CLFSeg: A Fuzzy-Logic based Solution for Boundary Clarity and Uncertainty Reduction in Medical Image Segmentation

Anshul Kaushal[*, †, 1]
kaushalanshul.itz@gmail.com
Kunal Jangid[*,2]
kunal24@iiserb.ac.in
Vinod K. Kurmi[2]
vinodkk@iiserb.ac.in

[1] UIET-H, Panjab University, Hoshiarpur, India

[2] Department of Data Science and Engineering
Indian Institute of Science Education and Research
Bhopal, India

Abstract

Accurate polyp and cardiac segmentation for early detection and treatment is essential for the diagnosis and treatment planning of cancer-like diseases. Traditional convolutional neural network (CNN) based models have represented limited generalizability, robustness, and inability to handle uncertainty, which affects the segmentation performance. To solve these problems, this paper introduces CLFSeg, an encoder-decoder based framework that aggregates the Fuzzy-Convolutional (FC) module leveraging convolutional layers and fuzzy logic. This module enhances the segmentation performance by identifying local and global features while minimizing the uncertainty, noise, and ambiguity in boundary regions, ensuring computing efficiency. In order to handle class imbalance problem while focusing on the areas of interest with tiny and boundary regions, binary cross-entropy (BCE) with dice loss is incorporated. Our proposed model exhibits exceptional performance on four publicly available datasets, including CVC-ColonDB, CVC-ClinicDB, EtisLaribPolypDB, and ACDC. Extensive experiments and visual studies show CLFSeg surpasses the existing SOTA performance and focuses on relevant regions of interest in anatomical structures. The proposed CLFSeg improves performance while ensuring computing efficiency, which makes it a potential solution for real-world medical diagnostic scenarios. Project page is available at https://visdomlab.github.io/CLFSeg/.

1 Introduction

The advent of deep learning algorithms has significantly impacted biomedical research, particularly in polyp and cardiac image segmentation. Polyps are growths of abnormal tissue in the colon that can be signs of colorectal cancer. It is important to accurately segment polyps to lower the death rate from cancer. So, polyp segmentation is a vital problem in medical

[*] Equal contribution.
[†] Work done at IISER Bhopal.

imaging, especially for the early identification and prevention of colorectal cancer. The manual analysis of these images can be labor-intensive and prone to inaccuracies due to intricate borders, and diversity of polyps in size, shape, appearance, and low-contrast areas.

Traditional segmentation methods like convolutional neural networks (CNNs) are time-consuming and lack precision. Recent advances have shown promising results, especially deep learning-based architectures such as ResNet [11], U-Net [30], MSU-Net [37], Resunet++ [13] and UcUNet [49] effectively capture local information from nearby pixels, allowing for fine detail extraction to segment area of interest. However, these methods often struggle with robustness. In contrast, transformers-based models - Swin-UNet [2], Transunet [6], TransUNet+ [21] and UGCANet [12], excel at capturing global details through attention mechanisms, making them more robust and generalizable to unseen data. However, attention mechanisms are challenging to train due to their quadratic complexity [10].

The hybrid models that combine CNNs and transformer architecture like PolySeg Plus [31], UPolySeg [22], RAPUNet [17], RaBiT [39], and EMCAD [27] show improvements in results, enhance feature extraction and global contextual understanding. DUCK-Net [8] is a U-shaped network having an encode-decoder structure and custom convolutional block designed to process images, capture low-level and high-level features, and demonstrate robustness and generalizability. However, challenges remain the same due to variations in polyps shape, low contrast with neighboring tissues, and the occurrence of image abnormalities.

This paper proposed a CLFSeg method that includes an encoder-decoder architecture with different convolution-based blocks and fuzzy modules to make the boundary fuzzier and to tackle the uncertainty and ambiguous data for robustness and precise segmentation. The CLFSeg model addresses key constraints in biomedical image segmentation by achieving performance comparable to self-attention methods while notably decreasing computing complexity by approximately 30%. This discourages crisp mask predictions, boosting adaptability in the segmented masks to more effectively address ambiguous and low-contrast areas and facilitates both single-class and multi-class predictions. This paper presents advanced research on CNNs that achieve remarkable performance. Our contributions are as follows:

- We propose a CLFSeg model, which consists of several convolutional and fuzzy modules to extract local and global features with fuzzy boundaries for precise, robust segmentation with computational efficiency on the multi-domain dataset.
- We propose a Fuzzy-Convolutional Module (FCM) consisting of parallel Resnet, Midscope, Widescope, Separable, and Fuzzy Module, then ConvGLU to extract local and global information for noise and uncertainty.
- Extensive experiments and ablation are performed to validate that our method achieves remarkable performance on four publicly available benchmark datasets.

2 Related Work

Classical CNN architectures such as U-Net [30] and its variants (e.g., U-Net++ [51], ResUNet [45], and ResUNet++ [13]) have been the foundation for many polyp and cardiac segmentation methods. These models effectively capture local features via encoder-decoder structures with skip connections, preserving spatial details necessary for fine-grained segmentation. Enhancements like HarDNet-DFUS [18] improved the backbone and decoder,

UPolySeg [22] incorporates dilated convolutions, and PolySeg Plus [31] implements efficient active learning to improve segmentation performance in scenarios with limited labeled data to better handle the diversity of polyp appearances and sizes. The Duck-Net [8] framework extended U-Net with customized convolutional blocks and residual downsampling, providing multi-resolution features and improving robustness to polyp variability. Similarly, LSSNet [42] addressed semantic gaps between layers by supplementing local and shallow features, effectively preserving multi-scale information to handle small and fuzzy polyps.

Attention-based models [16, 25, 44] have gained traction for their ability to capture long-range dependencies and focus on salient features. PraNet [9] uses reverse attention, ColonFormer [7], DUSFormer [43], and SwinPA-Net [6] employ Swin Transformer variants to extract global context with different modules, to precisely refine the polyp boundaries with noise suppression. Polyp-Mamba [47] utilizes state space models to extract long-range dependencies with reduced computational complexity. QueryNet [3] brings together polyp segmentation and detection into a single model, and ADSNet [24] introduces a complementary trilateral decoder and continuous attention modules to recover weak features and refine uncertain semantic areas. Recent work also explores frequency domain integration [38, 46] and hybrid architectures to enhance segmentation quality. MetaFormer [17] and MSRF-Net [35] blend convolutional and transformer modules to achieve a balance between local and global feature extraction at multiple scales. RaBiT [39] adopts a bidirectional feature pyramid network with reverse attention, optimizing the segmentation of polyp borders by refining feature maps across different scales.

Addressing the ambiguity and uncertainty inherent in medical images, especially at boundaries, fuzzy logic has been introduced into segmentation models. FuzzyNet [26] and FuzzyTransNet [19] combine fuzzy attention mechanisms with CNNs and transformers to better handle uncertain and noisy features. FANN [23] incorporates fuzzy membership functions within attention layers to efficiently manage ambiguous regions. Inspired by biological vision, TransNeXt [34] blends pixel-focused global context with convolutional gating to improve robustness and sensitivity to fine details.

3 Methodology

The proposed CLFSeg model illustrated in Figure 1, built on the baseline architecture [8], with key modifications like incorporating the Fuzzy-Convolution (FC) module as Figure 2 and reducing the complexity of the ResNet block. The FC module has a fuzzy logic module and ConvGLU for handling uncertainty and the ambiguous boundary of the polyp and cardiac structures while lowering computational complexity. Additionally, we provide interpretability using Grad-CAM++ [4], demonstrating the effectiveness of our methodology.

In the following subsection, we provide a detailed explanation of the logic and implementation of the FC module, emphasizing its relevance and benefits for medical image segmentation. Furthermore, we outline the approach used to significantly reduce computational complexity while maintaining network performance.

3.1 CLFSeg Model

Our model follows an encoder-decoder structure with skip and downscaling connections, along with an FC module. The encoder extracts features and converts them into intermediate representations, while the decoder combines the encoder's intermediate representations to

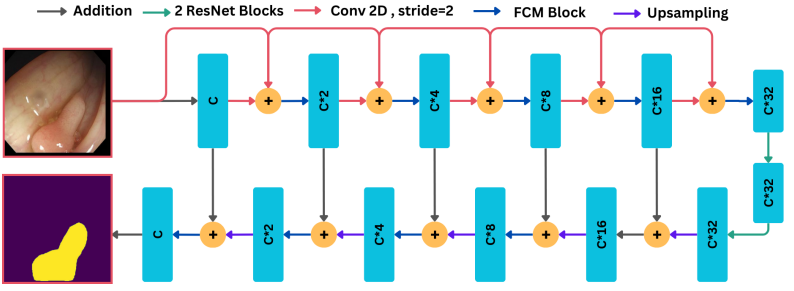

Figure 1: Overview of the CLFSeg model, showing the encoder-decoder structure with skip connections and downscaling layers. It highlights the integration of ResNet blocks, convolutional layers, and the Fuzzy-Convolution (FC) module designed to handle uncertainty and ambiguity in medical image segmentation, especially for polyp and cardiac structures. (Best view in color).

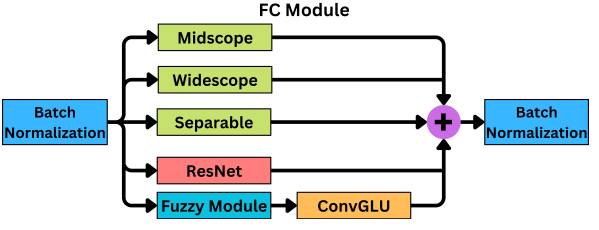

Figure 2: Overview of FC Module within the CLFSeg architecture, composed of five parallel branches: Midscope, Widescope, Separable, ResNet, and Fuzzy Module, followed by ConvGLU and batch normalization layers. (Best view in color).

generate the segmentation mask using skip connections. Downscaling connections transfer context to high-resolution layers [30], and the FC module, present in both the encoder and decoder, enables feature extraction via five parallel connections. These include:

1. Midscope block: The Midscope block mainly consists of two sequential convolution layers: the first employs a 3×3 kernel, while the second simulates a 7×7 kernel by using a dilation parameter of 2. This approach allows for capturing a larger global context with reduced computational cost.

2. Widescope block: Similar to Midscope, Widescope block applies three consecutive convolutions with a 3×3 kernel each and a dilation rate of $1, 2$, and 3, respectively. This configuration simulates kernel sizes of 3×3, 7×7, and 15×15.

3. Separable block: This block, which simulates the largest kernel, uses $1 \times N$ and $N \times 1$ convolutions to mimic an $N \times N$ kernel. This results in the largest local feature representation but may lead to the loss of diagonal features, which can be useful for detecting certain patterns essential for feature extraction.

4. ResNet Block: The baseline model [8] employs 6 ResNet layers arranged in 3 parallel connections: the first with a single ResNet block, the second with two serial ResNet blocks, and the third with three serial ResNet blocks. This configuration reportedly enhanced segmentation accuracy by enabling the model to capture finer details. However, in our experiments, we modified this setup to include only a single ResNet connection path, which resulted in: a) **Improved prediction masks**: This improvement is attributed to en-

Algorithm 1 Fuzzy Module Algorithm

Require: Input tensor $x \in \mathbb{R}^{B\times H\times W\times C}$, number of fuzzy sets n

Ensure: Output tensor after fuzzy module processing

1: Initialize trainable parameters: $w \in \mathbb{R}^{H\times W\times C}$, $b \in \mathbb{R}^{H\times W\times C}$, $\mu \in \mathbb{R}^{1\times 1\times 1\times n\times 1}$, and $\sigma \in \mathbb{R}^{1\times 1\times 1\times n\times 1}$

2: Apply element-wise multiplication to the input as $x \leftarrow w \odot x + b$

3: Perform layer normalization on x as $x \leftarrow LayerNormalization(x)$

4: Apply LeakyReLU activation function as $x \leftarrow LeakyReLU(x)$

5: Reshape the tensor to $x \in \mathbb{R}^{B\times H\times W\times n\times C}$, where n is the number of fuzzy sets

6: Apply Gaussian membership function for each fuzzy set as $x \leftarrow \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

7: Compute the mean across fuzzy set dimension n as $x \leftarrow \frac{1}{n}\sum_{i=1}^{n} x_i \quad \in \mathbb{R}^{B\times H\times W\times C}$

8: **return** x

hanced gradient propagation and better gradient convergence with a larger area that can be segmented out. b) **Significant reduction in computation**: We observed a reduction of approximately **30%** in FLOPs across various filter sizes as viewed in Table 2.

5. ConvGLU: The ConvGLU [34] layer, is designed to enhance feature learning capabilities. It consists of dual linear projections and a final projection after feature concatenation. The layer introduces a refined gating mechanism that allows each channel to be influenced by neighboring channels through depthwise convolution. This gating mechanism refines coarse-grained features and suppresses unwanted excitations.

6. Fuzzy Module: The Fuzzy Module enhances the model's ability to manage uncertainty and ambiguity in data, particularly in complex tasks like medical image segmentation, where features and boundaries are often unclear [50]. Traditional attention mechanisms and activation functions, such as *sigmoid*, *tanh*, and *LeakyReLU*, struggle with ambiguous data and applying nuanced focus across feature channels. Due to these shortcomings, the Fuzzy Module is introduced within the model. The module combines fuzzy logic with attention mechanisms, enabling the network to assign channel-specific attention through learnable Gaussian membership functions [23]. This approach reduces uncertainty in feature representations and improves the network's focus on relevant features. Additionally, it addresses monotonicity by accounting for a range of variations within data, considering both positive and negative representations.

Our objective was to effectively reduce uncertainty by generating fuzzy representations of the input image, resulting in a robust segmentation mask as summarized in Algorithm 1. For this, our method uses the Gaussian membership function (Equation 1). By working similarly to a weighted average Gaussian filter, it reduces image noise by assigning a degree of membership to each feature. Furthermore, the mean and variance of parameters defined as μ and σ, respectively, are trainable, allowing the model to extract features of varying importance. These fuzzy sets are then summed together, capturing complex patterns and relationships within the data and averaging these membership functions (Equation 2), which helps in summarizing the overall values into a single channel. Here, C is the number of fuzzy sets present in the input image x.

$$f(x,\mu,\sigma) = e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1}$$

$$\overline{f}(x,\mu,\sigma) = \frac{1}{C}\sum_{i=0}^{c} e^{\frac{-(x_i-\mu_i)^2}{2\sigma_i^2}} \tag{2}$$

7. Fuzzy-ConvGLU Module: Incorporating the ConvGLU layer after the Fuzzy Module (as illustrated in Figure 2) in a neural network architecture significantly enhances its feature-learning capabilities. The Fuzzy Module captures and encodes complex relationships from the input features, producing a rich feature map. When the ConvGLU layer processes this output, the model's ability to extract meaningful information from these detailed features is further enhanced.

Although the ConvGLU layer utilizes a gating mechanism, in the context of our 2D image dataset, it primarily acts as a transformation gating process. The output from the Fuzzy Module, which consists of a single channel, is fed into this layer. ConvGLU then focuses on spatial filtering and gating, allowing it to perform more precise spatial feature selection even with limited channels. Additionally, the gating mechanism in the ConvGLU layer controls the flow of processed features, ensuring that only the most relevant information is passed forward. This helps the model avoid learning overly generalized features, preserving its sensitivity to finer details in the input data. Each position or token receives a unique gating signal based on local features, helping prevent the model from becoming too coarse in its feature interpretation. To support our explanations, we have done an extensive ablation study, as shown in Table 3. Additionally, as demonstrated in Table 2, this configuration significantly reduces computational complexity while offering a more effective attention mechanism.

4 Experimental Results

Dataset: *CVC-ColonDB* consists of 380 images and their corresponding ground truth labels with 574×500 resolution. *CVC-ClinicDB* consists of 612 images and corresponding pixel-wise annotations with 384×288 resolution from 31 colonoscopy sequences. *ETIS-LaribPolyp* consists of 196 images and respective ground-truth labels with 1255×966 resolution. *Automated Cardiac Diagnosis Challenge (ACDC)* is an MRI image dataset containing 100 cases and 3 labels.

Implementation Setup: The CVC-ColonDB, ETIS-LaribPolypDB, and CVC-ClinicDB datasets are split into 80:10:10 for training, validation, and test sets, while for ACDC, the ratio is 70:10:20, i.e., 1312:380:210 images. We apply data augmentation using standard flips on the X and Y axes, color jitter, and affine transformations. The model is designed to handle both binary and multi-class segmentation tasks, generating reliable segmentation saliency maps in each case. The training process used a batch size of 4 over 1000 epochs, input image size is 352x352 pixels, the learning rate is 1e-4, and the optimizer is RMSprop [40]. The model is implemented with the TensorFlow [1] framework and trained on an NVIDIA A100 GPU. We employ a hybrid BCE and dice Loss in equal proportions, and the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), precision, recall, and accuracy are used for evaluation of the segmentation models. This approach encourages the model to achieve pixel-level performance while preserving global region consistency.

5 Comparison with SOTA

The proposed CLFSeg model achieves significant improvements over existing state-of-the-art methods on multiple polyp datasets shown in Table 1. On the CVC-ColonDB polyp dataset, CLFSeg attained a DSC of **0.9593**, outperforming prior leading methods like DUCKNet (DSC 0.9389), ASRDNet (DSC 0.9337), and PraNet (DSC 0.9131) by more than

Dataset	Methods	DSC (↑)	IoU (↑)	Precision	Recall	Accuracy
CVC-ColonDB	FuzzyTransNet [19]	0.7780	0.6890	-	-	-
	UNet [30]	0.8032	0.7037	0.8100	0.8274	0.9807
	FuzzyNet (PVT) [26]	0.8110	0.7280	-	-	-
	PSTNet [16]	0.8270	0.7480	-	-	-
	QueryNet [3]	0.8278	0.7593	0.8351	0.8526	-
	LSSNet [17]	0.8937	0.8221	-	-	-
	PraNet [2]	0.9131	0.8401	**0.9657**	0.8659	0.9901
	InCoLoTransNet [25]	0.9309	0.9013	0.9177	0.9156	-
	ASRDNet [14]	0.9337	0.8756	0.9327	0.9347	-
	DUCKNet (17 filter) [8]	0.9353	0.8785	0.9314	0.9392	0.9929
	DUCKNet (34 filter) [8]	0.9230	0.8571	0.9113	0.9351	0.9914
	DUCKNet (24 filter, our-training) [8]	0.9389	0.8848	0.9365	0.9413	0.9933
	CLFSeg (17 filters, OURS)	0.9460	0.8976	0.9582	0.9342	0.9941
	CLFSeg (24 filters, OURS)	0.9503	0.9053	0.9583	**0.9430**	0.9943
	CLFSeg (34 filters, OURS)	**0.9593**	**0.9218**	0.9634	0.9401	**0.9945**
CVC-ClinicDB	U-Net [30]	0.7631	0.6169	0.7989	0.7303	0.9599
	FAENet [28]	0.9330	0.8830	-	-	-
	DUSFormer-L [4]	0.9350	0.9020	0.9610	-	-
	FuzzyNet (PVT) [26]	0.9370	0.8890	-	-	-
	ADSNet [24]	0.9380	0.8900	-	-	-
	SwinPA-Net [6]	0.9410	0.8940	-	-	-
	FuzzyTransNet [19]	0.9420	0.8910	-	-	-
	PSTNet [16]	0.9450	0.9010	-	-	-
	CTHP [15]	0.9471	0.9021	0.9524	0.9444	-
	Duck-Net (17 filters) [8]	0.9450	0.8952	0.9488	0.9406	0.9903
	Duck-Net (34 filters) [8]	0.9478	0.9009	0.9468	**0.9489**	0.9907
	Duck-Net (24 filters, our-training) [8]	0.9430	0.8922	0.9539	0.9324	0.9900
	Polyp-Mamba [4]	0.9490	9070	-	-	-
	CLFSeg (17 filters, OURS)	**0.9533**	**0.9108**	**0.9636**	0.9432	**0.9918**
	CLFSeg (24 filters, OURS)	0.9530	0.9103	0.9594	0.9467	0.9917
	CLFSeg (34 filters, OURS)	0.9500	0.9048	0.9568	0.9433	0.9912
ETIS-LaribPolyp	FuzzyNet (PVT) [26]	0.7910	0.7020	-	-	-
	U-Net [30]	0.7984	0.6969	0.8322	0.7724	0.9734
	PSTNet [16]	0.8000	0.7260	-	-	-
	QueryNet [3]	0.8189	0.7399	0.7488	0.7740	-
	Polyp-Mamba [4]	0.8250	0.7470	-	-	-
	PraNet [2]	0.8827	0.7900	0.9825	0.8013	0.9877
	FCN-Transformer [31]	0.9163	0.8455	**0.9633**	0.8736	0.9915
	ASRDNet [14]	0.9313	0.8714	0.9055	0.9586	-
	InCoLoTransNet [25]	0.9316	0.9225	0.9283	0.9392	-
	Duck-Net (17 filters) [8]	0.9324	0.8734	0.9539	0.9118	0.9930
	Duck-Net (34 filters) [8]	0.9354	0.8788	0.9309	0.9400	0.9931
	Duck-Net (24 filters, our-training) [8]	0.9396	0.8861	0.9372	0.8861	0.9936
	CLFSeg (17 filters, OURS)	0.9292	0.8678	0.9189	0.9396	0.9924
	CLFSeg (24 filters, OURS)	0.9140	0.8416	0.8826	**0.9476**	0.9905
	CLFSeg (34 filters, OURS)	**0.9487**	**0.9024**	0.9596	0.9380	**0.9946**

Table 1: Comparing the performance of different segmentation methods on the CVC-ColonDB, CVC-ClinicDB, and ETIS-LaribPoly dataset. Bold shows best model results.

2%. Similarly, on the CVC-ClinicDB dataset, CLFSeg reached a DSC of **0.9533**, surpassing SOTA methods. On the ETIS-LaribPolyp dataset, CLFSeg also demonstrated remarkable gains, achieving a DSC of **0.9487**, outperforming earlier models by a margin of approximately **1%** compared to baseline DuckNet [8] architectures. This performance reflects the proposed CLFSeg's superior capability in capturing both local and global features with fuzzy boundary refinement.

In the cardiac segmentation task on the ACDC dataset, CLFSeg match the performance with baseline [8] for 17 and 34 filters. But exceeded the performance of recent models such

Methods	DSC (↑)	IoU (↑)	#FLOPs (M) ↓	#Params (M) ↓
U-Net [30]	0.8755	-	-	-
EM-ViT [14]	0.9029	-	-	-
CSWin-UNet [20]	0.9146	-	-	-
Lite-MixedNet [29]	0.9207	-	-	-
PVT-EMCAD-B2 [27]	0.9212			
RWKV-UNet [34]	0.9217	-	-	-
MIST [28]	0.9256	-	-	-
LHU-Net [32]	0.9266	-	-	-
FCT [41]	0.9302	-	-	-
Adaptive t-vMF [15]	0.9368	0.8851	-	-
Duck-Net (17 filters, our-training) [8]	0.9392	0.8853	38	38
Duck-Net (24 filters, our-training) [8]	0.9470	0.8993	77	77
Duck-Net (34 filters, our-training) [8]	0.9428	0.8917	155	155
CLFSeg (17 filters, OURS)	**0.9522**	**0.9087**	**26**	**38**
CLFSeg (24 filters, OURS)	0.9403	0.8873	**54**	**71**
CLFSeg (34 filters, OURS)	0.9498	0.9043	**106**	**132**

Table 2: Comparing the performance of different segmentation methods on the ACDC dataset. Bold shows the best model results.

Layers			DSC (↑)	IoU (↑)	Precision	Recall	Accuracy
Fuzzy Module	ConvGLU	1-ResNet					
			0.9389	0.8848	0.9365	0.9413	0.9933
✓			0.9374	0.8823	0.9279	0.9472	0.9930
	✓		0.9337	0.8757	**0.9657**	0.9287	0.9943
		✓	0.9269	0.8638	0.9489	0.9059	0.9921
✓	✓		0.9469	0.8991	0.9476	0.9202	0.9928
✓		✓	0.9491	0.9031	0.9517	**0.9520**	**0.9947**
	✓	✓	0.9277	0.8651	0.9627	0.8951	0.9923
✓	✓	✓	**0.9503**	**0.9053**	0.9583	0.9430	0.9943

Table 3: Ablation study based on different layers (Fuzzy module, ConGLU, and ResNet) on CVC-ColonDB dataset with 24 filters. Bold highlights the best results.

as Duck-Net and Adaptive t-vMF for 17 filters, reaching a DSC of **0.9522** and IoU of **0.9087** while reducing computational complexity by up to **30%** as shown in Table 2.

5.1 Ablation Study

1. Analysis of the Different Layers of the CLFSeg model: Three layers—Fuzzy Module, ConvGLU, and 1-ResNet are evaluated in the ablation studies, as shown in Table 3. With a fixed filter size of 24, these layers underwent extensive testing. Results indicate that the combination of Fuzzy Module with ConvGLU and 1-ResNet achieved optimal segmentation maps, reaching **0.9503** DSC and **0.9053** IoU. The findings demonstrate the effectiveness of our proposed approach, showing a **+2.46%** improvement over the single use of the 1-ResNet layer. From the Table 2, we conclude that 1-ResNet used in our proposed CLFSeg model achieves good performance with low computational cost compared to baseline Duck-Net [8].
2. Analysis of the effect of different filter sizes: In Table 1, we observe that increasing the filter size improved DSC of CVC-ColonDB dataset from **0.9460** to **0.9593**, marking a **+1.38%** increase. Our model, with a filter size of 34, delivered SOTA performance on CVC-

Filters	Layers			DSC (↑)	IoU (↑)	Precision	Recall	Accuracy
	1-resnet	2-resnet	3-resnet					
17 Filters	✓			**0.9460**	**0.8976**	**0.9582**	0.9342	**0.9941**
	✓	✓		0.9416	0.8897	0.9402	**0.9430**	0.9936
	✓	✓	✓	0.8799	0.7857	0.8507	0.9113	0.9864
24 Filters	✓			**0.9503**	**0.9053**	0.9583	**0.9430**	**0.9943**
	✓	✓		0.9451	0.8959	0.9522	0.9380	0.9940
	✓	✓	✓	0.9448	0.8953	**0.9608**	0.9293	0.9940
34 Filters	✓			**0.9495**	**0.9039**	**0.9624**	0.9370	**0.9945**
	✓	✓		0.9380	0.8832	0.9599	0.9170	0.9933
	✓	✓	✓	0.9455	0.8966	0.9422	**0.9488**	0.9940

Table 4: Ablation study based on different parallel ResNet blocks on CVC-ColonDB dataset with 24 filters. Bold highlights the best results.

Losses with 24 filters			DSC (↑)	IoU (↑)	Precision	Recall	Accuracy
BCE	Dice	Focal					
✓			0.9423	0.8909	**0.9623**	0.9231	0.9938
	✓		0.9379	0.8832	0.9525	0.9238	0.9933
✓	✓		**0.9503**	**0.9053**	0.9583	0.9430	**0.9943**
	✓	✓	0.9455	0.8966	0.9401	**0.9509**	0.9940

Table 5: Ablation study based on different loss functions on CVC-ColonDB dataset with 24 filters. Bold highlights the best results.

Dataset	CVC-ColonDB				CVC-ClinicDB				ACDC			
Filters	17		34		17		34		17		34	
Methods	DuckNet	CLFSeg	DuckNet	CLFSeg	DuckNet	CLFSeg	DuckNet	CLFSeg	DuckNet	CLFSeg	DuckNet	CLFSeg
HD95	1.97	**1.43**	2.00	**1.50**	4.44	**3.54**	6.39	**5.97**	0.90	**0.58**	1.36	**0.81**

Table 6: Analysis of boundary clarity on DuckNet and proposed CLFSeg methods across datasets and filter settings (lower HD95 is better).

ColonDB dataset, with configurations of 24 and 17 filters surpassing benchmarks [8].

3. Analyse parallel ResNet block: The impact of using parallel ResNet blocks with an increasing number of layers is analyzed, as outlined in Table 4. Results show that adding more parallel ResNet blocks negatively impacted model efficiency, despite thorough testing with all three filter sizes. A single ResNet block effectively captured fine details, while additional blocks merely extended training time and added complexity, leading to an average **4.65%** decrease in performance across all filter sizes when additional blocks are used.

4. Analysis of the boundary Clarity: Our evaluation considers HD95, a boundary-aware metric that measures the maximum distance between the true and predicted segmentation boundaries while discarding extreme outliers (95% confidence interval). Our model consistently achieves superior performance on this metric. Table 6 reports the improvements across 17- and 34-filter configurations for CVC-ColonDB, CVC-ClinicDB, and ACDC datasets. On average, CLFSeg reduces the HD95 score by 0.54 units compared to DUCKNet.

5. Analysis of the different loss functions: In Table 5, hybrid BCE-dice loss produced a DSC of **0.9503**, while BCE loss and dice loss alone resulted in **0.9423** and **0.9379**, respectively. The Focal and dice loss achieved a DSC of **0.9455**, demonstrating that an equal contribution of BCE and dice losses results in the best model performance.

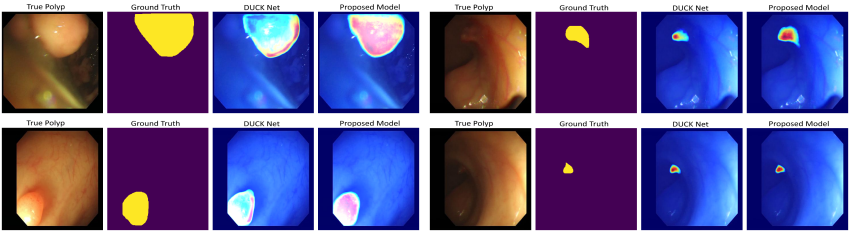

Figure 3: Comparison of GradCam++ Visualization between DuckNet and CLFSeg model on CVC-ColonDB dataset. (Best view in color).

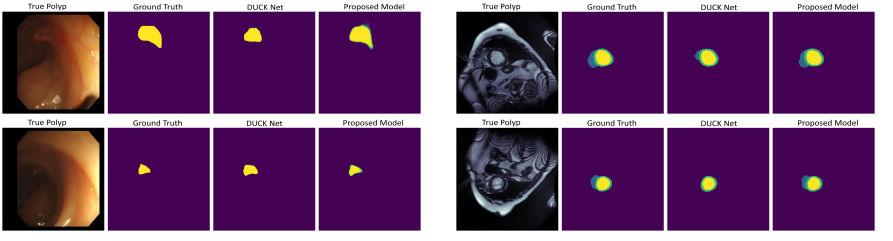

Figure 4: Comparison of Segmentation Map from proposed CLFSeg model for CVC-ColonDB (left side), and ACDC dataset (right side). (Best view in color).

6. Visualization of Activation Maps: Grad-CAM++ provides a visual representation of the regions focused on by the model. Figure 3 shows that CLFSeg more accurately captures the entire ground truth for both small and large regions compared to DuckNet, which often overlooks relevant polyp areas. The red-highlighted regions are more prominent in CLFSeg, indicating its ability to focus on clinically significant areas, particularly due to the FC module. CLFSeg effectively captures finer details while making confident predictions, as evidenced by the red regions in Grad-CAM++. This underscores the model's precision and interpretability, enabling improved boundary delineation across varying polyp sizes.

7. Visualization of Segmentation Map: Figure 4 compares **segmentation maps** across CVC-ColonDB and ACDC datasets. In CVC-ColonDB, the proposed CLFSeg model outperforms DuckNet, where DuckNet misclassifies relevant regions, while CLFSeg accurately highlights polyp boundaries and uncertainty areas. In the ACDC dataset, the proposed CLFSeg effectively segments and differentiates the different cardiac regions. It helps in conditions where a single image consists of multiple areas of interest and precise boundary delineation is important, like cancer treatment.

6 Conclusion

The proposed CLFSeg framework enhances automated medical image segmentation tasks with a fast, efficient segmentation mask, highlighting the FC module. Fuzzy logic and a refining layer reduce uncertainty and enable flexible boundary fuzzification for more adaptable segmentation. Achieving state-of-the-art DSC scores of 0.9593, 0.9533, 0.9487, and 0.9522 on CVC-ColonDB, CVC-ClinicDB, ETIS-LaribPolypDB, and ACDC demonstrates high effectiveness. This approach advances AI in healthcare and sets a foundation for future improvements, including broader datasets and privacy-focused solutions.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[2] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[3] Jiaxing Chai, Zhiming Luo, Jianzhe Gao, Licun Dai, Yingxin Lai, and Shaozi Li. QueryNet: A Unified Framework for Accurate Polyp Segmentation and Detection . In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15008, pages 544 – 554. Springer Nature Switzerland, October 2024.

[4] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 839–847. IEEE, 2018.

[5] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[6] Hao Du, Jiazheng Wang, Min Liu, Yaonan Wang, and Erik Meijering. Swinpa-net: Swin transformer-based multiscale feature pyramid aggregation network for medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 35(4):5355–5366, 2024. doi: 10.1109/TNNLS.2022.3204090.

[7] Nguyen Thanh Duc, Nguyen Thi Oanh, Nguyen Thi Thuy, Tran Minh Triet, and Viet Sang Dinh. Colonformer: An efficient transformer based method for colon polyp segmentation. *IEEE Access*, 10:80575–80586, 2022.

[8] Razvan-Gabriel Dumitru, Darius Peteleaza, and Catalin Craciun. Using duck-net for polyp image segmentation. *Scientific reports*, 13(1):9803, 2023.

[9] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 263–273. Springer, 2020.

[10] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, et al. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):87–110, 2022.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[12] Pham Vu Hung, Nguyen Duy Manh, Nguyen Thi Oanh, Nguyen Thi Thuy, and Dinh Viet Sang. Ugcanet: A unified global context-aware transformer-based network with feature alignment for endoscopic image analysis. *arXiv preprint arXiv:2307.06260*, 2023.

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

[14] Sanaz Karimijarbigloo, Sina Ghorbani Kolahi, Reza Azad, Ulas Bagci, and Dorit Merhof. Frequency-domain refinement of vision transformers for robust medical image segmentation under degradation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 9176–9185. IEEE, 2025.

[15] Sota Kato and Kazuhiro Hotta. Adaptive t-vmf dice loss for multi-class medical image segmentation. *arXiv preprint arXiv:2207.07842*, 2022.

[16] Vinod Kumar Kurmi, Shanu Kumar, and Vinay P. Namboodiri. Attending to discriminative certainty for domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[17] Hyunnam Lee and Juhan Yoo. Metaformer and cnn hybrid model for polyp image segmentation. *IEEE Access*, 2024.

[18] Ting-Yu Liao, Ching-Hui Yang, Yu-Wen Lo, Kuan-Ying Lai, Po-Huai Shen, and Youn-Long Lin. Hardnet-dfus: An enhanced harmonically-connected network for diabetic foot ulcer image segmentation and colonoscopy polyp segmentation. *arXiv preprint arXiv:2209.07313*, 2022.

[19] Ruihua Liu, Siyu Duan, Lihang Xu, Lingkun Liu, Jinshuang Li, and Yangyang Zou. A fuzzy transformer fusion network (fuzzytransnet) for medical image segmentation: The case of rectal polyps and skin lesions. *Applied Sciences*, 13(16):9121, 2023.

[20] Xiao Liu, Peng Gao, Tao Yu, Fei Wang, and Ru-Yue Yuan. Cswin-unet: Transformer unet with cross-shaped windows for medical image segmentation. *Information Fusion*, 113:102634, 2025.

[21] Yuhang Liu, Han Wang, Zugang Chen, Kehan Huangliang, and Haixian Zhang. Transunet+: Redesigning the skip connection to enhance features in medical image segmentation. *Knowledge-Based Systems*, 256:109859, 2022.

[22] Subhashree Mohapatra, Girish Kumar Pati, Manohar Mishra, and Tripti Swarnkar. Upolyseg: A u-net-based polyp segmentation network using colonoscopy images. *Gastroenterology Insights*, 13(3):264–274, 2022.

[23] Yang Nan, Javier Del Ser, Zeyu Tang, Peng Tang, Xiaodan Xing, Yingying Fang, Francisco Herrera, Witold Pedrycz, Simon Walsh, and Guang Yang. Fuzzy attention neural network to tackle discontinuity in airway segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[24] Quang Vinh Nguyen, Van Thong Huynh, and Soo-Hyung Kim. Adaptation of distinct semantics for uncertain areas in polyp segmentation. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. URL https://papers.bmvc2023.org/0806.pdf.

[25] Yassine Oukdach, Anass Garbaz, Zakaria Kerkaou, Mohamed El Ansari, Lahcen Koutti, Ahmed Fouad El Ouafdi, and Mouna Salihoun. Incolotransnet: An involution-convolution and locality attention-aware transformer for precise colorectal polyp segmentation in gi images. *Journal of Imaging Informatics in Medicine*, pages 1–21, 2025.

[26] Krushi Bharatbhai Patel, Fengjun Li, and Guanghui Wang. Fuzzynet: A fuzzy attention module for polyp segmentation. In *NeurIPS'22 Workshop on All Things Attention: Bridging Different Perspectives on Attention*, 2022.

[27] Md Mostafijur Rahman, Mustafa Munir, and Radu Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024.

[28] Md Motiur Rahman, Shiva Shokouhmand, Smriti Bhatt, and Miad Faezipour. Mist: Medical image segmentation transformer with convolutional attention mixing (cam) decoder. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 404–413, 2024.

[29] Zitong Ren, Yongming Li, Liejun Wang, and Lianghui Xu. Lite-mixednet: Lightweight and efficient hybrid network for medical image segmentation. *Pattern Recognition*, page 111378, 2025.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[31] Abdelrahman I Saad, Fahima A Maghraby, and Osama Badawy. Polyseg plus: Polyp segmentation using deep learning with cost effective active learning. *International Journal of Computational Intelligence Systems*, 16(1):148, 2023.

[32] Yousef Sadegheih, Afshin Bozorgpour, Pratibha Kumari, Reza Azad, and Dorit Merhof. Lhu-net: A light hybrid u-net for cost-efficient, high-performance volumetric medical image segmentation. *arXiv preprint arXiv:2404.05102*, 2024.

[33] Edward Sanderson and Bogdan J Matuszewski. Fcn-transformer feature fusion for polyp segmentation. In *Annual conference on medical image understanding and analysis*, pages 892–907. Springer, 2022.

[34] Dai Shi. Transnext: Robust foveal visual perception for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17773–17783, 2024.

[35] Abhishek Srivastava, Debesh Jha, Sukalpa Chanda, Umapada Pal, Håvard D Johansen, Dag Johansen, Michael A Riegler, Sharib Ali, and Pål Halvorsen. Msrf-net: a multi-scale residual fusion network for biomedical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2252–2263, 2021.

[36] Chenhong Su, Xuegang Luo, Shiqing Li, Li Chen, and Juan Wang. Vmkla-unet: vision mamba with kan linear attention u-net. *Scientific Reports*, 15(1):13258, 2025.

[37] Run Su, Deyun Zhang, Jinhuai Liu, and Chuandong Cheng. Msu-net: Multi-scale u-net for 2d medical image segmentation. *Frontiers in Genetics*, 12:639930, 2021.

[38] Rui Tang, Hejing Zhao, Yao Tong, Ruihui Mu, Yuqiang Wang, Shuhao Zhang, Yao Zhao, Weidong Wang, Min Zhang, Yilin Liu, et al. A frequency attention-embedded network for polyp segmentation. *Scientific Reports*, 15(1):4961, 2025.

[39] Nguyen Hoang Thuan, Nguyen Thi Oanh, Nguyen Thi Thuy, Stuart Perry, and Dinh Viet Sang. Rabit: An efficient transformer using bidirectional feature pyramid network with reverse attention for colon polyp segmentation. *arXiv preprint arXiv:2307.06420*, 2023.

[40] Tijmen Tieleman. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26, 2012.

[41] Athanasios Tragakis, Chaitanya Kaul, Roderick Murray-Smith, and Dirk Husmeier. The fully convolutional transformer for medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3660–3669, 2023.

[42] Wei Wang, Huiying Sun, and Xin Wang. Lssnet: A method for colon polyp segmentation based on local feature supplementation and shallow feature supplementation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 446–456. Springer, 2024.

[43] Zhangrun Xia, Jingliang Chen, and Chengzhun Lu. Dusformer: Dual-swin transformer v2 aggregate network for polyp segmentation. *IEEE Access*, 12:8822–8832, 2024. doi: 10.1109/ACCESS.2024.3352428.

[44] Leyi Xiao, Yinuo Liu, and Chaodong Fan. Attention-enhanced separable residual with dilation net for medical image segmentation. *Neurocomputing*, page 130434, 2025.

[45] Xiao Xiao, Shen Lian, Zhiming Luo, and Shaozi Li. Weighted res-unet for high-quality retina vessel segmentation. In *2018 9th international conference on information technology in medicine and education (ITME)*, pages 327–331. IEEE, 2018.

[46] Wenhao Xu, Rongtao Xu, Changwei Wang, Xiuli Li, Shibiao Xu, and Li Guo. Pst-net: Enhanced polyp segmentation with multi-scale alignment and frequency domain integration. *IEEE Journal of Biomedical and Health Informatics*, 28(10):6042–6053, 2024. doi: 10.1109/JBHI.2024.3421550.

[47] Zhongxing Xu, Feilong Tang, Zhe Chen, Zheng Zhou, Weishan Wu, Yuyao Yang, Yu Liang, Jiyu Jiang, Xuyue Cai, and Jionglong Su. Polyp-mamba: Polyp segmentation with visual mamba. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 510–521. Springer, 2024.

[48] He Xue, Luo Yonggang, Liu Min, and Li Lin. A lighter hybrid feature fusion framework for polyp segmentation. *Scientific Reports*, 14(1):23179, 2024.

[49] Shukai Yang, Xiaoqian Zhang, Yufeng Chen, Youtao Jiang, Quan Feng, Lei Pu, and Feng Sun. Ucunet: A lightweight and precise medical image segmentation network based on efficient large kernel u-shaped convolutional module design. *Knowledge-Based Systems*, 278:110868, 2023.

[50] Sheng Zhang, Yingying Fang, Yang Nan, Shiyi Wang, Weiping Ding, Yew-Soon Ong, Alejandro F Frangi, Witold Pedrycz, Simon Walsh, and Guang Yang. Fuzzy attention-based border rendering orthogonal network for lung organ segmentation. *IEEE Transactions on Fuzzy Systems*, 2024.

[51] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.