ISSN : 2455-135X

DANet: Attention-based Dilated Network for Medical Image Segmentation

Wangkheirakpam Reema Devi*, Sudipta Roy and Khelchandra Thongam

Abstract

Colorectal cancer is a significant public health problem worldwide, and early detection is very necessary. However, the miss rate during routine colonoscopy examinations is very high, leading to undiagnosed polyps that can develop into colorectal cancer. Therefore, we proposed a novel architecture called Attention-based Dilated Network (DANet) for automatic polyp segmentation using convolutional neural networks (CNN). DANet uses a pre-trained ResNet50 proposed by He, Kaiming, et al [8] as an encoder and a Dilated Attention Convolution (DAC) block in between the encoder and decoder to learn a more robust feature representation. We evaluated DANet's performance on four datasets namely KvasirSEG , CVC-ClinicDB , ETIS-Larib PolypDB and KvasirInstrument datasets and found that it outperformed state-of-the-art methods in terms of standard segmentation metrics, such as Jaccard score, F1 score, recall, accuracy, and F2 score. The superiority of DANet's performance can be attributed to the effective combination of the pre-trained ResNet50 proposed by He, Kaiming, et al., [8] and the DAC block, which enables the network to learn long-range dependencies and focus on informative regions in the input image. This feature enables the network to segment the polyps accurately. In conclusion, the proposed DANet architecture is a promising solution for automatic polyp segmentation in the medical domain. Automated polyp segmentation methods such as DANet have the potential to improve detection rates and reduce miss rates, facilitating early detection and prevention of colorectal cancer, which can have a significant impact on public health.

Index Terms : DANet, Colonoscopy, DAC, ResNet50, CNN.

I. INTRODUCTION

Colorectal cancer is a significant public health concern due to its high incidence and mortality rates. Early detection of colorectal polyps, which can potentially develop into cancer, is crucial for timely intervention and improved patient outcomes. Colonoscopy is one of the primary screening methods, but it relies on human interpretation, which can be error-prone. Detecting polyps during colonoscopy is challenging for several reasons. Polyps often resemble normal tissue, making them difficult to distinguish visually. Polyps come in various shapes, sizes, and colours, adding complexity to the detection process. Presence of stool, bubbles, and other objects in the colon can obscure polyps and hinder accurate detection. Deep learning, particularly convolutional neural networks (CNNs), has shown promise in addressing these challenges. The U-Net architecture, introduced by Ronneberger et al., is widely used for biomedical image segmentation. It leverages an encoder-decoder structure to learn high-level features from input images and generate segmentation maps. This architecture has been adapted for colonoscopy images to identify polyps accurately. Zhang et al.'s ResU-Net is an extension of the U-Net architecture that incorporates residual blocks. Residual blocks are introduced to mitigate the vanishing gradient problem, a common issue in training deep neural networks. This modification helps the model learn more effectively and improves its ability to capture intricate features in colonoscopy images. Deep learning models can achieve high accuracy in identifying polyps, reducing the miss rate associated with human-dependent colonoscopy. These models offer consistency in their analysis, reducing the potential for human error. Deep learningbased systems can provide real-time feedback during colonoscopy procedures, assisting clinicians in identifying polyps. The field of medical image analysis continues to evolve. Future research may focus on. Incorporating additional information such as texture, motion, or other imaging modalities to enhance detection accuracy. Developing systems that seamlessly integrate with colonoscopy equipment to provide instant feedback to clinicians. Conducting extensive clinical trials and validations to ensure the safety and effectiveness of deep learning models in real-world healthcare settings.

Deep learning, particularly architectures like U-Net and ResU-Net, has made significant strides in improving the early detection of colorectal polyps during colonoscopy examinations. These advancements hold great promise for reducing the burden of colorectal cancer and improving patient outcomes.

DeepLabv3+ extends DeepLabv3 with an efficient decoder module and an ASPP block to improve its contextual comprehension and feature map quality. Tomar et al. presented a feedback attention network that amalgamates the residual block, SENet, and a new MixPool block to iteratively enhance the predicted mask with prior epoch data. The DANet structure extends the U-Net foundation, employing a ResNet50 encoder, and introduces an innovative dilated attention convolution module to create robust feature maps. Precise and effective polyp segmentation is vital for the timely identification and management of colorectal cancer, ranking as the third most widespread cancer worldwide. ResUNet++ by Jha et al. enhances the ResU-Net architecture by integrating residual networks, the Squeeze and Excitation Network (SENet), Atrous Spatial Pyramid Pooling (ASPP), and attention mechanisms. The ASPP block comprises several parallel dilated convolution layers, which serve to expand the convolution kernel's field of view and contribute to the generation of more robust feature maps. The incorporation of an attention mechanism is pivotal as it aids in pinpointing and emphasizing relevant regions, thereby enhancing the quality of segmentation outcomes. ResUNet++[24] has exhibited superior performance when compared to other state-of-the-art architectures, particularly in the realm of polyp segmentation. Another well-regarded approach for biomedical image segmentation is U-Net++, as originally proposed by Zhou et al. [12]. U-Net++[12] harnesses the advantages of skip connections and takes a novel approach by integrating dense skip connections to bridge the semantic gap between the encoder and decoder feature representations. These dense skip connections facilitate smoother information flow between the two, thereby preserving intricate image details effectively. U-Net++ has consistently showcased improved performance across a spectrum of biomedical image segmentation tasks, including the challenging task of polyp segmentation. PraNet, as introduced by Fan et al. [13], represents a Res2Net-derived structure tailored for the specific challenge of polyp segmentation. PraNet capitalizes on the combined strengths of the parallel partial decoder and parallel reverse attention mechanisms, both of which prioritize the enhancement of polyp boundary delineation. The parallel partial decoder contributes to capturing intricate image details, while the parallel reverse attention mechanism directs focus towards crucial regions. Additionally, PraNet incorporates deep supervision, a technique that enables the network to assimilate insights from multiple scales of features, thus elevating the quality of segmentation outcomes. Another contender in the realm of polyp segmentation is PolypSeg+, as devised by Wu et al. [3]. PolypSeg+ takes a lightweight approach, meticulously designed to enable real-time polyp segmentation. Its arsenal includes an adaptive scale context module, complemented by an attention mechanism, meticulously tailored to tackle the substantial scale variations among polyps. The adaptive scale context module empowers the network to dynamically adjust feature scales according to polyp sizes, an asset in refining segmentation accuracy. Furthermore, PolypSeg+ leverages an efficient global context module to fuse low-level and high-level features, ensuring the preservation of fine details and global contextual information within the image. This fusion is further fine-tuned through a lightweight feature pyramid fusion module. In summation, the field of polyp segmentation is enriched with diverse deep learning-based architectures, each harnessing techniques like residual networks, attention mechanisms, ASPP, deep supervision, and dense skip connections to enhance segmentation performance. The continuous evolution of more effective and precise polyp segmentation frameworks remains pivotal, as it holds the potential to facilitate timely identification and intervention in cases of colorectal cancer, potentially leading to life-saving measures and improved patient outcomes.

Motivated by the triumphs of deep learning in biomedical image segmentation, we introduced an innovative architecture termed Attention-based Dilated Network (DANet) for the task of polyp segmentation. DANet adopts an encoder-decoder framework, integrating the widely utilized pretrained ResNet50 model alongside the Dilated Attention Convolution (DAC) block. ResNet50, a pre-trained network, has exhibited remarkable performance across various computer vision tasks, including biomedical image segmentation. The DAC block is an inventive attention mechanism that incorporates dilated convolution to acquire a more resilient feature representation. The primary objective of this architecture is to elevate the precision of polyp segmentation in colonoscopy images. The DANet architecture comprises two fundamental components: the encoder and the decoder. The encoder leverages the pre-trained ResNet50, renowned for its efficacy in acquiring deep image representations, enhancing the feature extraction process. Meanwhile, the decoder incorporates a blend of up-sampling layers and multiple residual blocks. Up-sampling layers are employed to elevate the feature map resolution to match that of the input image, while residual blocks refine these feature maps and establish crucial skip connections bridging the encoder and decoder. The distinguishing element of the DANet architecture is the Dilated Attention Convolution (DAC) block, strategically positioned between the encoder and decoder. The DAC block is a composite of dilated convolution and an attention mechanism, enriching the learning of robust and efficient feature maps. Dilated convolution, a convolution variant with an expanded receptive field but without an upsurge in parameters, captures broader global information from the image, enhancing its utility in polyp segmentation. The attention mechanism prioritizes pertinent image regions and suppresses irrelevant areas, thereby bolstering the accuracy of polyp segmentation. To assess the effectiveness of the novel DANet architecture, we conducted experiments using four publicly accessible biomedical image segmentation datasets: CVC-ColonDB, EndoScene, CVCClinicDB, and ETIS-Larib. These datasets contain diverse polyps with varying shapes, sizes, and colors, presenting a challenge due to the polyps' visual similarity to the surrounding tissue. We compared DANet's performance against four leading models in the field, namely UNet, DeepLabv3+, PraNet, and PolypSeg+. The outcomes revealed that DANet consistently outperforms these benchmark models across all four datasets. This superior performance is primarily attributed to DANet's utilization of ResNet50 as an encoder, the incorporation of DAC (Dilated Attention Convolution) blocks, and its comprehensive evaluation across multiple datasets. Our proposed approach effectively addresses the complexities of polyp segmentation in colonoscopy images, surpassing the current state-of-the-art methods on publicly available benchmark datasets. The key innovation in DANet lies in its application of dilated convolutions within the DAC block, enabling the capture of multi-scale information, a critical aspect for the precise identification of polyps with varying characteristics. Our method's success is further reinforced by the adoption of a pre-trained

ResNet50 as the underlying neural network. Our experiments have demonstrated that DANet excels at accurately delineating polyps in colonoscopy images, even when dealing with small or challenging-to-distinguish polyps embedded within surrounding tissue. The DAC block's ability to leverage multi-scale information contributes to the precise segmentation of polyps of diverse dimensions, while the attention mechanism highlights critical image regions, further enhancing segmentation precision. In summary, the Attention-based Dilated Network (DANet) architecture we propose is an innovative encoder-decoder design that leverages the power of pre-trained ResNet50 and the Dilated Attention Convolution (DAC) block for biomedical image segmentation, particularly in polyp detection in colonoscopy images. Our approach consistently outperforms existing methods on four widely recognized benchmark datasets, thanks to the efficient capture of multi-scale contextual information through dilated convolutions and the enhancement of segmentation accuracy via the attention mechanism. Overall, our method holds great promise for improving the accuracy of clinical diagnosis, facilitating more effective and timely treatment of polyps, and ultimately reducing the incidence of colorectal cancer.

III. METHOD

The Attention-based Dilated Network (DANet) is a neural network architecture that combines elements of the U-Net architecture with some unique adjustments. In the network's initial encoding phase, it makes use of the ResNet50 design, a widely utilized pre-trained convolutional neural network (CNN) primarily designed for image classification tasks. The ResNet50 model has undergone extensive training on diverse datasets like ImageNet, showcasing its exceptional ability to extract intricate and meaningful features from images. Following the encoder network, there's the dilated attention convolution module, which serves to enhance the feature maps while also capturing valuable global contextual information. In essence, DANet merges the strengths of U-Net and ResNet50, making it a potent tool for various image-related tasks, thanks to its ability to harness rich feature representations and context.

A. Dilated Attention Network

The input image is processed through the ResNet50 [8] encoder, which has been pre-trained on the ImageNet classification challenge dataset. ResNet50 [8] is a member of the ResNet family known for its utilization of residual blocks, which consist of two 3x3 convolutional layers and an identity mapping, also referred to as a shortcut connection. This identity mapping establishes an alternate path connecting the input and output of the convolutional layers, facilitating better gradient flow during back-propagation. The output from ResNet50 [8] then proceeds to the Dilated Attention Convolution block, incorporating multiple parallel convolution layers along with channel-wise and spatial attention mechanisms. The use of dilated convolution layers expands the receptive field of the convolutional kernel, enabling the capture of more information and thus enhancing performance. Subsequently, the output from the Dilated Attention Convolution block is directed to the decoder section of the network, which commences with an up-sampling layer employing bilinear interpolation to increase the feature map's dimensions by a factor of 4. Following this, a concatenation operation merges the up-sampled feature map with low-level information from the encoder through a skip connection. This skip connection is vital for retaining low-level features that may otherwise be lost due to the network's depth and serves as a shortcut for gradient flow during back-propagation. The concatenated feature map then traverses a residual block to learn essential semantic features crucial for generating a high-quality segmentation mask. Afterward, the output from the residual block undergoes another round of up-sampling through a bilinear up-sampling layer, followed by a residual block, a 1x1 convolution layer, and a sigmoid activation function. The output of the sigmoid activation function serves as the predicted mask corresponding to the input image.

B. Dilated Attention Convolution block

The Dilated Attention Convolution (DAC) block plays a pivotal role within the innovative Attentionbased Dilated Network (DANet) architecture. Its primary objective is to amplify the receptive field of convolutional kernels, enabling the capture of more extensive global information from input feature maps. To achieve this, the DAC block employs four parallel convolutional layers, with each layer having an incrementally higher dilation rate. The dilation rate dictates the spacing between values in the kernel, facilitating a broader coverage of the input feature map as it increases. Specifically, the first layer employs a 1x1 kernel size, while the subsequent three layers utilize a 3x3 kernel size with dilation rates of 1, 3, 5, and 7. These convolutional layers' output feature maps are concatenated and then passed through a 1x1 convolutional layer, which serves to reduce the number of feature maps and overall dimensionality. This concatenated output undergoes further refinement through batch normalization and a subsequent ReLU activation function. To further enhance DAC block performance, two essential attention mechanisms come into play. The channel attention mechanism selectively assigns weight to different channels in the output feature maps. This is achieved by first obtaining a channel descriptor vector through global average pooling and then passing it through two fully connected layers with ReLU activation. The resultant output from these layers is then multiplied with the input feature maps, yielding weighted feature maps. Additionally, a spatial attention mechanism contributes to refining feature representation by utilizing a 1x1 convolutional layer to obtain a spatial descriptor vector, which is subsequently applied to the input feature maps. In summation, the DAC block within the DANet architecture is meticulously designed to capture global features and enhance the feature representation of input feature maps. By integrating parallel dilated convolutional layers, batch normalization, ReLU activation, and attention mechanisms, the DAC block significantly augments network performance, leading to state-of-the-art results in tasks such as polyp segmentation.

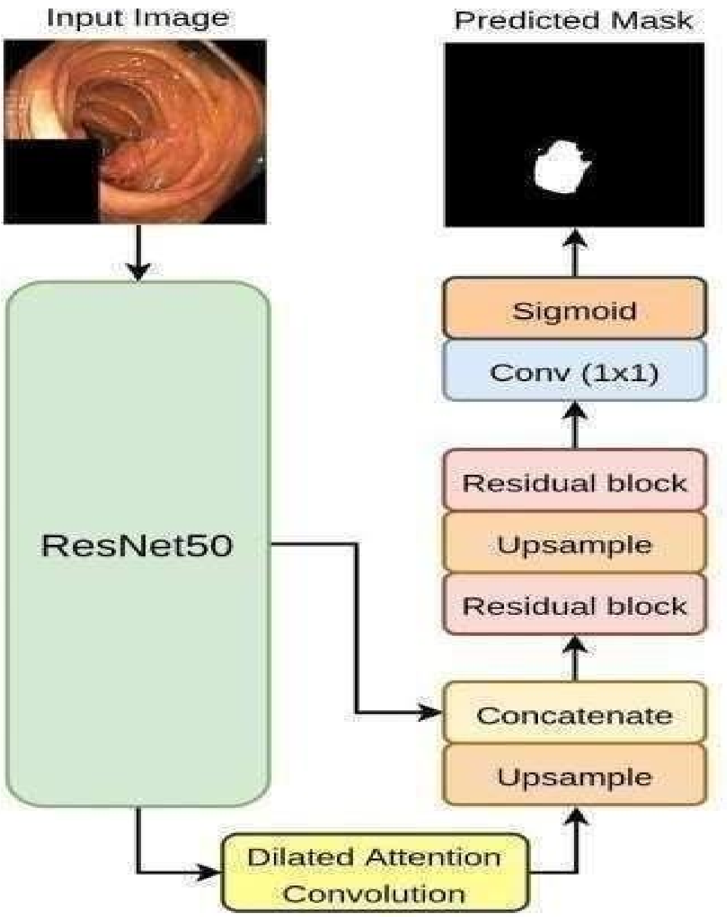

Fig. 1. The block diagram of the proposed Attention-based Dilated Network (DANet)

IV. EXPERIMENTAL SET UP

This section demonstrates the dataset used in the experimentation, implementation details and the performance metrics used to evaluate the performance of the different models.

A. Datasets and Evaluation Metrics

The CVC-ClinicDB dataset, as described in reference [23], comprises a total of 612 images captured during colonoscopy procedures. These images have a resolution of 576 x 720 pixels and encompass a diverse range of polyp types, including hyperplastic, adenomatous, sessile serrated adenoma, and traditional serrated adenoma. The dataset is partitioned into two subsets: CVCClinicDB-train, which consists of 450 training images, and CVCClinicDB-test, containing the remaining 162 images for testing. The masks for ground truth annotations have been meticulously crafted through manual annotation by an expert gastroenterologist.

Similarly, the Kvasir-SEG dataset, referenced as [16], consists of 1,000 colonoscopy images with a resolution of 640 x 480 pixels. These images encompass various polyp types, including hyperplastic, adenomatous, and serrated adenomas. The dataset is divided into Kvasir-SEGtrain with 800 training images and Kvasir-SEG-test with 200 testing images. Expert gastroenterologists have painstakingly annotated ground truth masks for accurate segmentation. Moreover, the ETIS-Larib PolypDB dataset, detailed in reference [17], comprises 300 colonoscopy images with a resolution of 576 x 720 pixels, containing different polyp types such as hyperplastic, adenomatous, and serrated adenomas. The dataset is split into three subsets: ETIS-Larib-train (240 images), ETIS-Larib-val (30 images), and ETIS-Larib-test (30 images).

The ground truth masks have been meticulously created by an expert gastroenterologist for precise evaluation.

Lastly, the Kvasir-Instrument dataset, referenced as [18], encompasses a vast collection of 5,000 colonoscopy images with a resolution of 1920 x 1080 pixels. These images showcase various endoscopic instruments, including biopsy forceps, coagulation forceps, and snare. The dataset is partitioned into Kvasir-Instrument-train, which contains 4,000 training images, and Kvasir-Instrument-test, comprising the remaining 1,000 images designated for testing. Expert gastroenterologists have carried out manual annotations to create accurate ground truth masks. To assess the performance of models on these datasets, standard segmentation metrics such as Jaccard (mIoU), F1 (Dice Coefficient), Recall, Precision, and F2-score are utilized. The Jaccard index (mIoU) quantifies the overlap between predicted and ground truth segmentation masks. The F1 score balances precision and recall, while recall measures the true positive rate and precision gauges positive predicted value. The F2-score is a weighted harmonic mean of precision and recall, with an emphasis on recall. These metrics are widely adopted in biomedical image segmentation to evaluate model performance comprehensively.

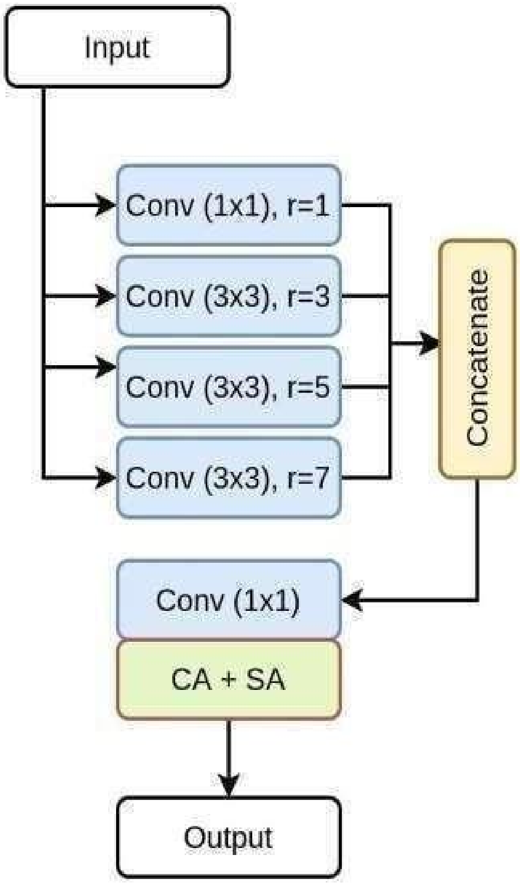

Fig. 2. The Dilated Attention Convolution (DAN) block diagram.

B. Implementation details

The proposed Attention-based Dilated Network (DANet) architecture's performance is assessed using three polyp datasets and one instrument dataset. The datasets are subjected to pre-processing and subsequently divided into training, validation, and testing subsets. The model is implemented utilizing the PyTorch framework and trained on an RTX 3090 GPU. The first pair of datasets, namely Kvasir-SEG and Kvasir-Instrument, serve as benchmarks for evaluating the model's performance in polyp segmentation and instrument detection tasks, respectively. To ensure uniformity, all images and masks are resized to 256 x 256 pixels. During training, a batch size of 16 is employed, and optimization is carried out using the Adam optimizer with a learning rate of 1e-4. The loss function comprises a combination of binary cross-entropy and dice loss. Moving on to the third and fourth datasets, CVC-ClinicDB and ETISLarib, they are used to assess the model's performance in polyp segmentation tasks. These datasets are partitioned into training, validation, and testing sets with an 80:10:10 ratio. Similar to the previous datasets, image and mask resizing to 256 x 256 pixels is applied, a batch size of 16 is utilized, and training employs the Adam optimizer with a learning rate of 1e-4, along with the same combination of binary cross-entropy and dice loss as the loss function. Additionally, an early stopping mechanism is integrated into the training process, automatically halting training if no improvement is observed over a continuous span of 50 epochs. This feature helps prevent overfitting and enhances the model's capacity to generalize effectively.

V. RESULTS

This section demonstrates the results of the proposed DANet and the three existing methods. Along with quantitative results, this section includes a discussion of quantitative results.

A. Quantitative Result

Table 1, 2, 3 and 4 in the paper present the quantitative evaluation results of the proposed Attentionbased Dilated Network (DANet) on the Kvasir-SEG, CVC-ClinicDB, ETIS-Larib PolypDB and Kvasir-Instrument datasets, respectively. The bold letter shows better performance than the others. On the KvasirSEG dataset [16], DANet achieved the best results on most of the evaluation metrics, including Jaccard score, F1 score, recall, accuracy, and F2 score. Specifically, DANet achieved a Jaccard score of 0.7966 and an F1 score of 0.8264, outperforming the most competitive UNet model with a margin of 4.94% in Jaccard score and 4.14% in F1 score. DANet also achieved higher recall, accuracy, and F2 score, as well as a competitive precision. According to Ayokunle O et al., [23],on the CVC-ClinicDB dataset, DANet outperformed all the other methods, achieving a Jaccard score of 0.8586, an F1 score of 0.9120, recall of 0.9479, an accuracy of 0.9890, and an F2 score of 0.9296. These results indicate that DANet is highly accurate in segmenting polyps in endoscopic images and can be used for computer-aided diagnosis. On the ETIS-Larib PolypDB dataset, DANet achieved a Jaccard score of 0.7272, which is 1.12% higher than UNet and 6.39% higher than U-Net++. Similarly, DANet achieved an F1 score of 0.8191, beating all the existing methods. These results demonstrate the effectiveness of the proposed Dilated Attention Network architecture in polyp segmentation tasks. On the Kvasir-Instrument dataset [18], DANet achieved the best results on most of the metrics, including Jaccard score, F1 score, recall, accuracy, and F2 score. Specifically, DANet achieved a Jaccard score of 0.9076, an F1 score of 0.9477, a recall of 0.9595, an accuracy of 0.9912, and an F2 score of 0.9533, indicating its high accuracy and reliability for instrument segmentation in endoscopic images. Overall, the quantitative evaluation results in Tables 1-4 demonstrate that the proposed Dilated Attention Network outperforms all the existing methods on the four publicly available polyp segmentation datasets and achieves state-of-the-art performance. These results suggest that the proposed architecture can be adapted for clinical applications on a computer-aided diagnosis (CADx) system.

TABLE I QUANTITATIVE RESULTS ON THE CVC-CLINICDB DATASETS.

Dataset: C-ClinicDB CV						
Model	**Jaccard**	**F1**	**Recall**	**Precision**	**Accuracy**	**F2**
U-Net[1]	0.84	0.89	0.90	0.87	0.98	0.89
ResU-Net [2]	0.78	0.86	0.88	0.88	0.97	0.87
U-Net++ [12]	0.83	0.89	0.91	0.89	0.98	0.90
ResUNet++ [10]	0.72	0.81	0.81	0.86	0.97	0.80
DANet	**0.85**	**0.91**	**0.94**	**0.93**	**0.98**	**0.92**

TABLE II QUANTITATIVE RESULTS ON THE KVASIR- SEG DATASETS

Dataset: Kvasir-SEG						
Model	**Jaccard**	**F1**	**Recall**	**Precision**	**Accuracy**	**F2**
U-Net[1]	0.74	0.82	0.85	0.87	0.95	0.83
ResU-Net [2]	0.66	0.76	0.80	0.82	0.93	0.77
U-Net++ [26]	0.74	0.82	0.84	0.86	0.94	0.82
ResUNet++ [26]	0.53	0.64	0.69	0.70	0.90	0.65
DANet	**0.79**	**0.89**	**0.91**	**0.92**	**0.96**	**0.90**

TABLE III
QUANTITATIVE RESULTS ON THE KVASIR-INSTRUMENT DATASETS.

Dataset:Kvasir- Instrument						
Model	**Jaccard**	**F1**	**Recall**	**Precision**	**Accuracy**	**F2**
U-Net[33]	0.87	0.92	0.92	0.93	0.98	0.92
ResU-Net [2]	0.87	0.92	0.92	0.94	0.98	0.92
U-Net++ [26]	0.88	0.93	0.92	0.94	0.98	0.92
ResUNet++ [26]	0.87	0.92	0.92	0.94	0.98	0.92
DANet	**0.90**	**0.94**	**0.95**	**0.97**	**0.99**	**0.95**

TABLE IV
QUANTITATIVE RESULTS ON THE ETIS-LARIB DATASETS.

Dataset: TIS-Larib E						
Model	**Jaccard**	**F1**	**Recall**	**Precision**	**Accuracy**	**F2**
U-Net[33]	0.71	0.80	0.80	0.83	0.98	0.80
ResU-Net [2]	0.44	0.53	0.46	0.90	0.97	0.48
U-Net++ [26]	0.66	0.75	0.78	0.81	0.98	0.76
ResUNet++ [26]	0.23	0.32	0.38	0.62	0.97	0.35
DANet	**0.76**	**0.81**	**0.92**	**0.94**	**0.99**	**0.87**

The tables mentioned in the previous response summarize the quantitative results of the proposed Attention-based Dilated Network (DANet) on four different datasets for polyp segmentation. The results demonstrate that DANet outperforms other state-of-the-art methods in terms of different evaluation metrics such as Jaccard score, F1 score, recall, accuracy, and F2 score. The Jaccard score (also known as Intersection over Union or IoU) measures the overlap between the predicted and ground truth segmentation masks, where a higher score indicates better segmentation accuracy. The F1 score is the harmonic mean of precision and recall, which balances both measures and is commonly used in classification tasks. The recall measures the proportion of true positives that were correctly identified, while the accuracy measures the proportion of correctly identified true and false positives. Finally, the F2 score is a weighted harmonic mean of precision and recall, where the recall is given more weight to prioritize sensitivity. In general, the proposed DANet achieves higher segmentation accuracy and better overall performance compared to other existing models, such as UNet and U-Net++. These results suggest that the Dilated Attention Network architecture, which combines the strengths of pre-trained ResNet50 and the novel Dilated Attention Convolution block, can be highly effective for polyp segmentation in clinical applications, such as computer-aided diagnosis (CADx).

B. Qualitative Results

Figure 3 in the paper presents qualitative results of the proposed Attention-based Dilated Network (DANet) on all four datasets, along with comparison with two other popular segmentation models, UNet[1] and U-Net++ [12]. The figure shows that DANet produces high-quality segmentation masks compared to the other two models. In particular, on the KvasirSEG and the ETISLarib PolypDB datasets, DANet shows superior performance to UNet and U-Net++, producing better segmentation masks. This suggests that DANet is better able to capture the important features and patterns in the images, leading to more accurate segmentation results. On the KvasirInstrument dataset [18], the proposed DANet model produces the most similar segmentation mask compared to the ground truth. This indicates that DANet is better able to accurately identify and segment the instruments in the

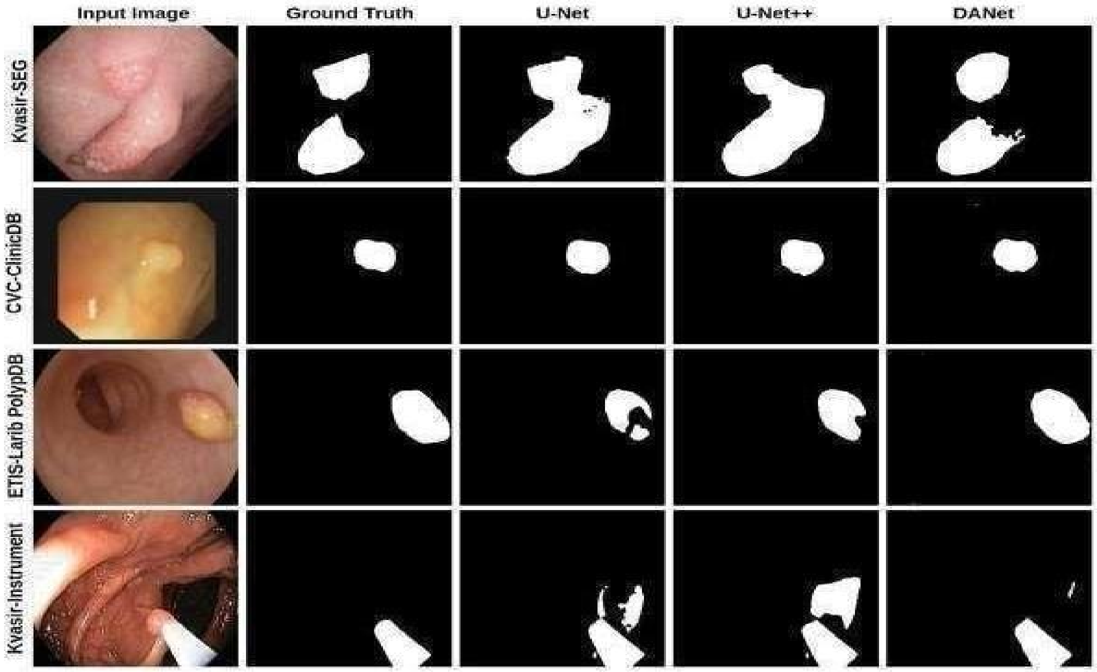

Fig. 3. Qualitative results comparison on our proposed DANet with the U-Net and U-Net++ on the Kvasir-SEG, CVC-ClinicDB, ETIS-Larib PolypDB and KvasirInstrument datasets.

images, which is important in medical applications where accurate identification of instruments can be crucial for diagnosis and treatment. Overall, the qualitative results in Figure 3 support the claim that the proposed DANet architecture achieves state-of-the-art performance on all four datasets, and is well-suited for use in computer-aided diagnosis (CADx) applications.

VI. CONCLUSION

The paper proposes a new architecture called Attention-based Dilated Network (DANet) that combines the strengths of pretrained ResNet50 [8] and a novel Dilated Attention Convolution (DAC) block. The DAC block employs multiple dilated convolution layers to expand the field of view, which allows the network to learn more global feature representations. The experiments conducted in the paper demonstrate that the proposed architecture achieves state-of-the-art performance on all four datasets used in the study. The high performance achieved on all three polyp segmentation datasets suggests that the proposed architecture can be adapted for clinical applications, particularly in computer-aided diagnosis (CADx). CADx is an area of research that aims to develop automated systems to aid medical professionals in making diagnostic decisions. The ability of the proposed architecture to accurately segment polyps indicates that it could be a useful tool for assisting in the detection and diagnosis of polyps in medical images. In the future, the authors suggest investigating unsupervised and semi-supervised learning approaches to better utilize the hidden capacity of unlabelled medical images. Unsupervised learning refers to the use of algorithms that can learn patterns in data without being explicitly trained on labelled data. Semi-supervised learning involves training a model using a combination of labelled and unlabelled data. By exploring these approaches, we will further improve the performance of the proposed architecture and expand its potential applications in clinical settings.

VII. CONFLICT OF INTEREST

The authors does not have any conflicts of interest to declare that they are relevant to the content of this article.

REFERENCES

[1] Ronneberger, Olaf et al. ”U-net: Convolutional networks for biomedical image segmentation.” International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[2] Zhang, Zhengxin, et al. ”Road extraction by deep residual u-net.” IEEE Geoscience and Remote Sensing Letters 15.5 ,2018: 749-753.

[3] Wu, H., Zhao et al.” A Lightweight Context-Aware Network for Real-Time Polyp Segmentation”. IEEE Transactions on Cybernetics. 2022. [4] Tomar, N. K., Jha, et al.”A feedback attention network for improved biomedical image segmentation”. IEEE Transactions on Neural Networks and Learning Systems. 2022. [5] Codella, Noel CF, et al. ”Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic).” IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018.

[6] Staal, Joes, et al. ”Ridge-based vessel segmentation in color images of the retina.” IEEE transactions on medical imaging 23.4 ,2004: 501-509.

[7] Leufkens, A. M., et al. ”Factors influencing the miss rate of polyps in a back to back colonoscopy study.” Endoscopy 44.05 ,2012: 470-475. [8] He, Kaiming, et al. ”Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[8] Chen, Liang-Chieh, et al. ”Encoder-decoder with atrous separable convolution for semantic image segmentation.” Proceedings of the European conference on computer vision (ECCV). 2018.

[9] Jha, Debesh, et al. ”Resunet++: An advanced architecture for medical image segmentation.” 2019 IEEE International Symposium on Multimedia (ISM). IEEE.

[10] Hu, Jie, Li Shen, and Gang Sun. ”Squeeze-and-excitation networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[11] Zhou, Zongwei, et al. ”Unet++: A nested u-net architecture for medical image segmentation.” Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2018. 3-11. [12] .Fan, Deng-Ping, et al. ”Pranet: Parallel reverse attention network for polyp segmentation.”
International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2020.

[13] Gao, Shang-Hua, et al. ”Res2net: A new multi-scale backbone architecture.” IEEE transactions on pattern analysis and machine intelligence 43.2 ,2019: 652-662.

[14] .Bernal, Jorge, et al. ”WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians.” Computerized medical imaging and graphics 43 ,2015: 99-111.

[15] Jha, Debesh, et al. ”Kvasir-seg: A segmented polyp dataset.” International Conference on Multimedia Modeling. Springer, Cham, 2020.

[16] Silva, Juan, et al. ”Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer.” International journal of computer assisted radiology and surgery 9.2 ,2014: 283293.

[17] Jha, Debesh, et al. ”Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy.” International Conference on Multimedia Modeling. Springer, Cham, 2021.

[18] Chen, Liang-Chieh, et al. ”Rethinking atrous convolution for semantic image segmentation.” arXiv preprint arXiv:1706.05587 ,2017.

[19] Jie Hu, Li Shen, et al. “ Squeeze-and-Excitation Networks “ IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132-7141

[20] Debesh Jha, et al. "A Comprehensive Study on Colorectal Polyp Segmentation With ResUNet++, Conditional Random Field and Test-Time Augmentation", IEEE Journal of Biomedical and Health Informatics, 2021

[21] Debesh Jha, Pia H. Smedsrud, et al. ,"ResUNet++: An Advanced Architecture for Medical Image Segmentation", 2019 IEEE International Symposium on Multimedia (ISM), 2019

[22] Ayokunle O. Ige, Nikhil Kumar Tomar, Felix O. Aranuwa, Oluwafemi Oriola et al. "ConvSegNet: Automated Polyp Segmentation from Colonoscopy using Context Feature Refinement with Multiple Convolutional Kernel Sizes", IEEE Access, 2023

[23] Nikhil Kumar Tomar, et al., "Automatic Polyp Segmentation with Multiple Kernel Dilated Convolution Network", 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), 2022

[24] Zaka-Ud-Din Muhammad, Zhangjin Huang, Naijie Gu, Usman Muhammad. "DCANet: deep context attention network for automatic polyp segmentation", The Visual Computer, 2022

[25] Yan Jin, Yibiao Hu, Zhiwei Jiang, Qiufu Zheng. "Polyp segmentation with convolutional MLP", The Visual Computer, 2022

[26] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Dag Johansen, Thomas De Lange, Pal Halvorsen, Havard D. Johansen. "ResUNet++: An Advanced Architecture for Medical Image Segmentation", 2019 IEEE International Symposium on Multimedia (ISM), 2019