

LIC-NET: LIGHTWEIGHT INTEGRATED CONVOLUTIONAL NETWORK FOR ACCURATE POLYP SEGMENTATION

SUCHITRA A. PATIL^{1,2}, CHANDRAKANT GAIKWAD³

¹Assistant Professor, Pillai College of Engineering, Department of Electronics and Telecommunication Engineering, India

²Research Scholar, Ramrao Adik Institute of Technology, Nerul, Department of Electronics and Telecommunication Engineering, India

³Professor, Ramrao Adik Institute of Technology, Nerul, Department of Computer Science and Engineering, India

E-mail: ¹spatil@mes.ac.in, ²chandrakant.gaikwad@dypatil.edu

ABSTRACT

The segmentation of a small polyp present in intestinal regions, which tend to be malignant, is a basic and essential task for the detection of colon cancer. Segmenting a small polyp is challenging due to the higher similarity of tissues. Inaccurate segmentation results in higher false positives for colorectal cancer classification. This work suggests a two-stage deep learning network for polyp segmentation. In the first stage, the colonoscopy image is preprocessed to generate salient regions. In the second stage, the salient regions are processed by a novel U-Net structure network called LIC-Net, which integrates transfer learning and multiscale feature extraction to increase the segmentation accuracy. Testing with Kvasir-Seg and CVC-ClinicDB, both in direct and cross-learning mode, the proposed solution achieved more than 90% accuracy. The false positives are at least 2% lower compared to the most recent deep learning based segmentation works.

Keywords: *Convolutional Neural Network, Fuzzy-TAN, Colonoscopy, Dilated Convolution, Polyp Segmentation, U-Shaped Model*

1. INTRODUCTION

Colorectal cancer (CRC) is a highly malignant cancer and the second biggest contributor to cancer-related fatalities all over the world [1]. Food consumption behaviors like processed meats, junk foods, obesity, smoking, and alcohol consumption are associated with the development of CRC. Worldwide, 9% of total cancer cases are CRC. If a CRC is found in the advanced stage, then there are very few options for treatment [2]. This leads to a critical need for timely detection and treatment initiation [3]. Manual analysis of a polyp is tedious due to blurry input complicated with different textures and slight differences in the background. The nature of the image acquired introduces higher error in manual analysis. Computer-assisted diagnosis is a solution to this problem. These methods extract various features from significant regions of the image and train machine learning classifiers to recognize cancer classes from the features. CRC is detected using the colonoscopy screening technique. In this method, the internal

colon and rectum regions are imaged with a camera, and this image is examined by a physician to detect CRC. This is the most adopted method for detecting CRC.

Despite its success, colonoscopy has certain drawbacks, ie. possibility of missing polyps, which raises the risk of cancer. Because polyps differ in size, shape, and appearance, it can be challenging to identify them apart from surrounding tissue and may be precursors to colorectal cancer (CRC), which increases the significance of identifying and removing them during a colonoscopy.

Much of the work has been done in solving these problems. Various automatic polyp segmentation approaches [4-6] based on active shape models, texture feature-based classification, graph-based methods, etc., have been proposed in earlier works with limited accuracy. Hand-crafted feature-based methods' segmentation accuracy is insufficient to meet clinical practice criteria.

Deep learning is a recent revolution that avoids the need for handcrafted features and processes the image as a whole to classify cancer

classes. Structural similarity to neighboring tissues and smaller granularity make segmenting polyp regions very challenging. The problem with training deep learning classifiers is that they need large training images. The models become overfit and have higher false positives when handling polyp regions with higher similarity to nearby tissues and backgrounds.

To enhance colonoscopy analysis, detection systems should ideally have the following essential features:

1. To handle patient variances, computer systems must constantly produce outcomes that can be relied upon and that are both flexible and trustworthy.
2. When it comes to clinical applications, real-time functionality is essential for rapid analysis and timely decision-making by medical experts during procedures such as colonoscopies.
3. Medical professionals must be able to readily navigate a system interface, which requires user-friendly design.
4. For this system to be widely used, cost and resource efficiency are essential, meaning that their economic viability and accessibility are required for widespread adoption.

To effectively contribute to the initial identification as well as the avoidance of colon cancer on a larger scale, the optimal system for colonoscopy analysis should combine reliability, high performance, real-time capability, user-friendly design, and cost-effectiveness.

In rural areas, there are various challenges, including the lack of proper healthcare monitoring systems. These areas lack skilled medical professionals, advanced diagnostic infrastructure, and timely access to screening technologies. Consequently, late detection of CRC and other critical diseases results in poor health outcomes. A lightweight and efficient solution that can be adapted for real-time and offline use is crucial to bridge this healthcare gap. This context points out that societies require automated, accurate, and computationally efficient frameworks like the one proposed in this work.

Addressing this problem, this work suggests a lightweight integrated convolutional network. The solution has two stages. At the first stage, topological active net (TAN) with energy

minimization is applied to remove similar tissue and background regions to polyp regions. This preprocessing allows learning more effective feature representations in the second stage. In the second stage a lightweight U-Net structure network with a novel encoder-decoder structure called LIC-Net is applied to the output of the first stage to get the polyp regions. With this two-stage processing, the polyp regions can be segmented accurately with lower false positives. This work suggests the following contributions.

- (i) A two-stage lightweight integrated deep learning network to segment polyp regions with higher accuracy. This network integrates a transfer learning encoder with a multi-scale feature extraction based decoder, which can segment the polyp regions with comparatively reduced training volume.
- (ii) Fuzzy energy minimization based topological active net segmentation model is proposed for first stage filtering which isolates similar tissue and background regions affecting the polyp segmentation accuracy.

The layout of this paper is as follows. The existing works on polyp segmentation and their issues are described in Section 2. Section 3 explains the proposed deep learning model for polyp segmentation. Following it, the results of the proposed segmentation model and its comparison to existing works are detailed in Section 4. Discussion about the results is in section 5. The final section, 6, summarizes the work and provides concluding remarks.

2. RELATED WORK

Yao et al [4] used flexible model in combination with clustering for polyp segmentation. The knowledge guided adjustment process used to identify potential polyp regions works only for certain polyp shapes and fails for highly irregular shapes. Gross et al [5] proposed a template matching approach to identify polyp contour in the image. The images were de-noised using linear diffusion filtering (LDF). Use of LDF subdued smaller blood vessels and emphasized major edges. The approach does not detect polyp contours with irregular shapes. Wang et al [6] did a pixel-level classification of polyp regions using a deep learning network. Local visual features extracted by small receptive fields are passed to convolutional layers and finally to classification. The false positives are higher in this method. Lee et al [7] used YOLOV2 deep learning algorithm for polyp segmentation. The volume of

training images was increased using data augmentation procedures. The images were downsized and processed by YOLOV2, due to this the small-sized polyps can be missed during detection. Also the approach has considered the problem of background similarity and polyp irregular shapes.

Ronneberger et al [8] segmented polyp regions using U-Net model. The architecture required less training volume. Though this architecture had good performance, it was not tested for the segmentation of smaller regions like a polyp. Zhang et al [9] extracted Densenet features and classified them for the presence of polyps using a softmax classifier. This method can only classify the image into two classes of polyp present or absent, and cannot provide the location of polyp. Livovsky et al [10] proposed two different deep learning architectures: RetinaNet and LSTM-SSD to segment polyp regions from videos. Though the polyp detection performance is higher, the volume of training data is higher. The method performs well for polyp lasting for a longer duration across video frames. Jha et al. [11-12] designed the ResUNet deep learning network for segmenting the polyps. This improved architecture provided better segmentation results with multiple residual and attention blocks. The improved ResUNet was able to achieve only 0.81 dice coefficient. Chen et al. [13] used convolutional neural networks for image segmentation. The network is made up of a deep convolutional network using convolution with upsampled filters, multiscale pyramid feature extraction and integration with probabilistic graphical networks. Though the method was able to achieve 79.7% mIOU, it was not tested against a complex environment like polyp segmentation. Srivastava et al [14] segmented polyps using deep learning network with a multi-scale residual architecture. Use of multi-scale features improved the segmentation performance. The feature fusion has allowed the solution to perform better even for small granular polyp regions. The method fails in the presence of low contrast and higher similarity to the background.

Dumitru et al [15] suggested DUCK-Net architecture, which performs accurate segmentation with lower training volume. DUCK-Net is an encoder-decoder structure with residual down-sampling and processing at multiple resolutions at the encoder side. The method can handle a polyp of various granularities. Fan et al [16] suggested PraNet, a deep learning architecture which uses two

stages of processing. At first stage, feature aggregation at higher layers with parallel partial decoding is done to establish a guidance area. In second stage, reverse attention is done to remove the background areas. Though this solution works for varied polyp size, homogeneous regions, and different kinds of texture, it was able to achieve Dice only in range of 0.7 to 0.87. Diakogiannis et al. [21] used residual connection architecture to enhance U-Net. But the method was not tested for polyp segmentation. Zhou et al [22] proposed UNet++, solving the problems in the previous Unet model. Skip connections were redesigned at the decoder end to aggregate multiscale features. Though this redesign improved the segmentation efficiency, it was not tested for a challenging environment like polyp segmentation. Huang et al [23] segmented polyp using encoder decoder architecture. Cascaded partial decoder is used at decoder end for multi scale feature aggregation. This is realized using skip connections. The method was not tested for case of higher similar background and irregular polyp shapes. Tomar et al [24] proposed feedback attention network. Recurrent learning is implemented at the encoder, decoder sides to improve the segmentation effectiveness. The feedback mechanism increases the strength of feature representation. Though the method was able to achieve mIoU of 0.8153, it was not tested for polyp segmentation. Valanarasu et al [25] segmented regions in the image using UNeXt network. It is an encoder-decoder structure with tokenized MLP generation at the encoder end. Feature resolution is increased at the decoder end with skip connections between the encoder and decoder. The effectiveness of this approach for polyp segmentation was not tested. Ige et al. [26] used ConvSegNet for segmenting images. It relied on context feature refinement with multiple kernel sizes to increase segmentation accuracy. But the method could not work for smaller granular polyp regions.

From the survey, it can be inferred that most deep learning solutions for segmentation are based on the encoder-decoder structure. Multi scale feature representations with residual learning were optimized in each solution to achieve effective segmentation. But the problem of similar background, irregular-shaped polyps occurring is close proximity to similar tissues, and low-contrast noises are effectively handled through preprocessing in existing works. As a result, these noises too get amplified in feature representation, resulting in reduced accuracy and higher false positives. This

problem of the impact of noise amplification factors on the segmentation accuracy and false positives is investigated in this work, and a solution is proposed for the same.

3. PROPOSED SOLUTION

The architecture of the suggested two-stage deep learning based technique for segmenting the polyp region is given in Figure 1. The background tissue similarity impact on segmentation is resolved in the first stage, and the output is given to the LIC-Net deep learning structure to localize the polyp segments.

3.1 System Methodology

In the first stage, TAN [27] is applied on the colonoscopy image. A mesh is placed on the image on size $L \times L$. For each node nx , energy is found in terms of external energy (E_e) and internal energy (E_i) as

$$e(nx) = \frac{E_e(nx)}{E_e(nx) + E_i(nx)} \quad (1)$$

External energy of node is computed in terms of average value of external energy of each pixel in its neighborhood region as

$$E_e(v(a, b)) = \omega f[I(v(a, b))] + \frac{\rho}{|Nk(a, b)|} \sum_{p \in N(a, b)} \frac{1}{\|v(a, b) - v(p)\|} f_i[I(v(p))] \quad (2)$$

Where $I(v(a, b))$ is the value of the intensity of the pixel at position $v(a, b)$. $Nk(a, b)$ represents the neighbors of the node at (a, b) . The function f_i divides the intensity values between the original and the gradient image at the position $v(a, b)$.

The internal energy is computed in terms of the average value of the internal energy of each pixel in its neighborhood region as

$$E_i(v(a, b)) = \alpha(|v_a(a, b)|^2 + |v_b(a, b)|^2) + \beta(|v_{aa}(a, b)|^2 + |v_{ab}(a, b)|^2 + |v_{bb}(a, b)|^2) \quad (3)$$

In the above equation α, β are the coefficients whose value is from 0 to 1.

The default TAN removes the mesh node by checking the energy of node against threshold which is not adaptive to image characteristics and background. So this work modifies it by applying fuzzy logic. The decision to remove the mesh node is made in terms of energy values of 8 neighboring nodes as shown in Figure 2 below. A training dataset of 9 feature vectors (energy values of corresponding nodes) and whether the node belongs to the foreground or background is created from the training images. The dataset is clustered using the fuzzy c-means clustering algorithm.

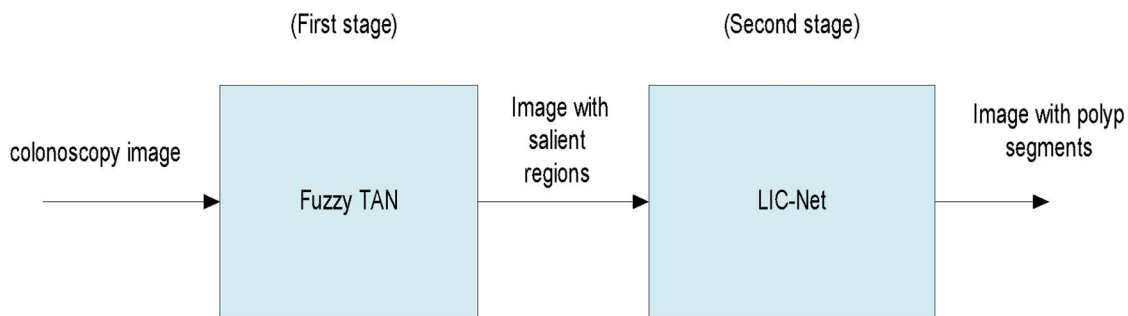


Figure 1: Two stage Solution Architecture

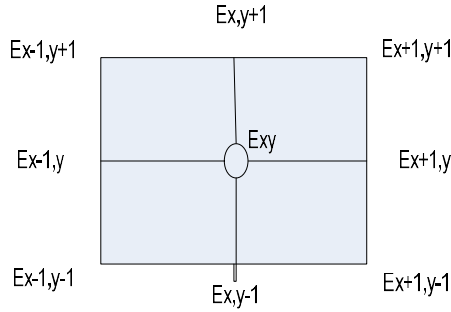


Figure 2: Eight Neighboring Nodes

The centroids of the cluster are defined in terms of coordinate (q) of the features belonging to cluster ($D_{e,q}$) as

$$D = \{D_{x,y}, x = 1,2 \text{ and } y = 1,2..9\} \quad (4)$$

Where x is cluster and y is feature coordinate.

The closeness of a feature point r to the cluster centroid $D_{x,y}$ is calculated in terms of Gaussian function to each feature coordinate as

$$G(f_{r,y}, D_{x,y}, \sigma_{x,y}) = e^{\frac{(f_{r,y} - D_{x,y})^2}{\sigma_{e,p}^2}} \quad (5)$$

Where

$$\sigma_{x,y} = \frac{1}{N_x} \sum_{r=1}^{N_x} (f_{r,y} - D_{x,y})^2 \quad (6)$$

Once the closeness to the feature coordinate is found, the closeness to the entire feature vector is calculated as the product of function on each coordinate as

$$\Psi_{r,x} = \prod_{y=1}^P G(f_{r,y}, D_{x,y}, \sigma_{x,y}) \quad (7)$$

$\Psi_{r,x}$ can also be calculated in terms of linear regression over feature coordinates as
In terms of fuzzy C mean clustering, the data point (r) membership to clusters is given as

$$\underline{N}(r) = \sum_{e=1}^P \Psi_{r,e} \Phi_{r,e} \quad (8)$$

Where

$$\Phi_{r,e} = W_{e,0} + \sum_{q=1}^P W_{e,q} f_{r,q} \quad (9)$$

In the above equation, W are the weights associated with feature coordinates in a linear regression setup.

Thus, two cluster membership functions are created, which take nine energy values as features and provide the membership value for the cluster (background or foreground). When the cluster membership value for the class background is higher compared to the foreground, the breakdown starts from this node. The node with the highest energy on its link to this node is removed. This process is repeated iteratively. A mask with 0 for the background area found by this method is created, and it is applied to the original image to get salient regions free from background for polyp segmentation.

The output of the first stage is passed to the second stage, where LIC-Net is invoked to segment the polyp. It uses pretrained ResNet50 at the encoder. It is used to extract numerous feature maps at various levels by using convolutions and downsampling processes. When the training samples are smaller, to get optimization without adding computational complexity, a pretrained network is used. Each feature map at four layers of ResNet50[11] is passed through skip connections having a basic 3X3 convolution. Following it, ReLU activation is done before decoder connectivity.

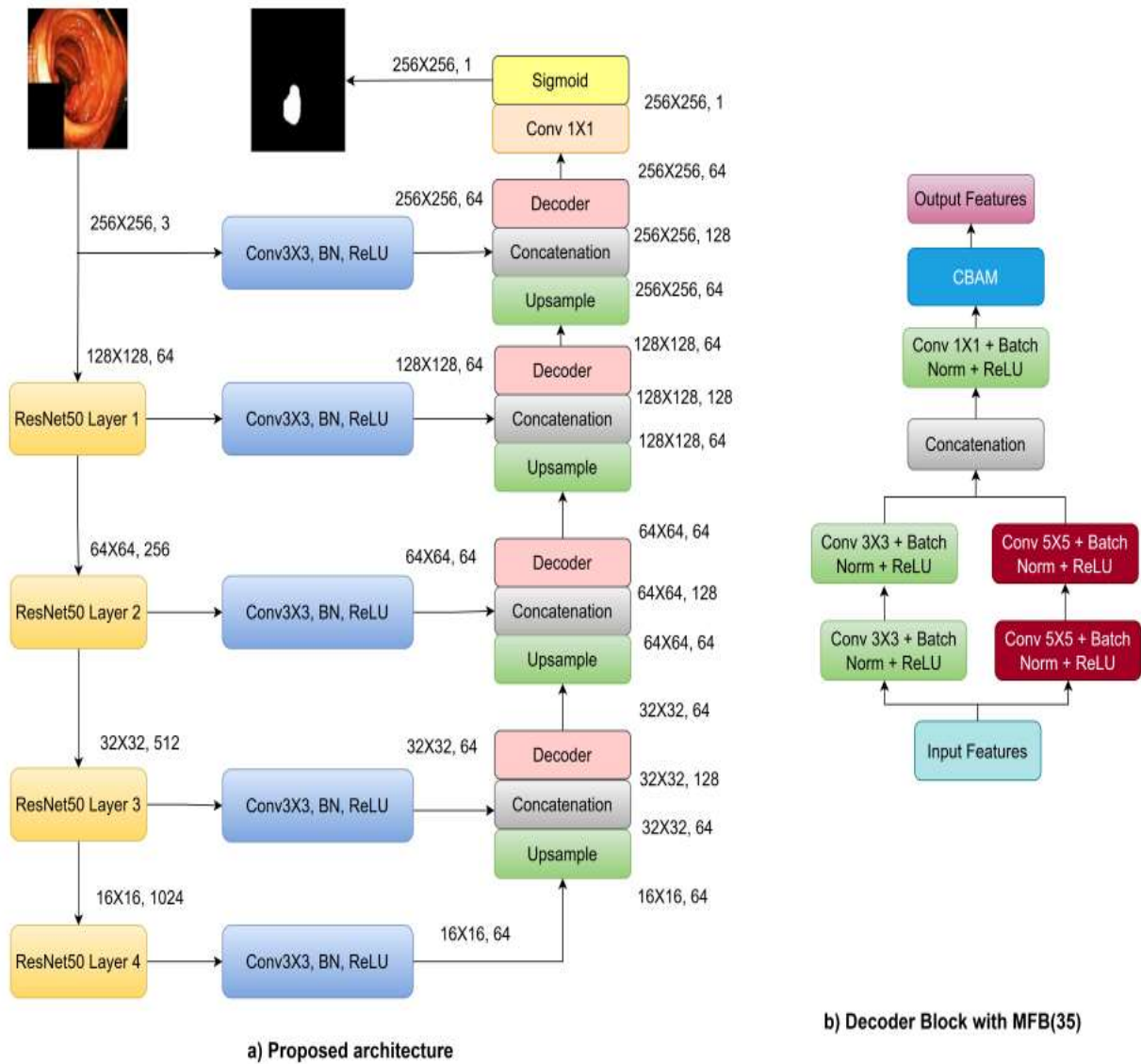


Figure 3: LIC-Net Architecture

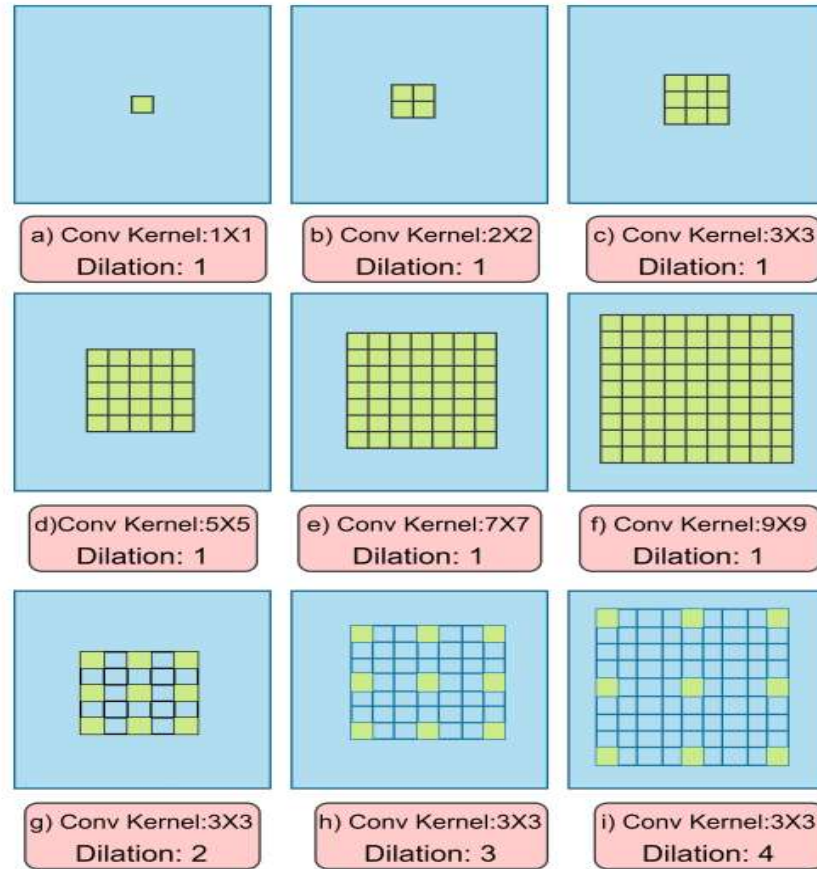


Figure 4: Various Convolutions with Variations of Kernel Sizes and Dilation Rates

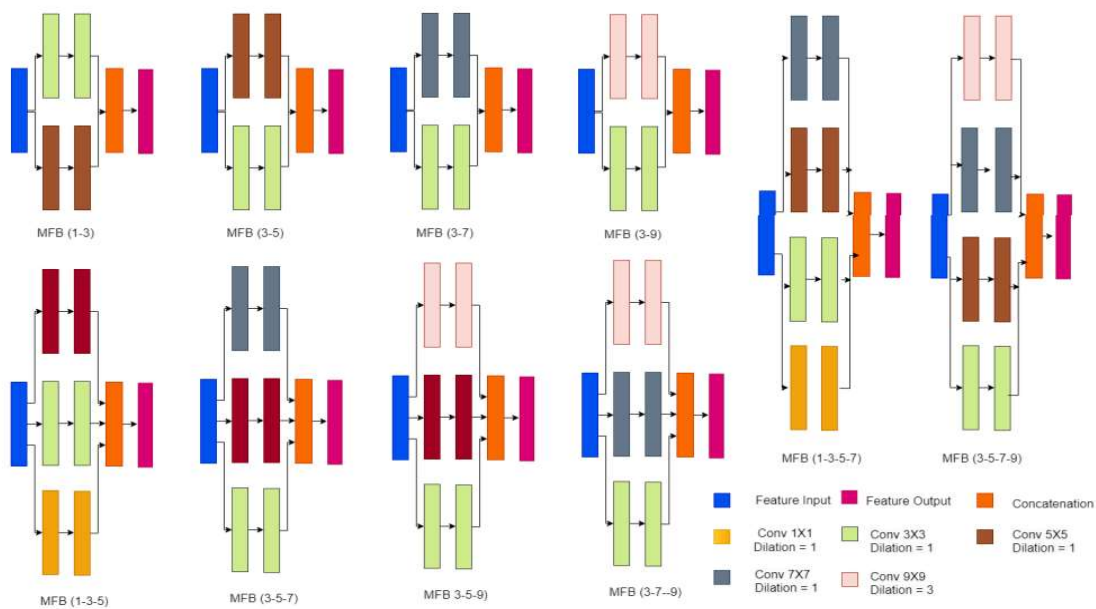


Figure 5: Summary of the Ten Multi-Feature Blocks (MFB) Used at the Decoder

Figure 3 summarizes the network structure. The preprocessed image of resized size (256X256, 3) is passed through the ResNet50 encoder. The initial layer sizes are (256X256, 3), (128X128, 64), (64X64, 126), (32X32, 512), and (16X16, 1024). These features are passed through a convolutional skip connection, producing (256 x 256, 64), (128 x 128, 64), (64 x 64, 64), (32 x 32, 64), and (16 x 16, 64), respectively. Due to the convolutional operation, every feature map at various levels of ResNet50 is converted to 64 numbers of channels at the decoder concatenation. Skip connections maintain the global parameters of polyps and usually handle the large-scale variation of polyps. A refinement module at the decoder with various combinations of convolutions and atrous dilation is used in this work. This lightweight refining module helps in the extraction of multi-scale information. In the network, the decoder MFB block is shown in Figure 3b. Input features from the prior decoder block are upsampled and concatenated with the skip connection of the respective layer of the encoder. Then these features in the decoder are convolved with different conv blocks in parallel. Conv blocks of dimension 3X3 and 5X5 are used, which achieved the maximum performance with less complexity. Then, further, these feature maps are joined and sent through a 1X1 conv layer, followed by BN and ReLU. To provide spatial and channel attention again, the convolution block attention module [18] is used. Combinations of convolutions (with and without dilation) in series and parallel are implemented to capture global and local semantic features. Further different kernel sizes are kept in the convolution layers to capture contextual information effectively.

Combining the convolution kernel with various receptive fields allowed for the development of ten different types of multi-feature blocks (MFB) with scale changes. The various MFBs are displayed in Figure 5. Different dilations are represented in Figure 4. a) to f) shows Convolution layers with dilation rate=1 and g) to i) Atrous Spatial Pyramids (ASPP) with dilation rate = 2,3 and 4. Via atrous convolution layers with various dilation rates, ASPP retrieves semantic information. Keeping the same complexity results in an increase in the receptive fields that stimulate convolution layers. As illustrated in Figure 4, kernels of various dimensions 1×1 , 3×3 , 5×5 , 7×7 , and 9×9 are implemented to filter the input channels to resolve this problem. The filtered channels are then concatenated. For MFB(1-3-3d2-5d2) it means it is a parallel network

consisting of 1X1, 3X3, and 3X3 with a dilation rate of 2 and 5X5 with a dilation rate of 2. Similarly, all other networks are designed. This design can help mitigate the issue of the undetermined structure of the segmented area. This method is suitable for demanding prediction tasks requiring precise spatial data. The model is trained using Kvasir-SEG and CVC-ClinicDB datasets, each of which divided into distinct subsets for testing, validation, and training. For training, augmented samples are used.

3.2 Algorithm of the Proposed LPC-Net Model

An algorithm describes the flow of the Polyp segmentation using LIC-Net. First it starts with input as a colonoscopy image of any size producing output image of 256X256X1 mask, indicating a polyp region.

Table 1: Algorithm for the Polyp Segmentation using LIC-Net

Input	Colonoscopy image (I)
Output	Binary segmentation mask (M) highlighting polyp region
Step1 : Preprocessing (Fuzzy TAN)	• Overlay mesh on image • Compute external & internal energy • Fuzzy c-means clustering • Remove high-energy background nodes • Generate saliency mask S and apply to image
Step 2 : Initialize Model, Load Pretrained ResNet-50 encoder	• Resize image to $256 \times 256 \times 3$ • Extract multiscale features at 4 levels
Step 3 : Skip Connections	• Apply 3×3 conv +BN+ ReLU on encoder outputs as skip connections. • Pass features to decoder via skip connections
Step 4 : Concatenation	• Concatenation of skip connections and lower branch after upsampling to match the Image size.
Step 5: Decoder (MFB Blocks)	• Concatenated branches to pass through Decoder • Use parallel convolutions (3×3, 5×5, etc.) in the decoder • Fuse with 1×1 conv $\rightarrow$ BN $\rightarrow$ ReLU

	• Add CBAM attention module
Step 6: Final Prediction	• Apply Conv 1X1 + sigmoid activation • Generate final binary mask (M) of size 256X256X1
<i>Training Details</i>	• Optimizer: Adam • Loss: Dice + BCE • LR: 1e-4 (with decay) • Epochs: 60 • Framework: PyTorch (Colab V100) • Batch Size: 16

collection range in size from 332×487 to 1920×1072 pixels.

CVC-ClinicDB [20] - It consists of 612 frame images extracted from colonoscopy videos. Each image has dimensions of 384×288 pixels and originates from 31 distinct colonoscopy sequences.

Annotating new images takes time and is costly, and creating it is a laborious process. High-quality annotations require expensive medical understanding. Privately generated datasets were used in some of the earlier research. Sharing medical data is challenging due to privacy and ethical concerns. There are not many publicly available medical image datasets. Due to the data hungry nature of CNN-based models, it is well acknowledged that larger data sets yield better results. Thus, we use data augmentation to make it more resilient. Data augmentation techniques such as coarse dropout, flipping in both directions, and random rotation were applied after resizing images and masks to 256×256 pixels, and the pixel values were normalized.

Six common metrics—Jaccard, Dice, Precision, Recall, Accuracy, and F2 measure—are used for performance measurement and comparison to existing works. In this work, the background area is denoted by the negative label of zero, and the polyp area is indicated by the positive label of one. Studies on ablation were performed on various combinations of MFB blocks that are used in the decoder. Also, a computational complexity comparison of the proposed model with benchmark architectures is stated at the end. A total of four Experiments were conducted. The datasets were divided in 80:10:10 and trained and tested for same as well as cross-validation to assess the proposed scheme.

3.3 Implementation details

The proposed system and state-of-the-art (SOTA) benchmark architectures were trained on Google Colab Pro with GPU V100 and implemented using the PyTorch framework. We trained these using the same hyperparameters to ensure a fair comparison. The Adam optimizer with a learning rate of 1×10^{-4} is used with an adaptive learning rate after every 10 epochs. The training process continued for 60 epochs for each model. For experimentation (which consists of a huge training process with various combinations of MFB blocks and different sets of datasets), we have trained only for 60 epochs. The loss function is dice loss and binary crossentropy, with a batch size of 16.

4. EXPERIMENTATION AND RESULTS

The proposed model's efficacy was demonstrated and then thoroughly tested using both qualitative and quantitative approaches. The characteristics of the two datasets used for performance evaluation are summarized in Table 2.

4.1 Datasets and preprocessing

We have used two datasets Kvasir-SEG[19] and CVC-ClinicDB [20] with a maximum number of images for training, to evaluate the proposed model.

Kvasir-SEG [19] -This dataset was acquired at the Norwegian Vestre Viken Health Trust and consists of endoscopic images that have been thoroughly annotated and verified. It contains the Kvasir-SEG subset, which concentrates on the polyp class. Kvasir-Seg offers bounding box information, matching masks, and 1000 polyp images obtained from electromagnetic imaging. The images in this

4.2 Quantitative Results

Table 3 shows the results with training and testing on the same datasets. On both datasets, the proposed model performed better than all the SOTA models. On Kvasir-Seg, Jaccard and Dice achieved 2.88% and 2.52%, respectively, more than the highest-performing network. On CVC-ClinicDB, 3.95% and 3.07%, which is much higher than ConvSegNet [26] and all other SOTA models.

Results for the cross datasets for both datasets are shown in Tables 4. On the cross dataset, the suggested model performed better than the

existing SOTA models in terms of all performance parameters.

4.3 Qualitative results:

Qualitative segmentation results for Kvasir-Seg are shown in Figure 6 and for CVC-ClinicDB in Figure 7 for the same as well as cross datasets. The test image, ground truth mask, U-Net [8] predicted output, and ConvSegNet [26] predicted output are shown in the first four columns, respectively. These models are compared with the proposed model's predicted mask, which is shown in the fifth column. The proposed model provides more accurate masks as ground truth compared to U-Net and ConvSegNet. A segmented polyp's green bounding box indicates the area that was accurately detected, and red indicates the area that is performing effectively. The results show that the model excels, particularly with small-size polyp datasets, highlighting its robustness. Qualitatively, our model's segmentation mask effectively captures fine details even when image quality is distorted. Although the segmentation may not be flawless, the mask's additional shape information enables potential correction using image post-processing techniques. When it comes to more challenging images, including flat and tiny polyps, which are typically overlooked during colonoscopy exams, the model performs rather well.

4.4 Ablation Study

The Kvasir-SEG dataset is used for an ablation analysis of MFB since it contains maximal polyp changes. Table 5 displays the experimental outcomes of various MFBs on the dataset. To construct the best network out of all of them, we also measured the number of trainable parameters for each of the ten networks. It has been observed following important conclusions: 1) Unnecessarily making the network more complex by increasing the number of convolutions in parallel doesn't improve network performance. 2) By adopting dilation, it also does not improve network performance. 3) MFB(3X5) performs better on all other configurations, and the number of trainable parameters have been reduced significantly over all other designs.

4.5 Computational Complexity

Table 6 provides a comparison of the suggested solution in comparison to SOTA models against the computational complexity metrics. The 11M parameter value obtained by the suggested model, when in comparison to SOTA models, makes it less complex than U-Net, HardDNet-MSEG, and

ConvSegNet. The suggested model has 43.37 Flops which is less than most of the SOTA models as stated in the table, which also proves it is reasonably better than these models. This number is also significantly better than almost all other models except HardNet-MSEG and U-Net++. The proposed model recorded 59 FPS, demonstrating here as well that it functions fairly well.

To illustrate the various network's learning properties, Figure 8 displays variations in training and validation loss, Jaccard, and F1 of two datasets. The effectiveness of the suggested model is noticeably better. These learning curves thus provide a summary of the fact that the proposed network converges quicker and approaches the optimal performing states in a significantly shorter amount of epochs.

5. DISCUSSIONS

The proposed LIC-Net framework demonstrates significant improvements in polyp segmentation, validated across two publicly available datasets: CVC-ClinicDB and Kvasir-SEG. These datasets offer diverse sets of polyp images with expert annotations, and the results confirm that the model achieves higher segmentation accuracy, reduced false positives, and improved computational efficiency compared to existing approaches.

Performance gains can be attributed to two main innovations. First, the fuzzy Topological Active Net (TAN) in the preprocessing stage effectively suppresses structurally similar background regions, enabling clearer polyp localization. Second, the LIC-Net architecture, integrating a ResNet50 encoder and Multi-Feature Block decoder with CBAM attention, enhances the model's ability to capture features across multiple scales and complexities.

The existing works against which the proposed model was compared were all based on the U-Net encoder-decoder structure. Prior efforts involved residual connections [21], multiscale aggregation [22, 23], recurrent feedback [24], MLP tokenization [25], and context enhancement [26]. The proposed model continues in this direction, incorporating ResNet50 at the encoder and a refined decoder, while uniquely combining it with fuzzy TAN as a preprocessing stage—an element overlooked in previous approaches.

Despite promising results, limitations exist. The reliance on public datasets may limit the generalization to diverse real-world clinical settings. Furthermore, the model has not been validated on embedded hardware platforms for real-time use.

Future research will focus on domain adaptation and semi-supervised learning to improve robustness in varied clinical environments. Additionally, efforts will be made to optimize the model for real-time deployment and validate its effectiveness through clinical trials.

6. CONCLUSION

A deep learning-based polyp segmentation technique is proposed in this work. The solution had two stages to improve segmentation efficiency. In the first stage, the fuzzy TAN model was used to remove the interference of background/ surrounding tissue similarities. In the second stage, LIC-Net is proposed to accurately segment the polyp. LIC-Net improved the U-Net architecture with ResNet50 at the encoder and MFB at the decoder end. As a result, the feature extraction from multiple receptive fields was effective. Performance comparison shows that

the proposed solution achieves 2% higher accuracy compared to existing deep learning models. In addition, the proposed solution has 20% lower computation complexity compared to U-Net structure.

Beyond colorectal polyp segmentation, this simple method could serve as a basis for various medical imaging tasks, including lesion or tumor segmentation in radiology and dermatology. Its low computational demand makes it suitable for real-time deployment on portable devices, especially in rural or resource-limited healthcare environments.

This work strives to set a first step towards the development of visually enabled AI based solutions for early diagnosis. Future work could be the extension of this framework to multi-class segmentation, validation on more datasets, and integration in embedded systems. An interdisciplinary cooperation between AI researchers and clinicians will be necessary to translate this solution into resource-limited health-care applications.

Table 2: The Specifics of the Datasets used in our Tests. The terms "Train," "Val," and "Test" Stand for the Quantity of Training, Validation and Testing Samples

Experiment	Training Dataset	Testing Dataset	Images	Size	Train	Val	Test
1	Kvasir-SEG	Kvasir-SEG	1000	720X576 to 1920X1072	792	88	120
2	CVC-ClinicDB	CVC-ClinicDB	612	384X288	490	61	61

Table 3: Comparison of Performance Metrics: Training and Testing on the Same Dataset (Kvasir-SEG and CVC-ClinicDB)

Kvasir-SEG						
Network	Dice	mIoU	Recall	Precision	Accuracy	F2
U-Net [8]	0.7350	0.8150	0.8340	0.8692	0.9465	0.8216
ResU-Net [21]	0.6634	0.7642	0.8025	0.8200	0.9341	0.7740
U-Net++ [22]	0.7419	0.8228	0.8437	0.8607	0.9491	0.8295
HardNet-MSEG [23]	0.7459	0.8260	0.8485	0.8652	0.9492	0.8358
FANet [24]	0.6941	0.7815	0.8452	0.8159	0.9220	0.8002
UNeXt [25]	0.6284	0.7318	0.7840	0.7656	0.9208	0.7507
ConvSegNet [26]	0.7987	0.8665	0.8922	0.8924	0.9637	0.8776
Proposed model	0.8275	0.8917	0.9261	0.8937	0.9685	0.9079

CVC-ClinicDB						
Network	Dice	mIoU	Recall	Precision	Accuracy	F2
U-Net [8]	0.8072	0.8734	0.8939	0.8840	0.9829	0.8834
ResU-Net [21]	0.7892	0.8648	0.8836	0.8804	0.9793	0.8722

ISSN: 1992-8645

www.jatit.org

E-ISSN: 1817-3195

U-Net++ [22]	0.8337	0.8913	0.9129	0.8988	0.9859	0.9026
HardNet-MSEG [23]	0.8388	0.8967	0.8929	0.9216	0.9871	0.8938
FANet [24]	0.7958	0.8625	0.8570	0.9151	0.9772	0.8569
UNeXt [25]	0.6676	0.7673	0.7546	0.8167	0.9722	0.7563
ConvSegNet [26]	0.8490	0.9083	0.9354	0.8980	0.9883	0.9205
Proposed model	0.8885	0.9390	0.9585	0.9245	0.9890	0.9501

Table 4: Comparison of Performance Metrics on Cross Dataset

Training: Kvasir-SEG and Testing : CVC-ClinicDB

Network	Dice	mIoU	Recall	Precision	Accuracy	F2
U-Net [8]	0.5514	0.6382	0.6888	0.8039	0.9546	0.6571
ResU-Net [21]	0.4967	0.5970	0.6210	0.8005	0.9465	0.5991
U-Net++ [22]	0.5475	0.6350	0.6933	0.7967	0.9504	0.6556
HardNet-MSEG [23]	0.6057	0.6960	0.7173	0.8528	0.9592	0.7010
FANet [24]	0.5345	0.6306	0.7707	0.6957	0.9283	0.6762
UNeXt [25]	0.3901	0.4915	0.6125	0.6609	0.9216	0.5318
ConvSegNet [26]	0.7178	0.7960	0.8220	0.8124	0.9737	0.8070
Proposed model	0.7208	0.7983	0.8184	0.8563	0.9623	0.8055

Training: CVC-ClinicDB and Testing: Kvasir-SEG

Network	Dice	mIoU	Recall	Precision	Accuracy	F2
U-Net [8]	0.3330	0.4592	0.8051	0.4095	0.7327	0.5648
ResU-Net [21]	0.2789	0.4000	0.8801	0.3087	0.6293	0.5348
U-Net++ [22]	0.3489	0.4692	0.8294	0.4095	0.7143	0.5772
HardNet-MSEG [23]	0.4338	0.5521	0.7585	0.5479	0.8142	0.6128
FANet [24]	0.4110	0.5189	0.8656	0.4762	0.7138	0.6163
UNeXt [25]	0.3163	0.4363	0.7203	0.4175	0.7475	0.5204
ConvSegNet [26]	0.6080	0.7156	0.9106	0.6664	0.8911	0.7835
Proposed model	0.6595	0.7573	0.8871	0.7344	0.9190	0.8013

Table 5: Multi-Feature Blocks(MFB) at the Decoder

Network	Trainable Parameters	Dice	mIoU	Recall	Precision	Accuracy	F2
MFB(1-3)	10,150,986	0.8054	0.8935	0.9209	0.8745	0.9623	0.8923
MFB(3-5)	11,330,634	0.8275	0.8917	0.9261	0.8937	0.9685	0.9079
MFB(3-7)	12,510,282	0.8205	0.8902	0.9266	0.8757	0.9642	0.9003
MFB(3-9)	14,083,146	0.8178	0.8830	0.9233	0.8838	0.9679	0.9010
MFB(1-3-5)	11,397,706	0.8048	0.8759	0.9170	0.8779	0.9650	0.8950
MFB(3-5-7)	13,876,298	0.8027	0.8752	0.9142	0.8739	0.9620	0.8902
MFB(3-5-9)	15,449,162	0.7948	0.8670	0.9102	0.8629	0.9608	0.8901
MFB(3-7-9)	16,509,514	0.8016	0.8715	0.9197	0.8715	0.9634	0.8919
MFB(1-3-5-7)	13,824,074	0.8111	0.8819	0.9233	0.8791	0.9669	0.9002
MFB(3-5-7-9)	17,756,234	0.7873	0.8627	0.9105	0.8592	0.9595	0.8838
MFB	11,857,994	0.8068	0.8765	0.9234	0.8727	0.9626	0.8966
MFB(1-3d2-5d2)	11,397,706	0.8025	0.8734	0.9266	0.8675	0.9648	0.8964
MFB(3-7-9d2)	16,509,514	0.8054	0.8762	0.9258	0.8756	0.9639	0.8988

MFB(1-3-5d4)	11,397,706	0.8030	0.8732	0.9164	0.8756	0.9638	0.8909
--------------	------------	--------	--------	--------	--------	--------	--------

Table 6: Computational Complexity Comparison of the Proposed Model with state-of-the-art Models

Network	Dice	mIoU	Recall
U-Net [8]	31.04	54.75	156.83
ResU-Net [21]	8.22	45.42	196.85
U-Net++ [22]	9.16	34.65	126.14
HardNet-MSEG [23]	33.34	6.02	42
FANet [24]	7.72	94.75	44
UNeXt [25]	1.47	569.56	88.89
ConvSegNet [26]	15.58	135.98	64
Proposed method	11.45	43.37	59.82

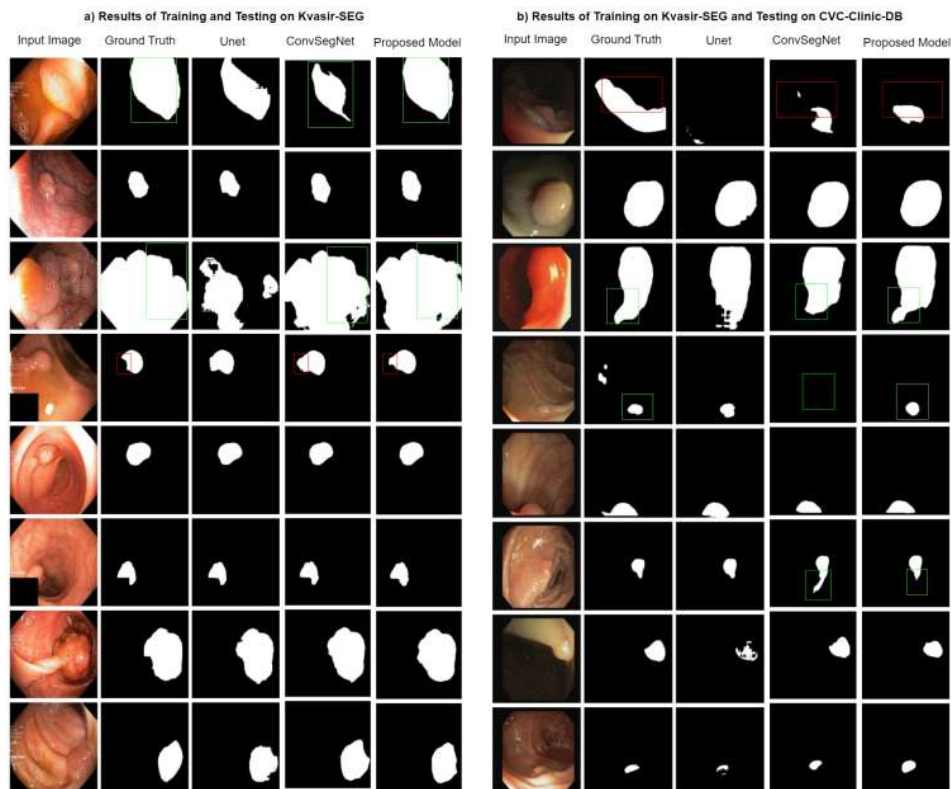


Figure 6: Qualitative Analysis of the Input Colonoscopic Images on Kvasir-SEG

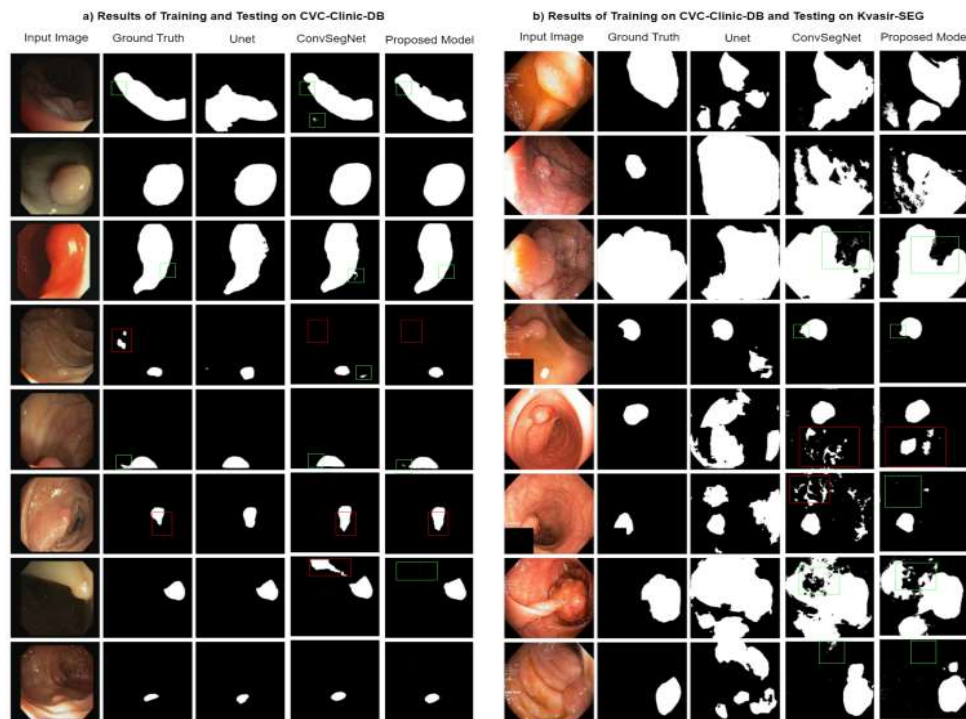


Figure 7: Qualitative Analysis of the Input Colonoscopic Images on CVC-ClinicDB

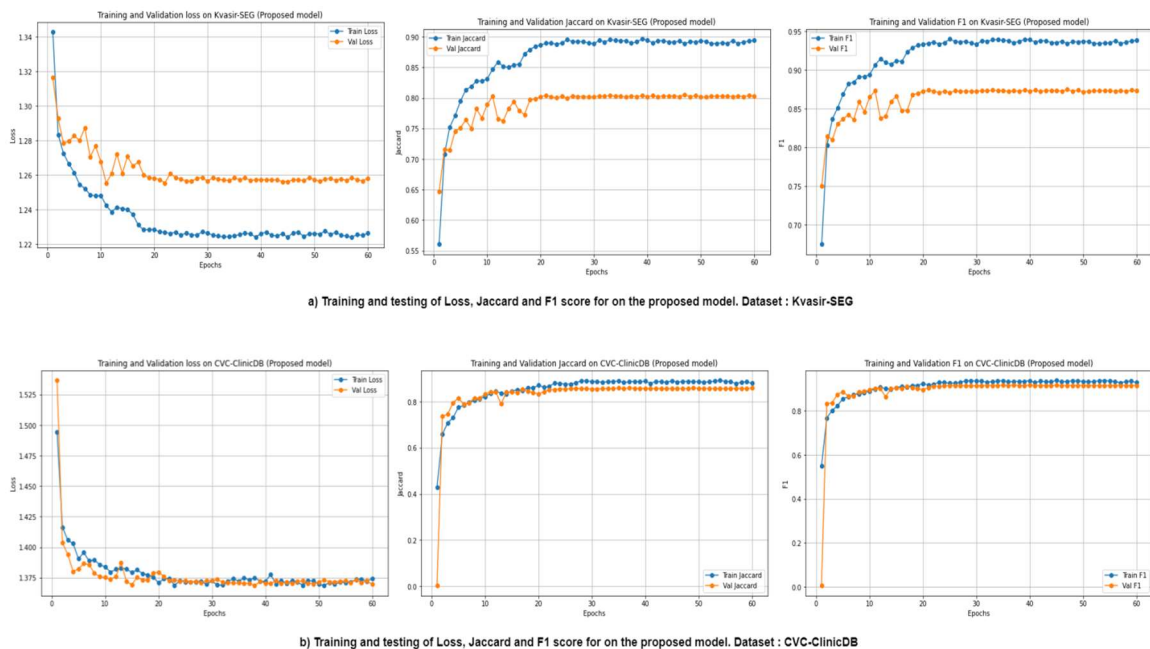


Figure 8: Training and Testing Curves a) Kvasir-SEG b) CVC-ClinicDB

ACKNOWLEDGMENT

The authors sincerely thank the individual who created the publicly available dataset for any

insightful conversations and informative assistance in the expanding field of medical image analysis.

DECLARATIONS

- Funding: No funding from any institutes.
- Conflict of interest: The authors declare that they have no competing interests.
- Ethics approval and consent to participate: Ethical approval Not applicable
- Author contributions: All authors have made contributions to the conception and design, analysis, and interpretation of data and they have been involved in drafting the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT (DAS)

- All the datasets used are publicly available at the following links.

<https://www.kaggle.com/datasets/balraj98/cvcclinicdb>

<https://datasets.simula.no/kvasir-seg/>

- Code availability: Code can be made available on request.

REFERENCES

- [1]. R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal, Cancer statistics, 2023, *Ca Cancer J Clin* 73 (1) (2023) 17–48.
- [2]. J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, F. Bray, Cancer incidence and mortality patterns in europe: Estimates for 40 countries and 25 major cancers in 2018, *European journal of cancer* 103 (2018) 356–387.
- [3]. D. A. Corley, C. D. Jensen, A. R. Marks, W. K. Zhao, J. K. Lee, C. A. Doubeni, A. G. Zauber, J. de Boer, B. H. Fireman, J. E. Schottinger, et al., Adenoma detection rate and risk of colorectal cancer and death, *New england journal of medicine* 370 (14) (2014) 1298–1306.
- [4]. J.Yao, M. Miller, M. Franaszek, R. M. Summers, Colonic polyp segmentation in ct colonography-based on fuzzy clustering and deformable models, *IEEE Transactions on Medical Imaging* 23 (11) (2004) 1344–1352.
- [5]. S. Gross, M. Kennel, T. Stehle, J. Wulff, J. Tischendorf, C. Trautwein, T. Aach, Polyp segmentation in nbi colonoscopy, in: *Bildverarbeitung für die Medizin 2009: Algorithmen—Systeme—Anwendungen Proceedings des Workshops vom 22. bis 25. März 2009 in Heidelberg*, Springer, 2009, pp. 252–256.
- [6]. P. Wang, X. Xiao, J. R. Glissen Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, et al., Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy, *Nature biomedical engineering* 2 (10) (2018) 741–748.
- [7]. Lee, J.Y., Jeong, J., Song, E.M. et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. *Sci Rep* 10, 8379 (2020)
- [8]. O. Ronneberger, P. Fischer, T. 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, Springer, 2015, pp. 234–241
- [9]. Zhang QW, Zhang Z, Xu J, Dai ZH, Zhao R, Huang J, Qiu H, Tang ZR, Niu B, Zhang XB, Wang PF, Yang M, Deng WY, Lin YS, Xiang S, Ge ZZ, Qian D, Li XB. Multi-step validation of a deep learning-based system with visual explanations for optical diagnosis of polyps with advanced features. *iScience*. 2024 Mar 8;27(4):109461.
- [10]. Livovsky DM, Veikherman D, Golany T, Aides A, Dashinsky V, Rabani N, Ben Shimol D, Blau Y, Katzir L, Shimshoni I, Liu Y, Segol O, Goldin E, Corrado G, Lachter J, Matias Y, Rivlin E, Freedman D. Detection of elusive polyps using a large-scale artificial intelligence system (with videos). *Gastrointest Endosc*. 2021 Dec;94(6):1099-1109.e10.
- [11]. D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H. D. Johansen, Resunet++: An advanced architecture for medical image segmentation, in: *2019 IEEE international symposium on multimedia (ISM)*, IEEE, 2019, pp. 225–2255.
- [12]. D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, H. D. Johansen, Doubleu-net: A deep convolutional neural network for medical image segmentation, in: *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*, IEEE, 2020, pp. 558–564.
- [13]. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE transactions on pattern analysis and*

- machine intelligence 40 (4) (2017) 834–848.
- [14]. A. Srivastava, D. Jha, S. Chanda, U. Pal, H. D. Johansen, D. Johansen, M. A. Riegler, S. Ali, P. Halvorsen, Msrf-net: a multi-scale residual fusion network for biomedical image segmentation, *IEEE Journal of Biomedical and Health Informatics* 26 (5) (2021) 2252–2263.
- [15]. R.-G. Dumitru, D. Peteleaza, C. Craciun, Using duck-net for polyp image segmentation, *Scientific Reports* 13 (1) (2023) 9803.
- [16]. D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Pranet: Parallel reverse attention network for polyp segmentation, in: *International conference on medical image computing and computer-assisted intervention*, Springer, 2020, pp. 263–273.
- [17]. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.
- [18]. S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, in: *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [19]. D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H. D. Johansen, Kvasir-seg: A segmented polyp dataset, in: *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*, Springer, 2020, pp. 451–462.
- [20]. J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, *Computerized medical imaging and graphics* 43 (2015) 99–111.
- [21]. F. I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu, Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data, *ISPRS Journal of Photogrammetry and Remote Sensing* 162 (2020) 94–114.
- [22]. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, *IEEE transactions on medical imaging* 39 (6) (2019) 1856–1867.
- [23]. C.-H. Huang, H.-Y. Wu, Y.-L. Lin, Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps, *arXiv preprint arXiv:2101.07172* (2021).
- [24]. N. K. Tomar, D. Jha, M. A. Riegler, H. D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, S. Ali, Fanet: A feedback attention network for improved biomedical image segmentation, *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [25]. J. M. J. Valanarasu, V. M. Patel, Unext: Mlp-based rapid medical image segmentation network, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 23–33.
- [26]. A. O. Ige, N. K. Tomar, F. O. Aranuwa, O. Oriola, A. O. Akingbesote, M. H. M. Noor, M. Mazzara, B. S. Aribisala, Convsegnet: Automated polyp segmentation from colonoscopy using context feature refinement with multiple convolutional kernel sizes, *IEEE Access* 11 (2023) 16142–16155.
- [27]. M. Bro-Nielsen, “Active nets and cubes,” IMM, Tech. Rep., 1994.