MSRD-Unet: Multiscale Residual Dilated U-Net for Medical Image Segmentation

Muna Khalaf[1*] iD   *Ban N. Dhannoon*[2] iD

[1]Department of Computer Science, College of Science for Women, University of Baghdad, Baghdad, Iraq.
[2]Department of Computer Science, College of Science, Al- Nahrain University, Baghdad, Iraq.
[*]Corresponding author: munakd_comp@csw.uobaghdad.edu.iq
E-mails address: ban.n.dhannoon@nahrainuniv.edu.iq.

Abstract:

Semantic segmentation is an exciting research topic in medical image analysis because it aims to detect objects in medical images. In recent years, approaches based on deep learning have shown a more reliable performance than traditional approaches in medical image segmentation. The U-Net network is one of the most successful end-to-end convolutional neural networks (CNNs) presented for medical image segmentation. This paper proposes a multiscale Residual Dilated convolution neural network (MSRD-UNet) based on U-Net. MSRD-UNet replaced the traditional convolution block with a novel deeper block that fuses multi-layer features using dilated and residual convolution. In addition, the squeeze and execution attention mechanism (SE) and the skip connections are redesigned to give a more reliable fusion of features. MSRD-UNet allows aggregation of contextual information, and the network goes without needing to increase the number of parameters or required floating-point operations (FLOPS). The proposed model was evaluated on three multimodal datasets: polyp, skin lesion, and nuclei segmentation. The obtained results proved that the MSDR-Unet model outperforms several state-of-the-art U-Net-based methods.

Keywords: Attention, Deep Learning, Dilated Convolution, Medical Image Segmentation, U-Net.

Introduction:

With the development of medical imaging technologies, medical images have become essential to medical research and clinical diagnosis. Medical image segmentation is indispensable in automated medical image analysis and understanding because it usually means capturing the region of interest in the medical image analyzed. More accurate segmentation leads to a more accurate analysis and understanding of a medical image.

Deep learning approaches based on convolutional neural networks (CNN) outperform traditional approaches in many applications [1, 2, 3]. CNN positively influenced medical image segmentation, especially after the proposal of the U-Net[4] model. U-Net has achieved a leap in performance in medical image segmentation. The improved performance of U-Net comes from its symmetric structure and skip connections. Its structure consists of an encoder that captures low-level features, a corresponding decoder that captures semantic features, and a skip connection allowing it to fuse low-level features with semantic features that give more informative features. In addition, U-Net could achieve good segmentation performance with a relatively small data training set [5, 6]. Hence, it has become the pioneer model in the field and inspired many modern models that modified it from different perspectives.

Residual convolutions improve feature utilization in classification problems[7]; many researchers adopted residual convolution into U-Net in medical image segmentation problems[8, 9]. The replacement of the traditional convolution block with a residual convolution block in U-Net increases the model's depth and alleviates the gradient vanishing problem[8, 9, 10].

Attention mechanisms have also succeeded in natural language processing [11] and many computer vision problems[12] and have been adopted in medical image segmentation. For instance, Ozan et al.[13] proposed a novel self-attention gate as an extension of the standard U-Net; this allows the model to

automatically learn to focus on salient features that are moved throw skip connection. Ashish and Jose[14] developed a semantic-guided attention module for medical image segmentation. This model integrates a multi-scale technique for combining semantic information at different levels with self-attention modules to accumulate relevant contextual features. In addition, the guided attention model was integrated with the U-Net model to get more optimal discriminative features and reduce redundant low-level features that may occur from the encoder-decoder approach. Jianhui et al.[15] adopted a u-shaped network with residual convolutions and used a Squeeze-and-Excitation (SE) block to recalibrate the importance of different channels. SE is an attention mechanism proposed in[16] that automatically learns the weight of each feature channel, then emphasizes valuable characteristics and suppresses the unhelpful ones.

Zongwei et al.[17] replaced direct skip connections used in U-Net with nested and dense skip connections that add more accurate results. They argued that the network with a direct skip connection from the encoder to the decoder fuse semantically dissimilar feature maps while a nested skip connection fuses more semantically similar feature maps, which is easier in the learning process. Huimin et al.[18] presented another design for skip connection with fewer parameters and more accurate segmentation results than the nested skip connection. They argued that capturing low-level and semantic features needed the fusing of smaller and same-scale feature maps from the encoder and larger-scale feature maps from the decoder.

The dilated convolution is a technique presented first for semantic segmentation here[19]. The dilated convolution allowed the same kernel to have a wider reception field commensurate with the dilated parameter chosen. Many researchers adopted this technique for medical image segmentation and reported good performances [20, 21, 22].

Recently, many researchers have integrated previously reviewed techniques to overcome the limitation of the convolution's fixed receptive field, gradient vanishing problem, and redundancy of feature maps. These integrations usually require more parameters and flop [8,22,23]. This paper proposes a new model that integrates these techniques in a new structure to achieve good segmentation performance using fewer parameters and flops.

In the proposed MSRD-Unet, the traditional convolution block is replaced with a novel Multiscale Residual Dilated Block (MSRDB), allowing the fusion of multi-scaled feature maps captured by different dilated kernels, as well as using residual style to get deeper to avoid vanishing

problems. In addition, the skip connections were redesigned to enhance fusion for the feature maps; a sequence and excitation attention mechanism [16] was used to recalibrate channel-wise attention. The contributions in this model can be summarized as follows:

- Multiscale Residual Dilated Block (MSRDB) aggregates two parallel dilation convolutions with the base convolution that feeds them. In addition, that base convolution is directly used as a residual connection. It allows aggregating different scale convolutions and alleviates the gradient vanishing without using more parameters.

- Skip connection is redesigned by fusing feature maps from different layers in the encoder before feeding them into the decoder layers. It helps to enhance the integration between the encoder feature map and the decoder maps.in addition, using SE block after skip connection allows to capture the most critical features and reduce redundancy.

The proposed model is evaluated using three medical datasets(Data Science Bowl[24], ISIC-2016[25], and CVC-ClinicDB[26]). The performance was compared with the U-net and three other states of art methods based on the U-shaped approach. The contributions of the paper can be summarized as follows:

The Proposed Method

This section presents the primary building blocks necessary for the proposed network MSRD-Unet and how the skip connection is redesigned to help the segmentation result. Fig.1 depicts the overall structure of the proposed MSRD-UNet.
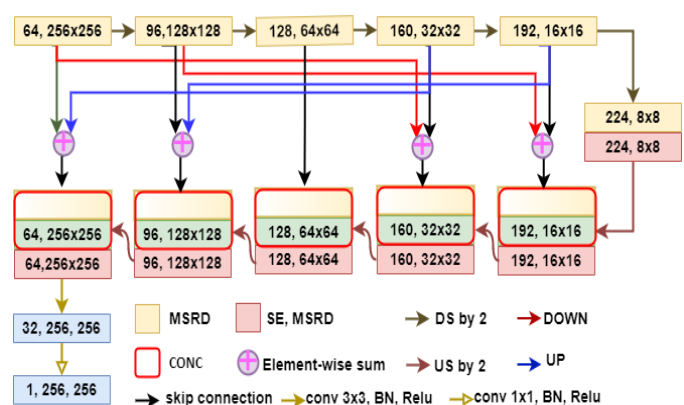

Figure 1. MSRD-Unet structure; each box represents a multi-channel feature map representing the feature map size and resolution.

1.Multiscale Residual Dilated Block (MSRDB)

In the proposed approach, the traditional convolution block was replaced with a multiscale residual dilated convolution block, as shown in Fig.2 (c). The convolution with default dilation 1 is passed to two parallel convolutions with different dilations 2 and 3, where all convolutions use the same kernel size (3 x 3). These three multiscale convolutions concatenate and then pass to a 1×1 convolution to compress these features. Each

Convolution (Conv) is followed by Batch Normalization (BN) and Rectified Linear Unit (ReLU). Extracting features at varying scales gives the network the potential to learn more while using dilation convolution reduces the number of parameters needed. At last, the residual connection is passed from base convolution to allow that deeper structure without degradation performance problems. Fig.2, shows the difference between MSRDB, traditional convolution, and residual convolution.
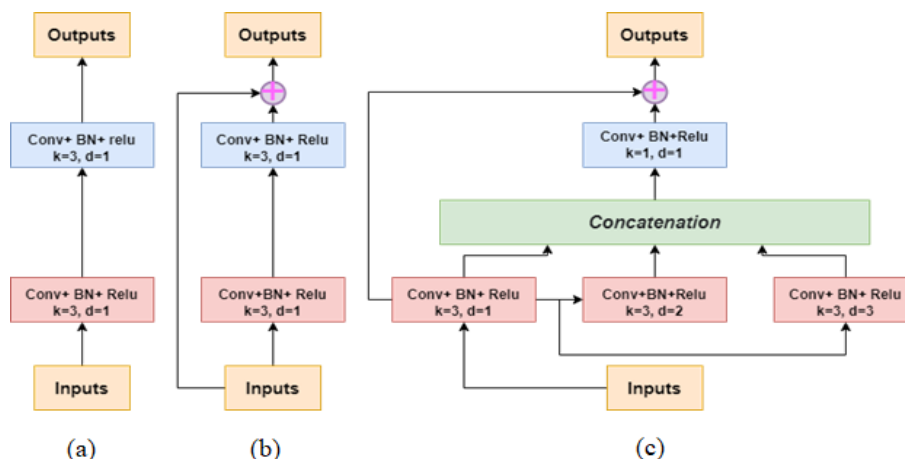

Figure 2. (a) Traditional convolution, (b) Residual convolution, and (c) Multiscale residual dilated convolution, where k represents kernel size and d represents the dilation.

2.U-Shaped Structure

MSRD-Unet, as illustrated in Fig. 1, takes a U-shaped structure. It consists of an encoder path (on the above) and a decoder path (on the down). Each of them consists of five layers. Each layer in the encoder path used MSRDB block to option features, followed by a 2x2 max pooling operation with stride 2 downsampling (DS) the resolution. At each DS, the number of channels increases by 32, and the feature map's resolution is downscale in half. Every layer in

the decoder path consists of a transposed convolution that upsamples (US) the feature map. At each US, the number of channels is compressed by 32, and the feature map's resolution is upscaled to double. Concatenation (CONC) is done between features map comes from encoder layer to its corresponding decoder layer followed by SE and MSRDB block. SE block is used for the dynamic channel-wise feature recalibration and to reduce redundant features before feeding them into the MSRDB block.
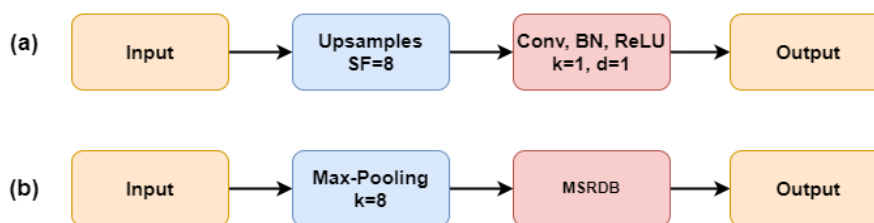

Figure 3. (a) Upsampling block (UP) applies upsamples with scale factor (SF) = 8, (b) downsampling block (DOWN) applies 2D max_ pooling with kernel size (k) = 8.

3.Skip Connection

The skip connection was redesigned to fuse features from different layers of the encoder before passing it to the decoder path. Feature fusion helps to integrate different scales of features where It

aggregates spatial features to high-level features within encoder layers. In the encoder path, the first layer was aggregated with the third layer, and the second layer was aggregated with the fifth layer using element-wise summation. After that, it passes

them to the corresponding decoder layer, as shown in Fig.1. to summation between two different feature map shapes need to use Upsampling block (UP) and downsampling block (DOWN) (detailed in Fig.3). Redesigned skips connection enhances segmentation results because it decreases the scale gaps of features between encoder and decoder layers simultaneously without losing spatial features.

Dataset

Three different medical image datasets with different requirements were selected to perform a convincing evaluation in this work. The first dataset is the 2018 Data Science Bowl for nuclei segmentation (containing 650 images with pixel-level annotations); 80% of the images were used as a training set, while the remaining 20% served as the testing set. The second dataset is the ISIC-2016 data set provided by the International Skin Imaging Collaboration for skin lesion segmentation. It consists of 900 images in the training set and 379 in the testing set with annotation masks. The last dataset is CVC-ClinicDB for polyp segmentation in Endoscopic Colonoscopy Frames. It contains 612 images with an annotation mask and training set; 80% of the images were used for the training set and the rest for the testing set. Table. 1 summarizes all information on the datasets used for the evaluation.

Table 1. Datasets information.

DataSets	Training Set	Test Set	Input Size	Application
Data Science Bowl	520	130	256×256	Nuclei Segmentation
ISIC-2016	900	379	Variable	Skin lesion Segmentation
CVC-ClinicDB	498	123	384×288	Colon Polyp Segmentation

Evaluation Metrics

Jaccard Similarity (JS) and Dice Score (DS) are essential metrics to evaluate medical image segmentation performance. They are suitable for medical datasets because of their imbalance issues [27]. As described in Eq. 1, Dice Score evaluates the spatial overlap between the ground truth and the predicted mask.

$$DS = (2 \times TP)/(2 \times TP + FN + TN) \qquad 1$$

As described in Eq. 2, Jaccard Similarity evaluates the overlap between the ground truth and the predicted mask. DS and JS evaluate the similarities between prediction and ground-truth masks.

$$JS = TP/(TP + FN + TN) \qquad 2$$

TP, FP, TN, and FN denote True Positive, False Positive, True Negative, and Negative, respectively. In addition, this paper also employed three measures commonly used in evaluating semantics segmentation; these are Accuracy (AC), Sensitivity (SE), and Specificity (SP), as described in Eqs. 3, 4, and 5, respectively[28].

$$AC = (TP + TN)/(TP + FP + TN + FN) \qquad 3$$
$$SE = TP/(TP + FN) \qquad 4$$
$$SP \ TN/(FP + TN) \qquad 5$$

Experiment and Results

In this part, the implementation details of the proposed model with other models are presented, and the performance of the proposed model comparing with other state-of-the-art models.

1.Implementation Details

all the models are implemented on a GPU machine in Google's Colaboratory using the PyTorch framework[29]. Dice loss[30] is used as a standard segmentation loss function because it is suitable for imbalance issues in the datasets. Adam optimizer[31] was employed with a learning rate starting at 1e-4, and ReduceLROnPlateau [32] was used to optimize it. All the models were trained for 100 epochs with a batch size of 8 and an image resolution of 256×256 pixels. During training, random vertical and horizontal data flipping was used as data augmentation. Fig.4 shows the curves, with red and blue colors depicting the training and testing losses of the proposed model, respectively, for each dataset.

Results and Discussion:

An ablation study is performed to detect the influence of different modifications on MSRD-Unet performance. as previously declared, MSRD-Unet has three modifications: the MSRDB and the redesigning skip connection and adding SE block. So it trained and tested all datasets with and without these modifications, as illustrated in Table. 2.

The same hypermeter setting is applied to each experiment during the training and testing phases. DS was used to evaluate the results of the testing set for each dataset listed in Table.1, as it is a fundamental metric in segmentation performance. In the First experiment, the datasets training and testing by the model used only MSRDB with a direct skip connection and without adding SE. It also compares that model if replaced the MSRDB with traditional and residual convolution blocks, which is illustrated in Fig. 2. The experiment showed that MSRDB improves the model's performance in all datasets, as shown in Table 2.

After that, the model MSRD-Unet was tested without SE block to detect the effect of redesigning the skip connection. It enhances the result relative to the

model with a direct skip connection tested before, as illustrated in Table 2. At last, the MSRD-Unet is tested with all components to detect the influence of SE on performance. It enhances the result, as illustrated in Table. 2.

Table 2. Ablation study of MSDR-Unet on three datasets.

DataSet	DS			DS	
	Traditional-B	Residual-B	MSDRB	Without SE	MSRD-UNet
Data Science Bowl	0.84	0.83	0.90	0.911	0.923
CVC-ClinicDB	0.83	0.845	0.88	0.89	0.902
ISIC-2016	0.84	0.86	0.89	0.90	0.911

The three datasets listed in Table. 1 were evaluated using five performance criteria (DS, JS, AC, SE, and SP). The proposed MSRD-Unet was compared to state-of-the-art approaches, such as U-Net, Unet 3+ [18], Att-UNet[13], and RU-Net[8]. Table. 3 compares our method and other state-of-the-art methods from the number of parameters and computational complexity illustrated by floating-point operations (FLOPs). MSDR-Unet requires approximately two fewer FLOPs and parameters than U-Net, Att-UNe, and RU-Net. However, it requires a quarter more parameters than Unet3+ and approximately 3% more FLOPs. The structure of the proposed model requires more memory than other models, while its inference time on a single test image is close to others models. So MSDR-Unet consumes acceptable computation resources. MSDR-Unet outperformed other state-of-art methods in three different datasets, as illustrated in Table. 4.
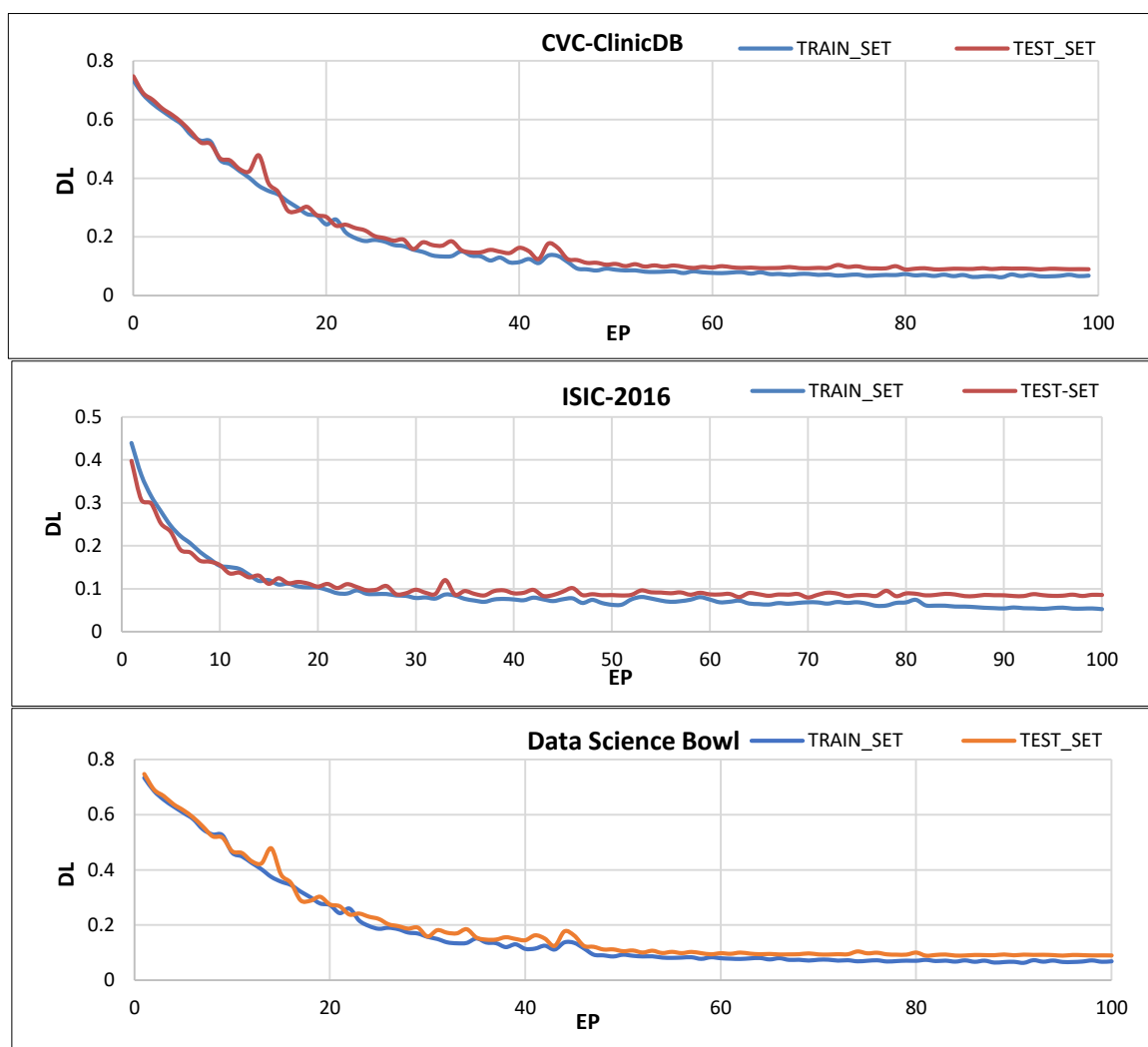

Figure 4. Dice loss (DL) curves for ISIC datasets (CVC-ClinicDB, Data Science Bowl, and ISIC-2016) training on the proposed network (MSRD-Unet), where TRAIN-SET represents the training set, TEST-SET represents the testing set, and EP represents the number of epochs.

Table 3. The resources used in the proposed networks are compared to state-of-the-art models.

Model	Unet 3+	MSRD-Unet (proposed)	U-Net	Att-UNet	RU-Net
Number of parameters	9,163,329	13,522,065	34,527,041	35,588,341	35,922,817
Number of FLOPS	34,757,410,816	35,548,439,552	65,562,214,400	66,679,645,184	68,259,151,872
Size in Memory	1142.71MB	1732.76MB	1361.96MB	1846.30MB	1611.28MB
Inference (s)	0.21	0.21	0.22	0.21	0.21

Table 4. Quantitative experimental results of the proposed MSRD-Unet and other state-of-the-art models on three different medical images on segmentation task.

DataSet	model	DS	JS	AC	SE	SP
Data Science Bowl	U-Net	0.903	0.854	0.968	0.914	0.980
	Unet 3+	0.903	0.840	0.968	0.902	0.976
	Att-UNet	0.911	0.852	0.969	0.913	0.973
	RU-Net	0.890	0.834	0.968	0.884	0.977
	MSRD-Unet	**0.923**	**0.862**	**0.979**	**0.918**	**0.990**
CVC-ClinicDB	U-Net	0.856	0.784	0.978	0.857	0.980
	Unet 3+	0.863	0.794	0.980	0.873	0.981
	Att-UNet	0.852	0.781	0.980	0.862	0.981
	RU-Net	0.859	0.791	0.970	0.861	0.980
	MSRD-Unet	**0.902**	**0.838**	**0.986**	**0.895**	**0.994**
ISIC-2016	U-Net	0.889	0.812	0.943	0.907	0.962
	Unet 3+	0.89	0.82	0.95	0.903	**0.965**
	Att-UNet	0.886	0.811	0.94	0.908	0.963
	RU-Net	0.89	0.83	0.94	0.921	0.961
	MSRD-Unet	**0.911**	**0.849**	**0.95**	**0.929**	0.951

On the Data Science Bowl dataset, the proposed model outperformed other state-of-the-art methods by all metrics. It improved the results of its closest competitor method (Att-UNet) by 1.31%, 1.17%, 1.03%, 0.55%, and 1.75% for DS, JS, AC, SE, and SP, respectively.

On the CVC-ClinicDB dataset, the proposed model outperformed other state-of-the-art methods by all metrics and improved the results of its closest competitor method (Unet 3+) by 4.52%, 5.54%, 0.61%, 2.52%, and 1.33% for DS, JS, AC, SE, and SP, respectively.

On the ISIC-2016 dataset, the proposed model outperformed other state-of-the-art methods by all metrics except for Specificity compared to Unet 3+. It improved the results of its closest competitor method (RU-Net) by 1.25%, 1.89%, 0.07%, 0.86%, and 1.33% for DS, JS, AC, and, SE, respectively.

MSRD-Unet achieved good performance in three datasets that have different requirements. Its structure allows it to go deeper and wider at each layer. It captures feature maps that integrate the low-level features with more semantic features. This integration is essential to getting good semantic segmentation performance. In addition, MSRD-Unet is generalized testing set in all datasets, as explained in loss curves in Fig. 4, where underfitting or overfitting problems do not appear. Fig.5 illustrates some qualitative examples of the performance of the proposed Networks in DataScience Bowl, CVC-ClinicDB, and ISIC-2016. It demonstrates their ability to overcome challenges found in each dataset. Each dataset has its challenges, as illustrated in Fig.5 (A, B, and C), including the variety in colors, sizes, shapes, and texture.
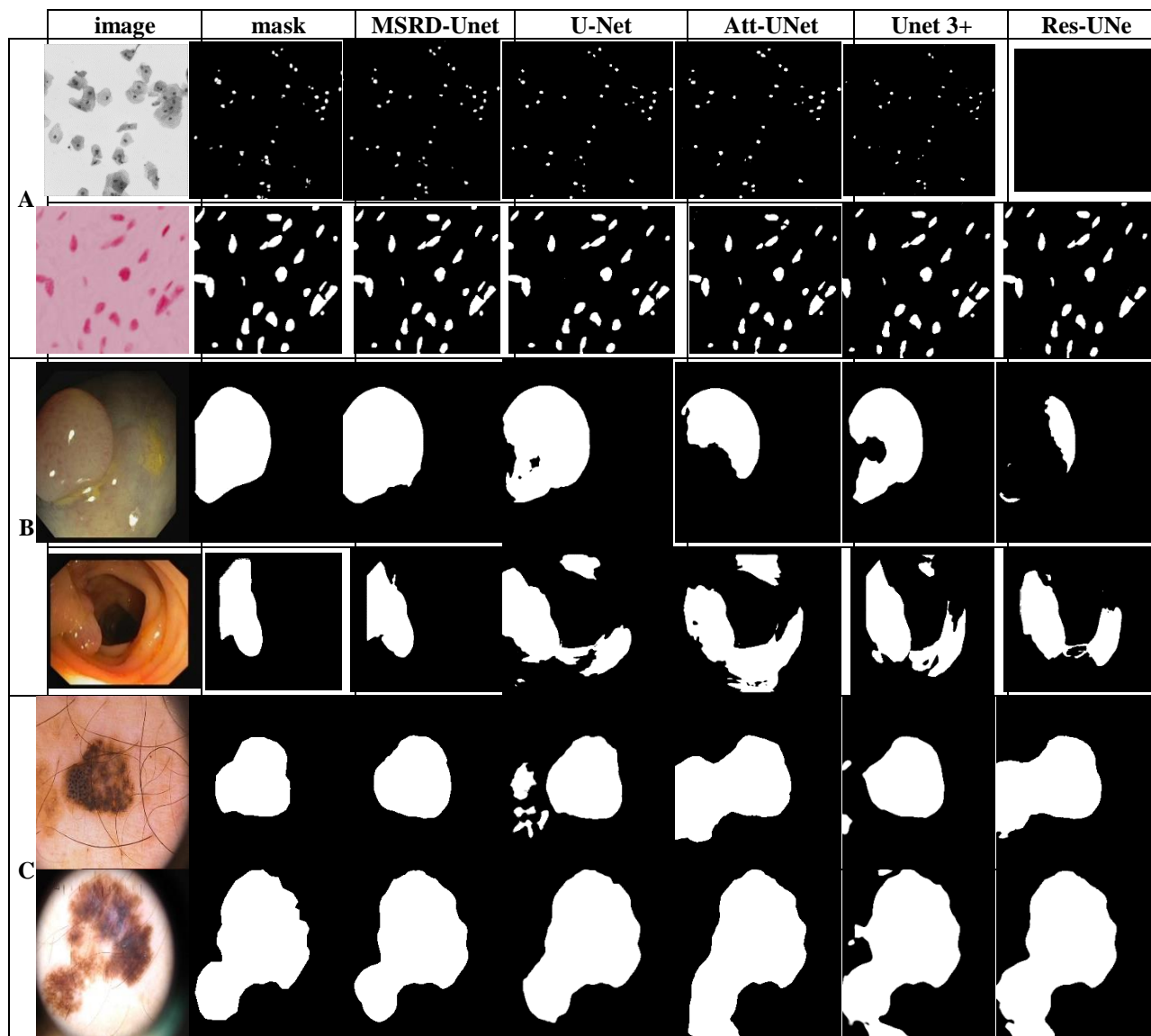
Figure 5. Segmentation Results produced by MSRD-Unet of some examples (A) from Data Science Bowl test set, (B) CVC-ClinicDB test set, and (C) ISIC-2016 test set.

Conclusion:

This paper presented an end-to-end network based on U-shaped deep learning for medical image segmentation called MSRD-Unet. This approach captures different levels of features by combining multiscale residual dilated convolution and redesigning skip connections to fuse different levels of features before passing them to the decoder. In addition, the SE block was used in the decoder path to recalibrate the dynamic channel-wise feature. Although the proposed method requires a few parameters and flops, it consumes more memory and has no improvement in time. Three different datasets were chosen, and five metrics were used to evaluate the proposed method's performance compared to the state-of-the-art methods. The quantitative and qualitative results showed that the proposed approach outperformed other methods and still

involved acceptable computation costs. In future work, we intend to continue investigating the MSDRB and redesigning the skip connection to improve accuracy and efficiency.

.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Baghdad.

Authors' contributions statement:

M.K. and B.N. Contributed to the research analysis, design, implementation of the research, analysis of the results, and the writing of the manuscript.

References

1. Asroni A, Ku-Mahamud KR, Damarjati C, Slamat HB. Arabic speech classification method based on padding and deep learning neural network. Baghdad Sci J. 2021; 18(2): 925–36.

2. Harba ES, Harba HS, Abdulmunem IA. Advanced Intelligent Data Hiding Using Video Stego and Convolutional Neural Networks. Baghdad Sci J. 2021; 18(4): 1317–27.

3. Dong S, Wang P, Abbas K. A survey on deep learning and its applications. Comput Sci Rev [Internet]. 2021; 40: 100379. https://doi.org/10.1016/j.cosrev.2021.100379

4. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Cham: Springer; 2015. p. 234–41.

5. Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J. A survey on U-shaped networks in medical image segmentations. Neurocomputing [Internet]. 2020; 409: 244–58. https://doi.org/10.1016/j.neucom.2020.05.070

6. Khalaf M, Dhannoon BN. Skin Lesion Segmentation based on U-Shaped Network. Karbala Int J Mod Sci. 2022; 8(3): 493-502.

7. Sarwinda D, Paradisa RH, Bustamam A, Anggia P. Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. Procedia Comput Sci [Internet]. 2021; 179(2019): 423–31. https://doi.org/10.1016/j.procs.2021.01.025

8. Zuo Q, Chen S, Wang Z. R2AU-Net: attention recurrent residual convolutional neural network for multimodal medical image segmentation. Secur. Commun. Netw. 2021 Jun 10;2021. https://doi.org/10.1155/2021/6625688

9. Zhuang J. LadderNet: Multi-path networks based on U-Net for medical image segmentation. arXiv Prepr arXiv181007810 [Internet]. 2018; 2–5. http://arxiv.org/abs/1810.07810

10. Huang Z, Zhao Y, Liu Y, Song G. GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation. Biomed Signal Process Control [Internet]. 2021; 70(June): 102958. https://doi.org/10.1016/j.bspc.2021.102958

11. Vashishth S, Upadhyay S, Tomar GS, Faruqui M. Attention Interpretability Across NLP Tasks. 2019; 1–10. Available from: http://arxiv.org/abs/1909.11218

12. Wang F, Tax DMJ. Survey on the attention based RNN model and its applications in computer vision. 2016; Available from: http://arxiv.org/abs/1601.06823

13. Oktay O, Schlemper J, Folgoc L Le, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: Learning Where to Look for the Pancreas. arXiv Prepr arXiv180403999 [Internet]. 2018; (Midl). Available from: http://arxiv.org/abs/1804.03999

14. Sinha A, Dolz J. Multi-scale self-guided attention for medical image segmentation. IEEE J Biomed Health Inform. 2020 Apr 14;25(1):121-30.

15. Wen J, Li Z, Shen Z, Zheng Y, Zheng S. Squeeze-and-Excitation Encoder-Decoder Network for Kidney and Kidney Tumor Segmentation in CT Images. In: International Challenge on Kidney and Kidney Tumor Segmentation. Cham: Springer; 2022. p. 71–9.

16. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 7132-7141).

17. Zhou Z, Siddiquee MR, Tajbakhsh N. UNet ++ : A Nested U-Net Architecture for Medical Image Segmentation Zongwei. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham: Springer International Publishing; 2018. p. 3–11.

18. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen YW, Wu J. Unet 3+: A full-scale connected unet for medical image segmentation. InICASSP. IEEE Int Conf Acoust Speech Signal Process 2020 May 4 : 1055-1059. https://doi.org/10.1109/ICASSP40776.2020.9053405.

19. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv Prepr. 2016; 1412.7062.

20. Wang S, Hu S-Y, Cheah E, Wang X, Wang J, Chen L, et al. U-Net Using Stacked Dilated Convolutions for Medical Image Segmentation. arXiv Prepr arXiv200403466 [Internet]. 2020; Available from: http://arxiv.org/abs/2004.03466

21. Moreno Lopez M, Ventura J. Dilated convolutions for brain tumor segmentation in MRI scans. In: nternational MICCAI brainlesion workshop. Cham: Springer; 2018. p. 253–62.

22. Yang J, Zhu J, Wang H, Yang X. Dilated MultiResUNet: Dilated multiresidual blocks network based on U-Net for biomedical image segmentation. Biomed Signal Process Control [Internet]. 2021; 68(January):102643. https://doi.org/10.1016/j.bspc.2021.102643

23. Su R, Zhang D, Liu J, Cheng C. MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation. Front Genet. 2021; 12(February): 1–14.

24. Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat Methods [Internet]. 2019; 16(12): 1247–53. http://dx.doi.org/10.1038/s41592-019-0612-7

25. Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, et al. Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). eprint arXiv160501397 2016 [Internet]. 2016; 3–7. Available from: http://arxiv.org/abs/1605.01397

26. Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D,

Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput Med Imaging Graph. 2015; 43: 99–111.

27. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: A survey. IET Image Process. 2022; 16(5): 1243–67.

28. Yacin Sikkandar M, Alrasheadi BA, Prakash NB, Hemalakshmi GR, Mohanarathinam A, Shankar K. Deep learning based an automated skin lesion segmentation and intelligent classification model. J Ambient Intell Humaniz Comput. 2021; 12(3): 3245–55.

29. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019; 32.

30. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep Learn Med image Anal multimodal Learn Clin Decis Support. 2017; 10553 LNCS(2017): 240–8.

31. Kingma DP, Ba JL. Adam: A method for stochastic optimization. arXiv Prepr. 2015; 1–15.

32. Al-Kababji A, Bensaali F, Prasad S. Scheduling Techniques for Liver Segmentation: ReduceLRonPlateau Vs OneCycleLR. J Imaging [Internet]. 2022; 8(3): 55. Available from: http://www.qf.org.qa

شبكة ال MSRD-Unet : شبكة U-Net مطورة لتقسيم الصور الطبية بالاعتماد على التلافيف المتوسعة المتعددة ذات وصلات التخطي المعتمدة على الانتباه

منى خلف[1]   بان نديم ذنون[2]

[1] قسم علم الحاسوب، كلية العلوم للبنات، جامعة بغداد، بغداد، العراق.
[2] قسم علوم الحاسبات، كلية العلوم، جامعة النهرين، بغداد، العراق.

الخلاصة:

زيادة حجم البيانات الطبية الصورية بسبب التطور الحاصل باجهزة التصوير الطبية من جهة وتطور اجهزة الكومبيوتر وزيادة قدرتها الحسابية وقدرتها الخزنية من جهة اخرى دفع بعجلة اتمتة تحليل وتشخيص الصور الطبيةحيث اصبحت عملية مجهدة ومكلفة لاطباء التشخيص كذلك تساعد الاتمته على تشارك الخبرات على نطاق اوسع . تعد تجزئة الصور الطبية من اول واهم مراحل التي تؤثر بشكل اساسي على اتمتة تحليل الصور الطبية وبالتالي تؤثر على مدى دقة التشخيص . اظهرت مناهج التعلم العميق في تجزئة الصورالطبية موثوقية اكثربالنتائج من المناهج التقليدية المتبعة سابقا. وخصوصا مع ضهور شبكة U-Net التي احدثت طفرة في نتائج التقسيم للصور الطبية باستخدام التعلم العميق.تتكون الشبكة من جزئين جزء يعمل على تصغير الصورة واستخراج الخصائص الاولية لها والجزء الثاني يعمل على اعادة تكوين الصورة واستخدام الخصائص الاولية التي نتجت من الجزء الاول بالاعتماد على وصلات التخطي. في هذا البحث قدمنا طريقة مطورة مبنية على تركيبة الU-Net تدعى شبكة MSDR-Unet. انها تستبدل البنية الاساسية لكل جزء في شبكة اليونت ببنية مطورة تدمج ابعاد مختلفة للخصائص المستخلصة من البيانات بالاضافة الى اعتمادها على تقنيات التخطي والانتباه والتلافيف المتوسعة للتعليم العميق. ولزيادة الدمج بين الخصائص تم اعادة تصميم وصلات التخطي . تم تقييم النموذج بثلاث مجاميع بينات للتقسيم مختلفة المتطلبات حيث المجموعة الاولى تخص تقسيم الافات الجلدية والمجموعة الثانية تخص تقسيم الزوائد اللحمية في القولون والاخيرة تخص تقسيم نوى الخلايا حيث كل مجموعة لها خصائص وابعاد والوان مختلفة. استخدمنا للتقييم خمس مقاييس معتمدة لتقييم جودة التقسيم في الصور الطبية. تم مقارنة طريقتنا مع طريقة اليو وثلاث طرق حديثة قياسية مطورة من اليو ومعتمدة كطرق تقسيم للصور الطبية حيث اثبتت طريقتنا تحسن ملحوض بالنتائج.

الكلمات المفتاحية: التعلم العميق, التلافيف المتوسعة, تقنية الانتباه, تقسيم الصور الطبية، شبكة ال U-Net .