Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

RA-UNet: A Deep Learning Approach for Precise Colorectal Polyp Segmentation

MD. Rakibul Hossain
Dept. of ICE
Pabna University of Science and Technology
Pabna-6600, Pabna, Bangladesh
Email: mdrakibulhossain7869@gmail.com

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

Sumayta Shama
Dept. of ICE
Pabna University of Science and Technology
Pabna-6600, Pabna, Bangladesh
Email: sumaytashama@gmail.com

Md. Rashed
Dept. of ICE
Pabna University of Science and Technology
Pabna-6600, Pabna, Bangladesh
Email: rashedulislam.ice.pust@gmail.com

Akif Mahdi
Dept. of ICE
Pabna University of Science and Technology
Pabna-6600, Pabna, Bangladesh
Email: akif2100@gmail.com

***Abstract*— Colorectal cancer remains a significant health concern, contributing to a high number of cancers diagnoses and mortality worldwide. Early detection through colonoscopy is critical for identifying and removing precancerous polyps before they progress to malignancy. However, the manual examination has inherent limitations, necessitating automated methods for improving operational effectiveness and finding accuracy. Here a framework using deep-learning (DL) is suggested for detecting and segmenting colorectal polyps, leveraging an enhanced UNet architecture integrated with attention gates and residual blocks. Residual blocks mitigate the challenges of training deep networks by enabling the straight flow of gradients over skip connections, preserving fine details, and capturing complex features. Attention gates amplify informative regions in feature maps, improving model focus on relevant structures while suppressing noise. These architectural enhancements allow for precise delineation of polyps in colonoscopy images. Our approach demonstrates unmatched performance on standard datasets, with a Dice coefficient of 85.2% and IoU of 79.2% on Kvasir-SEG and a Dice coefficient of 92.3% and IoU of 84.12% on CVC-ClinicDB. These results emphasize our approach's probability of improving colorectal cancer screening by enabling accurate, efficient, and early detection, ultimately contributing to better patient outcomes and reduced mortality rates.**

Keywords*— *Colorectal polyps, U-net, Residual block, Attention gate, Polyp segmentation.

I. INTRODUCTION

Cancer of the colorectal region is the most critical and deadly cancers globally and it stands as the third most common malignancy in prevalence. Timely identification of polyps is pivotal in lowering the risk of colorectal cancer and ensuring successful treatment outcomes [1]. Perhaps the most reliable techniques for finding polyps early are colonoscopy [2]. The structure, figure, and color of the polyps seen in colonoscopy pictures are frequently different. So, it is extremely difficult to recognize and differentiate the polyps in the colonoscopy pictures [3]. Accurate and timely recognition of colon polyps is essential for enhancing the effectiveness of this cancer treatment and improving treatment outcomes. Therefore, the detection and monitoring of polyp size play a crucial role in ensuring effective analysis and treatment [4]. Evolvement in machine learning, especially deep learning, has opened new areas of medical image analysis which also includes polyp detection and segmentation. Models such as UNet, ResUNet++, and other encoder-decoder architectures have successfully captured complex patterns and have shown promising results in this field [5]. Convolutional Neural Networks (CNNs) are employed to extract relevant features, capturing both local and global patterns in the data [6]. Many modern deep-learning architectures, playing a crucial role in the effective extraction of features essential for accurate model predictions[7]. Here, we propose colorectal polyp segmentation research using an RA-UNet version of Unet architecture that includes residual blocks and attention gates to improve feature learning and concentrate on relevant regions. The most important elements of our design are the use of residual blocks to facilitate gradient flow and feature representation, as well as attention gates to suppress irrelevant characteristics and sharpen the focus on polyps. Our model showed better segmentation accuracy and robustness when tried on the Kvasir-SEG dataset and CVC-clinicDB. Our approach shows notable improvements in accuracy. This method may reduce the workload on medical practitioners, help diagnose colorectal cancer early, and set the way for further study on automated polyp diagnosis.

This paper contributes in the following way:

1. The study presents an RA-UNet for improved feature extraction and noise reduction in colorectal polyp segmentation.
2. The proposed model achieves dice scores of 85.2% (Kvasir-SEG) and 92.3% (CVC-ClinicDB), outperforming standard UNet and other segmentation methods in medical image analysis.
3. The research highlights hybrid deep-learning models' potential to automate polyp detection, reducing diagnostic workloads and improving early colorectal cancer detection.

II. RELATED WORKS

Research on automated colorectal polyp segmentation focuses on improving early colorectal cancer (CRC) diagnosis and treatment by addressing challenges such as polyp morphology variation and noise in medical images. Convolutional Neural Networks (CNNs) are commonly used for medical image segmentation, with various UNet-based designs enhancing performance through feature aggregation and attention modules. Tashk, Herp, and Nadimi [8] improved UNet with an adaptive CNN architecture using symmetric skip connections for efficient segmentation. Zhang, Liu, and Wang [9] integrated residual units and skip connections within UNet's encoder-decoder structure for effective semantic segmentation. Safarov and Whangbo [10] introduced A-DenseUNet, enhancing UNet++ with DenseNet encoders, attention blocks, and restrictive skip connections. Nguyen et al. [11] proposed MED-Net, a multimodal deep encoder-decoder model capturing diverse image characteristics. Mahmud et al. [12] developed a model featuring deep reconstruction, fused skip modules, and depth-dilated inception blocks. Yeung et al. [13] presented Focus UNet, improving segmentation with double attention-gated deep networks, while Mohapatra et al. [14] introduced a U-shaped encoder-decoder with leaky ReLU, max pooling, and transposed convolutions. Ahamed et al. [15] combined inception and residual blocks with skip connections in an encoder-decoder framework. Despite advancements, challenges remain, including limited model generalizability across datasets, difficulty detecting small or flat polyps, and increased computational demands from attention mechanisms, hindering real-time clinical applications.

III. Proposed Methodology

This study proposes RA-UNet, an advanced DL model designed for colorectal polyp segmentation with architectural enhancements. RA-UNet incorporates Residual Blocks and Attention Gates into the UNet architecture and optimizes it for precise feature extraction and efficient noise suppression.

A. Dataset Description

This study utilized two freely offered datasets, Kvasir-SEG [16] and CVC-ClinicDB [17], to estimate the result of the proposed model. These datasets are widely recognized for benchmarking segmentation algorithms in the context of colorectal polyp detection. The Kvasir-SEG dataset, derived from the original Kvasir database, contains 1,000 images captured during colonoscopy, each accompanied by a corresponding segmentation mask. These images and masks are organized in separate folders and are identically named for easy matching. Additionally, the dataset includes a JSON file containing bounding box annotations for the images. The images in Kvasir-SEG are JPEG-compressed and vary in resolution which ranges from 332×487 to 1920×1072 pixels, reflecting the variability encountered in real-world colonoscopy scenarios.The CVC-ClinicDB dataset which is introduced by Bernal et al, is composed of video sequences specifically curated for polyp segmentation tasks. It includes 23 separate videos captured under white-light conditions, yielding 31 unique sequences, each featuring a distinct polyp. Frames with poor quality due to visual blurring or inadequate patient preparation were excluded, ensuring high data reliability. This resulted in a dataset of 612 pictures. Each one has resolution of 576×768 pixels. These frames provide a comprehensive representation of polyps with varying appearances, facilitating robust segmentation model evaluation. The dataset was split into training and testing sets with a 67% to 33% ratio to ensure a balanced evaluation. This division helps train the model effectively while reserving sufficient data for performance assessment.

B. Proposed Network Architecture

The proposed architecture in fig. 1 is a deep learning-based UNet model designed for precise polyp segmentation in colonoscopy images. It starts with an input layer that processes raw images, followed by an encoding path with 3×3 convolutional layers activated by ReLU. The model integrates residual blocks and attention gates to enhance performance in binary segmentation tasks. Residual blocks, with their skip connections, address the vanishing gradient

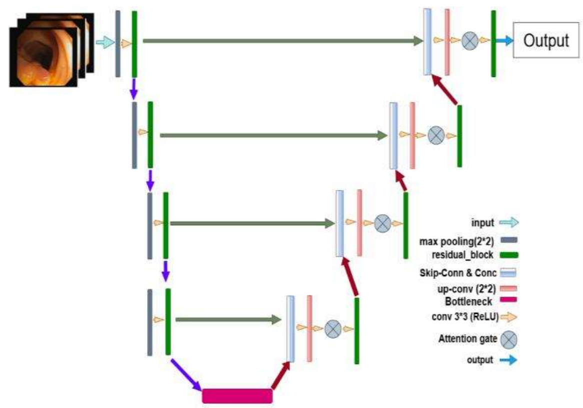

Fig. 1. Proposed model RA-UNet Architecture.

problem, enabling efficient learning in deeper networks, preserving critical low-level features, and supporting the learning of complex representations. This approach improves model robustness, reduces overfitting, and accelerates convergence during training. Attention gates are incorporated into the decoder to help the model focus on the most relevant regions of the input image. They dynamically highlight important features while suppressing irrelevant or noisy areas, which is crucial for accurately identifying small or subtle structures like polyps. This selective attention improves segmentation accuracy and reduces false positives in less critical regions. The combination of residual blocks and attention gates creates a synergistic effect: residual blocks ensure stable learning and feature retention, while attention gates enhance spatial focus and precision. Together, they improve both quantitative metrics (accuracy, Dice coefficient, IoU) and qualitative segmentation outcomes.

C. Implementation details

The implementation starts by loading image and mask data from specified directories, ensuring uniform resizing and normalization for compatibility with the model input requirements. A customized model structure version of the U Net, incorporating residual blocks for improved gradient flow and feature preservation throughout the network layers. These blocks consist of two convolutional layers with skip connections via identity mappings, aiding in the efficient learning of residual functions. Additionally, attention gates are integrated into the decoder path to selectively highlight informative features, enhancing that model's ability to concentrate on pertinent areas of the image during training and inference. The optimizer known as Adam is used to generate the model applying a cross-entropy loss function which is binary suited for Segmentation of binary tasks and includes the IoU coefficient as an evaluation metric to measure segmentation accuracy. The training uses a fraction of the data for validation and employs callbacks such as model checkpointing to preserve the top-performing weights and early halting to avoid overfitting. Valuation on a separate test set assesses the model's generalization ability, providing quantitative metrics on accuracy and IoU. For qualitative analysis, random images from the test set are chosen to generate predictions. These are visually compared alongside ground truth masks using Matplotlib, facilitating visual validation of polyp segmentation from colonoscopy images.

IV. RESULT AND DISCUSSION

Table 1 presents a comparative analysis of evaluation indexes for the traditional UNet, the ResNet, and our projected model RA-U Net for segmenting polyp automatically. The proposed RA-U Net outperformed the traditional UNet and ResUNet models in all metrics. Regarding the Dice Coefficient on the Kvasir-SEG dataset, UNet scored 82.1%, and ResUnet achieved 79.1%. At the same time, RA-UNet outperformed all by 85.1%. Using the CVC-ClinicDB dataset, Unet scored 85.1%, and ResUNet achieved 82.1%. Here also, our model got a high score of 92.24%. The dice score proves the ability to segment polyps with high overlap accuracy. IoU score using Kvasir-Seg dataset was 75.6% for Unet; for ResUnet, it was 73.1%. RA-U net achieved a high score of 79.2%. IoU score ensures better coverage for polyp regions. Table 1 highlights the enhanced functionality of the suggested model in both datasets with all the metrics.

TABLE I. PERFORMANCE EVALUATION OF THE PROPOSED RA-UNET MODEL USING KVASIR-SEG & CVC- CLINICDB DATASET.

Model	Dataset	Dice	IoU	Recall	Precision
Traditional UNet	Kvasir-SEG	0.821	0.756	0.702	0.851
	CVC-ClinicDB	0.851	0.789	0.724	0.882
ResUNet	Kvasir-SEG	0.791	0.731	0.776	0.801
	CVC-ClinicDB	0.821	0.762	0.791	0.824
RA-UNet	Kvasir-SEG	0.85	0.79	0.821	0.812
	CVC-ClinicDB	0.92	0.84	0.832	0.851

Fig 2 exemplifies the performance evaluation of three segmentation models—Traditional UNet, ResUNet, and RA-UNet—on two datasets: Kvasir-SEG and CVC-ClinicDB. The metrics evaluated include Dice coefficient, Intersection over Union (IoU), Recall, and Precision. The Traditional UNet proves reasonably strong performance, achieving a Dice coefficient of 0.821 and 0.851 on the Kvasir-SEG and CVC-ClinicDB datasets, respectively. However, the IoU and Recall values indicate room for improvement in capturing fine-grained details. ResUNet, with its incorporation of residual connections, shows competitive results, particularly in Recall (0.776 on Kvasir-SEG and 0.791 on CVC-ClinicDB), suggesting better sensitivity in identifying polyps. Nonetheless, its overall segmentation accuracy slightly lags compared to Traditional UNet on certain metrics. The proposed RA-UNet outperforms both baseline models across nearly all metrics. On Kvasir-SEG, it achieves a Dice coefficient of 0.85 and an IoU of 0.79, indicating improved segmentation precision. On CVC-ClinicDB, RA-UNet attains the highest Dice coefficient (0.92) and IoU (0.84), demonstrating its robust ability to delineate polyp boundaries effectively. Furthermore, its higher Precision values (0.812 on Kvasir-SEG and 0.851 on CVC-ClinicDB) confirm its reduced rate of false positives, making it a more reliable tool for accurate polyp segmentation. These results affirm that the enhancements in RA-UNet, particularly the integration of residual blocks and attention gates, significantly improve its capability to segment complex and diverse polyp structures in medical images.

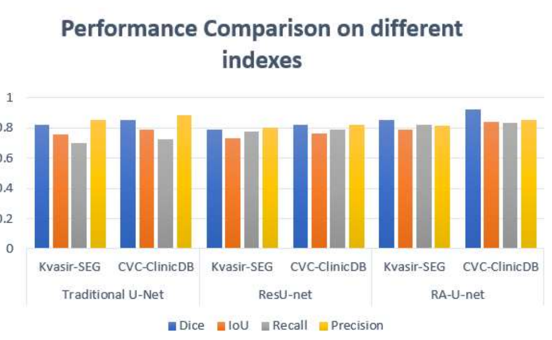

Fig. 2. Performance visualization of different models.

Table 2 summarizes the relative results of segmentation models tried on the Kvasir-SEG and CVC-ClicinsDB datasets. The ResUNet++ model incorporated with residual blocks, squeeze-and-excitation units, ASPP, and attention blocks, has achieved a DSC of 81.3%, while UNet++ alone scored 82.1%. UNet scored 81.8%, respectively. The projected RA-UNet model achieved the highest DSC of 85.2% on Kvasir-SEG and 92.3% on the CVC-ClinicDB Dataset, outperforming all other models. This improvement was possible by including residual blocks and attention gates. These metrics confirm that the proposed RA-UNet balances training optimization and generalization to unseen data. The RA-UNet model has achieved outstanding outcomes on both Kvasir-SEG and CVC-ClinicDB datasets. These outcomes prove that the model is accurate and efficient for detecting polyps. It is necessary for initial diagnosis and colorectal cancer therapy.

TABLE II. COMPARISON OF OUR PROPOSED APPROACH WITH EXISTING METHODS.

Ref.	**Dataset**	**Model**	**Dice coefficient**	**IoU**
Zhang et al. [9]	Kvasir-SEG	ResUnet	0.79	0.73
Jha et al. [18]	Kvasir-SEG	ResUNet++	0.81	0.79
	CVC-ClinicDB		0.79	0.79
Ronneberger et al. [19]	Kvasir-SEG	UNet	0.81	0.74
Zhou et al. [20]	Kvasir-SEG	UNet++	0.82	0.74
Tashk et al. [8]	CVC-ClinicDB	FCN + U Net	0.83	0.74
Fan et al. [21]	CVC-ClinicDB	Pra Net	0.89	0.84
Safarov and Whangbo [10]	CVC-ClinicDB	A-DenseU-Net	0.89	0.81
Our model	**Kvasir-SEG**	**RA-UNet**	**0.85**	**0.79**
	CVC-ClinicDB	**RA-UNet**	**0.92**	**0.84**

V. CONCLUSION

This study proposes an enhanced U-Net architecture, RA-UNet, for colorectal polyp segmentation in medical images. By integrating residual blocks and attention gates, the model improves feature extraction and focus, enabling accurate polyp delineation despite variations in size, shape, and appearance. Residual blocks enhance training efficiency by improving gradient flow, while attention gates prioritize regions of interest and suppress irrelevant features, making the model robust to noise. RA-UNet offers an effective approach for early detection and segmentation of polyps, with potential applications in other medical imaging tasks. Future work could explore integrating advanced techniques like transformers or self-supervised learning to improve generalization and computational efficiency for real-time inference in clinical settings.

REFERENCES

[1] M. Gupta and A. Mishra, "A systematic review of deep learning-based image segmentation to detect polyp," Artificial Intelligence Review, vol. 57, no. 1, 2024. Available: https://doi.org/10.1007/s10462-023-10621-1.

[2] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," Artificial Intelligence in Medicine, vol. 108, Aug. 2020.

[3] J. Bernal, N. Tajkbaksh, F. J. Sanchez, B. J. Matuszewski, H. Chen, L. Yu, et al., "Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge," IEEE Transactions on Medical Imaging, vol. 36, no. 6, pp. 1231–1249, Jun. 2017.

[4] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, et al., "A benchmark for endoluminal scene segmentation of colonoscopy images," Journal of Healthcare Engineering, vol. 2017, pp. 1–9, Jul. 2017.

[5] IEEE Xplore, "A Comprehensive Study on Colorectal Polyp Segmentation with ReSUNET++, Conditional Random Field and Test-Time Augmentation," IEEE Journals & Magazines, Jun. 2021. Available: https://ieeexplore.ieee.org/abstract/document/9314114.

[6] Md. Imran Hossain, Md. Mojahidul Islam, Tania Nahrin, Md. Rashed, and Md. Atiqur Rahman, "Improving Speech Emotion Recognition and Classification Accuracy Using Hybrid CNN-LSTM-KNN Model," Int. J. Res. Publ. Rev., vol. 5, no. 8, pp. 4164-4173, Aug. 2024.

[7] M. I. Hossain, S. Ahmad, S. Hasan, M. A. Rahim, and J. Shin, "Lung cancer detection and classification using LeNet-LSTM model on computed tomography images," Multidiscip. Sci. J., vol. 7, no. 4, p. 2025188, 2024. [Online]. Available: https://doi.org/10.31893/multiscience.2025188.

[8] A. Tashk, J. Herp, and E. Nadimi, "Fully automatic polyp detection based on a novel U-Net architecture and morphological post-process," in Proceedings of the 2019 3rd International Conference on Control, Artificial Intelligence, Robotics and Optimization (ICCAIRO), 2019, pp.37–41.Available: https://doi.org/10.1109/ICCAIRO47923.2019.00015.

[9] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749–753, 2018.

[10] S. Safarov and T. K. Whangbo, "A-DenseUNet: Adaptive Densely Connected UNet for Polyp Segmentation in Colonoscopy Images with Atrous Convolution," Sensors, vol. 21, no. 4, p. 1441, 2021. Available: https://doi.org/10.3390/s21041441.

[11] N. Q. Nguyen, D. M. Vo, and S. W. Lee, "Contour-aware polyp segmentation in colonoscopy images using detailed upsampling encoder–decoder networks," IEEE Access, vol. 8, pp. 99495–99508, 2020.

[12] T. Mahmud, B. Paul, and S. A. Fattah, "PolypSegNet: A modified encoder–decoder architecture for automated polyp segmentation from colonoscopy images," Computers in Biology and Medicine, vol. 128, p. 104119, 2021.

[13] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy," Computers in Biology and Medicine, vol. 137, p. 104815, Oct. 2021. Available: https://doi.org/10.1016/j.compbiomed.2021.104815.

[14] [14] S. Mohapatra et al., "UPOLYSEG: A U-Net-Based polyp segmentation network using colonoscopy images," Gastroenterology Insights, vol. 13, no. 3, pp. 264–274, 2022. Available: https://doi.org/10.3390/gastroent13030027.

[15] Md. F. Ahamed et al., "IRv2-Net: A Deep Learning Framework for Enhanced Polyp Segmentation Performance Integrating InceptionResNetV2 and UNet Architecture with Test Time Augmentation Techniques," Sensors, vol. 23, no. 18, p. 7724, 2023. Available: https://doi.org/10.3390/s23187724.

[16] Jha D, Smedsrud PH, Riegler MA, Halvorsen P, de Lange T, Johansen D, Johansen HD (2020) KvasirSEG: a segmented polyp dataset. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11962 LNCS, pp 451–462. https://doi.org/10.1007/978-3-030-37734-2_37

[17] Balraj Ashwath, "CVC-ClinicDB," Kaggle.com, 2015. https://www.kaggle.com/datasets/balraj98/cvcclinicdb/data (accessed Feb. 06, 2025).

[18] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM), 2019, pp. 225–2255.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA), 2018, pp. 3–11.

[21] D. P. Fan, G. P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel reverse attention network for polyp segmentation," in Lecture Notes in Computer Science (LNCS), vol. 12266, 2020, pp. 263–273. Available: https://doi.org/10.1007/978-3-030-59725-2_26.