8092	0.8331	0.6218
Baseline+ATE	0.7069	0.6166	0.6487	0.8255	0.8357	0.6386
Baseline+EFD	0.6383	0.6619	0.6817	0.8408	0.8459	0.6524
Baseline+ATE+EFD	**0.7723**	**0.6989**	**0.7316**	**0.8613**	**0.8868**	**0.7256**

F. Complexity

To demonstrate the computional complexity of ATENet, we count the number of parameters and floating point operators (GFLOPs) of the above models, as shown in Table 5. In Table 5, the computational cost of ATENet is only 14.489 GFLOPs, which is superior to the other models. Meanwhile, ATENet increases only a small amount of parameters to achieve the trade-off between segmentation accuracy and computational costs.

TABLE V
COMPARISON OF THE PARAMETERS AND COMPUTATIONAL COST OF DIFFERENT MODELS.

Models	GFLOPs	Params(M)
PraNet [9]	21.185	30.498
HardNet-MSEG [12]	18.427	**17.424**
MSNet [20]	27.427	27.692
SANet [16]	18.262	23.899
BDGNet [21]	17,572	32.729
CaraNet [30]	35.074	44.593
LDNet [18]	184.121	40.311
TGANet [22]	128.519	19.937
ATENet	**14.489**	29.816

V. CONCLUSION

In this work, we have mainly investigated the deep neural network for tiny polyps segmentation. We proposed an Adaptive Tiny-object Enhanced Network, which includes two important modules: ATE and EFD. The ATE module extracts and fuses multi-scale features, which can improve the segmentation accuracy of tiny polyps and polyp boundaries. The EFD module enhances the feature maps by suppressing the background noise and supplement rich details. The experimental results demonstrate that the proposed ATENet can accurately segment tiny polyps while maintaining low computational complexity. Therefore, the proposed ATENet is more suitable for clinical practice.

In the future, we will reduce the parameters of ATENet while maintaining the segmentation accuracy. Besides, we will refer to Transformer to redesign the backbone of ATENet to further improve the segmentation accuracy.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/ICME55011.2023.00389, 2023 IEEE International Conference on Multimedia and Expo (ICME)

ACKNOWLEDGMENT

This work is partly supported by National Natural Science Foundation of China (Nos. 61861024, 62271296, and 61871259), Natural Science Basic Research Program of Shaanxi(No. 2021JC-47), and Key Research and Development Program of Shaanxi (Nos.2022GY-436, and 2021ZDLGY08-07).

REFERENCES

[1] L. F. Sanchez-Peralta, L. Bote-Curiel, and A. Picon, et al., "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," Artificial intelligence in medicine, vol. 108, pp. 101923, 2020.

[2] J. G. B. Puyal, K. K. Bhatia, and P. Brandao, et al., "Endoscopic polyp segmentation using a hybrid 2D/3D CNN," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 295-305.

[3] D. Banik, K. Roy, and D. Bhattacharjee, et al., "Polyp-Net: A multi-model fusion network for polyp segmentation," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2020.

[4] O. Ronneberger, P. Fischer, and 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.

[5] Z. Zhou, M. M. R. Siddiquee, and N. Tajbakhsh, et al., "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1856-1867, 2019.

[6] X. Qin, Z. Zhang, and C. Huang, et al., "U2-Net: Going deeper with nested U-structure for salient object detection," Pattern Recognition, vol. 106, pp. 107404, 2020.

[7] F. Isensee, P. F. Jaeger, and S. A. A. Kohl, et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature methods, vol. 18, no. 2, pp. 203-211, 2021.

[8] T. Lei, R. Wang, and Y. Zhang, et al., "DefED-Net: deformable encoding context network for liver and liver-Tumor segmentation," IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 6, no. 1, pp. 68-78, 2022.

[9] D. Fan, G. Ji, and T. Zhou, et al., "Pranet: Parallel reverse attention network for polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 263-273.

[10] Y. Fang, C. Chen, and Y. Yuan, et al., "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.

[11] R. Zhang, G. Li, and Z. Li, et al., "Adaptive context selection for polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 253-262.

[12] C. Huang, H. Wu, and Y. Lin, "Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps," arXiv preprint, arXiv: 2101.07172, 2021.

[13] B. Dong, W. Wang, and D. P. Fan, et al., "Polyp-pvt: Polyp segmentation with pyramid vision transformers," arXiv preprint, arXiv: 2108.06932, 2021.

[14] J. Wang, Q. Huang, and F. Tang, et al., "Stepwise feature fusion: local guides global," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022, pp. 110-120.

[15] Y. Zhang, H. Liu, and Q. Hu, "Transfuse: Fusing transformers and cnns for medical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 14-24.

[16] J. Wei, Y. Hu, and R. Zhang, et al., "Shallow attention network for polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 699-708.

[17] Y. Shen, X.Jia, and M. Q. H. Meng, "Hrenet: A hard region enhancement network for polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 559-568.

[18] R. Zhang, P. Lai, and X. Wan, et al., "Lesion-aware dynamic kernel for polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022, pp. 99-109.

[19] M. Cheng, Z. Kong, and G. Song, et al., "Learnable oriented-derivative network for polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 720-730.

[20] X. Zhao, L. Zhang, and H. Lu, "Automatic polyp segmentation via multi-scale subtraction network," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 120-130.

[21] Z. Qiu, Z. Wang, and M. Zhang, et al., "BDG-Net: boundary distribution guided network for accurate polyp segmentation," in Medical Imaging: Image Processing, 2022, pp. 792-799.

[22] N. K. Tomar, D. Jha, and U. Bagci, et al., "TGANet: Text-guided attention for improved polyp segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022, pp. 151-160.

[23] M. Tan, and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in International Conference on Machine Learning, 2019, pp. 6105-6114.

[24] Z. Wu, L. Su, and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3907-3916.

[25] D. Jha, P. H. Smedsrud, and M. A. Riegler, et al., "Kvasir-seg: A segmented polyp dataset," in International Conference on Multimedia Modeling, 2020, pp. 451-462.

[26] J. Bernal, F. J. Snchez, and G. Fernndez-Esparrach, et al., "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," Comput. Med. Imag. and Grap., vol.43, pp. 99-111, 2015.

[27] J. Silva, A. Histace, and O. Romain, et al., "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer," International Journal of Computer Assisted Radiology and Surgery, vol.9, no.2, pp.283-293, 2014.

[28] J. Bernal, J. Snchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," Pattern Recognition, vol.45, no.9, pp.3166-3182, 2012.

[29] D. Vzquez, J. Bernal, and F. J. Snchez, et al., "A benchmark for endoluminal scene segmentation of colonoscopy images," Journal of Healthcare Engineering, vol. 2017, Article ID: 4037190.

[30] A. Lou, S. Guan, and H. Ko, et al., "CaraNet: context axial reverse attention network for segmentation of small medical objects," in Medical Imaging: Image Processing, 2022, pp. 81-92.