

Multiscale Hybrid Attention Network for Automated Segmentation of Colorectal Polyps

Ersan YAZAN^{1*} 

Adiyaman University, Faculty of Engineering, Department of Computer Engineering, Adiyaman, Türkiye

**Corresponding author: eyazan@adiyaman.edu.tr*

Abstract

Accurate and reliable segmentation of colorectal polyps plays a critical role in the early diagnosis of colorectal cancer. Although deep learning-based methods have shown significant progress in recent years, the variations in polyp size, shape, color, and texture can adversely affect segmentation performance. In this study, a hybrid network model is proposed for automatic polyp segmentation. The proposed approach utilizes multi-level single-layer feature maps extracted from a DAF3D-based structure, which has demonstrated strong performance in deep feature extraction. These feature maps are first processed with a channel attention module and then with a reverse attention module to generate a multi-layer feature representation. While the channel attention module emphasizes the most informative channels, the reverse attention module enhances boundary regions, thereby improving segmentation accuracy. To evaluate the effect of these modules, scenarios using only channel attention, only reverse attention, and both modules together were compared, revealing that the integrated use of both modules achieved the highest performance. Additionally, the proposed model was compared with several polyp segmentation networks from the literature under the same backbone architecture and identical training conditions. Experiments conducted on multiple datasets demonstrate that the hybrid architecture provides significantly superior performance in mean Dice and mean IoU metrics compared to other methods. The findings indicate that the proposed approach is an effective, stable, and highly accurate method for polyp segmentation.

Keywords

Polyp
segmentation,
Channel attention,
Reverse attention

Kolorektal Poliplerin Otomatik Segmentasyonu için Çok Ölçekli Hibrit Dikkat Ağı

Ersan YAZAN^{1*} 

¹ Adıyaman Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Adıyaman, Türkiye

*Sorumlu yazar: eyazan@adiyaman.edu.tr

Özet

Kolorektal poliplerin doğru ve güvenilir bir şekilde bölütlenmesi, kolorektal kanserin erken tanısında kritik bir öneme sahiptir. Son yıllarda derin öğrenme tabanlı yöntemlerle önemli ilerlemeler kaydedilmiş olsa da poliplerin boyut, şekil, renk ve doku açısından gösterdiği çeşitlilik, segmentasyon başarımını olumsuz etkileyebilmektedir. Bu çalışmada, otomatik polip segmentasyonu için hibrit bir ağ modeli önerilmektedir. Önerilen yaklaşımda, derin özellik çıkarımında başarılı olduğu bilinen DAF3D tabanlı yapıda elde edilen çok seviyeli tek katmanlı özellik haritaları kullanılmaktadır. Bu özellik haritaları önce kanal dikkat modülü, ardından ters dikkat modülü ile işlenerek çok katmanlı bir özellik temsiline dönüştürülmektedir. Kanal dikkat modülü en bilgilendirici kanalları öne çıkarırken, ters dikkat modülü sınır bölgelerini belirginleştirerek segmentasyon doğruluğunu artırmaktadır. Modüllerin etkisini değerlendirmek amacıyla yalnızca kanal dikkat, yalnızca ters dikkat ve her iki modülün birlikte kullanıldığı senaryolar karşılaştırılmış; her iki modülün birlikte kullanıldığı durumda en yüksek başarımın elde edildiği görülmüştür. Ayrıca önerilen model, literatürde yer alan çeşitli polip segmentasyon ağlarıyla aynı omurga mimarisi ve eşit eğitim koşulları altında karşılaştırılmıştır. Birden fazla veri seti üzerinde gerçekleştirilen deneyler, hibrit yapının mean Dice ve mean IoU metriklerinde diğer yöntemlere kıyasla anlamlı derecede üstün performans gösterdiğini ortaya koymaktadır. Elde edilen bulgular, önerilen yaklaşımın polip segmentasyonu için etkili, kararlı ve yüksek doğruluk sağlayan bir yöntem olduğunu göstermektedir.

Anahtar kelimeler

Polip bölütlemesi,
Kanal dikkat,
Ters dikkat

1. INTRODUCTION

Colorectal cancer (CRC) ranks as the third most commonly diagnosed cancer worldwide [1]. As with other cancer types, early diagnosis is critically important in CRC. Colonoscopy is the primary clinical method used to identify colorectal polyps. Research has shown that early colonoscopic examinations significantly reduce CRC incidence [2]. However, polyp detection presents technical challenges, as polyps of the same type can vary in size, color, and texture, and many polyps can be difficult to distinguish from the intestinal mucosa. Moreover, colonoscopy is highly dependent on operator experience, and studies report that clinicians may miss approximately 25–28% of polyps during the procedure [3, 4]. Therefore, accurate and reliable segmentation of polyps is vital for the prevention of CRC.

A variety of studies have been conducted on the classification and segmentation of polyps in colonoscopy images. Approaches based on handcrafted features such as color, texture, and shape are predominantly classification-oriented [5–8], yet their limited ability to represent different types of polyps has prevented them from achieving high performance [9]. In recent years, deep learning methods have become widely used for polyp segmentation. Brando et al. [10] proposed an FCN-based model in which three well-known networks—AlexNet, GoogLeNet, and VGG—were converted into fully convolutional architectures for polyp segmentation. Akbari et al. [11] reported improved segmentation accuracy using an enhanced FCN structure incorporating several additional processing steps. Another FCN-based polyp segmentation study was conducted by Zhang et al. [12]. Jha et al. [13] introduced the ResUNet++ model, which integrates characteristics from multiple architectures, and demonstrated that it performs better than both ResNet and U-Net. With the success of edge-guided networks used in salient object detection (SOD) [14, 15], similar strategies have been applied to polyp segmentation as well. Fang et al. [16] proposed the Selective Feature Aggregation Network, which leverages both region and boundary information for polyp segmentation. Another approach utilizing region and boundary cues is Psi-Net [17]; however, it has been noted that the relationship between region and boundary features is not fully captured in this model. Moreover, edge-guided networks present notable drawbacks, including the high cost of generating additional edge-specific datasets and the possibility of irrelevant edges being included in the data [18].

One of the widely adopted approaches in recent polyp segmentation research is the use of attention mechanisms. Due to their structural characteristics, CNNs tend to focus primarily on local information and often fail to adequately capture large-scale contextual features [19]. The main objective of attention mechanisms is to jointly evaluate local and global context, enhancing the relevant regions while suppressing the background. This enables each pixel in the image to be interpreted in a more comprehensive and detailed manner. The effectiveness of attention-enhanced architectures has also been

demonstrated in other medical image segmentation tasks, where models such as Attention-UNet and Attention-ResUNet have been investigated for the extraction and segmentation of clinically relevant structures [20]. In this context, various attention modules with different operational principles have been developed for polyp segmentation.

Tomar et al. [21] first introduced DDANet, a model based on a dual-decoder attention architecture, and reported that the network achieved a Dice score of 0.7874 on the test set after being trained for 200 epochs on the Kvasir-SEG dataset. They later proposed FANet [22], which incorporates a feedback mechanism. In this model, the output generated at the end of each epoch is used as an input for the next epoch through a dedicated attention module. After 100 epochs of training on the Kvasir-SEG dataset, FANet achieved a Dice score of 0.8803 on the test data. Additionally, Tomar et al. introduced TGANet [23], which integrates a text-guided attention module designed to detect polyps of varying sizes and quantities. Following 500 epochs of training on the Kvasir-SEG dataset, TGANet achieved a test Dice score of 0.8982.

Xu et al. proposed the DCSAU-Net model, which incorporates a Compact Split Attention (CSA) module designed to strengthen feature relationships between convolutional layers with different kernel sizes [24]. Trained for 200 epochs on the CVC-ClinicDB [25] dataset, the model achieved a Dice score of 0.9084 during testing. Huang et al. developed HarDNet-MSEG, an encoder-decoder architecture [26]. In this approach, HarDNet68 [27], known for its reduced memory access cost, is used as the backbone and forms the encoder component of the network. On the decoder side, a parallel partial decoder structure [28] is employed. To more effectively process feature maps of different resolutions, a Receptive Field Block (RFB) [29] is integrated into the architecture. The model was trained for 100 epochs using the Adam optimizer with a learning rate of $1e-4$, and the results were reported to outperform existing methods in the literature both in terms of speed and segmentation accuracy.

Several studies that employ different backbone networks or integrate various modules have utilized the Reverse Attention (RA) approach. Fan et al. [30] emphasized that region and boundary information are two essential components in distinguishing polyps. In their work, they first generated a coarse region prediction and then modeled the boundaries using a reverse attention mechanism. Their architecture, named PraNet, adopts Res2Net [31] as the backbone network and employs a parallel partial decoder to fuse low-level feature maps. The authors state that the proposed network offers three key advantages in terms of learning capacity, generalization ability, and training efficiency.

Kim et al. developed UACANet, which uses a modified version of the U-Net architecture incorporating an additional encoder and decoder [18]. In their work, a Parallel Axial Attention Encoder (PAA-e) was designed to effectively process low-level feature maps, and the

outputs of this encoder were fed both into the decoder module (PAA-d) and the bottom-up information flow. Additionally, the authors proposed the Uncertainty Augmented Context Attention (UACA) module, which was reported to produce clearer boundary predictions and to outperform the reverse attention approach.

CaraNet, proposed by Lou et al. [32], consists of three main components: a Partial Decoder (PD) that coarsely identifies polyp regions, a Channel-wise Feature Pyramid module (CFP) that processes multi-scale features, and an Axial Reverse Attention module that jointly evaluates local information and global context to produce more accurate feature representations. In this study, axial reverse attention is used instead of traditional reverse attention. Axial attention is based on the self-attention mechanism, which maps a query to a set of key/value pairs. Since self-attention incurs high computational cost

on large spatial dimensions, the attention mechanism in this approach is applied separately along the height and width axes, and the two resulting attention components are then combined to obtain the final axial attention.

In studies utilizing the reverse attention approach, experiments were conducted on five publicly available polyp datasets: CVC-ClinicDB [25], CVC-ColonDB [33], ETIS [1], Kvasir [34], and CVC300 [35]. Among these datasets, Kvasir is designed for competition purposes, while the others are widely used benchmark datasets. During the training process, the Adam optimization method was employed, the batch size was set to 16, and the learning rate was maintained at $1e-4$. The test performances of the models trained with different numbers of epochs are presented in Table 1.

Table 1. Comparison of the prediction performance of the methods

	epoch	CVC-300		CVC-ClinicDB		CVC-ColonDB		ETIS-LaribPolypDB		Kvasir	
		mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU
PraNet	20	0.871	0.797	0.899	0.849	0.709	0.640	0.628	0.567	0.898	0.840
UACANet	200	0.910	0.849	0.926	0.880	0.783	0.704	0.766	0.689	0.912	0.859
CaraNet	-	0.903	0.838	0.936	0.887	0.773	0.689	0.747	0.672	0.918	0.865
HarDNet	100	0.887	0.821	0.932	0.882	0.731	0.660	0.677	0.613	0.912	0.857

Examining Table 1, it is observed that PraNet yields the lowest performance across all datasets; however, this result corresponds to training the model for only 20 epochs. In the related study, it was stated that the network reached its optimal training level after 20 epochs. In contrast, HarDNet was trained for 100 epochs, and UACANet for 200 epochs. In the case of CaraNet, the number of epochs used was not specified.

In this study, a hybrid network model is proposed for automatic polyp segmentation. The feature extraction layers of the DAF3D [36] network, which has demonstrated strong performance in deep feature extraction, are utilized. Since the original DAF3D architecture was designed for volumetric 3D medical image analysis, several architectural modifications were introduced to adapt it for 2D image segmentation. Specifically, all 3D convolutional layers (Conv3d) were replaced with their 2D counterparts (Conv2d), trilinear upsampling operations were substituted with bilinear upsampling, and Group Normalization layers were replaced with Batch Normalization to better suit the 2D training paradigm. The extracted feature maps are first enhanced using the Channel Attention Module (CAM) [37], and then fused through the application of a Reverse Attention (RA) module. Finally, a sigmoid function is applied to the resulting feature map to produce the network output. The next section provides a detailed explanation of the proposed network architecture, followed by the presentation of experimental results

2. MATERIAL AND METHOD

Based on extensive experimental analysis, a hybrid network model has been developed for automatic polyp segmentation. This model is based on extracting feature maps at different levels and subsequently fusing these single-layer feature (SLF) maps. The SLF maps contain the deep features of the input image. While generating these feature maps through a pyramid network structure, information from the preceding layer is also utilized. Each SLF is enriched using the CAM. This design enables adaptive weighting of inter-channel dependencies across different resolution levels and enhances the discriminative representation capacity of each layer.

The multi-scale features processed by CAM are then progressively integrated through Reverse Attention (RA) blocks following a hierarchical top-down refinement strategy. As shown in Figure 1, the multi-scale features obtained from the FPN are first arranged as SLFs, and each SLF is strengthened at the channel level via CAM. By combining all CAM outputs, a multi-layer feature map (MLF) is generated, which serves as the input to the first RA block. Subsequently, the RA output from the highest level is fused with the CAM output of the next lower level within sequential RA blocks. This process is repeated across all layers, forming a hierarchical reverse-attention refinement chain.

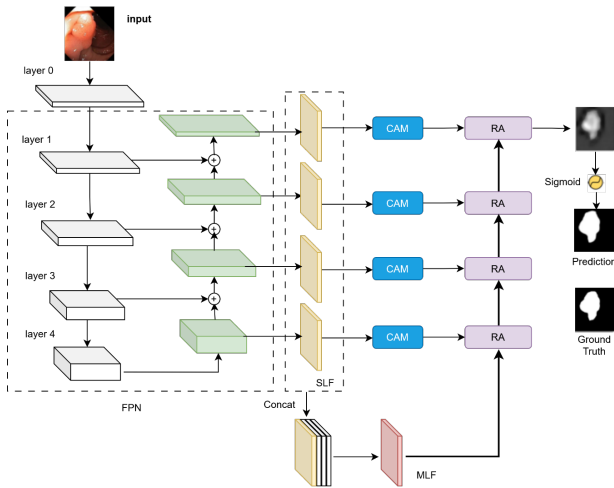


Figure 1. Proposed Network Architecture

Through this architecture, contextual information and fine-grained details across layers are jointly leveraged, enabling the segmentation mask to be progressively refined.

As seen in the architecture, the SLF blocks—each containing the CAM outputs from different resolution levels—are fused to form the Multi-Layer Features (MLF) structure. This stage aims to aggregate cross-scale information into a unified representation, thereby combining high-level semantic cues, mid-level structural information, and low-level fine details. The MLF module processes this multi-layer context and generates a high-level guidance signal that is subsequently passed to the RA pathway. In this way, the attention maps used by the RA modules are not derived from a single scale but are instead informed by an integrated multi-scale representation.

At this stage, the prediction process of the network operates as follows:

- The global context map generated by the MLF is processed in the first RA block together with the CAM output of the highest-level SLF. This RA unit applies the reverse attention mechanism to refine coarse localization and produce a sharper mask of the target region (e.g. polyps).
- The output of the RA block is passed to the next lower level and fused with the CAM output of the

corresponding lower-level SLF in another RA block. Thus, high-level, semantically rich but low-resolution information interacts with the higher-resolution details found in lower layers.

- This process is repeated for each layer. At every step, the CAM output of the current SLF is combined with the RA output from the previous level within a dedicated RA block.

Thus, the RA modules form a progressive refinement chain across the multi-scale feature maps enhanced by CAM. Each RA unit further refines the attention map received from the preceding level, revealing the segmentation target with increasing precision.

In the final stage, the output of the entire RA chain is passed through a sigmoid activation function to generate the final prediction mask. This design preserves high-level semantic context while preventing the loss of fine details at lower levels, enabling the network to achieve highly accurate segmentation results.

3. RESULTS

The network model proposed for automatic polyp segmentation is fundamentally based on extracting single-layer feature maps from the ResNet backbone and then applying attention mechanisms to these maps to generate a single multi-layer feature representation.

All input images are uniformly resized to 352×352 pixels. Instead of data augmentation, a multi-scale training strategy $\{0.75, 1, 1.25\}$ is employed during training. Adam is used for parameter optimization with a learning rate of $1e-4$. The batch size is set to 8, and training is conducted for 20 epochs

To evaluate the impact of the two attention modules added to the network, several experiments were conducted. First, the network was trained using only the channel attention module, then using only the reverse attention module. Finally, both modules were used together during training. The results obtained from these experiments are presented in Table 2.

Table 2. Ablation study for proposed network on the dataset

	CVC-300		CVC-ClinicDB		CVC-ColonDB		ETIS-LaribPolypDB		Kvasir	
	mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU
Backbone	0.753	0.691	0.759	0.704	0.578	0.512	0.527	0.471	0.759	0.699
BackBone + CAM	0.767	0.704	0.777	0.723	0.617	0.548	0.558	0.494	0.779	0.716
Backbone + RA	0.783	0.724	0.789	0.739	0.631	0.589	0.592	0.523	0.801	0.736
Backbone + CAM + RA	0.891	0.829	0.927	0.883	0.759	0.684	0.723	0.668	0.911	0.855

As shown in Table 2, both the channel attention module and the reverse attention module individually improved the performance of the network. However, the best results were achieved when both modules were used together.

To further validate the proposed method against existing approaches in the literature, two distinct comparative evaluation protocols were employed.

Attention Mechanism Comparison: To isolate and assess the individual contributions of different attention mechanisms, the candidate approaches were systematically evaluated on an identical backbone architecture under controlled and uniform conditions. All

models were trained for 20 epochs, and performance was measured using mean Dice coefficient (mDice) and mean Intersection over Union (mIoU) metrics. As presented in Table 3, the proposed method achieved the best performance among all compared attention mechanisms.

Table 3. Comparison of attention mechanisms on the same backbone architecture.

	CVC-300		CVC-ClinicDB		CVC-ColonDB		ETIS-LaribPolypDB		Kvasir	
	mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU
PraNet	0.759	0.699	0.766	0.711	0.583	0.519	0.533	0.477	0.770	0.705
CaraNet	0.751	0.690	0.787	0.736	0.633	0.563	0.573	0.504	0.785	0.727
UACANet	0.700	0.648	0.760	0.710	0.524	0.471	0.470	0.423	0.769	0.709
HarDNet	0.692	0.601	0.750	0.690	0.584	0.499	0.405	0.337	0.711	0.629
Proposed	0.891	0.829	0.927	0.883	0.759	0.684	0.723	0.668	0.911	0.855

Comparison with State-of-the-Art Methods: To enable a fair and rigorous comparison, the baseline models were evaluated using the results reported in their respective publications, without retraining. The proposed method, on the other hand, was trained for 50 epochs with a batch size of 8 due to GPU memory constraints, as the originally employed batch size of 16 could not be accommodated by the available hardware. The learning curves indicated a gradual deceleration in performance improvement beyond the 40th epoch, with only marginal gains observed

thereafter; consequently, 50 epochs were deemed sufficient for convergence, as no statistically meaningful improvement in validation metrics was observed beyond this point. Despite these relatively constrained training conditions, the proposed method achieved superior performance across all evaluated datasets, as presented in Table 4, outperforming all state-of-the-art approaches in terms of both mDice and mIoU and demonstrating the effectiveness of the proposed architecture for polyp segmentation.

Table 4. Comparison with state-of-the-art methods for polyp segmentation.

	epoch	CVC-300		CVC-ClinicDB		CVC-ColonDB		ETIS-LaribPolypDB		Kvasir	
		mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU	mDice	mIOU
PraNet	20	0.871	0.797	0.899	0.849	0.709	0.640	0.628	0.567	0.898	0.840
UACANet	200	0.910	0.849	0.926	0.880	0.783	0.704	0.766	0.689	0.912	0.859
CaraNet	-	0.903	0.838	0.936	0.887	0.773	0.689	0.747	0.672	0.918	0.865
HarDNet	100	0.887	0.821	0.932	0.882	0.731	0.660	0.677	0.613	0.912	0.857
Proposed	50	0.921	0.853	0.951	0.897	0.796	0.714	0.781	0.694	0.932	0.887

Figure 2 presents a qualitative comparison of the segmentation outputs generated by the compared methods across representative samples from each dataset. As illustrated, PraNet tends to produce noisy and spatially incoherent predictions, particularly on challenging datasets such as ETIS and CVC-ColonDB, and exhibits over-segmentation artifacts in certain cases, as observed in the CVC-300 sample. UACANet and CaraNet yield comparatively cleaner outputs; however, they occasionally fail to accurately delineate polyp boundaries. The proposed method, in contrast, consistently produces segmentation masks that closely conform to the ground truth across all datasets, demonstrating superior boundary accuracy and spatial consistency. These qualitative observations are in agreement with the quantitative results reported in Table 4, further validating the effectiveness of the proposed architecture.

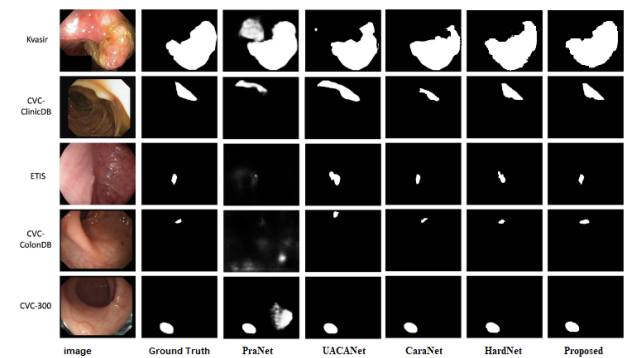


Figure 2. Segmentation outputs across five benchmark datasets.

Overall, the experimental results consistently demonstrate that the proposed method achieves superior segmentation performance across both the attention mechanism comparison and the state-of-the-art evaluation, validating the effectiveness of the proposed architecture for polyp segmentation.

4. DISCUSSION AND CONCLUSION

In this study, a hybrid deep learning architecture that combines multi-scale feature extraction with two different attention mechanisms is proposed for automatic polyp segmentation in colonoscopy images. The primary motivation behind the approach is the inability of conventional deep features to provide sufficient discriminative power on their own, particularly due to the low contrast and irregular morphology of polyps, which often leads to significant information loss in boundary regions. To address this issue, the multi-level feature maps extracted from the ResNet backbone are first enhanced using the Channel Attention Module (CAM) and then fused through the Reverse Attention Module (RA) to generate a multi-layer feature representation with high boundary sensitivity. The main contribution of this study is the demonstration that the sequential use of these two attention mechanisms produces a strong synergistic effect, improving boundary clarity, contextual representation, and overall segmentation accuracy.

Ablation studies clearly demonstrate that both modules contribute significantly to segmentation performance. In particular, when RA is used alone, a notable performance improvement is achieved across all datasets due to its ability to make boundary regions more prominent. The CAM module, by emphasizing informative features at the channel level, enables the network to learn more consistently in images with low contrast and high texture similarity. However, the most striking finding is that the combination of CAM and RA yields by far the highest performance across all datasets. This indicates that jointly addressing channel-wise focus and boundary sensitivity provides a more accurate representation of polyp structure, offering a distinct advantage especially for small or irregularly shaped polyps with complex boundary characteristics.

Another important outcome of this study is that the proposed architecture achieves higher mean Dice and mean IoU scores across all datasets when compared with competitive state-of-the-art methods in the literature such as PraNet, CaraNet, HarDNet-MSEG, and UACANet. Notably, even on challenging datasets such as ETIS-LaribPolypDB, the proposed method demonstrates a more balanced and stable performance than the other models, indicating strong generalizability. As supported by the ablation study results, this success can be largely attributed to the multi-scale representation capability of ResNet combined with the complementary nature of the proposed attention modules.

Considering the findings as a whole, the proposed architecture provides a comprehensive improvement in polyp segmentation by enhancing boundary sensitivity and enriching the level of detail. However, several limitations should also be acknowledged. Due to hardware constraints, the batch size had to be kept small, which may have introduced some variability into the training dynamics. Additionally, the study is limited to an image-based segmentation approach; incorporating video-based polyp detection and temporal consistency

methods could further improve performance in future work. Likewise, thoroughly evaluating the speed–accuracy trade-off in real-time applications would be important for better assessing the potential clinical integration of the model.

In conclusion, this study presents an effective deep learning approach for polyp segmentation in colonoscopy images, enriched with hybrid attention mechanisms and designed to leverage multi-scale feature information. The combined use of CAM and RA not only introduces a novel architectural contribution to the literature but also delivers strong performance even on challenging datasets. These findings indicate that the proposed model holds significant potential for integration into clinical decision-support systems. Future work focused on optimizing model speed, performing video-based validation, and evaluating the method on larger datasets could further expand its applicability and impact.

REFERENCES

- [1] Silva J, Histace A, Romain O, Dray X, Granado B. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*. 2014;9(2):283-293.
- [2] Hagggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in colon and rectal surgery*. 2009;22(04):191-197
- [3] Heresbach D, Barrioz T, Lapalus M, Coumaros D, Bauret P, Potier P, et. Al. Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies, *Endoscopy*. 2008;40:284–290.
- [4] Leufkens A, Oijen van M, Vleggaar F, Siersema P. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study, *Endoscopy*. 2012;44:470-475
- [5] Karkanis SA, Iakovidis DK, Maroulis DE, Karras DA, Tzivras M. Computer-aided tumor detection in endoscopic video using color wavelet features, *IEEE Trans. Inf. Technol. Biomed*. 2003;7:141–152.
- [6] Nawarathna R, Oh JH, Muthukudage J, Tavanapong W, Wong J, Groen PC, et. al. Abnormal image detection in endoscopy videos using a filter bank and local binary patterns, *Neurocomputing*. 2014;144:70–91
- [7] Mamonov AV, Figueiredo IN, Figueiredo PN, Tsai YHR. Automated Polyp Detection in Colon Capsule Endoscopy, *IEEE Trans. Med. Imaging*. 2014;33:1488–1502.
- [8] Tajbakhsh N, Gurudu SR, Liang J. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information, *IEEE Trans. Med. Imaging*. 2016;35:630–644.
- [9] Yu L, Chen H, Dou Q, Qin J, Heng PA. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE journal of biomedical and health informatics*. 2016;21(1):65-75

- [10] Brandao P, Mazomenos E, Ciuti G, Calìo R, Bianchi F, Mencias A, et. al. Fully convolutional neural networks for polyp segmentation in colonoscopy. In Medical Imaging 2017: Computer-Aided Diagnosis. 2017;10134:101-107
- [11] Akbari M, Mohrekesh M, Nasr-Esfahani E, Soroushmehr SR, Karimi N, Samavi S, 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 (EMBC). 2018;69-72
- [12] Zhang L, Dolwani S, Ye X. Automated polyp segmentation in colonoscopy frames using fully convolutional neural network and textons. In Annual Conference on Medical Image Understanding and Analysis. 2017;707-717. Springer, Cham.
- [13] Jha D, Smedsrud PH, Riegler MA, Johansen D, De Lange T, Halvorsen P, et. al. Resunet++: An advanced architecture for medical image segmentation. IEEE International Symposium on Multimedia (ISM) 2019. San Diego, IEEE;2019. p. 225-2255.
- [14] Su J, Li J, Zhang Y, Xia C, Tian Y. Selectivity or invariance: Boundary-aware salient object detection. International Conference on Computer Vision 2019. Seoul, IEEE/CVF; 2019. p. 3799-3808
- [15] Yang B, Zhang X, Chen L, Yang H, Gao Z. Edge guided salient object detection. Neurocomputing. 2017;221:60-71.
- [16] Fang, Y., Chen, C., Yuan, Y., & Tong, K. Y. Selective feature aggregation network with area-boundary constraints for polyp segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 2019. China, Springer, Cham; 2019 p. 302-310.
- [17] Murugesan B, Sarveswaran K, Shankaranarayana SM, Ram K, Joseph J, Sivaprakasam M. Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019. Strasbourg, IEEE; 2019 p. 7223-7226.
- [18] Kim T, Lee H, Kim D. Uacnet: Uncertainty augmented context attention for polyp segmentation. 29th ACM International Conference on Multimedia 2021. Chengdu, ACM; 2021 p. 2167-2175.
- [19] Bozdağ Z. Histopatolojik görüntülerde tümör bölütlenmesi. [Phd. thesis], Malatya: İnönü Üniversitesi; 2021.
- [20] Kayaoğlu M, Şengür A, Çiğirim SÇ, Bor S. Detection of Cervical Vertebrae Using Object Detection and Semantic Segmentation Methods in Lateral Cephalometric Radiographs, Turkish Journal of Nature and Science. 2025;14(2): 26–36.
- [21] Tomar NK, Jha D, Ali S, Johansen HD, Johansen D, Riegler MA, et. al. DDANet: Dual decoder attention network for automatic polyp segmentation. International Conference on Pattern Recognition 2021. Kolkata, Springer, Cham; p. 307-314.
- [22] Tomar NK, Jha D, Riegler MA, Johansen HD, Johansen D, Rittscher J, et. al. Fanet: A feedback attention network for improved biomedical image segmentation. IEEE Transactions on Neural Networks and Learning Systems. 2022;34(11): 9375-9388.
- [23] Tomar NK, Jha D, Bağcı U, Ali S. TGANet: Text-guided attention for improved polyp segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention 2022. Switzerland, Cham: Springer Nature; p.151-160.
- [24] Xu Q, Duan W, He N. DCSAU-Net: A Deeper and More Compact Split-Attention U-Net for Medical Image Segmentation. Computers in Biology and Medicine. 2023;154: 106626.
- [25] Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño, F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics. 2015;43:99-111.
- [26] Huang CH, Wu HY, Lin YL. 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.
- [27] Chao P, Kao CY, Ruan YS, Huang CH, Lin YL. Hardnet: A low memory traffic network. International Conference on Computer Vision 2019. Seoul, IEEE/CVF; 2019. p. 3552-3561.
- [28] Wu Z, Su L, Huang Q. Cascaded partial decoder for fast and accurate salient object detection. International Conference on Computer Vision and Pattern Recognition 2019. Seoul, IEEE/CVF; 2019. p. 3907-3916.
- [29] Chen S, Tan X, Wang B, Hu X. Reverse attention for salient object detection. In Proceedings of the European conference on computer vision (ECCV). 2018. p. 234-250.
- [30] Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J. et al. Pranet: Parallel reverse attention network for polyp segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 263-273). 2020; Springer, Cham.
- [31] Gao SH, Cheng MM, Zhao K, Zhang XY, Yang MH, Torr P. Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence. 2019; 43(2): 652-662.
- [32] Lou A, Guan S, Loew M. Caranet: Context axial reverse attention network for segmentation of small medical objects. Journal of Medical Imaging. 2023;10(1):14005-14005.
- [33] Bernal J, Sánchez J, Vilarino F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition. 2012;45(9): 3166-3182.
- [34] Jha D, Smedsrud PH, Riegler MA, Halvorsen P, Lange TD, Johansen D, et. al. Kvasir-seg: A segmented polyp dataset. International Conference on Multimedia Modeling. 2020;pp. 451-462. Springer, Cham.

- [35] Vazquez D, Bernal J, S´anchez FJ, Fern´andez-Esparrach GL´opez, AM, Romero A, Drozdal M, et. al. A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*. 2017; 2017(1): 4037190.
- [36] Wang Y, Dou H, Hu X, Zhu L, Yang X, Xu M, et. al. Deep attentive features for prostate segmentation in 3D transrectal ultrasound. *IEEE transactions on medical imaging*. 2019;38(12): 2768-2778.
- [37] Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et. al. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3146-3154, 2019.