Advances In Multidisciplinary & Scientific Research
A Multidisciplinary & Interdisciplinary Journal
Volume 9, No 3, September 2023 Series

Polyp Segmentation from Colonoscopy images using MultiResUNet with Depthwise Convolutions

Akingbesote, Alaba Olu
Department of Computer Science
Adekunle Ajasin University
P.M.B. 001, Akungba-Akoko, Ondo State, Nigeria
E-mails: oluwamodimu2012@gmail.com

ABSTRACT

Polyp segmentation is a critical task in medical imaging, specifically in colonoscopy. Colonoscopy is a widely used diagnostic procedure for detecting and removing polyps in the colon. In order to identify polyps that pose a major threat to human life by growing into colorectal cancer, accurate polyp segmentation from colonoscopy images can give crucial information. Since polyps come in different shapes and sizes, existing biomedical image segmentation architectures often suffer from under or over-segmentation issues when used for polyp segmentation. Additionally, low-level features in the encoder block are frequently fused with high-level features in the decoder block, which results in feature mismatch problems in existing polyp segmentation models. To address these challenges, this work proposes MultiResUNET with depthwise convolutions. A deep learning architecture that combines Residual U-Net and depthwise convolutional layers for accurate polyp segmentation from colonoscopy images. The architecture's multi-resolution feature extraction and depthwise convolution pathways allowed the model to effectively handle complex structures and variations in polyps' shapes and sizes. The proposed architecture captured polyp features using multi-resolution blocks consisting of three parallel convolutional blocks with different kernel sizes, which are then concatenated along the channel axis to form a fused feature map. Depthwise convolution is then used in the skip connections to capture quality contextual information while reducing the computational cost. Experiments on the benchmark dataset showed that the model outperformed existing methods and achieved 0.9638 mIOU and 0.9815 dice score on the Kvasir-SEG dataset.

Keywords: Polyp, Colonoscopy Convolution, Depthwise, MultiResUNet

Aims Research Journal Reference Format:
Akingbesote, A.O. (2023): Polyp Segmentation from Colonoscopy images using MultiResUNet with Depthwise Convolutions. Advances in Multidisciplinary and Scientific Research Journal Vol. 9. No. 3. Pp 9–20
www.isteams.net/aimsjournal. dx.doi.org/10.22624/AIMS/V9N3P2

INTRODUCTION

Polyp segmentation is a critical task in medical imaging, specifically in colonoscopy. Colonoscopy is a widely used diagnostic procedure for detecting and removing polyps in the colon. The third most prevalent cancer type overall for both men and women is colorectal cancer [1], and accurate segmentation of polyps is crucial for the timely identification and treatment of potentially cancerous growths [2].

On the other hand, artificial intelligence and image segmentation have shown to be useful in segmenting colorectal polyps, and this may help endoscopists to detect polyps that may otherwise be overlooked [3]. Detection of colorectal polyps and endoscopic tools may also play a role in developing robotic-assisted surgical systems [4].

In order to identify polyps that pose a major threat to human life by growing into colorectal cancer, accurate polyp segmentation from colonoscopy images can give crucial information. Early polyp segmentation techniques depended on manually extracted features, which were unable to adequately capture the information necessary for polyp segmentation. Recent studies have demonstrated the effectiveness of deep convolutional neural networks in identifying numerous illnesses. For instance, U-Net [5] was one of the first successful CNNs applied in medical imaging. The architecture uses an encoder-decoder architecture, combining low-level encoder branch features with high-level decoder branch features. U-Net have received significant attention for segmentation tasks due to the extraordinary abilities in utilizing multi-level features that help to reconstruct high-resolution feature maps, and it is based on the upsampling of feature maps concatenated with feature maps skipped from the encoder [5]. Based on this, several segmentation architectures have been proposed for polyp segmentation from colonoscopy, as seen in [6], [7], and [8], among many others.

In the work of Debesh *et al.* [9], based on channel attention, the squeeze-and-excitation block and Atrous Spatial Pyramidal Pooling (ASPP) were combined in the ResUNet++ model. In Kim *et al.* [10], the authors introduced the UACA-Net design, which is based on U-Net but includes additional encoder and decoder modules. Foreground, background, and uncertain region maps are calculated for each representation using saliency maps computed by a prediction module in the UACA-Net. With the development of deep learning in medical image analysis, polyp segmentation has achieved promising progress.

These deep learning-based solutions are designed to automatically predict segmentation maps for colonoscopy images, thereby assisting clinicians in performing the procedure. However, existing polyp segmentation models often suffer from feature mismatch issues, which can be attributed to the low-level features of the encoder being fused with the high-level features of the decoder blocks. This feature mismatch issue directly affects polyp segmentation architectures' segmentation and generalization performance. To contribute, this work proposes MultiResUNet with depthwise convolutions (MultiResUNet-DC). A novel deep learning architecture that combines Residual U-Net and depthwise convolutional layers for accurate polyp segmentation.

Specifically, the main contributions of this work can be summarized in three folds:

i. Firstly, the MultiRes block, which uses concatenated feature maps derived from convolution blocks of various filter sizes, is used to collect multi-scale information.
ii. Secondly, residual pathways are added with skip connections to encourage improved feature representation. These pathways incorporate both depthwise convolution blocks for precise localization of segmented features.
iii. Lastly, experiments on a publicly available dataset in Kvasir-SEG show that the proposed model outperformed the existing polyp segmentation architectures.

The remainder of this paper is organized as follows: Section II presents the existing related works, Section III presents the research methodology, Section IV presents the results with detailed discussions, and Section V concludes.

2. RELATED WORKS

Polyp segmentation from colonoscopy images is a critical task in the field of medical image analysis. It aims to assist clinicians in the early detection and diagnosis of colorectal cancer [1]. With the increasing prevalence of colorectal diseases, accurate and efficient polyp segmentation techniques have gained significant attention in recent years [11]. For clinical monitoring and surgical operations, polyp segmentation offers an accurate polyp border, which is of great importance. For precise and automated polyp segmentation, researchers have created effective techniques during the past two decades [12]. Generally, biomedical images segmentation approach can be classified into handcrafted methods and the ones based on deep learning, which are automatic [13]. The traditional methods mainly exploit handcrafted features to train a classifier to extract polyp regions from colonoscopy images [14][15][16]. However, deep learning approach automatically extracts feature and ignores the handcrafted method to achieve better results [17].

The most popular network for medical image segmentation is U-Net [5], which has an encoder-decoder design and has taken over as the standard network. As was already noted, the original U-shaped encoder-decoder architecture had significant drawbacks. Yamada *et al.*[18] used a two-stage model for segmentation. The architecture combines the strengths of Faster R-CNN. However, while the two-stage approach enhances segmentation accuracy, it introduces increased complexity and computational overhead. However, the need for separate U-Net and Mask R-CNN components hinders real-time applications and deployment on resource-constrained devices.

In HardDNet-MSEG [6], a receptive field block is proposed to capture more global context information, and experiments showed that the model had increased segmentation performance. However, the inclusion of the receptive field came at a higher computational cost. In PraNet [19], a parallel partial decoder is added to the bottom of the encoder to generate a global map, and reverse attention was used to account for the connection between polyp border cues and polyp areas, which sharpened the polyp boundary. However, since polyps come in different shapes and sizes, sharpening the polyp boundary can lead to over segmentation in the architecture, as evident in the visualization results presented in the work.

In Zhao *et al.* [20], In order to segregate polyps from colonoscopy pictures, MSNet architecture was presented. By pyramidally concatenating many subtraction units, they increased the representation of multi-scale features by combining lower-order and higher-order cross-level complementary information with level-specific information. However, the quality of features learnt from the benchmark dataset was low. The spatial-channel Attention Gate (scAG) was proposed by Khanh et al. [21] and included to the U-Net for the purpose of segmenting diseased tissues. By including contextual information into the encoder's low-level features, the scAG can direct the network model to concentrate more on the specific details of the key areas, narrowing the semantic gap between encoder and decoder features.

A local context attention module is used in ASCNet [22] to direct the network's attention to the more complicated and uncertain areas for prediction refining. Through dynamic receptive field update, Ji *et al*., [23] demonstrated gradually regulated self-attention. On the benchmark dataset, better segmentation performance was attained thanks to the network's utilisation of both spatial and temporal cues. However, the model had high computational cost. In Ige et al. [13], ConvSegNet was proposed, which used a context feature refinement module designed using progressively increasing kernels. Experiments on the benchmark datasets showed that the model had improved performance on the benchmark datasets. However, the rate at which the model segmented and processed individual frames of polyp image sequence was low.

Generally, different levels of features in the encoder block have different characteristics, which are often low-level. Also, fusing the low-level features with the high-level features of the decoder block often leads to feature mismatch. The feature mismatch issue directly affects the segmentation and generalization performance of polyp segmentation architectures. Existing systems that have attempted to solve this often propose attention mechanisms and other modules. However, such modules often come with increased computational cost. To contribute, this research proposes MultiResUNet-DC which uses MultiRes block to capture multi-scale information by utilizing concatenated feature maps obtained from convolution blocks of varying filter sizes, to improve the model's ability to handle diverse polyp sizes and shapes. A more detailed discussion on the proposed architecture is presented in the following section.

3. METHODOLOGY

The MultiResUNET architecture is designed with a focus on accurate segmentation while efficiently handling variations in polyp shapes and sizes. The concept behind a residual block is rooted in the idea of learning residual functions. A collection of transformations are used to produce the output in a typical neural network layer such as convolution, activation function, etc. from the input. The network learns residual mapping rather than the direct mapping from input to output in a residual block. This means that the network aims to learn the difference between the input and the desired output, which is often easier to learn than the entire mapping. For a residual block given an input x, the output y is computed as:

$$y = F(x) + x \qquad (1)$$

Where $F(x)$ is the transformation applied to the input x to model the residual, $+$ operation is the elementwise addition, and x is the original input. The architecture of the residual block is presented in Fig. 1.

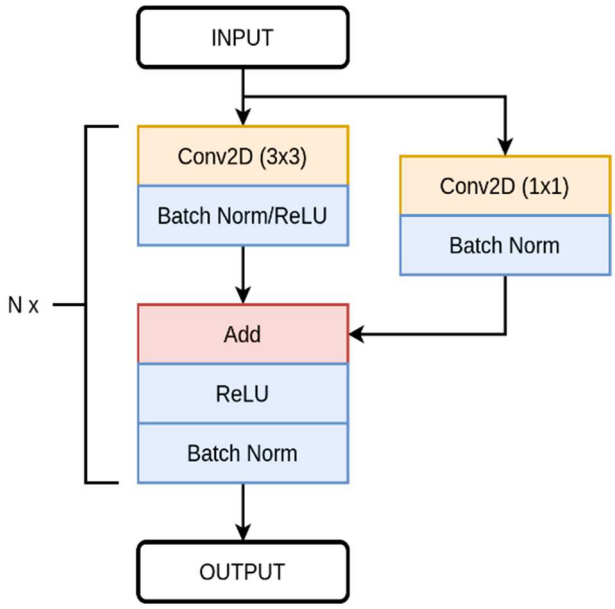


Fig. 1. Residual Block

In the proposed model, we begin by introducing the depthwise convolution block. In a depthwise convolution, each input channel is convolved separately with its corresponding filter. This means that each channel's information is processed independently. The building block in the proposed MultiResUNet-DC employs DepthwiseConv2D layers, followed by batch normalization and a Rectified Linear Unit (ReLU) activation function. This block is used to enhance feature extraction within the network. The convolution block employs standard Conv2D layers instead of depthwise convolutions. Similar to the depthwise convolution block, it includes batch normalization and ReLU activation, contributing to feature enrichment, and it is given as:

$$f(x) = \max(0, I_x) \quad (2)$$

To encourage information flow and improve feature representation, we introduce residual pathways within the network. These pathways incorporate both depthwise convolution blocks and convolution blocks to capture and integrate distinctive features from different layers. The MultiRes block, a key component of the architecture, is designed to capture multi-scale information. It utilizes concatenated feature maps obtained from convolution blocks of varying filter sizes. Batch normalization is applied to enhance the stability of the concatenated features. This block contributes to the network's ability to handle diverse polyp sizes and shapes. Then, the encoder employs the MultiRes block and a combination of depthwise convolutional and max-pooling layers to capture and downsample feature information, while the decoder block uses Conv2DTranspose layers for upsampling and integrates skip connections to enable precise localization of segmented features. The architecture of the proposed MultiResUNet-DC model is presented in

Fig. 2.

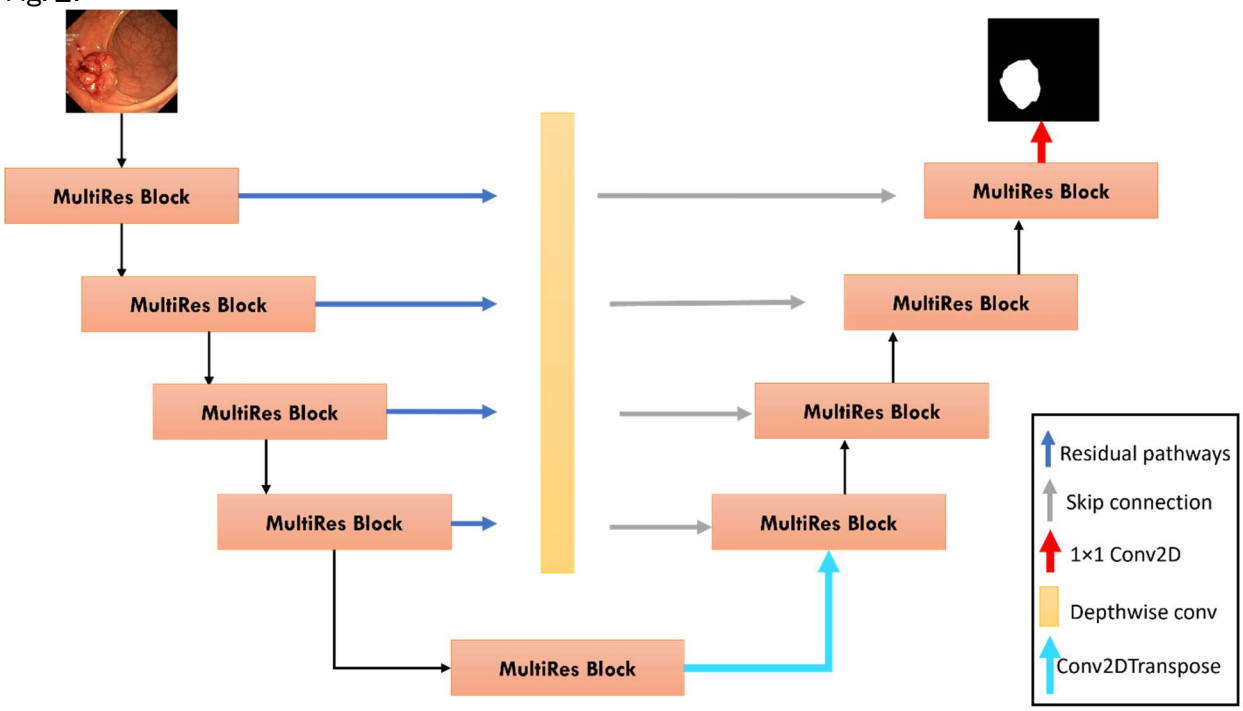


Fig. 2. Architecture of the Proposed MultiResUNet-DC

The MultiResUNET-DC architecture is constructed by sequentially connecting encoder and decoder blocks. The encoder starts with the input layer and progressively downsamples the feature maps. The decoder, in turn, takes these feature maps and performs upsampling while integrating skip connections from the residual block.

4. RESULTS

This section presents the dataset used for model evaluation, the performance metrics, the results of experiments on the proposed MultiResUNet-DC architecture, and the discussions. The model was implemented using Python 3.9, Tensorflow 2.9 on a workstation equipped with NVIDIA GeForce RTX 3050Ti, CUDA version 12.1 and 4GB GPU.

4.1 Dataset

One thousand (1000) polyp images, their corresponding masks, and bounding box data were captured by electromagnetic imaging equipment and are included in Kvasir-SEG dataset [24]. The bounding box data may be utilized for detection, while the images and their ground truths can be used for segmentation. The resolution of the images in this collection varies from 332 x 487 to 1920 x 1072 pixels. Samples of the images and the annotated masks from this dataset are shown in

Fig. 3.

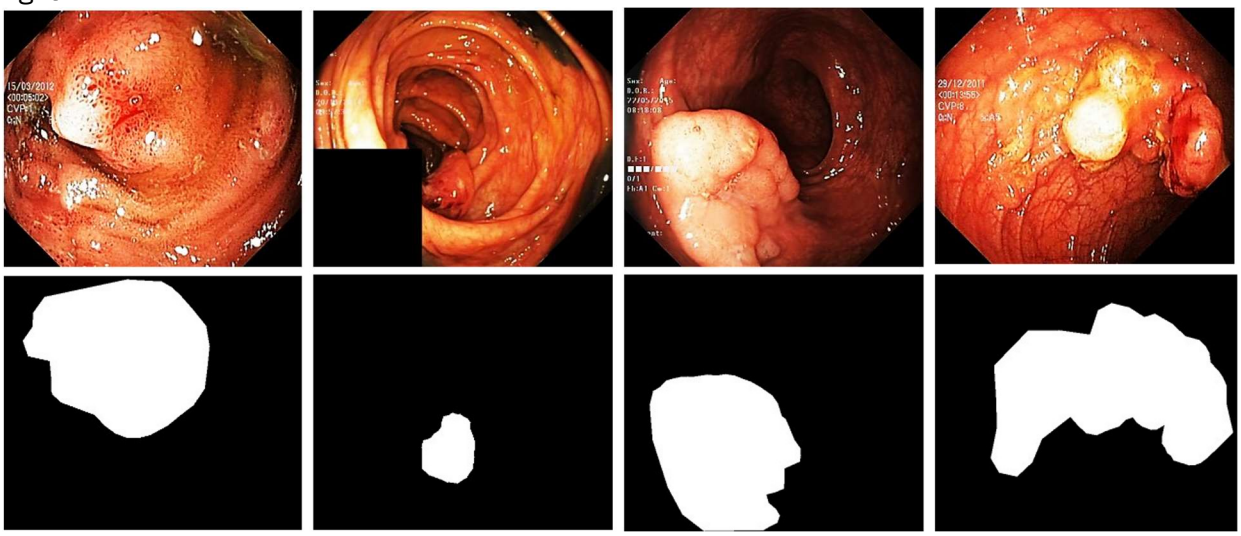


Fig. 3: Sample Images and masks in the Kvasir-SEG dataset

4.2 Performance Metrics

Five (5) common metrics are utilized to assess the performance of the proposed MultiResUNet-DC model in comparison to the existing approaches. We took into account Jaccard, Dice score, Recall, Precision, and F2-Measure.The Jaccard similarity index (IoU) is the ratio of the overlapping area between the predicted and ground truth to the area of union between the predicted and ground truth segmentation.

It is calculated as shown in equation 1, and the Dice Score measures the boundary matching between predicted and ground truth segmentation, as shown in equation 4.

$$Jaccard\ Similarity\ Index = \frac{TP}{TP+FP+FN} \quad (3)$$

$$DSC = \frac{2\times TP}{2\times TP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F2_{measure} = \frac{5\times Precision\times Recall}{4\times Precision+Rec} \quad (7)$$

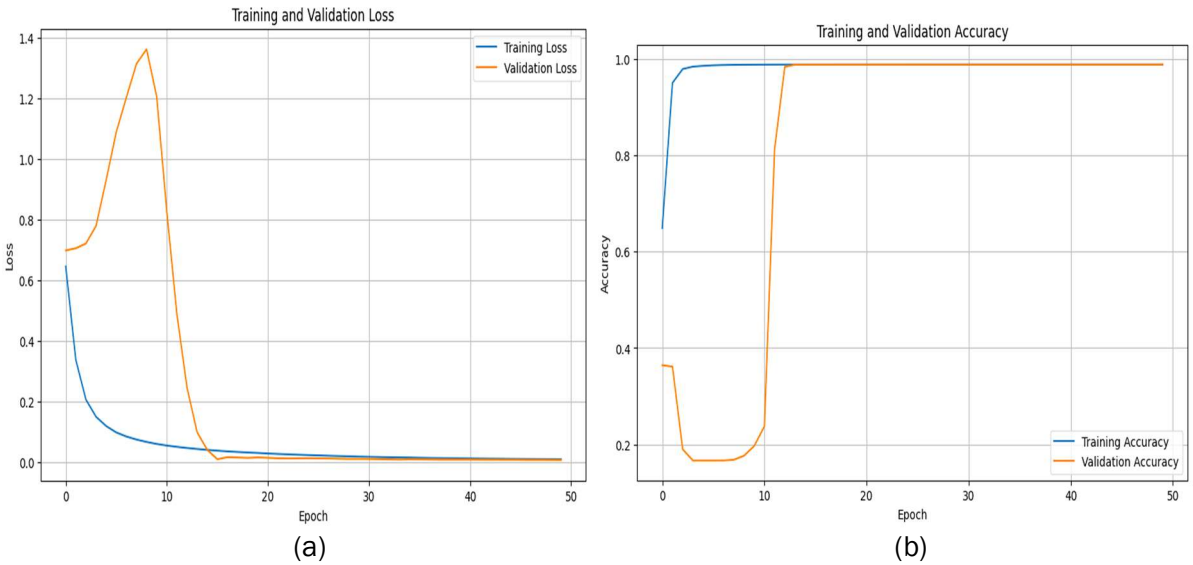
4.3 Results

The following hyperparameters were employed to train the proposed model, as presented in Table 1.

Table 1. Model Hyperparameters

S/N	Hyperparameter	Value(s)
1	Channel dimension (Encoder block)	32,64,128,256
2	Channel dimension (Decoder block)	256, 128, 64, 32
3.	Bridge dimension	512
4.	Maxpool size	2, 2
5.	Loss	Binary Crossentropy
6.	Epochs	50
7.	Optimizer	Adam
8.	Learning rate	$1e-4$

After 50 epochs, the model training accuracy and loss is presented in Fig. 4 (a) and (b)



(a) (b)

Figure 4. Model training (a) loss (b) accuracy after 50 epochs

The results of the experiments on the benchmark Kvasir-SEG dataset and comparison with existing segmentation architecture are presented in Table 2.

Table 2. Performance Comparison of the Proposed MultiResNet-DC on the Kvasir-SEG Dataset

Model	mIOU	Dice	F2	Precision	Recall
U-Net [5]	0.7472	0.8264	0.8353	0.8703	0.8504
U-Net++ [8]	0.7419	0.8228	0.8295	0.8607	0.8437
ResU-Net [25]	0.6634	0.7642	0.7740	0.8200	0.8025
ColonSegNet [26]	0.7240	0.8200	-	0.8430	0.8490
HarDNet-MSEG [6]	0.8480	0.9040	0.9150	0.9070	0.9230
ConvSegNet [13]	0.7936	0.8618	0.8855	0.7656	0.7840
MultiResNet [27]	0.9490	0.9050	-	0.9470	0.9540
MultiResUNet-DC	**0.9638**	**0.9815**	**0.9876**	**0.9716**	**0.9815**

As shown, the proposed MultiResUNet-DC achieved a mIOU score of 0.9638, which outperformed the mIOU recorded on the existing architectures of U-Net, U-Net++, ResU-Net, ConvSegNet and MultiResNet,. The Dice score achieved by the MultiResUNet-DC architecture on the benchmark dataset also outperformed the existing systems used for model comparison. Similarly, the F2 score, Precision and Recall saw significant improvements when Kvasir-SEG dataset was segmented using the proposed MultiResUNet-DC architecture. The visualization of the segmentation results of thirty random samples (group in 3) from the test masks was done, and this is presented in Fig. 5(a) to 5(c).

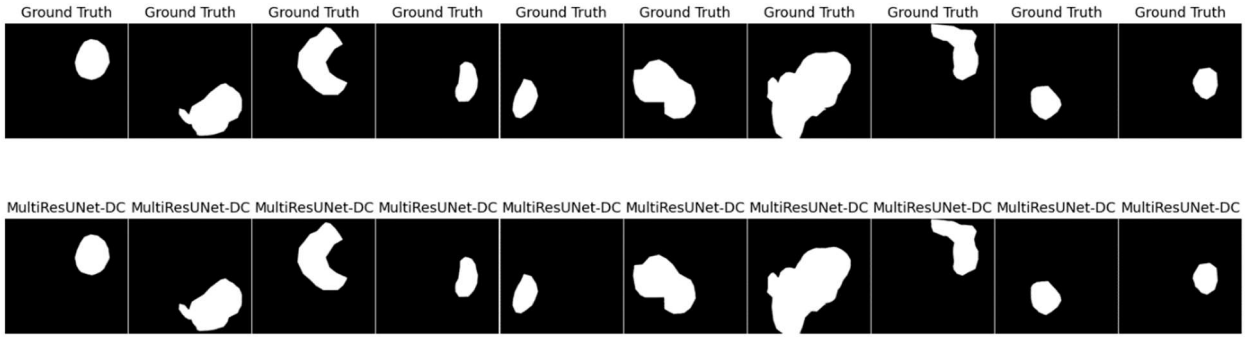


Fig. 5(a). Visualization of first ten samples (Ground truth vs MultiResUNet-DC output)

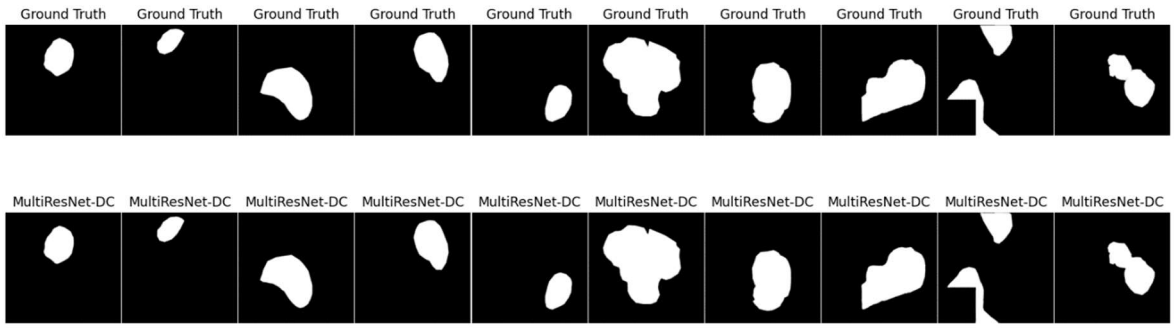


Fig. 5(b). Visualization of second ten samples (Ground truth vs MultiResUNet-DC output)

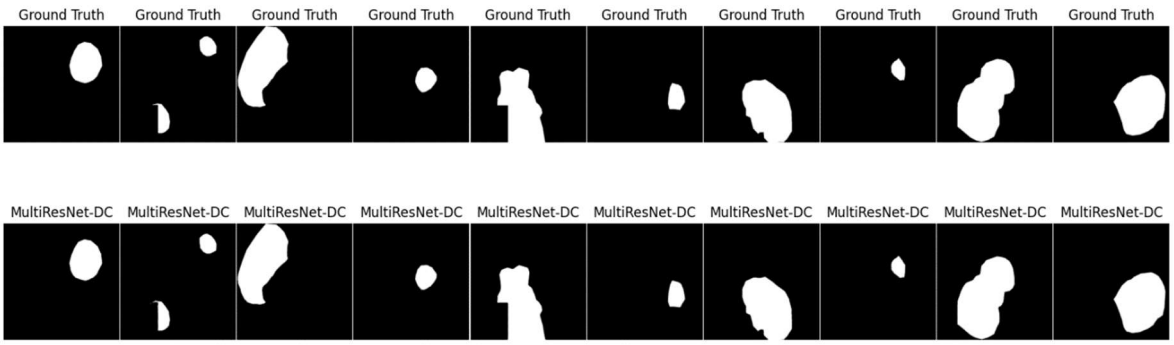


Fig. 5(c). Visualization of third ten samples (Ground truth vs MultiResUNet-DC output)

As shown in Fig. 5(a) to (c), it can be observed that the output of the proposed MultiResUNet-DC is closely similar to the ground truth of the polyp masks, regardless of the shape and size of the polyp. Therefore, proving the effectiveness of the proposed MultiResUNet-DC in accurately segmenting polyps from colonoscopy images.

Also, a comparison of the models' computational costs in terms of frames per seconds (FPS) and number of model parameters was done and presented in Table 3.

Table 3: Comparison of the Models' Computational Costs

Model	FPS	Parameters (Million)
U-Net [5]	156.83	31.04M
ResU-Net [25]	196.85	8.22M
FANet [28]	44.00	7.72M
ConvSegNet [13]	64.00	15.58M
MultiResUNet [27]	29.41	7.269M
MultiResUNet-DC	37.99	**5.805M**

As shown in Table 3, the proposed MultiResNet-DC is a lightweight and efficient segmentation architecture, as the model only has 5.805 million model parameters, compared to MultiResUNet which has 7.269 million, ConvSegNet with 15.58 million, FANet with 7.72 million and U-Net with 31.04 million. Also, the frames per second achieved is comparable with the state-of-the-art, as the proposed MultiResUNet-DC recorded 97.99 FPS.

5. CONCLUSION

In this paper, we introduced MultiResUNET-DC, a novel deep learning architecture that combines Multiple Residual U-Net and depthwise convolutional layers for accurate biomedical image segmentation. The model was introduced to capture multi-scale information by utilizing concatenated feature maps obtained from convolution blocks of varying filter sizes to improve the model's ability to handle diverse polyp sizes and shapes.

The architecture's multi-resolution feature extraction and depthwise convolution pathways contribute to its ability to effectively handle complex cell structures and variations in image quality. Our experimental results on the Kvasir-SEG dataset, which is a publicly available dataset, demonstrate the superior performance of MultiResUNet in comparison to existing methods, achieving 0.9638 mIOU 0.9815 dice, 0.9876 F2, 0.9716 precision and 0.9815 recall. Therefore, showcasing its potential to advance segmentation in colonoscopy images. Future work will investigate ways the architecture can be improved to capture more quality features with increased Frames per second.

REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA. Cancer J. Clin.*, vol. 65, no. 2, pp. 87–108, 2015, doi: 10.3322/caac.21262.

[2] L. Zhang, S. Dolwani, and X. Ye, "Automated polyp segmentation in colonoscopy frames using fully convolutional neural network and textons," *Commun. Comput. Inf. Sci.*, vol. 723, pp. 707–717, 2017, doi: 10.1007/978-3-319-60964-5_62.

[3] R. Feng *et al.*, "SSN: A Stair-Shape Network for Real-Time Polyp Segmentation in Colonoscopy Images," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2020-April, no. March, pp. 225–229, 2020, doi: 10.1109/ISBI45749.2020.9098492.

[4] S. Hosseinzadeh Kassani, P. Hosseinzadeh Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Automatic Polyp Segmentation Using Convolutional Neural Networks," in *Advances in Artificial Intelligence*, 2020, pp. 290–301.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 9351, no. Cvd, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[6] 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," pp. 1–13, 2021, [Online]. Available: http://arxiv.org/abs/2101.07172

[7] Y. Fang, D. Zhu, J. Yao, Y. Yuan, and K. Y. Tong, "ABC-Net: Area-Boundary Constraint Network with Dynamical Feature Selection for Colorectal Polyp Segmentation," *IEEE Sens. J.*, vol. 21, no. 10, pp. 11799–11809, 2021, doi: 10.1109/JSEN.2020.3015831.

[8] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11045 LNCS, pp. 3–11, 2018, doi: 10.1007/978-3-030-00889-5_1.

[9] D. Jha *et al.*, "ResUNet++: An Advanced Architecture for Medical Image Segmentation," in *Proceedings - 2019 IEEE International Symposium on Multimedia, ISM 2019*, 2019, pp. 225–230. doi: 10.1109/ISM46123.2019.00049.

[10] T. Kim, H. Lee, and D. Kim, "UACANet: Uncertainty Augmented Context Attention for Polyp Segmentation," *MM 2021 - Proc. 29th ACM Int. Conf. Multimed.*, pp. 2167–2175, 2021, doi: 10.1145/3474085.3475375.

[11] J. Zhong, W. Wang, H. Wu, Z. Wen, and J. Qin, "PolypSeg: An Efficient Context-Aware Network for Polyp Segmentation from Colonoscopy Videos," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12266 LNCS, pp. 285–294, 2020, doi: 10.1007/978-3-030-59725-2_28.

[12] X. Huang *et al.*, "Polyp segmentation network with hybrid channel-spatial attention and pyramid global context guided feature fusion," *Comput. Med. Imaging Graph.*, vol. 98, no. May, p. 102072,

2022, doi: 10.1016/j.compmedimag.2022.102072.
[13] A. O. Ige *et al.*, "ConvSegNet: Automated Polyp Segmentation From Colonoscopy Using Context Feature Refinement With Multiple Convolutional Kernel Sizes," *IEEE Access*, vol. 11, no. February, pp. 16142–16155, 2023, doi: 10.1109/ACCESS.2023.3244789.
[14] D. You, S. Antani, D. Demner-Fushman, and G. R. Thoma, "An MRF model for biomedical image segmentation," *Proc. - IEEE Symp. Comput. Med. Syst.*, pp. 539–540, 2014, doi: 10.1109/CBMS.2014.128.
[15] A. Van Opbroek, M. A. Ikram, M. W. Vernooij, and M. De Bruijne, "Transfer learning improves supervised image segmentation across imaging protocols," *IEEE Trans. Med. Imaging*, vol. 34, no. 5, pp. 1018–1030, 2015, doi: 10.1109/TMI.2014.2366792.
[16] A. Norouzi *et al.*, "Medical image segmentation methods, algorithms, and applications," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 31, no. 3, pp. 199–213, 2014, doi: 10.1080/02564602.2014.906861.
[17] A. O. Ige and M. H. Mohd Noor, "A survey on unsupervised learning for wearable sensor-based activity recognition," *Appl. Soft Comput.*, p. 109363, 2022, doi: https://doi.org/10.1016/j.asoc.2022.109363.
[18] M. Yamada *et al.*, "Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy," *Sci. Rep.*, vol. 9, no. 1, p. 14465, 2019.
[19] D. P. Fan *et al.*, "PraNet: Parallel Reverse Attention Network for Polyp Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in AI and Lecture Notes in Bioinformatics)*, 2020, vol. 12266 LNCS, pp. 263–273. doi: 10.1007/978-3-030-59725-2_26.
[20] X. Zhao, L. Zhang, and H. Lu, "Automatic Polyp Segmentation via Multi-scale Subtraction Network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2021, vol. 12901 LNCS, pp. 120–130. doi: 10.1007/978-3-030-87193-2_12.
[21] T. L. B. Khanh *et al.*, "Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging," *Appl. Sci.*, vol. 10, no. 17, 2020, doi: 10.3390/app10175729.
[22] R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, and Y. Yu, "Adaptive Context Selection for Polyp Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 2020, pp. 253–262.
[23] G.-P. Ji *et al.*, "Progressively Normalized Self-Attention Network for Video Polyp Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021, pp. 142–152.
[24] K. Pogorelov *et al.*, "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," *Proc. 8th ACM Multimed. Syst. Conf. MMSys 2017*, pp. 164–169, 2017, doi: 10.1145/3083187.3083212.
[25] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, 2018, doi: 10.1109/LGRS.2018.2802944.
[26] D. Jha *et al.*, "Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning," *IEEE Access*, vol. 9, pp. 40496–40510, 2021, doi: 10.1109/ACCESS.2021.3063716.
[27] T. M. Khan, M. Arsalan, I. Razzak, and E. Meijering, "Simple and robust depth-wise cascaded network for polyp segmentation," *Eng. Appl. Artif. Intell.*, vol. 121, no. February, p. 106023, 2023, doi: 10.1016/j.engappai.2023.106023.
[28] N. K. Tomar *et al.*, "FANet: A Feedback Attention Network for Improved Biomedical Image Segmentation," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–14, 2022, doi: 10.1109/tnnls.2022.3159394.