Automatic Polyp Segmentation with Multiple Kernel Dilated Convolution Network

Nikhil Kumar Tomar*, Abhishek Srivastava†, Ulas Bagci‡, Debesh Jha‡
*School of Computer and Information Sciences, Indira Gandhi National Open University
†Computer Vision and Pattern Recognition Unit, Indian Statistical Institute
‡ Machine and Hybrid Intelligence Lab, Department of Radiology, Northwestern University, USA

***Abstract*—The detection and removal of precancerous polyps through colonoscopy is the primary technique for the prevention of colorectal cancer worldwide. However, the miss rate of colorectal polyp varies significantly among the endoscopists. It is well known that a computer-aided diagnosis (CAD) system can assist endoscopists in detecting colon polyps and minimize the variation among endoscopists. In this study, we introduce a novel deep learning architecture, named MKDCNet, for automatic polyp segmentation robust to significant changes in polyp data distribution. MKDCNet is simply an encoder-decoder neural network that uses the pre-trained *ResNet50* as the encoder and novel *multiple kernel dilated convolution (MKDC)* block that expands the field of view to learn more robust and heterogeneous representation. Extensive experiments on four publicly available polyp datasets and cell nuclei dataset show that the proposed MKDCNet outperforms the state-of-the-art methods when trained and tested on the same dataset as well when tested on unseen polyp datasets from different distributions. With rich results, we demonstrated the robustness of the proposed architecture. From an efficiency perspective, our algorithm can process at (≈ 45) frames per second on RTX 3090 GPU. MKDCNet can be a strong benchmark for building real-time systems for clinical colonoscopies. The code of the proposed MKDCNet is available at https://github.com/nikhilroxtomar/MKDCNet.**

***Index Terms*—Deep learning, polyp segmentation, colonoscopy, multi-scale fusion, dilated convolution**

I. Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related death and the third leading common cause of cancer worldwide [1]. The five-year survival rate is 90% for 39% of the patients that are diagnosed with localized stage disease but declines to 71% and 14% once diagnosed with regional and distant stage respectively [2]. Colonoscopy is considered the primary technique for colon cancer screening because it offers both detecting and removal of the polyp in a single operation. U.S. Preventive Services Task Force recommends forty-five to be considered as the new fifty for screening of CRC [3]. Colonoscopy can reduce the mortality through early detection at treatable stage and remove precancerous adenomas [4], [5].

During the colonoscopy operation, the average miss rate of the polyp is around 22-28% [6]. It is mainly because colonoscopy is an operator-dependent procedure and high inter-observer variations are seen in endoscopists' skills in detecting polyps [7]. During routine colonoscopy, the most frequently missed polyps are flat and smaller polyps [8]–[10]. Studies have shown that even a 1% increase in adenomas detection leads to a 3% decrease in the risk of interval colon cancer [11]. Therefore, it is highly critical to decrease the polyp miss-rate via an automated systems for CRC screening.

A Computer-Aided diagnosis (CADx) can highlight the suspicious frames and improve colonoscopy procedures. Jha et al. [12] proposed DoubleU-Net that used two U-Net's where the output of first U-Net acts as a soft-attention to the other. The network uses VGG-19 as an encoder and efficient blocks such as squeeze and excitation network [13] and atrous spatial pyradimal pooling [14] to capture some semantically meaningful information. DoubleU-Net showed state-of-the-art (SOTA) results on different biomedical image segmentation datasets. Wu et al. [15] proposed a lightweight context-aware network, PolypSeg+, for real-time polyp segmentation. The proposed architecture can capture distinguishable polyp features even with less trainable parameters and retain real-time speed. Tomar et al. [16] proposed a feedback attention network (FANet) for improved biomedical image segmentation, where they showed the SOTA performance on seven publicly available benchmark datasets. FANet unifies the mask of the previous epoch with the current training epoch and rectifies the prediction iteratively during the test time for improved performance. Ji et al. [17] proposed a progressively normalized self-attention network (PNS-Net) for video polyp segmentation. Shen et al. [18] proposed a hard region enhancement network (HRENet) for automatic polyp segmentation.

Despite the several automated methods proposed to improve the accuracy of polyp segmentation, further investigations are required to show the generalizability of the existing and the proposed method. Currently, most of the algorithms are only trained and tested on the same datasets [16], [17], [19]–[21]. Therefore, we aim to develop a novel deep learning algorithm to work well on varying distribution datasets coming from different institutions across different countries. To this end, we introduce a multiple kernel dilated convolution network (MKDCNet) architecture and test its performance on four still image datasets (polyps) and one cell nuclei dataset.

The main contribution of our work can be summarized as follows:

1) We present a novel deep learning architecture, MKDCNet, that utilizes novel multiple kernel dilated convolution block to increase the field of view of convolution kernel in order to capture local and global features. The multi-scale feature fusion block fuses different decoder blocks

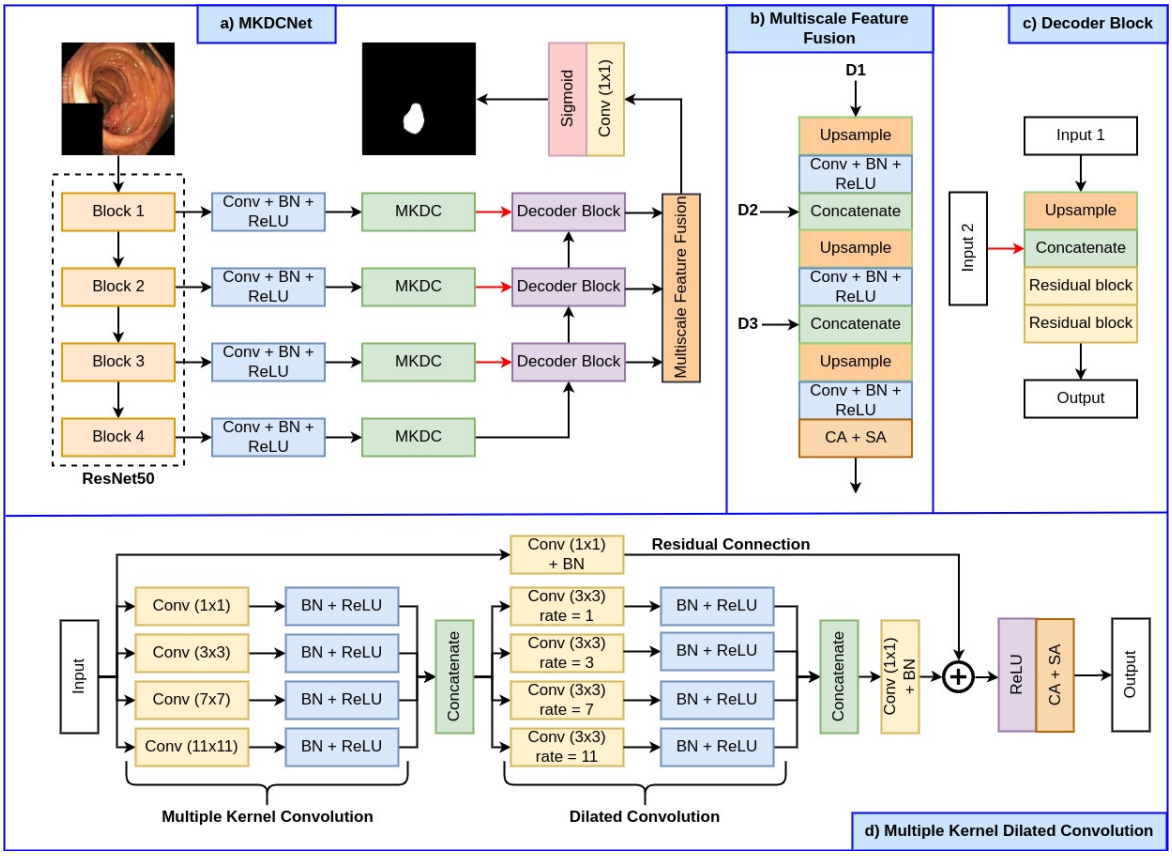

Fig. 1: Block diagram of the proposed MKDCNet along with its building blocks.

output for more robust feature representation that helps in accurate polyp segmentation.

2) We obtained SOTA results on four publicly available polyp datasets (same train-test set), and a nuclei segmentation dataset. Similarly, the proposed method outperformed other methods on three cross-center polyp dataset. Extensive experimental results shows the strong learning and generalization ability of MKDCNet.

II. Method

The proposed MKDCNet architecture is illustrated in Figure 1. The architecture begins with a pre-trained ResNet50 [22] as the encoder from which we extract four different feature maps. Each of these feature map is then passed through a sequence of 3×3 convolution layer, batch normalization, and a ReLU activation function. The output from the ReLU activation function is then passed through our novel Multiple Kernel Dilated Convolution (MKDC) block, which consists of multiple parallel convolution layers with different kernel sizes and dilation rates. After that, we have three decoder blocks, the output from all the three decoder blocks is passed through a Multiscale Feature Fusion (MSFF) block where we upsample and fuse the feature map to produce a more robust semantic representation. Finally, this feature map is then passed through a 1×1 convolution followed by a sigmoid activation function generating a binary segmentation mask.

A. Multiple kernel dilated convolution (MKDC) block

The *MKDC* block begins with four parallel convolution layers with a kernel size of 1×1, 3×3, 7×7 and 11×11 respectively. The kernel size's progressive increase helps capture a broad range of features, allowing the network to learn a more robust representation. Each convolution layer is then followed by batch normalization and a ReLU activation function. Next, each of these feature maps are then concatenated and passed through four parallel convolution layer, each having a dilation rate of 1, 3, 7 and 11, respectively. The use of different dilated convolutions helps to further expand the field of view and allows the network to capture more details and refine the significant features. In this sense, the *MKDC* is similar to multi-resolution strategies but in our case we capture rich details with convolutional kernels instead of using multiple parallel architectures or iterative and simultaneous connection from each resolutions. Each of the convolution layer is then followed by batch normalization and ReLU activation function. After that, we perform a concatenation over these features and feed them to a 1×1 convolution followed by a residual connection. Finally, the generated feature maps are passed through a channel and spatial attention mechanism which further highlight the significant features.

B. Decoder block

The decoder block begins with a bilinear upsampling which increases the spatial dimensions (height and width) of the input

TABLE I: Details of the datasets used in our experiments.

Dataset	Images	Size	Application
Kvasir-SEG [23]	1000	Variable	Colonoscopy
BKAI-IGH [24]	1000	1280×995	Colonoscopy
CVC-ClinicDB [25]	612	384×288	Colonoscopy
MedAI Challenge test set [26]	200	Variable	Colonoscopy
2018 Data Science Bowl [27]	670	256×256	Nuclie

feature map by a factor of two. After that, the upsampled feature map is then concatenated with the output of another *MKDC* block, that brings more semantic information to the decoder increasing its feature representation. Next, we have two residual block, where each residual block consists of a convolutional block and an identity mapping connecting the input and output of the convolutional block. The convolutional block begins with two 3×3 convolution layer, where each is followed by a batch normalization and a ReLU activation function.

C. Multiscale feature fusion (MSFF) block

We use the proposed *MSFF* block to enhance the feature at different scales by aggregating them to produce a more robust feature representation. The *MSFF* block takes the output from the first decoder block and passes it through a bilinear upsampling layer to increase its spatial dimensions by a factor of two. After that, it is followed by a 3×3 convolution layer, batch normalization and a ReLU activation function. The output of the ReLU activation function is then concatenated with the output from the second decoder block. Next, we again follow a bilinear upsampling layer where the concatenated feature map is upsampled by a factor of two and then followed by a 3×3 convolution layer, batch normalization and a ReLU activation function. The output from the ReLU activation function is then concatenated with the output from the third decoder block. After this, the feature map is again upsampled and passed through a 3×3 convolution layer, batch normalization and a ReLU activation function. The feature map is then passed through channel and spatial attention mechanism that focus on significant features and thus improve the feature representation and its robustness.

III. Experimental setup

In this section, we will present the datasets, evaluation metrics, and implementation details used in this study.

A. Datasets and evaluation

For this study, we have select four publicly available polyp datasets and a nuclie segmentation dataset. The details about the number of images, their size, and their application can be found in Table I. We have utilized Kvasir-SEG [23], BKAI-IGH [24], CVC-ClinicDB [25], and MedAI challenge test set [26] datasets for the polyp segmentation task. For the cell nuclei segmentation task, we have used the 2018 Data Science Bowl [27] dataset. To evaluate the performance of all the models, we have used metrics such as Dice Coefficient (DSC), mean Intersection over Union (mIoU), precision, recall, accuracy, F2-score, and Frame Per Second (FPS).

B. Implementation details

We have implemented the proposed MKDCNet and the SOTA methods using the PyTorch framework. For a fair comparison, we have used the same set of hyperparameters for all models used in this study. All models were trained on NVIDIA RTX 3090 GPU, where both the images and masks were first resized to 256×256 pixels for better utilization of GPU. The datasets were then split into training, validation and testing in the ratio of 80:10:10, except for Kvasir-SEG, where a split of 880/120 was used for training and testing respectively. An online data augmentation strategy was used on the training dataset which includes random rotation, horizontal flipping, vertical flipping and coarse dropout. The data augmentation helped to increase the robustness of the model. All the models were trained with an Adam optimizer having a learning rate of $1e^{-4}$ with a batch size of 16. A combination of binary cross-entropy loss and dice loss was used. ReduceLROnPlateau was used while training to reduce the learning rate for better performance. An early stopping criterion was also used to stop the training when the model stops improving.

IV. Result

At first, we performed validation of the algorithms on same datasets (same distribution). Next, we tested the trained model on completely unseen polyp datasets from different medical centers (different distribution).

A. Performance test on the same dataset

Table II shows the result of the MKDCNet and SOTA methods. On the Kvasir-SEG dataset, MKDCNet achieved a DSC of 0.8887 and mIoU of 0.8267 and outperformed the most competitive benchmarking method DeepLabv3+ with ResNet50 encoder with a margin of 0.5% in DSC and 0.94% in mIoU. Similarly, MKDCNet had a higher recall, precision, F2-score and nearly equal accuracy. Both DeepLabv3+ and MKDCNet had a real-time speed. Similarly, with the BKAI-IGH [24], our method outperformed DeepLabv3+ with a margin of 0.41% in DSC and 0.78% in mIoU. Additionally, we performed experiments on the 2018 Data Science Bowl [27] dataset, where we showed that our method consistently outperforms all other baseline methods. Figure 2 showed the example of qualitative results along with the heatmap. The qualitative results indicated that MKDCNet had better segmentation as compared to the UNet [28] and DeepLabv3+ [31].

B. Performance test on completely unseen dataset

Table III shows the results on the unseen dataset. For the unseen CVC-ClinicDB [25], our MKDCNet outperformed DeepLabv3+ with 1.01% in DSC and 0.78% in mIoU showing the superior generalization capability of our proposed method compared to others. Similarly, for the unseen BKAI-IGH dataset [24], our method outperformed best performing DeepLabv3+ by 1.97% in DSC and 1.93% in mIoU. For MedAI challenge test dataset, we only evaluated the performance on 200 positive polyp images provided by the task organizers. The models trained on Kvasir-SEG obtained better

TABLE II: Quantitative results on the experimented datasets.

Method	DSC	mIoU	Rec.	Prec.	Acc.	F2	FPS
Dataset: Kvasir-SEG [23]							
U-Net [28]	0.8264	0.7472	0.8504	0.8703	0.9510	0.8353	156.83
ResU-Net [29]	0.7642	0.6634	0.8025	0.8200	0.9341	0.7740	**196.85**
U-Net++ [20]	0.8228	0.7419	0.8437	0.8607	0.9491	0.8295	126.14
ResU-Net++ [19]	0.6453	0.5341	0.6964	0.7080	0.9044	0.6575	57.99
HarDNet-MSEG [30]	0.8260	0.7459	0.8485	0.8652	0.9492	0.8358	42.00
DeepLabV3+ (ResNet50) [31]	0.8837	0.8173	0.9014	0.9028	**0.9679**	0.8904	102.62
DDANet [32]	0.7415	0.6448	0.7953	0.7670	0.9326	0.7640	88.70
MKDCNet (Ours)	**0.8887**	**0.8267**	**0.9076**	**0.9088**	0.9677	**0.8954**	47.54
Dataset: BKAI-IGH [24]							
U-Net [28]	0.8286	0.7599	0.8295	0.8999	0.9903	0.8264	**160.27**
ResU-Net [29]	0.7433	0.6580	0.7447	0.8711	0.9843	0.7387	128.93
U-Net++ [20]	0.8275	0.7563	0.8388	0.8942	0.9895	0.8308	123.45
ResU-Net++ [19]	0.7130	0.6280	0.7240	0.8578	0.9832	0.7132	55.86
HarDNet-MSEG [30]	0.7627	0.6734	0.7532	0.8344	0.9863	0.7528	41.20
DeepLabV3+ (ResNet50) [31]	0.8937	0.8314	0.8870	0.9333	**0.9937**	0.8882	99.16
DDANet [32]	0.7269	0.6507	0.7454	0.7575	0.9851	0.7335	86.46
MKDCNet (Ours)	**0.8978**	**0.8392**	**0.8955**	**0.9365**	0.9934	**0.8947**	45.98
Dataset: 2018 Data Science Bowl [27]							
U-Net [28]	0.9122	0.8476	0.9021	**0.9339**	0.9799	0.9052	160.53
ResU-Net [29]	0.9183	0.8546	0.9236	0.9198	0.9809	0.9207	**188.74**
U-Net++ [20]	0.9114	0.8479	0.9107	0.9269	0.9799	0.9101	119.45
ResU-Net++ [19]	0.9157	0.8508	0.9162	0.9211	0.9798	0.9153	55.91
HarDNet-MSEG [30]	0.8344	0.7327	0.8686	0.8251	0.9640	0.8538	40.53
DeepLabV3+ (ResNet50) [31]	0.9027	0.8306	0.9220	0.8902	0.9774	0.9134	98.53
DDANet [32]	0.9117	0.8452	0.8452	0.9297	0.9792	0.9053	90.33
MKDCNet (Ours)	**0.9204**	**0.8586**	**0.9270**	0.9194	**0.9815**	**0.9237**	46.56

TABLE III: Quantitative results on the unseen polyp dataset.

Method	DSC	mIoU	Rec.	Prec.	Acc.	F2	FPS
Train Dataset: Kvasir-SEG [23], Test Data: Unseen CVC-ClinicDB [25]							
U-Net [28]	0.6336	0.5433	0.6982	0.7891	0.9484	0.6563	166.05
ResU-Net [29]	0.5970	0.4967	0.6210	0.8005	0.9465	0.5991	**195.38**
U-Net++ [20]	0.6350	0.5475	0.6933	0.7967	0.9504	0.6556	127.80
ResU-Net++ [19]	0.4642	0.3585	0.5880	0.5770	0.9159	0.5084	57.96
HarDNet-MSEG [30]	0.6960	0.6058	0.7173	0.8528	0.9592	0.7010	42.38
DeepLabV3+ (ResNet50) [31]	0.8142	0.7388	0.8331	**0.8735**	**0.9717**	0.8198	103.17
DDANet [32]	0.5234	0.4183	0.6502	0.5935	0.9275	0.5718	91.32
MKDCNet (Ours)	**0.8243**	**0.7466**	**0.8494**	0.8637	0.9709	**0.8325**	46.71
Train Dataset: Kvasir-SEG [23], Test Data: Unseen BKAI-IGH [24]							
U-Net [28]	0.6347	0.5686	0.6986	0.7882	0.9753	0.6591	162.60
ResU-Net [29]	0.5836	0.4931	0.6716	0.6549	0.9671	0.6177	**199.02**
U-Net++ [20]	0.6269	0.5592	0.6900	0.7968	0.9741	0.6493	128.59
ResU-Net++ [19]	0.4166	0.3204	0.6979	0.3922	0.9061	0.5019	57.22
HarDNet-MSEG [30]	0.6502	0.5711	0.7420	0.7469	0.9713	0.6830	42.44
DeepLabV3+ (ResNet50) [31]	0.7286	0.6589	0.7919	0.8123	**0.9787**	0.7493	103.25
DDANet [32]	0.5006	0.4115	0.6612	0.4825	0.9507	0.5592	91.73
MKDCNet (Ours)	**0.7483**	**0.6782**	**0.8087**	**0.8155**	0.9756	**0.7651**	42.741
Train Dataset: Kvasir-SEG [23], Test Data: MedAI Challenge test data (polyp) [26]							
U-Net [28]	0.6716	0.5725	0.7462	0.7438	0.9279	0.6957	159.90
ResU-Net [29]	0.6165	0.4991	0.6726	0.6977	0.9139	0.6315	**192.90**
U-Net++ [20]	0.6638	0.5702	0.7258	0.7594	0.9333	0.6845	128.64
ResU-Net++ [19]	0.4306	0.3246	0.5865	0.4677	0.8629	0.4793	60.20
HarDNet-MSEG [30]	0.6821	0.5877	0.756	0.7689	0.9271	0.7006	43.91
DeepLabV3+ (ResNet50) [31]	0.7784	0.6875	0.8332	0.8054	**0.9544**	0.7989	106.77
DDANet [32]	0.5738	0.4643	0.6638	0.6131	0.9141	0.6058	90.22
MKDCNet (Ours)	**0.7961**	**0.7054**	**0.8397**	**0.8151**	0.9532	**0.8103**	46.59
Train Dataset: BKAI-IGH [24], Test Data: MedAI Challenge test data (polyp) [26]							
U-Net [28]	0.5840	0.4837	0.5925	0.8147	0.9155	0.5726	166.94
ResU-Net [29]	0.4620	0.3605	0.4822	0.6989	0.8930	0.4525	**196.09**
U-Net++ [20]	0.5554	0.4530	0.6037	0.7475	0.8941	0.5591	126.01
ResU-Net++ [19]	0.3288	0.2419	0.3560	0.4779	0.8666	0.3313	59.25
HarDNet-MSEG [30]	0.4466	0.3550	0.4204	0.7427	0.9017	0.4210	42.84
DeepLabV3+ (ResNet50) [31]	0.6541	0.5675	0.6711	0.8535	0.9284	0.6552	100.86
DDANet [32]	0.5322	0.4281	0.5764	0.6547	0.8952	0.5351	91.02
MKDCNet (Ours)	**0.6985**	**0.6078**	**0.7210**	**0.8360**	**0.9366**	**0.7009**	48.05

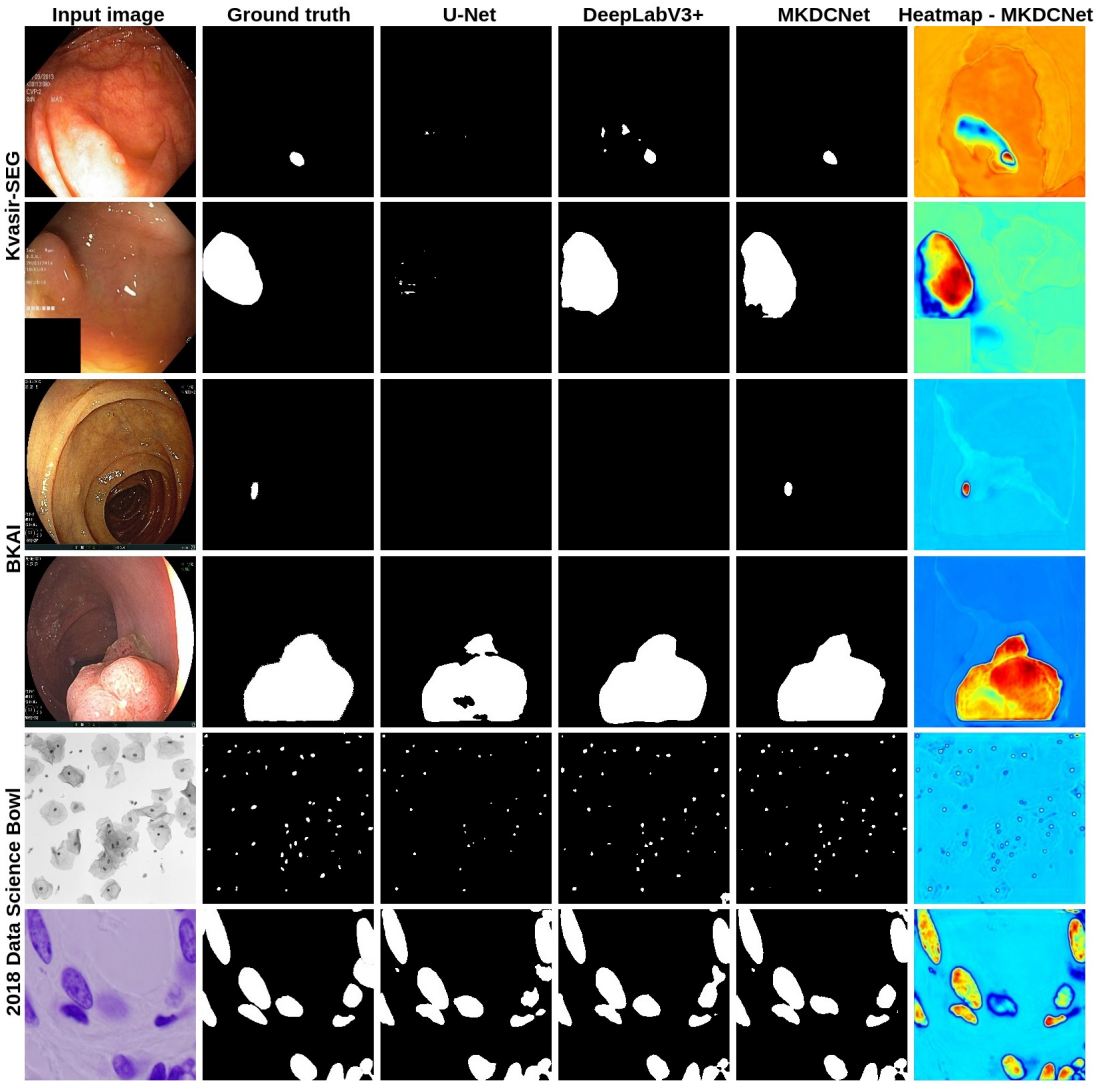

Fig. 2: Qualitative results comparison along with the heatmap on the Kvasir-SEG [23], BKAI-IGH [24], and 2018 Data Science Bowl [27] datasets. The heatmaps provide insight into the intermediate feature maps from the multi scale feature fusion block. The heatmap shows the *region of interest* and its statistical significance and the color intensity shows the *effect*. The *red* and *yellow* colors denote the most significant feature and the *blue* color denote the least significance feature.

TABLE IV: Ablation study of the proposed MKDCNet on the Kvasir-SEG [23].

No.	Method	DSC	mIoU	Recall	Precision
#1	MKDCNet w/o Multiple Kernel Dilated Convolution	0.8763	0.8138	0.8997	0.9071
#2	MKDCNet w/o Multiscale Feature Fusion	0.8720	0.8045	0.8974	0.8931
#3	MKDCNet w/o Multiple Kernel Dilated Convolution & Multiscale Feature Fusion	0.8785	0.8073	0.9003	0.8953
#4	MKDCNet	**0.8887**	**0.8267**	**0.9076**	**0.9088**

performance on the MedAI challenge dataset and slightly weaker performance with the BKAI datasets, which might be because BKAI-IGH dataset was captured at a different hospital (Institute of Gastroenterology and Hepatology (IGH), Vietnam), whereas the MedAI challenge dataset came from the HyperKvasir [33] whose distribution was similar to Kvasir-SEG (as both of them are captured at Vestre Viken Hospital Trust, Norway), despite the image frames being different. For both models trained on Kvasir-SEG and BKAI-IGH, proposed MKDCNet outperformed DeepLabv3+ by 1.77% and 4.4% in DSC, respectively.

C. Ablation study

In Table IV, we presented the ablation study on Kvasir-SEG dataset. When we compared setting #3 and setting #4, there was a 1.02% improvement in DSC and a 1.94% in mIoU with the multiple Kernel dilated convolution and multiscale feature fusion block in the network. Similarly, the Table IV also showed an improvement over both of the individual blocks.

V. Conclusion

We presented a novel architecture, MKDCNet, that utilizes ResNet50 as an encoder and the novel multiple kernel dilated convolution block to learn more robust representation

to automatically segment polyps from colonoscopy images with high performance. Extensive experimental results on four publicly available datasets (both on the same set as well as on completely unseen datasets) consistently showed that MKDCNet has the promising capability to improve the segmentation accuracy. With MKDCNet, we obtained a real-time processing speed of nearly 45 frames per second. Our results exhibited that MKDCNet has a better generalizability, accuracy, and real-time speed. Thus, MKDCNet can be a strong new baseline for developing artificial intelligence-based support to improve the traditional colonoscopy procedure. In the future work, we plan to exploit MKDCNet under federated learning settings where we can train multiple institute datasets and minimize the privacy concerns raised by each center.

Acknowledgement: This project is supported by the NIH funding: R01-CA246704 and R01-CA240639.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] A. C. Society, "Colorectal cancer facts & figures 2020–2022," *Published online*, p. 48, 2020.

[3] K. W. Davidson, M. J. Barry, C. M. Mangione, M. Cabana, A. B. Caughey, E. M. Davis, K. E. Donahue, C. A. Doubeni, A. H. Krist, M. Kubik *et al.*, "Screening for colorectal cancer: Us preventive services task force recommendation statement," *Jama*, vol. 325, no. 19, pp. 1965–1977, 2021.

[4] A. G. Zauber, S. J. Winawer, M. J. O'Brien, I. Lansdorp-Vogelaar, M. van Ballegooijen, B. F. Hankey, W. Shi, J. H. Bond, M. Schapiro, J. F. Panish *et al.*, "Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths," *N Engl J Med*, vol. 366, pp. 687–696, 2012.

[5] H. Brenner, J. Chang-Claude, C. M. Seiler, A. Rickert, and M. Hoffmeister, "Protection from colorectal cancer after colonoscopy: a population-based, case–control study," *Annals of internal medicine*, vol. 154, no. 1, pp. 22–30, 2011.

[6] A. Leufkens, M. Van Oijen, F. Vleggaar, and P. Siersema, "Factors influencing the miss rate of polyps in a back-to-back colonoscopy study," *Endoscopy*, vol. 44, no. 05, pp. 470–475, 2012.

[7] J. T. Hetzel, C. S. Huang, J. A. Coukos, K. Omstead, S. R. Cerda, S. Yang, M. J. O'brien, and F. A. Farraye, "Variation in the detection of serrated polyps in an average risk colorectal cancer screening cohort," *American Journal of Gastroenterology*, vol. 105, no. 12, pp. 2656–2664, 2010.

[8] D. Heresbach, T. Barrioz, M. Lapalus, D. Coumaros, P. Bauret, P. Potier, D. Sautereau, C. Boustière, J. Grimaud, C. Barthélémy *et al.*, "Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies," *Endoscopy*, vol. 40, no. 04, pp. 284–290, 2008.

[9] M. W. Short, M. C. Layton, B. N. Teer, and J. E. Domagalski, "Colorectal cancer screening and surveillance," *American family physician*, vol. 91, no. 2, pp. 93–100, 2015.

[10] 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*, vol. 2, no. 10, pp. 741–748, 2018.

[11] 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*, vol. 370, no. 14, pp. 1298–1306, 2014.

[12] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "Doubleu-net: A deep convolutional neural network for medical image segmentation," in *Proceedings of the International symposium on computer-based medical systems (CBMS)*, 2020, pp. 558–564.

[13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2018, pp. 7132–7141.

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and 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*, vol. 40, no. 4, pp. 834–848, 2017.

[15] H. Wu, Z. Zhao, J. Zhong, W. Wang, Z. Wen, and J. Qin, "Polypseg+: A lightweight context-aware network for real-time polyp segmentation," *IEEE Transactions on Cybernetics*, 2022.

[16] N. K. Tomar, D. Jha, M. A. Riegler, H. D. Johansen, D. Johansen, J. Rittscher, P. Halvorsen, and S. Ali, "Fanet: A feedback attention network for improved biomedical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[17] G.-P. Ji, Y.-C. Chou, D.-P. Fan, G. Chen, H. Fu, D. Jha, and L. Shao, "Progressively normalized self-attention network for video polyp segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2021, pp. 142–152.

[18] Y. Shen, X. Jia, and M. Q.-H. Meng, "Hrenet: A hard region enhancement network for polyp segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2021, pp. 559–568.

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

[20] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. 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*, 2018, pp. 3–11.

[21] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *Proceedings of the International conference on medical image computing and computer-assisted intervention (MICCAI)*, 2020, pp. 263–273.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 770–778.

[23] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. d. Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," in *Proceedings of the International Conference on Multimedia Modeling (MMM)*, 2020, pp. 451–462.

[24] P. N. Lan, N. S. An, D. V. Hang, D. Van Long, T. Q. Trung, N. T. Thuy, and D. V. Sang, "NeoUNet: Towards accurate colon polyp segmentation and neoplasm detection," *arXiv preprint arXiv:2107.05023*, 2021.

[25] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.

[26] S. Hicks, D. Jha, V. Thambawita, P. Halvorsen, B.-J. Singstad, S. Gaur, K. Pettersen, M. Goodwin, S. Parasa, T. de Lange *et al.*, "Medai: Transparency in medical image segmentation," *Nordic Machine Intelligence*, vol. 1, no. 1, pp. 1–4, 2021.

[27] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin *et al.*, "Nucleus segmentation across imaging experiments: the 2018 data science bowl," *Nature methods*, vol. 16, no. 12, pp. 1247–1253, 2019.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the International Conference on Medical image computing and computer-assisted intervention (MICCAI)*, 2015, pp. 234–241.

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

[30] C.-H. Huang, H.-Y. Wu, and 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.

[31] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[32] N. K. Tomar, D. Jha, S. Ali, H. D. Johansen, D. Johansen, M. A. Riegler, and P. Halvorsen, "DDANet: Dual decoder attention network for automatic polyp segmentation," in *Proceedigns of the International Conference on Pattern Recognition workshop*, 2021, pp. 307–314.

[33] H. Borgli *et al.*, "Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific data*, vol. 7, no. 1, pp. 1–14, 2020.