Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 680-691 e-ISSN: 2656-8632
RESEARCH ARTICLE OPEN ACCESS

Dual Attention and Channel Atrous Spatial Pyramid Pooling Half-UNet for Polyp Segmentation

Beatrix Datu Sarira, and Heri Prasetyo

Department of Informatics, Faculty of Information Technology and Data Science, Universitas Sebelas Maret, Indonesia

Corresponding author: Heri Prasetyo (e-mail: heri.prasetyo@staff.uns.ac.id)

Abstract **Colorectal cancer (CRC) is a leading cause of cancer-related deaths, with two million cases detected in 2020 and causing one million deaths annually. Approximately 95% of CRC cases originate from colorectal adenomatous polyps. Early detection through accurate polyp segmentation is crucial for preventing and treating CRC effectively. While colonoscopy screening remains the primary detection method, its limitations have prompted the development of Computer-Aided Diagnostic (CAD) systems enhanced by deep learning models. This study proposes a novel neural network architecture called Dual Attention and Channel Atrous Spatial Pyramid Pooling Half-UNet (DACHalf-UNet) for medical polyp image segmentation that balances optimal performance with computational efficiency. The proposed model builds upon the U-Net framework by integrating Double Squeeze-and-Excitation (DSE) blocks in the encoder after the Ghost Module, Channel Atrous Spatial Pyramid Pooling (CASPP) in the bottleneck and decoder, and Attention Gate (AG) mechanisms within the architecture. DACHalf-UNet was trained and evaluated on the CVC-ClinicDB and Kvasir-SEG datasets for 70 epochs. Evaluations demonstrated superior performance with F1-Score and IoU values of 94.23% and 89.28% on CVC-ClinicDB, and 88.40% and 81.47% on Kvasir-SEG, respectively. Comparative analysis showed that DACHalf-UNet outperforms existing architectures including U-Net, U-Net++, ResU-Net, AGU-Net, CSAP-UNet, PRCNet, UNeXt, and UNeSt. Notably, the model achieves this performance with only 0.56 million trainable parameters and 30.29 GFLOPs, significantly reducing computational complexity compared to previous methods. These results demonstrate that DACHalf-UNet effectively addresses the need for accurate and efficient polyp segmentation, potentially enhancing CAD systems and contributing to improved CRC detection and treatment outcomes.**

Keywords **Colorectal Cancer; Polyp Segmentation; Deep Learning; Computational Efficiency; DACHalf-UNet**.

I. Introduction

Colorectal cancer (CRC) represents a significant threat to human health, with two million cases detected in 2020 and causing one million deaths annually [1], [2]. Studies have shown that around 95% of CRC cases are caused by colorectal adenomatous polyps that grow on the surface of the colon, rectum, stomach, and pharynx [3], [4]. Therefore, there is an urgent clinical need for the early detection of polyps to enable effective CRC prevention and treatment [5], [6].

Currently, colonoscopy screening is the most effective medical procedure to detect polyps and reduce the risk of CRC [7]. However, this method can only detect about 60% of CRC cases and is relatively time-consuming [8]. To address these limitations, Computer-Aided Diagnostic (CAD) technology is being utilized to assist healthcare professionals in detecting and analyzing CRC more effectively and efficiently [8]. Accurate, automatic, and efficient polyp segmentation can enhance the performance of CAD systems, thus encouraging the advancement of deep learning-driven models for automated polyp segmentation [8], [9].

The development of deep learning models, particularly Convolutional Neural Networks (CNN), has resulted in significant advancements in automated image segmentation techniques [10], [11]. CNN can improve polyp segmentation accuracy by integrating multi-scale image features to enhance segmentation performance [12], [13]. Although deep learning methods in previous research have produced good performance in segmentation tasks, few have considered important factors such as computational

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

complexity, number of parameters, and memory efficiency that are crucial in Point-of-Care (PoC) application scenarios [10].

One of the CNN architectures developed for image segmentation tasks is U-Net. The U-Net architecture consists of a downsampling encoder, upsampling decoder, and skip connections, which integrate local and global information during the encoding and decoding processes [14]. Several U-Net variants, such as UNet++ [15], have proposed using nested and dense skip connections to reduce the semantic gap between the encoder and decoder. The performance of UNet++ showed F1-Score and IoU results of 79.40% and 72.90% on the CVC-ClinicDB dataset and 82.10% and 74.30% on the Kvasir-Seg dataset, with a total of 9.04 million parameters. The study by Li et al. [16] proposed the UNeSt architecture for polyp segmentation, using U-Net as the backbone and integrating deeply separable convolutional layers and a multi-layer perceptron to reduce the number of parameters and computational complexity. UNeSt produced computational parameters of 0.92 million with F1-Score and IoU values on the CVC-ClinicDB dataset of 86.50% and 79.09%. The UNet++ and UNeSt architectures have proven capable of performing polyp segmentation effectively but still result in many computational parameters and complex network structures.

To address these issues, a more efficient Half-UNet architecture was proposed [17]. Half-UNet involves three main strategies to reduce network complexity: unifying the number of channels, full-scale feature fusion, and ghost modules. Evaluation of Half-UNet on the DDSM, LIDC-IDRI, and MICCAI 2009 datasets reduced the number of parameters by up to 98.6% without reducing medical image segmentation performance [17].

Research by Shu et al. [18] proposed a U-Net architecture integrated with Double Squeeze-and-Excitation (DSE) in the bottleneck layer to enhance feature extraction. The DSE structure first extracts channel kernels using Global Average Pooling (GAP), followed by Global Max Pooling (GMP) extraction in the second stage. Evaluation of the CVC-ClinicDB dataset showed that the U-Net architecture with DSE outperformed the UNet++ model, with F1-Score and IoU values of 91.5% and 86.4%. Oktay et al. [19] introduced the Attention Gate (AG) mechanism, which focuses on generating adaptive feature maps and identifying feature responses to retain only the most relevant features. The integration of AG within the U-Net architecture enhances the model's attention to target regions [19].

Xiong et al. [20] designed a segmentation architecture using the U-Net backbone with the addition of Channel Atrous Spatial Pyramid Pooling (CASPP). CASPP is a development of ASPP with the addition of the CBAM module [21] at the beginning of the module to pay more attention to important features. It effectively combines channel and spatial attention mechanisms while maintaining low computational overhead. This architecture improved the precision, F1-Score, and IoU values by 0.32%, 1.42%, and 1.27%, respectively, compared to the highest values from other architectures in performing crack segmentation tasks on the DeepCrack537 dataset.

Based on the related research above, this research focuses on developing a model that reduces computational parameters while improving the accuracy of polyp segmentation. The proposed architecture is the Dual Attention and Channel Atrous Spatial Pyramid Pooling Half-UNet (DACHalf-UNet), with the following main contributions: (1) Integration of Double Squeeze-and-Excitation (DSE) in the encoder after the Ghost Module; (2) Integration of Channel Atrous Spatial Pyramid Pooling (CASPP) in the bottleneck and decoder; and (3) Implementation of Attention Gate within the architecture. Integrating these methods in DACHalf-UNet is expected to enable accurate and efficient polyp medical image segmentation and can contribute to successful treatment outcomes.

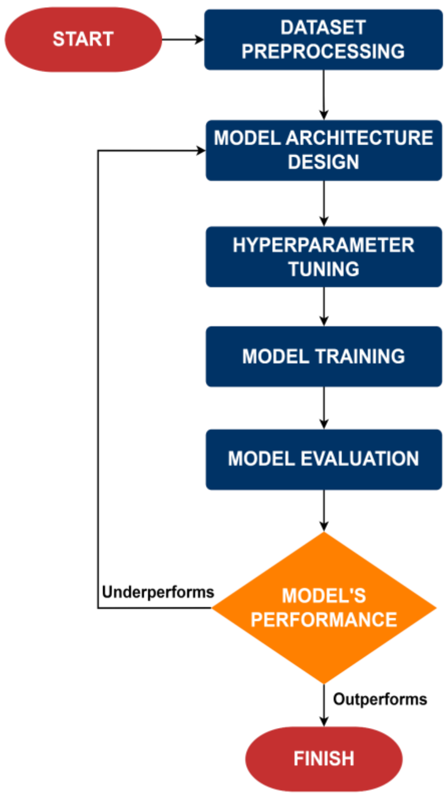


Fig. 1. **Stages of Research Methods**

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

II. Method

This research uses the methodology shown in Fig. 1. The research stages are divided into five main phases: (1) Dataset Preprocessing, (2) Model Architecture Design, (3) Hyperparameter Tuning, (4) Model Training. and (5) Model Evaluation.

A. Dataset Preprocessing

This research uses the CVC-ClinicDB [22] and Kvasir-SEG [23] datasets. The CVC-ClinicDB contains 612 polyp images and ground truth with a 288 x 368 pixels resolution. Kvasir-SEG contains 1000 polyp images and ground truth with varying resolutions ranging from 332 x 487 to 1920 x 1072 pixels. The first step was resizing the images to 256 x 256 pixels. The datasets were then divided into training, validation, and test sets with a ratio of 80%, 10%, and 10%. Data augmentation was subsequently applied to the training set to increase the number of data and prevent overfitting. The augmentation techniques included horizontal flip, vertical flip, and rotations of 90° and 180°. Examples of both original and augmented images are shown in Fig. 2.

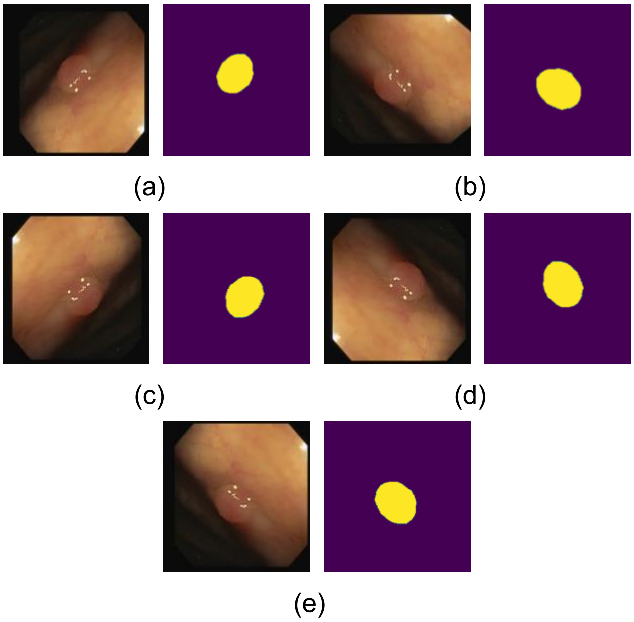


Fig. 2. Augmentation Results of (a) Original image, (b) 90 rotation, (c) 180 rotation, (d) Horizontal flip, and (e) Vertical flip.

Table 1. Number of Dataset Splits

Dataset	Train Set	Train Set (augmented)	Validation Set	Test Set
CVC-ClinicDB	490	2450	61	61
Kvasir-SEG	800	4000	100	100

Table 1 presents the distribution of the training, validation, and test sets used in this study. The augmentation process is applied only to the training set in order to increase the variability of the training data and prevent overfitting. As a result, the number of training images in the CVC-ClinicDB dataset increased to 2,450, and in the Kvasir-SEG dataset to 4,000 images.

B. Model Architecture Design

The overall architecture of DACHalf-UNet is shown in Fig. 3, which has five network depth levels with U-Net as the backbone. DACHalf-UNet implements the concepts of unified the number of filters at each level and full-scale feature fusion (FSFF) from Half-UNet. FSFF differs from traditional methods, such as the concatenation operations in U-Net and UNet3+, which have high computational complexity and memory usage [17]. In contrast, FSFF uses addition operations to maintain computational efficiency and reduce architectural complexity. In DACHalf-UNet, the number of filters for each block is the same.

This model uses ghost modules as in Half-UNet to reduce the complexity and number of parameters while maintaining model performance. DSE block is placed after ghost module to adaptively enhance important image features. At the bottleneck, the CASPP block is applied to improve detailed feature extraction by suppressing redundant information and refining edge features by combining channel and spatial attention. This model also uses attention gates (AG) to help improve its focus on target regions.

Ghost Module is a component within Convolutional Neural Network (CNN) architectures designed to reduce computational complexity and parameter count while maintaining model performance. Using convolutional layers to extract features from input images often results in redundant feature maps, increasing the number of parameters. These redundant features can lead to longer training times and excessive memory usage [24]. The number of parameters used in conventional convolution and the Ghost Module can be calculated using Eq. (1) and (2) [24], where K is the kernel size, C_{in} is the number of input channels, and C_{out} is the number of output channels.

$$params = \left(K^2 * C_{in} + 1\right) * C_{out} \quad (1)$$

$$params = \left(K^2 * (C_{in} + 1) + 2\right) * C_{out} / 2 \quad (2)$$

As shown in Fig. 4a, ghost module implements depth-wise and point-wise convolution, followed by batch normalization and an activation function. Point-wise convolution uses 1x1 convolution to combine information between channels. Meanwhile, depth-wise convolution uses 3×3 convolution to extract important

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

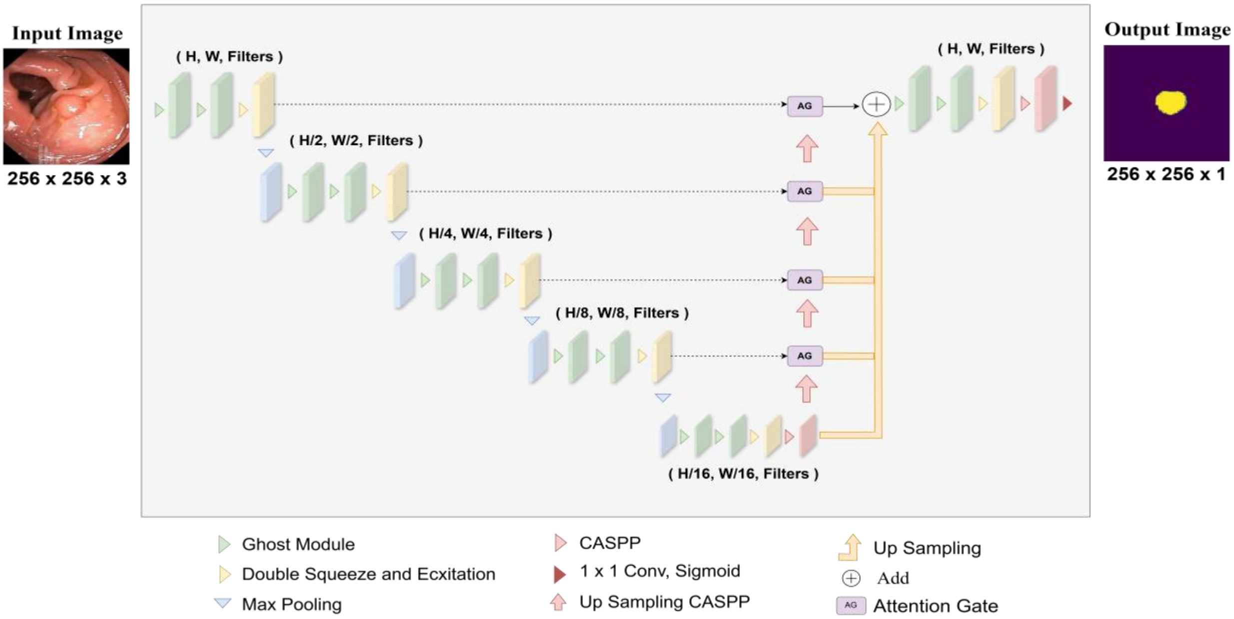
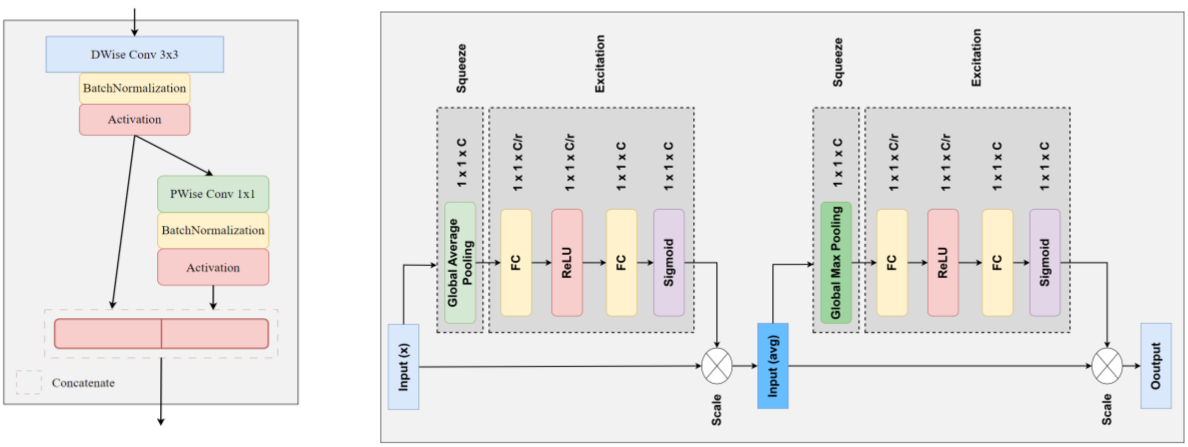


Fig. 3. Architecture of DACHalf-UNet

spatial features from the input, thereby reducing the number of parameters required by the model.

Fig. 4(b) illustrates the Double Squeeze-and-Excitation (DSE). DSE In this research consists of two stages. The first stage transforms the input into a weight vector using global average pooling (GAP). Then, the convolution block functions perform the non-linear transformation. The output of this block is normalized using ReLU and sigmoid functions to generate a weight vector. At the end of the first stage, the model will produce a weight vector representing important information from the input used in the second stage. In the second stage, the output from the first stage becomes the input for the second stage. The input in the second stage is processed using global max pooling (GMP). GMP aims to identify the maximum value across all elements of the input. This maximum value is processed in the convolution block. The output from the convolution block is normalized using the sigmoid function to constrain the output values between 0 and 1. This second stage will produce enhanced output by combining information from previously obtained weight vectors [19].

Channel Atrous Spatial Pyramid Pooling (CASPP) is a block arrangement consisting of a convolutional block attention module (CBAM) and atrous spatial pyramid pooling (ASPP). In the CASPP structure, the CBAM block is placed at the beginning to emphasize important features in the image. There are two types of attention in CBAM: channel attention and spatial attention. As illustrated in Fig. 5(a), channel attention uses average pooling and max pooling to capture

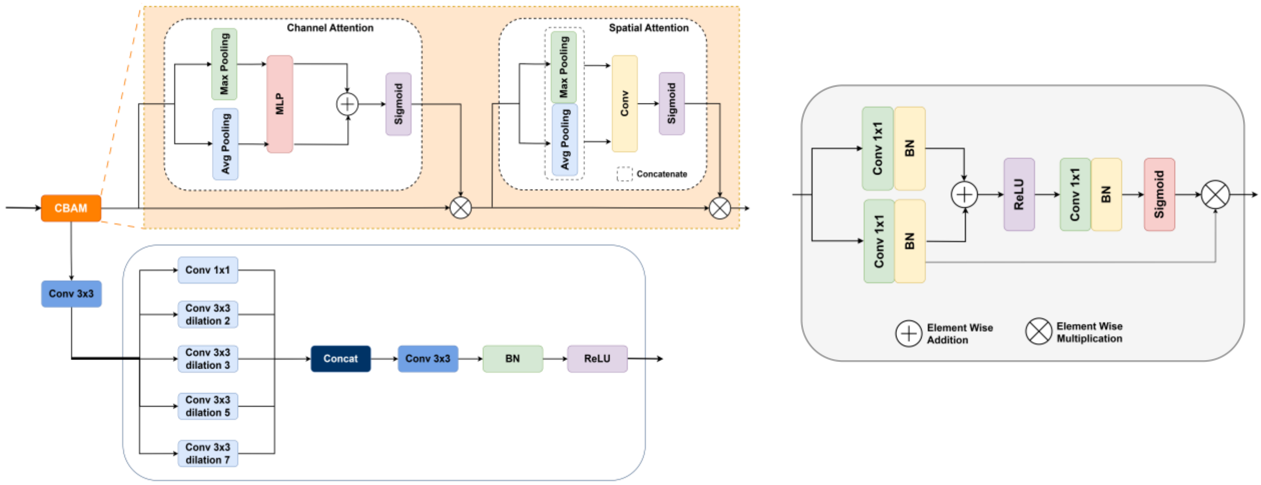


(a) (b)

Fig. 4. Architecture of (a) Ghost Module and (b) DSE

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 680-691 e-ISSN: 2656-8632



(a) (b)

Fig. 5. Architecture of (a) CASPP and (b) Attention Gate

important information from feature maps. The output of the channel attention is then used as the input for the spatial attention, which focuses on highlighting the most relevant areas within the feature maps. The features extracted by CBAM are subsequently passed into a 3×3 convolution block. These features are forwarded to a dilated convolution block with 2, 3, 5, and 7 dilation rates. The outputs from the dilated convolution are concatenated and further processed using a 3x3 convolution block, batch normalization, and activation function.

The Attention gate (AG) is applied to features transmitted through skip connections in the encoder stage to eliminate noise and irrelevant responses. The outputs from DSE and CASPP will become inputs to the AG. As shown in Fig. 5(b), the input passes through a 1x1 convolution block and batch normalization. The resulting output is then passed through an activation function, followed by another 1×1 convolution block and batch normalization, and finally processed with a sigmoid activation function.

In its implementation, the feature x_l and the gating signal g are convolved using a linear transformation and converted into attention coefficients using a sigmoid function. This transformation process is formulated in Eq. (3) [19]. Subsequently, the q_{att} value is converted into the attention coefficient a_l through the sigmoid activation function in Eq. (4) [19]. The attention coefficient is then used to adjust the input feature by performing element-wise multiplication, as described in Eq. (5) [19].

$$q_{att} = \psi^T (\sigma_1(W_x x_l + W_g g + b_g)) + b_\psi \quad (3)$$

$$\alpha_l = \sigma_2(q_{att}) \quad (4)$$

$$\widehat{x_l} = x_l . \alpha_l \quad (5)$$

where q_{att} is the attention score used to determine feature relevance, ψ is the parameter used in linear transformation, σ is the activation function used, W_x is the weight matrix on the input feature x_l, W_g is the weight matrix on the gating signal, b_g is the bias value in the linear operation on the gating signal, and b_ψ is the bias value after transformation with ψ.

C. Hyperparameter Tuning

The next step is adjusting hyperparameter tuning to obtain optimal results and efficient parameters. Selecting appropriate hyperparameters for the model can directly impact model performance. As tabulated in Table 2, the initial hyperparameters used in this paper include a batch size of 16, 64 filters, [1]s, Adam optimizer, initial learning rate of 0.001 (divided by ten if the validation loss does not decrease for 10 epochs), ReLU activation function, attention ratio of 8, BCE loss function, and network depth of 5.

Table 2. Initial Hyperparameters

Hyperparameter	Value
Batch Size	16
Filters	64
Epoch	70
Optimizer	Adam
Learning Rate	0.001 (divided by ten if the validation loss does not decrease for 10 epochs)
Activation Function	ReLU
Ratio Attention	8
Loss Function	Binary Cross-Entropy
Depth	5

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

D. Model Training

After obtaining the optimal hyperparameters, the training process is carried out on the proposed model. The use of these parameters is expected to have a significant impact on the model's performance. The training is conducted using the training set of the selected dataset. The input consists of RGB medical images, and the output is a segmented binary image. Weights were initialized using the He normal method, which is suitable for ReLU activations and helps improve training stability. Model training is performed on Google Colaboratory using the A100 GPU accelerator and TensorFlow framework.

E. Model Evaluation

Evaluation metrics are used to assess a model's performance. Common metrics for evaluating polyp segmentation results include Intersection over Union (IoU) and F1-Score. The formulas for calculating IoU and F1-Score are presented in Eq. (6) and (7) [25].

$$F1-Score = \frac{2TP}{2TP+FP+FN} \quad (6)$$

$$IoU = \frac{TP}{TP+FP+FN} \quad (7)$$

TP (true positive) represents positive data that is correctly predicted, TN (true negative) represents negative data that is correctly predicted, FP (false positive) is negative data that is predicted as positive data, and FN (false negative) is positive data that is predicted as negative data.

III. Result

A. Hyperparameter Tuning

The hyperparameter tuning process in this study was conducted by testing various types of optimizers, activation functions, attention ratios, loss functions, network depths, and filter sizes. For time efficiency, the hyperparameter tuning test in this research only used the CVC-ClinicDB dataset, and it was performed sequentially so that the initial hyperparameter values in Table 2 would change during the testing process.

Optimizer, activation function, attention ratio, loss function, network depth, and number of filters were systematically tested to determine the best configurations for the proposed model, with all results are summarized in Table 3 and Fig. 6 displays the graphs of the hyperparameter tuning results. The optimizer evaluation used Adam and RMSprop. Based on the results, Adam optimizer achieved the highest F1-Score of 93.95% and an IoU of 89.00%, outperforming RMSprop. As for activation functions, GELU achieved the highest F1-Score of 94.23% and IoU of 89.28%. Attention ratio experiments showed that a ratio of 8 yielded the best results compared to 4 and 16. In terms of loss functions, BCE outperformed Dice Loss and Tversky Loss, achieving the highest F1-Score

Table 3. Hyperparameter Tuning Test Results

Hyperparameter		F1-Score	IoU	Trainable Parameter
Optimizer	**Adam**	**93.95%**	**89.00%**	**568,387**
	RMSProp	92.99%	87.51%	568,387
Activation Function	ReLU	93.95%	89%	568,387
	GELU	**94.23%**	**89.28%**	**568,387**
	SELU	92.08%	86.13%	568,387
Ratio Attention	4	93.37%	88.08%	582,835
	8	**94.23%**	**89.28%**	**568,387**
	16	92.08%	86.13%	561,163
Loss Function	Dile Loss	92.08%	86.13%	568,387
	BCE	**94.23%**	**89.28%**	**568,387**
	Treversky Loss	93.46%	88.58%	568,387
Depth	3	92.32%	86.46%	467,229
	4	93.17%	87.73%	517,808
	5	**94.23%**	**89.28%**	**568,387**
Filter	16	92.32%	86.46%	65,887
	32	93.17%	87.73%	256,763
	64	**94.23%**	**89.28%**	**568,387**

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

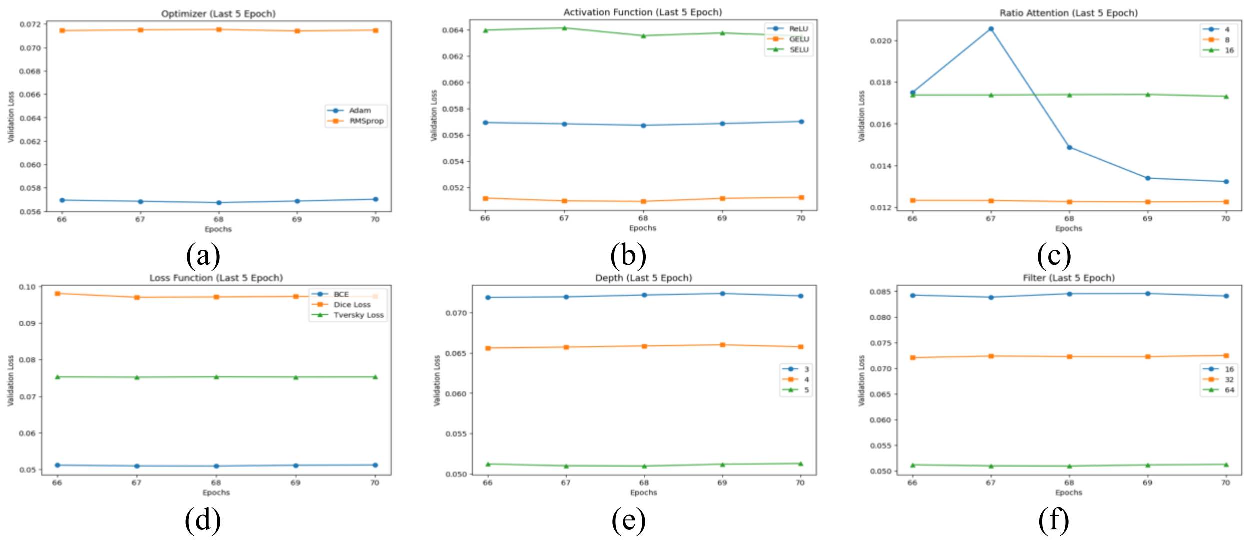


Fig. 6. Hyperparameter Tuning Graph Results: (a) Optimizer, (b) Activation Function, (c) Ratio Attention, (d) Loss Function, (e) Depth, and (f) Filter

and IoU. In terms of network depth, depth of 5 achieved the highest F1-Score and IoU values. Furthermore, the results demonstrate that increasing the network depth leads to a corresponding increase in the number of trainable parameters, rising from 467,229 at depth 3 to 568,387 at depth 5, thereby highlighting a trade-off between model complexity and performance.The number of filters 64 achieved the highest F1-Score and IoU values. Therefore, filter 64 was selected for use in subsequent experiments.

B. Model Ablation

To comprehensively evaluate the effectiveness of each block in the proposed method, this study conducted several ablation experiments on two public polyp segmentation datasets. The ablation experiments conducted in this study are as follows: (1) Ablation 1: Half-UNet, (2) Ablation 2: DACHalf-UNet without AG, (3) Ablation 3: DACHalf-UNet without DSE, (4) Ablation 4: DACHalf-UNet without CASPP, and (5) Ablation 5: DACHalf-UNet (proposed method).

As tabulated in Table 4, the proposed DACHalf-UNet yields the best performance over the CVC dataset, achieving 94.23% and 89.28% for the F1 score and IoU, respectively. The Kvasir-SEG dataset achieved an F1-Score of 88.40% and IoU of 81.47%. The total number of trainable parameters in the proposed model is 568,387. Specifically, the introduction of CASPP block in the ablation experiments led to a gradual increase in the number of trainable parameters compared to other configurations. However, it provided the most significant improvement in segmentation performance among the ablation experiment. As shown in Fig. 7, the proposed method segments polyp regions more completely than other ablation experiments. The first input image in the left column for example, although other methods can also segment the target area, the proposed method provides a much more refined segmentation result. The segmentation results of Half-UNet in Fig. 7b are incomplete, however our proposed method can obtain the segmentation result that is much closer to the ground truth.

Table 4. Ablation Experiment Results on CVC-ClinicDB and Kvasir-SEG

Method	CVC-ClinicDB		Kvasir-SEG		Trainable Parameter
	F1-Score	IoU	F1-Score	IoU	
Ablation 1	87.08%	77.12%	84.55%	73.23%	**0.22 M**
Ablation 2	93.82%	89%	86.01%	78.67%	0.53 M
Ablation 3	93.47%	88.09%	87.42%	80.22%	0.55 M
Ablation 4	92.97%	88%	86.05%	78.08%	0.26 M
Ablation 5	**94.23%**	**89.28%**	**88.40%**	**81.47%**	0.56 M

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

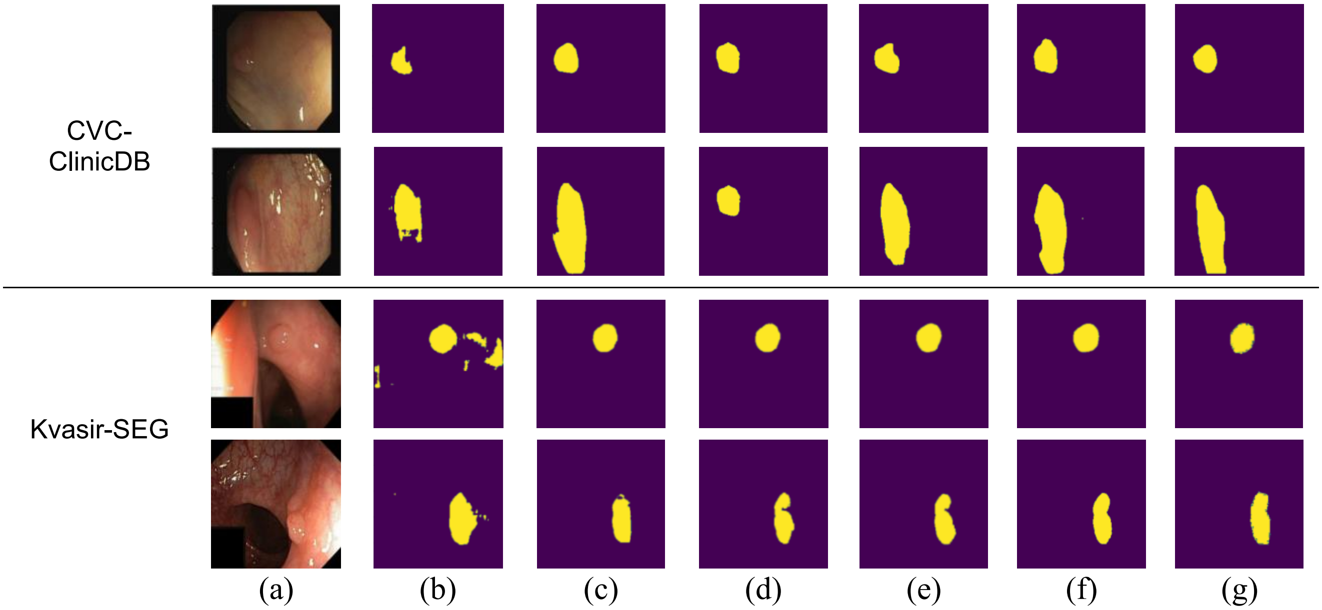


Fig. 7. Qualitative Comparison of Model Ablation (a) Input Image, (b) Ablation 1, (c) Ablation 2, (d) Ablation 3, (e) Ablation 4, (f) Ablation 5, and (g) Ground Truth

C. Experimental Results

The hyperparameter tuning and model ablation processes produced the optimal parameters and the best-performing model. The optimal hyperparameters were then used to train the model on the CVC-ClinicDB and Kvasir-SEG datasets. Table 5 presents the optimal hyperparameters values during the tuning process.

Table 5. Optimal Hyperparameters

Hyperparameter	Value
Batch Size	16
Filters	64
Epoch	70
Optimizer	Adam
Learning Rate	0.001 (divided by ten if the validation loss does not decrease for 10 epochs)
Activation Function	GELU
Ratio Attention	8
Loss Function	Binary Cross-Entropy
Depth	5

Table 6. Experiment Results

Dataset	F1-Score	IoU	Trainable Parameter
CVC-ClinicDB	94.23%	89.28%	568,387
Kvasir-SEG	88.40%	81.47%	

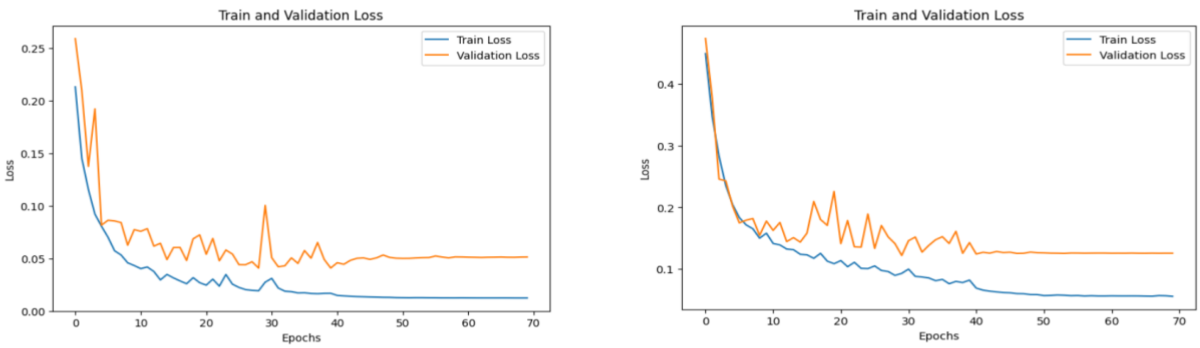
After the training process, the next step is to evaluate the model using the testing set from both datasets to assess the model's performance in segmentation tasks. Table 6 presents the F1-Score and IoU values obtained during the evaluation process, along with comparisons between the ground truth images and the predicted images in this study, as shown in Table 7 for the CVC-ClinicDB dataset and the Kvasir-SEG dataset.

Training and validation losses were monitored to assess model convergence and detect overfitting. Training required 17 seconds per epoch for CVC-ClinicDB and 31 seconds per epoch for Kvasir-SEG. As shown in Fig. 8a and 8b, both training and validation losses stabilized after the 40th epoch for both datasets. Final training and validation losses were 0.012 and 0.051 for CVC-ClinicDB, and 0.055 and 0.125 for Kvasir-SEG, respectively.

IV. Discussion

This study introduces DACHalf-UNet, a novel deep learning polyp segmentation architecture that balances high segmentation accuracy and computational efficiency. The performance of DACHalf-UNet will be compared with U-Net [26], U-Net++ [15], ResU-Net [27], AGU-Net [28], CSAP-UNet [29], PRCNet [10], UNeXt [30], and UNeSt [16]. The results presented in Table 8 demonstrate the model's effectiveness on the CVC-ClinicDB and Kvasir-SEG datasets. The DACHalf-UNet achieved an F1-Score of 94.23% and IoU of 89.28% on CVC-ClinicDB, and an F1-Score of 88.40% and IoU of 81.47% on Kvasir-SEG, demonstrating high accuracy in polyp region

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893



(a) (b)

Fig. 8. Training and Validation Loss on: (a) CVC-ClinicDB and (b) Kvasir-SEG

Table 7. Experiment Results on CVC-ClinicDB and Kvasir-SEG

Dataset	Input Image	Ground Truth	Predicted Image	F1-Score/IoU (%)
CVC-ClinicDB				98.12/96.31
				98.26/96.58
				97.95/95.98
Kvasir-SEG				98.44/96.93
				98.10/96.27
				97.77/95.64

segmentation. These results indicate that DACHalf-UNet is capable of accurately segmenting polyps, even in challenging conditions such as low contrast or irregular shapes.

The enhanced segmentation performance demonstrates substantial clinical relevance. Improved IoU values facilitate precise polyp boundary identification, which is essential for accurate size assessment and complete endoscopic removal, consequently reducing recurrence rates. Enhanced F1-scores, achieved through better recall, enable identification of small or inconspicuous polyps frequently overlooked in clinical practice. Furthermore, increased precision minimizes false-positive results, reducing unnecessary follow-up examinations and clinical ambiguity. These advancements collectively strengthen diagnostic reliability in colorectal cancer screening, support informed surgical decision-making, and contribute to improved patient care outcomes.

The architectural components of DACHalf-UNet significantly contributed to its performance. The DSE block, placed after the Ghost Modules in the encoder, enhanced feature recalibration by emphasizing important features. The CASPP module, used in the

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

Table 8. Quantitative Comparison with Previous Methods

Method	CVC-ClinicDB		Kvasir-SEG		Trainable Parameter
	F1-Score	IoU	F1-Score	IoU	
U-Net	85.53%	77.26%	82.69%	73.24%	34.53 M
U-Net++	86.19%	78.19%	86.05%	77.01%	9.16 M
ResUNet-a	81.33%	70.66%	78.59%	67.09%	13.17 M
AGU-Net	92.13%	86.50%	85.55%	78.87%	1.17 M
CSAP-UNet	88.61%	81.54%	84.28%	76.53%	139.70 M
PRCNet	92.50%	86.90%	79.90%	71.60%	31.17 M
UNeXt	82.81%	73.85%	81.58%	71.53%	1.47 M
UNeSt	86.50%	79.09%	82.69%	72.85%	0.92 M
DACHalf-UNet (Proposed Method)	**94.23%**	**89.28%**	**88.40%**	**81.47%**	**0.56 M**

bottleneck and decoder, improved multi-scale feature extraction and edge refinement through combined channel and spatial attention. Attention Gates in the skip connections further refined segmentation by focusing on relevant polyp regions and suppressing irrelevant features.

Ablation studies provided valuable insights into the contributions of these individual components. The proposed method consistently outperformed other configurations, highlighting the synergistic effect of integrating DSE, CASPP, and attention gate. For instance, the comparison between Half-UNet and DACHalf-UNet shows substantial improvements in F1-Score and IoU on both datasets. The qualitative results in Fig. 7 further illustrate that the proposed method produces more complete and refined segmentation masks compared to the ablated versions, aligning more closely with the ground truth.

As shown in Table 8, DACHalf-UNet outperforms several state-of-the-art polyp segmentation methods on both CVC-ClinicDB and Kvasir-SEG datasets, achieving a strong balance of segmentation accuracy and efficiency with only 0.56 million parameters and 30.29 GFLOPs.

Despite its promising results, this study has several limitations. The evaluation was conducted on only two public datasets, which may not fully capture the diversity of polyp appearances, sizes, and imaging conditions encountered in real-world clinical settings. Additionally, the use of fixed train, validation, and test splits may not adequately reflect performance variability across different data distributions. The model’s reliance on 2D image segmentation may also limit its ability to capture spatial context inherent in 3D or video-based data.

To address these limitations, future research should focus on validating the DACHalf-UNet model across multiple clinical centers and varied imaging systems to ensure its reliability in diverse clinical scenarios. Prospective clinical trials are necessary to assess its real-world diagnostic value and integration potential. Expanding the application of DACHalf-UNet to other medical imaging tasks could further demonstrate its versatility.

V. Conclusion

This study proposes a novel neural network model for polyp medical image segmentation with optimal performance and improved efficiency. The proposed method, Dual Attention and Channel Atrous Spatial Pyramid Pooling Half-UNet (DACHalf-UNet) builds upon the U-Net architecture by integrating DSE block, Attention Gate, CASPP, and ghost module. The integration of these components has been shown to enhance model performance while maintaining low computational cost. The DACHalf-UNet model was trained and evaluated on the CVC-ClinicDB and Kvasir-SEG datasets over 70 epochs. Evaluation on the CVC-ClinicDB dataset yielded an F1-score of 94.23% and IoU of 89.28%, while the Kvasir-SEG dataset achieved an F1-score of 88.40% and IoU of 81.47%. DACHalf-UNet outperformed U-Net, U-Net++, ResU-Net, AGU-Net, CSAP-UNet, PRCNet, UNeXt, and UNeSt in polyp segmentation, achieving higher F1-Score and IoU metrics. The proposed method achieved computational efficiency with only 0.56 million trainable parameters and 30.29 GFLOPs, demonstrating that DACHalf-UNet delivers optimal segmentation performance while maintaining low computational complexity. For future work, DACHalf-UNet can be extended to other medical image segmentation tasks such as liver, brain, or lung segmentation, explored for real-time deployment in clinical settings, optimized through model compression techniques, and adapted for 3D volumetric image analysis.

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

Acknowledgment

This work was fully supported and funded by the RKAT Universitas Sebelas Maret (UNS) of the year 2025 under the research grant PENELITIAN PENGUATAN KAPASITAS GRUP RISET (PKGR-UNS) A with the contract 371/UN27.22/PT.01.03/2025.

References

[1] L. Lu, S. Chen, H. Tang, X. Zhang, and X. Hu, "A multi-scale perceptual polyp segmentation network based on boundary guidance," *Image Vis. Comput.*, vol. 138, p. 104811, Oct. 2023, doi: 10.1016/j.imavis.2023.104811.

[2] H. Xia, Y. Qin, Y. Tan, and S. Song, "BA-Net: Brightness prior guided attention network for colonic polyp segmentation," *Biocybern. Biomed. Eng.*, vol. 43, no. 3, pp. 603–615, Jul. 2023, doi: 10.1016/j.bbe.2023.08.001.

[3] S. Ghosh, A. Bandyopadhyay, S. Sahay, R. Ghosh, I. Kundu, and K. C. Santosh, "Colorectal Histology Tumor Detection Using Ensemble Deep Neural Network," *Eng. Appl. Artif. Intell.*, vol. 100, p. 104202, Apr. 2021, doi: 10.1016/j.engappai.2021.104202.

[4] C. Wang, R. Xu, S. Xu, W. Meng, and X. Zhang, "Automatic polyp segmentation via image-level and surrounding-level context fusion deep neural network," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106168, Aug. 2023, doi: 10.1016/j.engappai.2023.106168.

[5] L. Yang, C. Zhai, Y. Liu, and H. Yu, "CFHA-Net: A polyp segmentation method with cross-scale fusion strategy and hybrid attention," *Comput. Biol. Med.*, vol. 164, p. 107301, Sep. 2023, doi: 10.1016/j.compbiomed.2023.107301.

[6] G. Liu *et al.*, "CAFE-Net: Cross-Attention and Feature Exploration Network for polyp segmentation," *Expert Syst. Appl.*, vol. 238, p. 121754, Mar. 2024, doi: 10.1016/j.eswa.2023.121754.

[7] L. Meng, Y. Li, and W. Duan, "Three-stage polyp segmentation network based on reverse attention feature purification with Pyramid Vision Transformer," *Comput. Biol. Med.*, vol. 179, p. 108930, Sep. 2024, doi: 10.1016/j.compbiomed.2024.108930.

[8] W. Zhang, F. Lu, H. Su, and Y. Hu, "Dual-branch multi-information aggregation network with transformer and convolution for polyp segmentation," *Comput. Biol. Med.*, vol. 168, p. 107760, Jan. 2024, doi: 10.1016/j.compbiomed.2023.107760.

[9] H. Prasetyo, M. A. F. Rohman, A. W. H. Prayuda, and J.-M. Guo, "Enhancing Polyp Segmentation Efficiency Using Pixel Channel Attention HalfU-Net," in *2024 IEEE 10th Information Technology International Seminar (ITIS)*, Surabaya, Indonesia: IEEE, Nov. 2024, pp. 381–386. doi: 10.1109/itis64716.2024.10845590.

[10] J. Li *et al.*, "PRCNet: A parallel reverse convolutional attention network for colorectal polyp segmentation," *Biomed. Signal Process. Control*, vol. 95, p. 106336, Sep. 2024, doi: 10.1016/j.bspc.2024.106336.

[11] G. Yue, W. Han, S. Li, T. Zhou, J. Lv, and T. Wang, "Automated polyp segmentation in colonoscopy images via deep network with lesion-aware feature selection and refinement," *Biomed. Signal Process. Control*, vol. 78, p. 103846, Sep. 2022, doi: 10.1016/j.bspc.2022.103846.

[12] Y. Liu, Y. Yang, Y. Jiang, and Z. Xie, "Multi-view orientational attention network combining point-based affinity for polyp segmentation," *Expert Syst. Appl.*, vol. 249, p. 123663, Sep. 2024, doi: 10.1016/j.eswa.2024.123663.

[13] D. Shao, H. Yang, C. Liu, and L. Ma, "AFANet: Adaptive feature aggregation for polyp segmentation," *Med. Eng. Phys.*, vol. 125, p. 104118, Mar. 2024, doi: 10.1016/j.medengphy.2024.104118.

[14] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy: IEEE, Jan. 2021, pp. 1236–1242. doi: 10.1109/ICPR48806.2021.9413346.

[15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609.

[16] J. Li, P. Ding, F. Lin, Z. Chen, A. A. Heidari, and H. Chen, "UNeSt: A fast segmentation network for colorectal polyps based on MLP and deep separable convolution," *Biomed. Signal Process. Control*, vol. 100, p. 107165, Feb. 2025, doi: 10.1016/j.bspc.2024.107165.

[17] H. Lu, Y. She, J. Tie, and S. Xu, "Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation," *Front. Neuroinformatics*, vol. 16, p. 911679, Jun. 2022, doi: 10.3389/fninf.2022.911679.

[18] X. Shu, J. Wang, A. Zhang, J. Shi, and X.-J. Wu, "CSCA U-Net: A channel and space compound attention CNN for medical image segmentation," *Artif. Intell. Med.*, vol. 150, p. 102800, Apr. 2024, doi: 10.1016/j.artmed.2024.102800.

[19] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," 2018, *arXiv*. doi: 10.48550/ARXIV.1804.03999.

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

[20] B. Xiong *et al.*, "FCT-Net: A dual-encoding-path network fusing atrous spatial pyramid pooling and transformer for pavement crack detection," *Eng. Appl. Artif. Intell.*, vol. 137, p. 109190, Nov. 2024, doi: 10.1016/j.engappai.2024.109190.

[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," 2018, *arXiv*. doi: 10.48550/ARXIV.1807.06521.

[22] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imaging Graph.*, vol. 43, pp. 99–111, Jul. 2015, doi: 10.1016/j.compmedimag.2015.02.007.

[23] D. Jha *et al.*, "Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning," *IEEE Access*, vol. 9, pp. 40496–40510, 2021, doi: 10.1109/ACCESS.2021.3063716.

[24] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features from Cheap Operations," Mar. 13, 2020, *arXiv*: arXiv:1911.11907. doi: 10.48550/arXiv.1911.11907.

[25] H. Al Jowair, M. Alsulaiman, and G. Muhammad, "Multi parallel U-net encoder network for effective polyp image segmentation," *Image Vis. Comput.*, vol. 137, p. 104767, Sep. 2023, doi: 10.1016/j.imavis.2023.104767.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[27] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020, doi: 10.1016/j.isprsjprs.2020.01.013.

[28] M. A. F. Rohman, H. M. Akbar, A. D. A. Firdaus, and H. Prasetyo, "AGU-NET: AttentionGhostU-NetUntuk Segmentasi Penyakit Polip Berbasis Citra Biomedis," vol. 1, 2023.

[29] X. Fan, J. Zhou, X. Jiang, M. Xin, and L. Hou, "CSAP-UNet: Convolution and self-attention paralleling network for medical image segmentation with edge enhancement," *Comput. Biol. Med.*, vol. 172, p. 108265, Apr. 2024, doi: 10.1016/j.compbiomed.2024.108265.

[30] J. M. J. Valanarasu and V. M. Patel, "UNeXt: MLP-Based Rapid Medical Image Segmentation Network," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, vol. 13435, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., in Lecture Notes in Computer Science, vol. 13435. , Cham: Springer Nature Switzerland, 2022, pp. 23–33. doi: 10.1007/978-3-031-16443-9_3.

Author Biography

Beatrix Datu Sarira is a final-year undergraduate student at Universitas Sebelas Maret (UNS), pursuing a degree in Informatics under the Faculty of Information Technology and Data Science. With a keen interest in data technologies and artificial intelligence, her academic journey has been deeply rooted in research and practical applications in the field of computer science. She is enthusiastic about applying machine learning and deep learning solutions to real-world problems. Beyond her academic endeavors, she actively seeks opportunities for collaboration, professional growth, and impactful research.

Heri Prasetyo received the doctoral degree from the Department of Electrical Engineering, National Taiwan University of Science and Technology (NTUST), Taiwan, in 2015. He received the Best Dissertation Award from the Taiwan Association for Consumer Electronics (TACE) in 2015, the Best Paper Awards from the International Symposium on Electronics and Smart Devices 2017 (ISESD 2017), ISESD 2019, the International Conference on Science in Information Technology (ICSITech 2019), the International Conference on Smart Technology, Applied Informatics, and Engineering (APICS 2022), the International Conference on Informatics and Computing (ICIC 2023), International Conference on Computer, Control, Informatics and its Applications (IC3INA 2024), and the Outstanding Faculty Award from Universitas Sebelas Maret (UNS) in 2019 and 2023. His research interests include multimedia signal processing, computational intelligence, pattern recognition, and machine learning. He can be contacted at email: heri.prasetyo@staff.uns.ac.id.

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893

Manuscript received March 8, 2025; Revised May 12, 2025; Accepted May 20, 2025; date of publication May 30, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.893