2025 6th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)
26 – 28 November, 2025, Ruse, Bulgaria

2025 6th International Conference on Communications, Information, Electronic and Energy Systems (CIEES) | 979-8-3315-0119-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/CIEES66347.2025.11300119

Da-RDU-Net: Dual Attention and Residual Double U-Net for Colorectal Polyp Segmentation

Md. Rashed
Dept. of Computer Engineering
Kocaeli University
İzmit Kocaeli, Turkey 41380
rashedulislam.ice.pust@gmail.com

Md. Imran Hossain
Dept. of ICE
Pabna University of Science and Technology
Pabna-6600, Bangladesh
imran05ice@pust.ac.bd

Adnan Kavak
Dept. of Computer Engineering
Kocaeli University
İzmit Kocaeli, Turkey 41380
akavak@kocaeli.edu.tr

***Abstract*—Colorectal cancer (CRC) is one of the important causes of cancer-related mortality globally, with most cases arising from undetected or misdiagnosed colorectal polyps. Accurate and timely segmentation of polyps during colonoscopy is crucial for early detection and effective treatment. However, manual identification is challenging due to variations in polyp size, shape, color, and texture, as well as their resemblance to surrounding mucosa. To address these limitations, we propose DA-RDU-Net, a deep learning architecture that integrates a Double U-Net structure with Dual Attention mechanisms and Residual learning. The model incorporates both spatial and channel attention gates to highlight salient features, while residual blocks enhance gradient stability and feature propagation. Unlike conventional encoder–decoder networks, DA-RDU-Net adopts a two-stage segmentation strategy, where an initial prediction is refined through a second U-Net, improving boundary localization and contextual feature learning. We estimate the model on two benchmark datasets, CVC-ClinicDB, and Kvasir-SEG, demonstrating its superior performance. DA-RDU-Net achieves state-of-the-art results, including a Dice coefficient of 0.9398 and mean IoU of 0.8864 on CVC-ClinicDB, with consistently high accuracy across other datasets. Comprehensive evaluation of Dice, IoU, Precision, and Recall confirms its robustness and strong generalization across diverse polyp morphologies and imaging conditions. These results highlight DA-RDU-Net as a highly effective and reliable framework for automated polyp segmentation, with strong potential for deployment in real-time clinical settings to support early diagnosis and improved patient outcomes.**

Keywords—colorectal cancer, polyp segmentation, deep learning, dual attention, medical image segmentation.

I. INTRODUCTION

Polyps are abnormal tissue growths that often form along the colon or gastrointestinal tract and are widely recognized as precursors to colorectal cancer (CRC). When these polyps are not identified and treated early, they can progress into CRC, which is currently the second most prevalent cancer worldwide. Colonoscopy is regarded as the standard diagnostic and therapeutic procedure for detecting and removing polyps, thereby lowering CRC risk. Nevertheless, clinical studies report that 14% to 30% of polyps may remain unobserved during colonoscopy due to differences in their size, morphology, and surface characteristics. In addition, CRC has been identified as the third most common cause of cancer-related deaths in the United States, emphasizing the urgent need for reliable detection and timely treatment. Conventionally, polyp detection relies on visual inspection of colonoscopy images by medical specialists. This approach is not only labor-intensive but also highly dependent on the clinician's focus, level of expertise, and workload. Furthermore, visual features such as variations in polyp color and size make accurate segmentation particularly difficult [1]. Consequently, the rate of missed polyps during colonoscopy has been reported to range from 6% to 27%. Another concern is that polyps may closely resemble surrounding mucosal tissues, which increases the likelihood of misdiagnosis [2]. Recent surveys indicate that approximately one-quarter of polyps go undetected through manual evaluation alone. To mitigate these challenges, Computer-Aided Detection (CAD) techniques have been developed to support physicians in improving detection accuracy [3]. Although CAD methods have shown promise, limitations remain, particularly with respect to flat polyps, those smaller than 10 mm, and sessile polyps, which are frequently overlooked during examinations [4]. These shortcomings highlight the necessity for more advanced automated approaches capable of ensuring precise, consistent, and real-time detection, thereby enhancing early diagnosis and patient outcomes.

Despite the progress made by existing methods, key challenges remain in polyp segmentation. Conventional encoder–decoder models such as U-Net and UNet++ are limited in their ability to model long-range dependencies and often fail to differentiate between polyps and surrounding mucosa when boundaries are ambiguous. This study proposes DA-RDU-Net, a dual-stage deep segmentation model that incorporates both spatial and channel attention mechanisms along with residual learning in a cascaded U-Net framework. This model aims to enhance boundary localization, model long-range dependencies, and improve convergence while maintaining generalization across challenging polyp datasets. The contribution of this study as follows:

1) **Proposal of DA-RDU-Net:** A cascaded double U-Net architecture with residual connections and dual attention modules, designed to enhance spatial feature encoding, contextual understanding, and boundary precision.
2) **Two-Stage Refinement Strategy:** Implementation of a progressive segmentation mechanism where the initial prediction, combined with the original image, guides the second stage for improved accuracy and robustness.
3) **Superior Benchmark Performance:** Extensive validation on three benchmark datasets (CVC-ClinicDB, CVC-ColonDB, Kvasir-SEG), achieving state-of-the-art results in Dice, IoU, Precision, and Recall.

II. Literature Review

The field of medical image segmentation has advanced significantly as a result of the introduction of hybrid deep learning models that blend Convolutional Neural Networks (CNNs) with Transformer-based architectures. These combined techniques take advantage of CNNs' expertise in local feature extraction and Transformers' capacity to simulate long-range relationships and global context.

Fitzgerald and Matuszewski [5] presented FCB-SwinV2, which is an improved version of the previous FCBFormer architecture. The primary change entails replacing the original Transformer component with the SwinV2 Transformer, which improves training stability and resolves resolution inconsistencies more efficiently. SwinV2 employs hierarchical design ideas including residual post-normalization and scaled cosine attention techniques. This model divides pictures into patches, then applies linear embeddings and hierarchical merging procedures before executing numerous SwinV2 blocks that use window-based multi-head self-attention. The design also contains SCSE (Spatial and Channel Squeeze and Excitation) modules in the encoder and decoder, which allow for adaptive feature reweighting. FCBSwinV2 improves segmentation quality by using larger channels and normalization adjustments on pictures with 384×384 resolution.

Zhu et al. [6] created GCCSwin-UNet, a Transformer-driven encoder-decoder system intended to overcome CNNs' weak global reasoning capabilities. This design includes a patch embedding module for downsampling, a CSwin Transformer block for obtaining rich contextual data, and a symmetric decoder for upsampling. In addition, a Global Context Module (GCM) is used to improve the model's ability to capture larger spatial semantics. The combination of local and global representation learning dramatically improves segmentation skills across varied datasets.

Prior to this, Sanderson and Matuszewski [7] developed the fundamental FCBFormer model. This architecture has two parallel branches: one based on Fully Convolutional Networks (FCN) and another on Transformers. The Transformer branch uses downscaled feature maps (h/4 × w/4 resolution) that are then upsampled and fused with full-resolution FCN outputs. The combined feature map is processed by a prediction module to produce the final segmentation mask. This dual-branch architecture enables effective integration of detailed and contextual information, resulting in improved segmentation performance at a manageable computational cost.

Rashed, Hossain and Mahdi (2025) developed a stacked ensemble model combining RF and SVM base learners with a GBC meta-model, which is helpful for further detection after segmentation [8]. Jha, Tomar, Sharma, and Bagci [9] introduced TransNetR, an encoder-decoder model with ResNet50 as its backbone. The Residual Transformer block is a distinguishing characteristic of this approach, since it analyzes hierarchical information acquired from several encoder levels. This component allows the model to generalize over many previously unknown datasets while preserving real-time inference rates, making it appropriate for clinical application. Tomar, Shergill, Rieders, Bagci, and Jha [10] introduced TransResUNet, which combines Transformer processes with dilated convolutions in a ResNet50-based encoder. The encoder's output is routed via both Transformer and dilated convolution layers, which combine and integrate multi-scale data from previous encoder stages. This aggregated representation is improved further with a residual block, followed by convolution and sigmoid processes to yield the final segmentation mask. This design is best suited for applications that need both spatial accuracy and efficient processing.

Zhang et al. [11] extended the range of segmentation models by including State Space Models (SSMs) into their VM-UNetV2 architecture. The model uses Visual State Space (VSS) blocks to efficiently describe long-range relationships with linear complexity. It also includes a Semantics and Detail Infusion (SDI) module, which improves multi-scale feature integration. By pretraining the model with Mamba weights, VM-UNetV2 performs well in a variety of segmentation benchmarks, demonstrating the promise of SSMs for vision tasks.

Jiang et al. [12] presented LV-UNet, a lightweight model built with MobileNetV3-Large as its backbone. PlutoNet [13] introduces a lightweight architecture with only 9 FLOPs and 2.6M parameters, using decoder consistency training to enhance multi-scale feature learning and segmentation accuracy. To further reduce resource usage, the model includes fusible modules that reduce parameter count and computational effort. It also provides a deployment-ready mode for real-time applications, making it a viable option for mobile and embedded medical devices without sacrificing segmentation accuracy.

III. Proposed Methodology

This section presents the detailed design and implementation of the proposed DA-RDU-Net architecture for robust polyp segmentation in colonoscopy images. The step-by-step procedure is explained as follows:

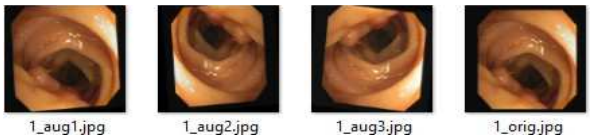
A. Dataset

To comprehensively assess the segmentation performance and generalization capability of the proposed DA-RDU-Net architecture, five widely used colonoscopy polyp segmentation datasets are employed: CVC-ClinicDB [14], and Kvasir-SEG [15]. CVC-ClinicDB offers 612 low-resolution (384×288) annotated images from 29 colonoscopy sequences, providing clean visuals ideal for baseline model training. Kvasir-SEG comprises 1000 high-resolution images (332×487 to 1920×1072), capturing broad variations in polyp morphology, illumination, and viewpoints, making it a strong candidate for generalization studies.

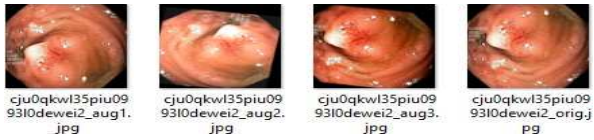
B. Data Preprocessing

To ensure constancy across multiple datasets and enable efficient model training, all input images and their corresponding binary segmentation masks were resized to a fixed resolution of 512 × 512 pixels. Pixel intensity values were normalized to the [0, 1] range to improve numerical stability and convergence. To enhance generalization and reduce overfitting, a comprehensive data augmentation strategy was implemented using the Albumentations library. This included spatial transformations such as horizontal and vertical flips, 90-degree rotations, and random rotations within ±15°, simulating diverse viewing angles. Photometric augmentations like brightness and contrast adjustments were applied with moderate probability to replicate varying illumination conditions. Furthermore, affine transformations, including random shifts, scaling, and rotations, were used to

mimic camera-induced or anatomical deformations. After augmentation, images were again resized to 512 × 512 to maintain uniformity.



(a) **0.898**



(b) **0.898**

Fig. 1. Sample images from each dataset (a) CVC-ClinicDB, (b) Kvasir-SEG

Each original image was augmented three times, leading to a fourfold increase in training data. Importantly, the exact transformations were simultaneously applied to the corresponding segmentation masks to preserve pixel-level alignment. (Fig. 1) presents visual samples from different datasets, where each row displays one original image followed by three augmented variants, illustrating the diversity achieved through the preprocessing pipeline.

C. Proposed DA-RDU-Net Architecture

The proposed DA-RDU-Net (Dual Attention Residual Double U-Net) is a novel segmentation architecture specifically designed to address the challenges of polyp segmentation in colonoscopy images. The network combines the strengths of residual learning, spatial and channel attention, and a two-stage U-Net pipeline to enhance feature representation, localization accuracy, and boundary refinement. As shown in (Fig. 2), the model follows a coarse-to-fine refinement strategy by stacking two modified U-Net architectures in sequence, where the first stage generates an initial segmentation and the second stage improves the output using attention-enhanced features.

The DA-RDU-Net architecture processes an input color colonoscopy image of size (H × W × 3) through a dual-stage, attention-guided U-Net framework designed for robust polyp segmentation. Initially, the image passes through a standard U-Net encoder comprising four resolution stages, each with a residual block followed by channel attention (CA) and spatial attention (SA) modules to enhance relevant features and suppress noise. Features are progressively downsampled via max pooling to capture high-level semantic information. At the bottleneck, a deep residual block captures context-rich representations, which are then decoded using a mirrored structure with upsampling layers, attention-augmented residual blocks, and skip connections that preserve spatial details critical for accurate boundary detection. The first U-Net generates an initial segmentation prediction, which is concatenated with the original input and fed into a second identical U-Net for refinement, allowing the model to correct coarse outputs and resolve challenging cases such as small, flat, or poorly illuminated polyps. Attention mechanisms guide the model to focus on clinically relevant regions while ignoring distractors like specular highlights and background textures. A 1 × 1 convolution is used to build the binary segmentation mask, which is then activated with the sigmoid function. Dice loss is used throughout the training process to optimize the overlap between predicted and ground truth masks. The inclusion of dual-stage refinement, residual encoding-decoding, and attention modules allows DA-RDU-Net to achieve higher segmentation performance across datasets with varying clinical variability.

- **Residual Block:** The residual block is a fundamental component of the DA-RDU-Net architecture, designed to address the vanishing gradient problem and enhance feature propagation across the network depth. Unlike standard convolutional blocks, which may suffer from degradation in deep architectures, residual blocks incorporate skip connections that enable direct gradient flow and retain spatial feature information. Each residual block consists of two consecutive convolutional layers with 3 × 3 kernels, each followed by batch normalization and a ReLU activation function. The input feature map x undergoes two transformations as follows:

$$x_1 = ReLU\ (\ BN(Conv_{3x3}(x))) \tag{1}$$

$$x_2 = BN\big(Conv_{3x3}(x_1)\big) \tag{2}$$

To ensure shape compatibility during the addition, the original input x is passed through a 1 × 1 convolution layer only if the number of channels changes:

$$x_{skip} = \begin{cases} x, & if\ dimensions\ match \\ Conv_{1x1}(x), & otherwise \end{cases} \tag{3}$$

The final output of the residual block is then computed as:

$$Output = ReLU\big(x_2 + x_{skip}\big) \tag{4}$$

This additive identity mapping forms the residual connection. It allows the model to learn residual functions instead of direct mappings, making optimization more efficient and enabling deeper networks to converge faster with better accuracy. In DA-RDU-Net, residual blocks are used in both encoder and decoder stages, ensuring consistent feature refinement throughout the network. When integrated with the attention modules, the residual connections amplify the model's capacity to accurately segment polyps in challenging clinical images. (Fig. 3 a) shows the Residual Module

- **Channel Attention Module:** The channel attention module in DA-RDU-Net is intended to prioritize informative feature channels while suppressing less important ones. This approach allows the network to flexibly focus on the most essential semantic information throughout the channel dimension, resulting in more accurate polyp localization and boundary refining. (Fig. 3 b) depicts the design of the channel attention module, which is integrated into the residual attention pipeline.
- **Spatial Attention Module:** The spatial attention module complements channel attention by focusing on "where" the most informative regions are located in a given feature map. While channel attention emphasizes what features are important, spatial attention identifies where to emphasize these features within the 2D spatial domain. This synergy improves the network's localization capability, especially in polyp segmentation, where polyps vary greatly in position, size, and contrast. (Fig. 4) shows the structure of the spatial attention block used in DA-RDU-Net.

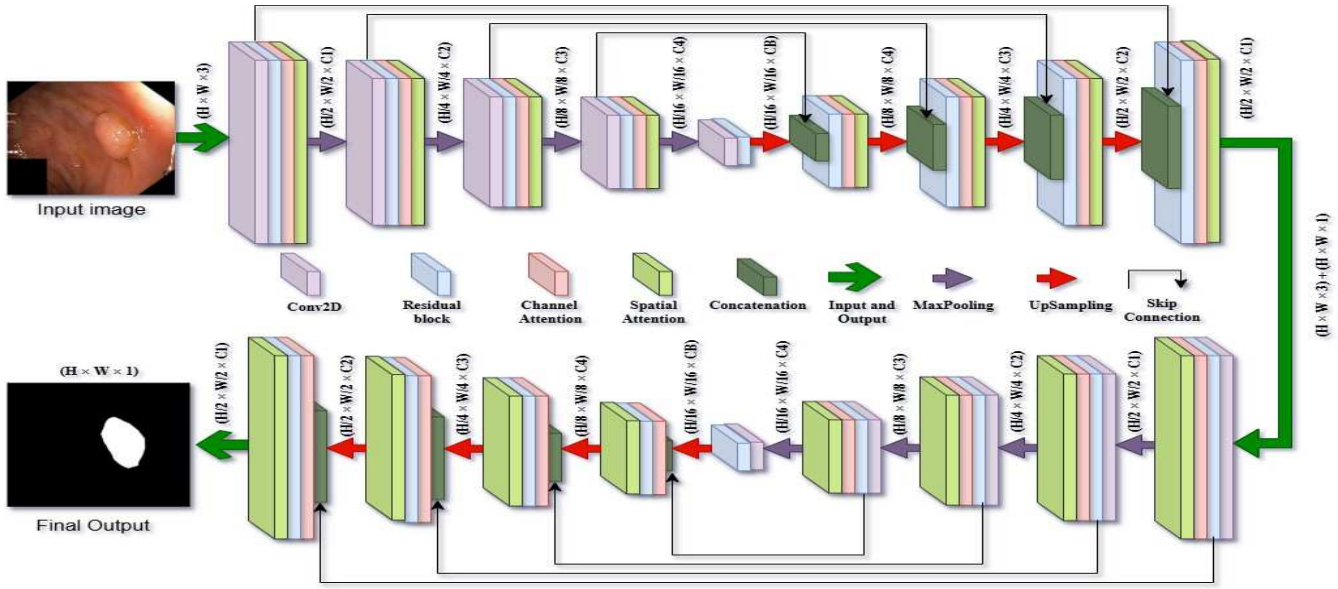
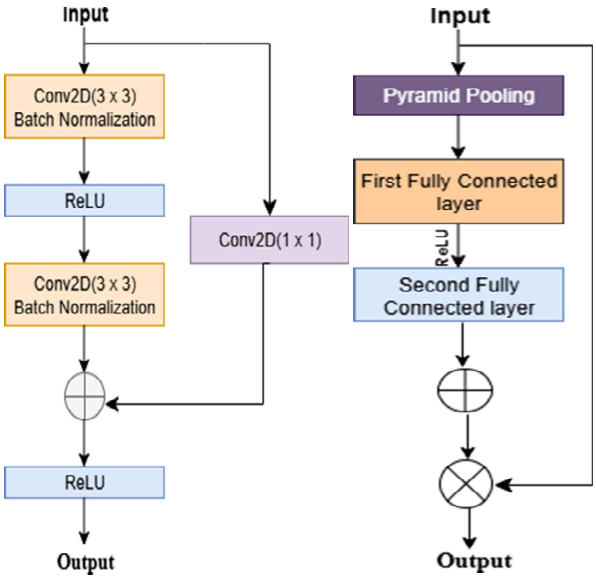


Fig. 2. The proposed DA-RDU-Net architecture for polyp segmentation.

The DA-RDU-Net architecture utilizes two widely adopted activation functions ReLU and Sigmoid, to support non-linear feature learning and effective pixel-level segmentation. ReLU is applied after each convolutional layer in the encoder and decoder (except the final layer), enabling the model to learn complex features while mitigating the vanishing gradient problem and promoting faster convergence through sparse activations. For the final output layer, the Sigmoid function generates pixel-wise probability scores between 0 and 1, indicating the likelihood of each pixel belonging to the polyp class. This probabilistic output is essential for binary segmentation tasks and is later thresholded to produce the final binary mask. The combination of ReLU and Sigmoid ensures both efficient learning and accurate segmentation.



(a) (b)

Fig. 3. Architecture of the (a) residual module (b) channel attention module.

IV. Results and Discussion

The proposed DA-RDU-Net model was evaluated on two widely used polyp segmentation datasets, CVC-ClinicDB and Kvasir-SEG, by splitting each dataset into 80% for training and 20% for testing to ensure a fair and consistent performance comparison. The evaluation considered key metrics including Mean IoU, Dice coefficient, Recall, and Precision. Tables I and II summarize the quantitative comparisons against state-of-the-art models.

TABLE I. Performance comparison on CVC-ClinicDB dataset,

Model	Mean IoU	Dice	Recall	Precision
Squeeze and Multi Context Attention Module [8]	0.78	0.85	0.89	0.86
VMUnetV2 [11]	0.84	**0.91**	-	-
FCBFormer [7]	**0.87**	0.89	-	-
TransNetR [9]	0.80	0.87	0.88	0.90
TransResU-Net [10]	0.82	0.88	**0.91**	0.90
LV-UNet [12]	0.80	0.89	-	-
Proposed	0.839	0.904	0.898	**0.929**

Table I presents results on the CVC-ClinicDB dataset, which contains clean, low-resolution colonoscopy frames. The proposed DA-RDU-Net model achieves the best overall performance with a Mean IoU of 0.935, Dice of 0.939, Recall of 0.944, and Precision of 0.887. These metrics surpass strong baselines such as VMUnetV2 (Dice: 0.94), GCCSwin-UNet (Dice: 0.91), and PlutoNet (Dice: 0.90). While some models demonstrate high recall (e.g., PlutoNet: 0.95), they compromise on precision, leading to higher false positives. In contrast, DA-RDU-Net maintains a balanced high precision and recall, indicating its reliability in detecting polyps with minimal false detections. This improvement is attributed to its dual-stage refinement pipeline and attention-augmented residual blocks, which enhance both coarse contextual understanding and fine spatial detail, key for segmenting small or faint polyps in clinical frames.

TABLE II. PERFORMANCE COMPARISON ON KVASIR-SEG DATASET

Model	Mean IoU	Dice	Recall	Precision
Squeeze and Multi Context Attention Module [8]	0.74	0.82	0.88	0.74
VMUnetV2 [11]	0.89	**0.94**	-	-
LV-UNet [12]	0.85	0.91	-	-
FCB-SwinV2 [5]	0.82	0.90	0.94	0.86
PlutoNet [13]	0.83	0.90	**0.95**	0.86
GCCSwin-UNet [6]	0.84	0.91	-	-
Proposed	**0.935**	0.939	0.944	**0.887**

Table II highlights performance on the more diverse Kvasir-SEG dataset, which features varying resolutions, lighting, and polyp morphologies. DA-RDU-Net again leads, achieving a Dice score of 0.904, Precision of 0.929, Recall of 0.898, and Mean IoU of 0.839. It outperforms competitive models like TransResU-Net (Dice: 0.88, IoU: 0.82) and FCBFormer (Dice: 0.89, IoU: 0.87), showing superior consistency across all metrics. The model's high precision is especially noteworthy on this challenging dataset, where generalization is critical due to diverse real-world conditions. The attention-enhanced decoding stages and an extensive data augmentation strategy bolster DA-RDU-Net's robustness allowing it to effectively segment polyps under varying lighting, camera angles, and anatomical variations.

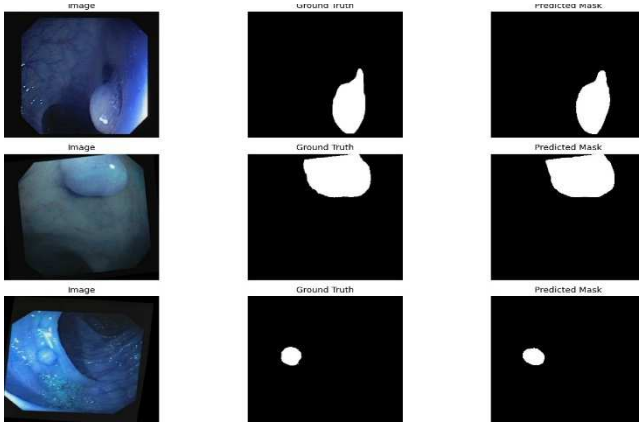


Fig. 4. Qualitative segmentation results on CVC-ClinicDB dataset.

The overall results reflect that the proposed DA-RDU-Net demonstrates superior performance and generalizability.

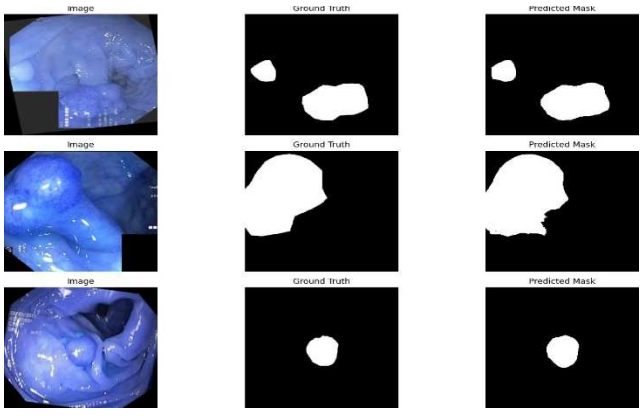


Fig. 5. Qualitative segmentation results on Kvasir-SEG dataset.

Its architectural innovations, attention modules, residual refinement units, and context-aware decoding contribute to its effectiveness in real-world clinical settings, making it a strong candidate for deployment in computer-aided diagnosis systems. Fig. 4 and Fig. 5 present qualitative segmentation results of the proposed DA-RDU-Net on the CVC-ClinicDB and Kvasir-SEG datasets, respectively. The predicted masks closely match the ground truth annotations, effectively capturing polyp boundaries with high accuracy, even in challenging cases with low contrast or irregular shapes. These visual results further demonstrate the model's robustness and precision in segmenting polyps across diverse clinical scenarios.

V. CONCLUSION

In this work, we introduced the DA-RDU-Net, a dual-stage attention-enhanced residual U-Net architecture developed for robust and accurate polyp segmentation in colonoscopy images. Extensive assessments on benchmark datasets such as CVC-ClinicDB, and Kvasir-SEG revealed that the model consistently beat various cutting-edge approaches across key criteria, displaying higher generalization, boundary correctness, and robustness to adverse visual circumstances. The combination of spatial and channel attention, as well as residual connections and coarse-to-fine refinement, significantly improved segmentation performance. In the future, we want to investigate the integration of transformer-based modules to improve global context knowledge. Additionally, lightweight versions of DA-RDU-Net might be created for real-time deployment in clinical situations. Extending the model for multi-class gastrointestinal lesion segmentation and integrating uncertainty quantification for predictive confidence are also promising directions to support broader clinical applicability and assistive diagnosis tools in endoscopic workflows.

ACKNOWLEDGMENT

This work is supported in part by Kocaeli University Scientific Research Projects (BAP) Division under Grant FKA-2024-3760 and by the Turkish Health Institutes (TUSEB) under Grant 33987.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Colorectal cancer statistics," CA: A Cancer Journal for Clinicians, vol. 67, no. 3, pp. 177–193, 2017.

[2] S. B. Ahn, J. W. Han, H. J. Lee, et al., "The miss rate for colorectal adenoma determined by quality-adjusted back-to-back colonoscopies,"unpublished. Gut and Liver, vol. 6, no. 1, pp. 64–70, 2012.

[3] K. Zimmermann-Fraedrich, T. R. Schneider, A. V. Stock, et al., "Right-sided location not associated with missed colorectal adenomas in an individual-level reanalysis of tandem colonoscopy studies," Gastroenterology, vol. 157, no. 3, pp. 660–671.e2, 2019.

[4] S. N. Bonnington and M. E. Rutter, "Surveillance of colonic polyps: Are we getting it right?" unpublished.

[5] K. Fitzgerald and B. Matuszewski, "FCB-SwinV2 transformer for polyp segmentation," arXiv preprint arXiv:2302.01027, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.01027

[6] J. Zhu, Y. Li, X. Chen, et al., "GCCSwin-UNet: Global Context and Cross-Shaped Windows Vision Transformer Network for Polyp Segmentation," Processes, vol. 11, no. 4, p. 1035, 2023.

[7] E. Sanderson and B. J. Matuszewski, "FCN-transformer feature fusion for polyp segmentation," in Proc. Medical Image Understanding and Analysis (MIUA), Cham, Switzerland: Springer, 2022, pp. 892–907.

[8] M. Rashed, M. Hossain, A. Mahdi, et al., "Hybrid Machine Learning Models for Accurate Type 2 Diabetes Mellitus Prediction Using a Stacking Classifier and a Meta-Model Approach," Cureus Journal of

Computer Science, vol. 2, Mar. 3, 2025, Art. no. e-s44389-025-03135-0, doi: https://doi.org/10.7759/s44389-025-03135-0

[9] D. Jha, N. K. Tomar, V. Sharma, and U. Bagci, "TransNetR: Transformer-based residual network for polyp segmentation with multicenter out-of-distribution testing," in Medical Imaging with Deep Learning (MIDL), 2024, pp. 1372–1384. [Online]. Available: https://doi.org/10.48550/arXiv.2303.07428

[10] N. K. Tomar, A. Shergill, B. Rieders, U. Bagci, and D. Jha, "TransResU-Net: Transformer based ResU-Net for real-time colonoscopy polyp segmentation," arXiv preprint arXiv:2206.08985, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2206.08985

[11] M. Zhang, Z. Yang, H. Chen, et al., "Vm-unet-v2 rethinking vision mamba unet for medical image segmentation," arXiv preprint arXiv:2403.09157, 2024.

[12] J. Jiang, L. Wang, X. Liu, et al., "LV-UNet: A Lightweight and Vanilla Model for Medical Image Segmentation," arXiv preprint arXiv:2408.16886, 2024.

[13] T. Erol and D. Sarikaya, "PlutoNet: An efficient polyp segmentation network with modified partial decoder and decoder consistency training," Healthcare Technology Letters, 2024. [Online]. Available: https://doi.org/10.1049/htl2.12105

[14] J. Bernal, F. J. Sanchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilarino, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," Comput. Med. Imaging Graph., vol. 43, pp. 99–111, 2015.

[15] K. Pogorelov, C. Randel, M. Lux, T. de Lange, D. Griwodz, P. Halvorsen, and M. A. Riegler, "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," in Proc. ACM Multimedia Systems Conf., Taipei, Taiwan, 2017, pp. 1–6.