Adaptation of Distinct Semantics for Uncertain Areas in Polyp Segmentation

Quang Vinh Nguyen
vinhbn28@jnu.ac.kr
Van Thong Huynh
vthuynh@jnu.ac.kr
Soo-Hyung Kim
shkim@jnu.ac.kr

Department of AI Convergence
Chonnam National University
Gwangju, South Korea

Abstract

Colonoscopy is a common and practical method for detecting and treating polyps. Segmenting polyps from colonoscopy image is useful for diagnosis and surgery progress. Nevertheless, achieving excellent segmentation performance is still difficult because of polyp characteristics like shape, color, condition, and obvious non-distinction from the surrounding context. This work presents a new novel architecture namely Adaptation of Distinct Semantics for Uncertain Areas in Polyp Segmentation (ADSNet), which modifies misclassified details and recovers weak features having the ability to vanish and not be detected at the final stage. The architecture consists of a complementary trilateral decoder to produce an early global map. A continuous attention module modifies semantics of high-level features to analyze two separate semantics of the early global map. The suggested method is experienced on polyp benchmarks in learning ability and generalization ability, experimental results demonstrate the great correction and recovery ability leading to better segmentation performance compared to the other state of the art in the polyp image segmentation task. Especially, the proposed architecture could be experimented flexibly for other CNN-based encoders, Transformer-based encoders, and decoder backbones.

1 Introduction

Image segmentation is a major and significant topic in computer vision. This task classifies each pixel of the incoming image into predetermined classifications. Medical image segmentation is one of the highlight applications of segmentation techniques such as polyp segmentation, brain tumor segmentation, skin lesion segmentation, or lung segmentation.

Due to its complexity, polyp segmentation has recently attracted a lot of interest. Polyps are abnormal tissue growth from the surface of internal organs and can be found in the colon, rectum, stomach, or even throat. In most cases, polyps are benign, which means that they do not indicate illness or maliciousness. However, since polyps are capable of developing into cancer, a long-term diagnostic is necessary to determine whether or not they have become malignant. Therefore, identifying polyps in the colonoscopy image is helpful in facilitating the early detection of polyp-related diseases. Colonoscopy is the primary method

to locate and remove polyps, but this procedure requires a significant amount of time. In addition, there are still some challenges in a practical setting: common polyps can differ in size, color, and shape. Besides, polyps can develop haphazardly or densely in numerous sites, and may be challenging to distinguish them from surrounding tissues. This requires a reliable endoscopic polyps segmentation approach with high segmentation efficiency in difficult contexts.

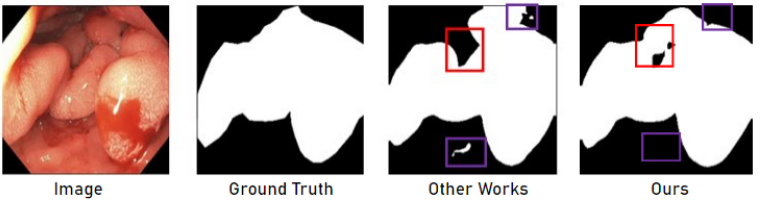

Figure 1: Segmentation example of our model. Red boxes refer to uncertain areas. Purple boxes stand for noise details.

In general, difficult details and hazy areas pose restrictions to medical image segmentation architectures. Previous deep learning-based works focused on the result at the final stage without paying attention to modifying and recovering these areas. Therefore, modern architecture systems struggle to identify and categorize challenging features and unclear areas, as shown in Figure 1. The area indicated in the red box of other works is fairly dark and separates from the nearby polyp items, in this case, the model could not identify sufficient desired objects. Similarly, areas in the purple box are confused with actual polyps, causing misclassification. Under our supervision, these areas might be correctable and recoverable, since they gradually weaken and are diluted with decoder progress. Motivated by this, we propose a new approach, Adaptation of Distinct Semantics for Uncertain Areas in Polyp Segmentation (ADSNet), that uses a complementary trilateral decoder to give a superior early global map. Proposing a new module: continuous attention, we fix inaccurate details and restore missed areas following two semantics: background semantic and object semantic. This strategy improves learning ability, generalization capability and overcomes the weaknesses of current models in challenging contexts.

The following sections are organized as follows: In section II, we present outstanding methods for salient object detection and associated reports to the polyp segmentation task. Section III explores the ADSNet components, while section IV executes experiments to observe the performance of the proposed solution. In the final section, we summarize the entire contributions of the paper.

2 Related Work

Salient Object Detection. SOD is a crucial preprocessing method for many computer vision applications including semantic segmentation, visual tracking, and image retrieval. Traditional SOD solutions are mainly based on heuristic priors (such as color, texture, and contrast) to construct saliency maps. With the advancement of Deep Neural Networks (DNNs), salient object detection (SOD) [3] [23] has made significant strides. First, several methods pay attention to improve the accuracy, Edge Guidance Network (EGNet) [29] combines each path from the top-down stream for the salient object branch and the edge detection branch.

BANet [19] employed side-out fusion and single-stream, respectively, for boundary and object branches. To forecast salient objects with complete structure and exquisite borders, AFNet [10] proposed a multi-scale attentive feedback model and Boundary-Enhanced Loss. Other approaches consider striking a balance between accuracy and efficiency, RAS [5] explicitly multiplies a reverse prediction region to acquire the remaining features for saliency refinement. CPD [25] introduced a cascade partial decoder that uses an attention strategy to enhance high-level features while discarding shallow layer features for acceleration. ITS-DNet [31] proposed an interactive two-stream decoder to investigate several cues, such as saliency, contour, and their interaction.

Polyp Segmentation. With the development of deep learning techniques, the polyp segmentation task is explored strongly. Brandao et al. [4] proposed a fully convolutional neural network to gain great polyp segmentation performance. The first U-shape architecture Unet [17] with two paths: an encoder to extract features and a decoder to educate the final map proved excellent performance in biomedical image segmentation. Motivated by the U-shape structure, several advanced versions were introduced. ResUnet++ [11] utilized the advantages of residual link, channel attention, and Atrous Spatial Pyramidal Pooling. Unet++ [32] kept rich information by dense concatenation via multi-stages. Several other variants use multi U-shape like DoubleUnet [12], and CUNet [21] to achieve better feature representation. Newer methods focus on the relationship between area and boundary, SFANet [9] considered area and border constraints. PraNet [8] applied Reverse Attention Module to correct the boundary area to boost the segmentation performance. Inspired by PraNet [8], UACANet [14] was designed with a new attention: parallel axial attention and augmenting uncertain area to model border information. SSFormer [22] used a pyramid Transformer encoder to improve the generalization ability of models and propose a new decoder to emphasize local features. Another transformer-based architecture, Polyp-PVT [7] proposed a similarity aggregation module to extract local pixels and global semantic cues from the polyp area, effectively suppressing noise in features and significantly improving their expressive capabilities.

3 Methodology

The overall proposed architecture is visualized in Figure 2. The network extracts four levels of spatial features $\{f_i : i = 1,...,4\}$. The complementary trilateral decoder [30] constructs an early global map from encoder features. The continuous attention produces two distinct semantic masks from weak and strong regions before concatenating with the early global map to give a superior final mask. Each component will be explained in further detail in the following subsections.

3.1 Encoder Backbone

In computer vision tasks in general, the encoder component extracts features that bring important semantics for analyzing objects. Recent works often used CNN-based, Transformer-based encoders or combined both in several particular situations. CNN-based approaches do well in extracting local features by using the local kernel, Transformer-based architectures are effective in grasping global relationships, improving generalization ability, and multi-scale feature processing ability, but they need more data or a strong pre-trained. In a differ-

ent approach without transformer, our work uses a CNN-based encoder to extract features and exploit the features obtained by analyzing separate semantics. This also brings many advances in generalization ability and multi-scale feature processing ability. In particular, we use Efficientnet-V2S [20] as an encoder backbone that produces multiple levels of spatial features $f_i \in R^{\frac{H}{2^{i+1}} * \frac{W}{2^{i+1}} * C_i}$ where $C_i \in \{256, 512, 1024, 2048\}$ and $i \in \{1, 2, 3, 4\}$.

3.2 Complementary Trilateral Decoder

Deep features offer important information for locating objects, while shallow features have high resolution and do not store many valuable textures. Therefore, we focus on utilizing high-level features with no regard for low-level features. In this work, we introduce a Complementary Trilateral Decoder [30] - a new SOTA decoder for silent object detection to address the loss of spatial structure, lack of boundary detail, and diluted semantic context. The early global map is obtained by $M = CTD(f_1, f_2, f_3, f_4)$. By estimating and using a decision parameter, we divide the early global map into two weak and strong regions to analyze separated semantics. The first component is strong (S) areas that hold clear object structure, and the second one is weak (W) areas that are determined by the decision parameter and refer to uncertain areas that could be reconstructed.

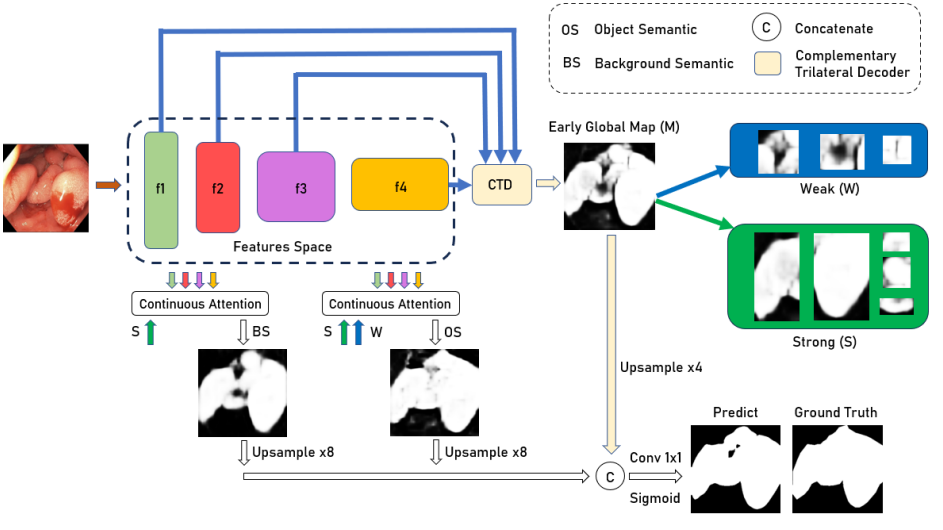

Figure 2: The proposed ADSNet architecture.

3.3 Continuous Attention for Background Semantic and Object Semantic

The multi-scale representation can assist in perceiving multi-scale objects for SOD. Particularly, polyps frequently come in a variety of sizes, hence we suggest a unique structure called Progress Atrous Spatial Pyramidal Pooling (PASPP) [27] to progressively capture multi-scale high-level features $\{f_2, f_3, f_4\}$. Lower-level features $\{f_1\}$ often hold rich detail

information in appearance, we adapt Camouflage Identification Module (CIM) [7] to represent better texture and edge information. The CIM [7] is operated by the following two consecutive attention mechanisms:

$$Att_c = \sigma(H_1(G_{max}(x)) + H_2(G_{avg}(x))) \otimes x \tag{1}$$

$$Att_s = \sigma(Conv3x3(Cat(R_m(x), R_a(x)))) \otimes x \tag{2}$$

Where x is the input feature. G_{max}, G_{avg} are global max pooling and global average pooling functions, respectively. H_1, H_2 are two convolutional layers to reduce the dimension and recover the original dimension. R_m, R_a are max pooling and average pooling following channel dimension. Cat is concatenate operation, while σ stands for Sigmoid function.

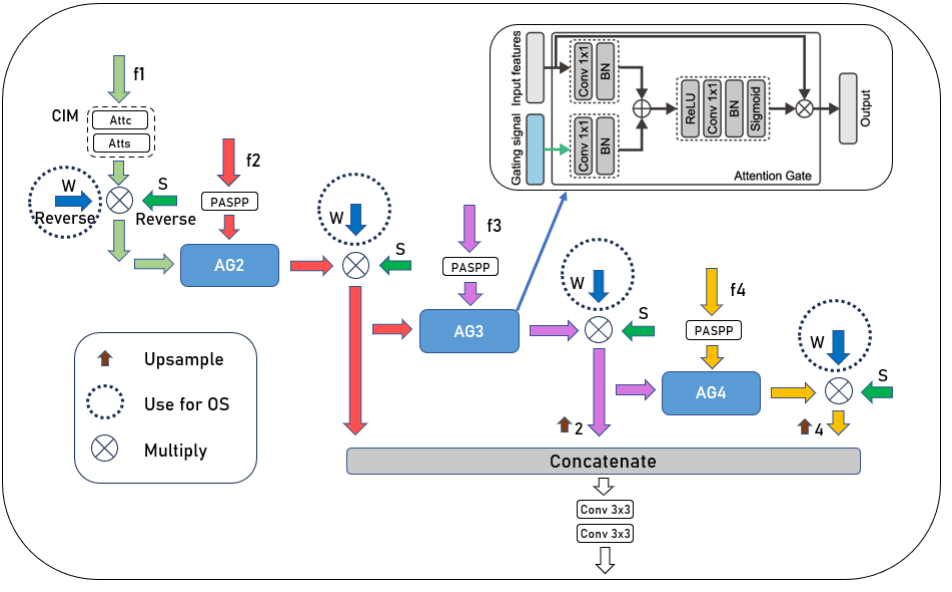

Figure 3: Continuous Attention.

Background Semantic. At the early stage, the beginning mask rough prediction objects without clear structural details. To boost the segmentation performance, the details around the objects need to be detected and classified more precisely. Several previous works used reverse attention[5] to highlight the outside area of objects to process the boundary constraint without correcting inaccurate details through texture information at low-level features. While low-level features often hold rich detail information, such as texture, color, edges, and polyps frequently resemble the background in appearance. Inspired by that motivation, we utilize f_1 to guide high-level features, modify inaccurate details around polyp objects and highlight strong components (S) to give a background semantic mask (BS).

$$BS = Conv3x3(Conv3x3(Cat(S \otimes AG_i), i = 2,3,4)) \tag{3}$$

Object Semantic. Recent architectures can not segment sufficiently recognized objects by creating a final mask only with an encoder and a decoder. Following our findings, sev-

eral areas the model can detect, but they are so weak and easy to vanish through decoder progress because of the highlight of strong details, or detected areas but very noisy with the surrounding space. In addition, polyp colonoscopy images are one of the most difficult objects to segment because of the similarity between objects and backgrounds. Discovering weak features or challenging objects will have the potential to give outstanding performance. Therefore, instead of only using the strong (S) component as in the background semantic, we combine weak (W) and strong (S) components to explore and recover uncertain areas.

$$OS = Conv3x3(Conv3x3(Cat((S,W) \otimes AG_i), i = 2,3,4)) \tag{4}$$

To implement the two ideas proposed above, we introduce Continuous Attention as depicted in Figure 3 including gate attention mechanisms that are executed consecutively to modify the semantics of input high-level features $\{f_2, f_3, f_4\}$ before concatenating them to provide OS and BS as described in Figure 2.

3.4 Loss Function

When analyzing polyp images, the consideration of boundary and area contributes greatly to segmentation performance, so we propose using a loss function described below. Whereas ACE loss [6] offers advantages in terms of geometrical limitations, compact curvature, length, and area of active contours with region similarity, whereas BCE loss [26] represents a pixel restriction. We define the loss function as follows:

$$Loss(y,\bar{y}) = ACE(y,\bar{y}) + BCE(y,\bar{y}) \tag{5}$$

Where $\bar{y}$, y denote the predicted mask and ground truth, respectively.

ACE Loss

$$\begin{aligned} ACE(y,\bar{y}) =& (\alpha + \beta\overline{K}^2)|\nabla\bar{y}| + \lambda y(c_1 - \bar{y})^2 \\ &+ \lambda(1-y)(c_2 - \bar{y})^2 \end{aligned} \tag{6}$$

Where α and β execute the trade-off of length and curvature. λ is used to balance two clauses. $\overline{K}$ is the curvature of y, c_1 and c_2 are mean intensities of the interior and outer regions.

BCE Loss

$$BCE(y,\bar{y}) = \sum_{i=1}^{n} y_i log(\overline{y_i}) + (1-y_i) log(1-\overline{y_i}) \tag{7}$$

4 Experiments

This section describes the dataset, evaluation metrics, experimental results, and further insights into the output. From experimental results, we evaluate and compare ADSNet with current cutting-edge techniques for the polyp segmentation challenge.

Dataset

- Kvasir-Seg [13] There are 1000 polyp images and related annotations. The sizes range from 332x487 to 1920x1072, and the polyps that appear in the images likewise have a range in size and shape.

- CVC-ClinicDB [2] consists 612 images with the size of 384 × 288 from 25 colonoscopy videos.
- ETIS [1] contains 196 images with the size of 1225 × 966 collected from 34 colonoscopy videos. Because polyps are frequently small and challenging to locate, this dataset is more challenging.
- CVC-ColonDB [18] chooses 380 images from 15 different colonoscopy sequences.

Evaluation Metrics We use some standard metrics for the purpose of segmenting medical images. The Dice Similarity Coefficient (DSC) [16] statistic and Intersection over Union (IoU) [15] term are used to quantify the degree of similarity between two samples. MAE [15] shows the average absolute error between the true mask and the anticipated mask. Weighted F-measure (F_β^w) [15] measures the effect of false negative, false positive, and true positive.

4.1 Results

We set up the same training and testing with other methods which are compared in the next section: 1450 (900 images from Kvasir-Seg [13] and 550 images from CVC-ClinicDB [2]) for training. ETIS [1], CVC-ColonDB [18] for generalization ability testing, and remaining images in Kvasir-Seg [13] and CVC-ClinicDB [2] for learning ability evaluation.

Methods	Kvasir-Seg				CVC-ClinicDB			
	Dice	***IoU***	F_β^w	***MAE***	***Dice***	***IoU***	F_β^w	***MAE***
DCRNet	0.886	0.825	0.868	0.035	0.896	0.844	0.890	0.010
PraNet	0.898	0.840	0.885	0.030	0.899	0.849	0.896	0.009
SANet	0.904	0.847	0.892	0.028	0.916	0.859	0.909	0.012
Polyp-PVT	0.917	0.864	0.911	0.023	0.937	0.889	0.936	**0.006**
Ours	**0.920**	**0.871**	**0.916**	**0.020**	**0.938**	**0.890**	**0.940**	**0.006**

Table 1: Quantitative evaluation of diverse models on Kvasir-Seg & CVC-ClinicDB Datasets. The best results are bolded.

Learning ability. In this experiment, the domain of the test and train set is similar. We compare ADSNet with recent SOTA methods including: PraNet [8], DCRNet [28], SANet [24], and Polyp-PVT [7]. For a fair comparison, all results of the models mentioned above are referenced from the original papers. The detailed results shown in Table 1 indicate that ADSNet improves recent approaches. 0.920 Dice score and 0.871 IoU score on Kvasir-Seg [13] and 0.938, 0.890 on CVC-ClinicDB [2], respectively. Besides, our model also is better on F_β^w and MAE indexes. For qualitative results, Figure 4 shows obtained segmentation performance by all models on Kvasir-Seg [13] and CVC-ClinicDB [2]. The segmentation performance of competitors is referenced completely from public sources in Polyp-PVT [7]. The great agreement between predicted samples by ADSNet and ground truths is illustrated. This proves the efficient and accurate segmentation ability of the proposed architecture in difficult and challenging situations with uncertain areas that previous approaches have not dealt with yet.

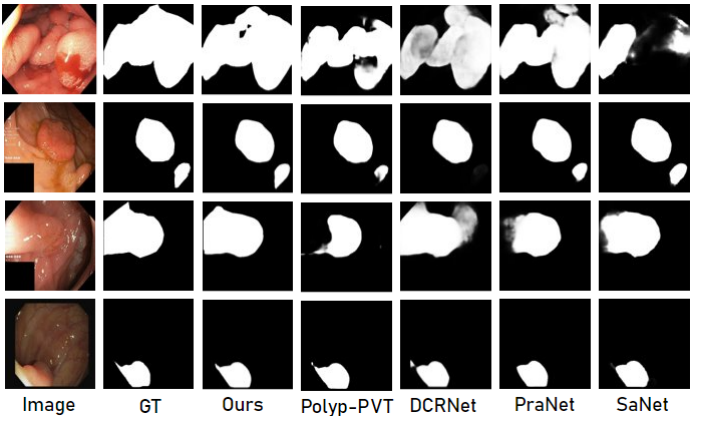

Figure 4: Qualitative analysis on Kvasir-Seg and CVC-ClinicDB dataset of different models in several noteworthy cases.

Generalization ability. The compared results are shown in Table 2. On ETIS [1] dataset, the performance obtain 0.798 Dice score and 0.715 IoU score improved by 1.39% and 1.27% over Polyp-PVT [7] respectively. ADSNet also shows great generalizability in CVC-ColonDB [18] when surpassing all other SOTA methods with the highest scores, 0.815 Dice and 0.730 IoU score. Although on CVC-ColonDB [18] the F_β^w reduces when compared with SANet [24], our proposal still improves on the remaining. Several noteworthy and difficult samples are shown in Figure 5. It is clear that ADSNet is capable of partitioning complex and minute details in obscure situations in addition to segmenting well in challenging cases with uncertain areas.

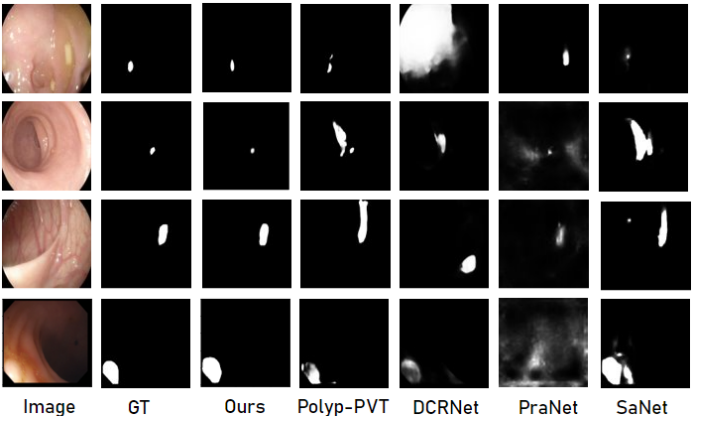

Figure 5: Qualitative analysis on ETIS and CVC-ColonDB dataset of different models in several noteworthy cases.

Train Set	Kvasir-Seg & CVC-ClinicDB							
Test Set	CVC-ColonDB				ETIS			
Methods	*Dice*	*IoU*	F_β^w	*MAE*	*Dice*	*IoU*	F_β^w	*MAE*
DCRNet	0.704	0.631	0.684	0.052	0.556	0.496	0.506	0.096
PraNet	0.712	0.640	0.699	0.043	0.628	0.567	0.600	0.031
SANet	0.753	0.670	**0.892**	0.043	0.750	0.654	0.685	0.015
Polyp-PVT	0.808	0.727	0.795	0.031	0.787	0.706	0.750	0.013
Ours	**0.815**	**0.730**	0.860	**0.029**	**0.798**	**0.715**	**0.792**	**0.012**

Table 2: Quantitative evaluation of diverse models on unseen Datasets ETIS & CVC-ColonDB. The best results are bolded.

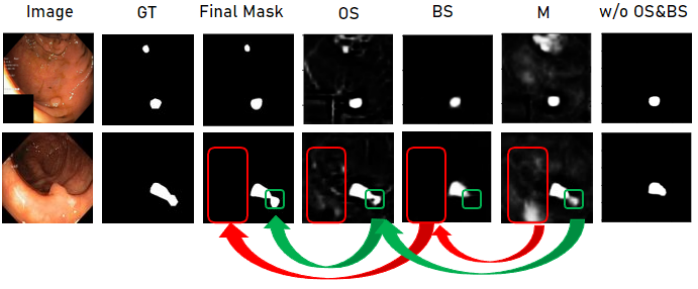

Figure 6: The contribution of OS, BS, and M.

4.2 Further insights

This part proves the effectiveness of the combination of OS, BS, and M in the final mask. When the background semantic (BS) modifies wrong details around and highlights the primary structure of objects. The object semantic (OS) explores ambiguous and challenging areas. We display several segmentation performances in Figure 6. It can be seen that OS and BS assist the model capture context information in both object and background semantics. Meanwhile, without OS or BS, the architecture has difficulty in detecting and categorizing uncertain details, causing misclassification as in the last column. We also visualize the output feature map of the early, background semantic, and object semantic in Figure 7 to verify the effectiveness of analyzing two distinct semantics, ADSNet offers significantly more sufficient features.

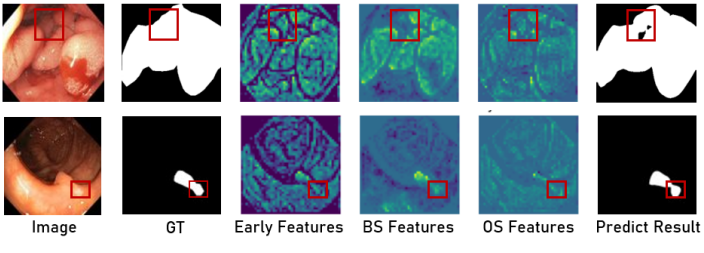

Figure 7: Feature map of the early, BS and OS.

5 Conclusion

In this paper, we have proposed a new novel architecture called ADSNet for exploring uncertain areas to improve polyp segmentation performance. We have introduced a new SOTA decoder that utilizes high-level features from the CNN-based encoder to produce an early global map. Finally, we have analyzed two separate semantics of the early global map: background semantic and object semantic using continuous attention to give an excellent map. The obtained results demonstrate the superiority of the proposed model in dealing with challenging cases with uncertain areas, from that, obtain better Dice, IoU, f_β^w, and MAE scores over recent state-of-the-art methods.

Acknowledgements This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) grant funded by the Korea government(MSIT). This work was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF2021R1I1A3A04036408). This work was supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2021-0-02068, Artificial Intelligence Innovation Hub). The corresponding author is Soo-Hyung Kim.

References

[1] Jorge Bernal, Javier Sánchez, and Fernando Vilarino. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, 45(9):3166–3182, 2012.

[2] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015.

[3] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. Salient object detection: A survey. *Computational visual media*, 5:117–150, 2019.

[4] Patrick Brandao, Evangelos Mazomenos, Gastone Ciuti, Renato Caliò, Federico Bianchi, Arianna Menciassi, Paolo Dario, Anastasios Koulaouzidis, Alberto Arezzo, and Danail Stoyanov. Fully convolutional neural networks for polyp segmentation in colonoscopy. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, pages 101–107. SPIE, 2017.

[5] Shuhan Chen, Xiuli Tan, Ben Wang, and Xuelong Hu. Reverse attention for salient object detection. In *Proceedings of the European conference on computer vision (ECCV)*, pages 234–250, 2018.

[6] Xu Chen, Xiangde Luo, Yitian Zhao, Shaoting Zhang, Guotai Wang, and Yalin Zheng. Learning euler's elastica model for medical image segmentation. *arXiv preprint arXiv:2011.00526*, 2020.

[7] Bo Dong, Wenhai Wang, Deng-Ping Fan, Jinpeng Li, Huazhu Fu, and Ling Shao. Polyp-pvt: Polyp segmentation with pyramid vision transformers. *arXiv preprint arXiv:2108.06932*, 2021.

[8] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23*, pages 263–273. Springer, 2020.

[9] Yuqi Fang, Cheng Chen, Yixuan Yuan, and Kai-yu Tong. Selective feature aggregation network with area-boundary constraints for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 302–310. Springer, 2019.

[10] Mengyang Feng, Huchuan Lu, and Errui Ding. Attentive feedback network for boundary-aware salient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1623–1632, 2019.

[11] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255. IEEE, 2019.

[12] Debesh Jha, Michael A Riegler, Dag Johansen, Pål Halvorsen, and Håvard D Johansen. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*, pages 558–564. IEEE, 2020.

[13] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*, pages 451–462. Springer, 2020.

[14] Taehun Kim, Hyemin Lee, and Daijin Kim. Uacanet: Uncertainty augmented context attention for polyp segmentation. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 2167–2175, 2021.

[15] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 248–255, 2014.

[16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[18] Juan Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International journal of computer assisted radiology and surgery*, 9:283–293, 2014.

[19] Jinming Su, Jia Li, Yu Zhang, Changqun Xia, and Yonghong Tian. Selectivity or invariance: Boundary-aware salient object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3799–3808, 2019.

[20] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[21] Zhiqiang Tang, Xi Peng, Shijie Geng, Yizhe Zhu, and Dimitris N Metaxas. Cu-net: Coupled u-nets. *arXiv preprint arXiv:1808.06521*, 2018.

[22] Jinfeng Wang, Qiming Huang, Feilong Tang, Jia Meng, Jionglong Su, and Sifan Song. Stepwise feature fusion: Local guides global. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part III*, pages 110–120. Springer, 2022.

[23] Wenguan Wang, Qiuxia Lai, Huazhu Fu, Jianbing Shen, Haibin Ling, and Ruigang Yang. Salient object detection in the deep learning era: An in-depth survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3239–3259, 2021.

[24] Jun Wei, Yiwen Hu, Ruimao Zhang, Zhen Li, S Kevin Zhou, and Shuguang Cui. Shallow attention network for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 699–708. Springer, 2021.

[25] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3907–3916, 2019.

[26] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[27] Qingsen Yan, Bo Wang, Dong Gong, Chuan Luo, Wei Zhao, Jianhu Shen, Qinfeng Shi, Shuo Jin, Liang Zhang, and Zheng You. Covid-19 chest ct image segmentation–a deep convolutional neural network solution. *arXiv preprint arXiv:2004.10987*, 2020.

[28] Zijin Yin, Kongming Liang, Zhanyu Ma, and Jun Guo. Duplex contextual relation network for polyp segmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022.

[29] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8779–8788, 2019.

[30] Zhirui Zhao, Changqun Xia, Chenxi Xie, and Jia Li. Complementary trilateral decoder for fast and accurate salient object detection. In *Proceedings of the 29th acm international conference on multimedia*, pages 4967–4975, 2021.

[31] Huajun Zhou, Xiaohua Xie, Jian-Huang Lai, Zixuan Chen, and Lingxiao Yang. Interactive two-stream decoder for accurate and fast saliency detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9141–9150, 2020.

[32] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming 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: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.