2024 13th Iranian/3rd International Machine Vision and Image Processing Conference (MVIP)

2024 13th Iranian/3rd International Machine Vision and Image Processing Conference (MVIP) | 979-8-3503-5049-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/MVIP62238.2024.10491156

Ovarian Tumor Ultrasound Image Segmentation with Deep Neural Networks

Mahnaz Siahpoosh
Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
m.siahpoosh@modares.ac.ir

Maryam Imani
Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
maryam.imani@modares.ac.ir

Hassan Ghassemian
Image processing and Information Analysis Lab, Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
ghassemi@modares.ac.ir

***Abstract*—The precise and automated segmentation of ovarian tumors in medical images plays a pivotal role in the treatment of ovarian cancer in women. U-Net has demonstrated remarkable success in the field of medical image segmentation. However, due to its small receptive field, U-Net faces challenges in extracting global context information. Moreover, due to the significant variation in scale and size among tumors, it is essential to employ a network capable of effectively extracting information at Multiple scales. In this study, we present a U-Net-based network named PCU-Net for the segmentation of ovarian tumors, incorporating ConvMixer and Pyramid Dilated Convolution (PDC) modules. The ConvMixer module captures global context information by utilizing large-size kernels. The PDC module integrates local and global contextual patterns through utilization of parallel dilated convolution with different dilation rate. Furthermore, our model has fewer parameters than U-Net. We assess the proposed method's performance using the Multi-Modality Ovarian Tumor Ultrasound (MMOTU) dataset. The results indicate that in comparison to U-Net, our proposed PCU-Net exhibits an improvement of 4.23% in terms of Intersection over Union (IoU) and 2.99% in terms of Dice Similarity Coefficient (DSC).**

***Index Terms*—Segmentation, Ultrasound Images, U-Net, ConvMixer, Pyramid Dilated Convolution**

I. Introduction

Ovarian cancer ranks as one of the most lethal diseases among women [1]. Early-stage diagnosis and detection of ovarian tumors can markedly reduce the mortality rate. Currently, mostly manual methods are used for ovarian tumor segmentation, which are time-consuming and sensitive to radiologist's experience and human error. In addressing the challenges associated with manual segmentation, computer-assisted diagnosis (CAD) has been extensively developed for the automated analysis and segmentation of images.

In the diagnosis of ovarian cancer, Ultrasonography serves as a commonly employed medical imaging tool which utilizes high-frequency sound waves (ultrasound) to generate images of internal organs. This imaging method is non-invasive, non-radiative, cost-effective, and provides real-time capabilities for disease detection. However, segmenting ovarian tumors from ultrasound images presents challenges. The considerable variability in size, shape, and tissue patterns among tumors in different cases, as well as low contrast, high speckle noise and shadow artifacts of ultrasound images are among these challenges.

Deep learning have recently demonstrated successful applications in medical images semantic segmentation [2]–[9], computer vision and natural language processing. Fully Convolutional Networks (FCN) is a neural architecture designed for semantic segmentation in computer vision by Long et al. FCN preserves spatial information through convolutional layers, making it a foundational model for pixel-wise image classification tasks [10], [11].

U-Net [2] is a well-known convolutional neural network architecture, primarily designed for biomedical image segmentation. It features a U-shaped architecture with a contracting path for context extraction and an expansive path for precise localization. In the past few years, several medical segmentation networks, such as Attention U-Net [3], U-Net++ [12], UNet3+ [13], and SK-U-Net [14], based on the U-Net architecture, have been proposed. Attention U-Net is an extension of the U-Net model that incorporates an attention mechanism. This enhancement improves the model's focus on important regions in images, leading to better results in image segmentation tasks. U-Net++ is an extension of the U-Net architecture that incorporates skip pathways and nested skip pathways to enhance accuracy in image segmentation. This modification improves the network's ability to capture more complex features in images. TransUnet [4] is a fusion of CNN and Transformer [15] architectures, capitalizing on high-resolution spatial information from CNN features, along with the global context encoded by transformers.

979-8-3503-5049-4/24/$31.00 ©2024 IEEE

U-Net is not able to see the whole image due to the use of standard convolution and cannot extract global context information. In scenarios where the segmentation of medical images is concerned, the inclusion of global context proves to be highly significant. To solve this limitation, we propose the lightweight network CU-Net, which is a combination of U-Net and ConvMixer [16] module. ConvMixer can combine distant spatial locations and extract global context information by using large-sized convolutional kernels. Moreover, as previously stated, the tumors exhibit significant size diversity in various cases, and U-Net is incapable of adapting to these changes with a single receptive field size. Hence, our objective has been to design a network capable of amalgamating global and local context information. To achieve this, we present the PCU-Net network, drawing inspiration from U-Net and ConvMixer, and incorporating the PDC module. The PDC module employs four dilated convolutions in parallel for improved extraction of multiscale features.

In summary, our paper provides the following contributions:

- We propose a strong segmentation network, named CU-Net, derived from U-Net and ConvMixer, which achieves a large view of the receptive field with small trainable parameter size.
- Finally, we proposed a segmentation network (PCU-Net), based on U-Net, ConvMixer block and PDC. Notably, PCU-Net features a reduced parameter size during training compared to U-Net, while can accurately capture tumors with different sizes.
- We effectively enhance the performance of ovarian ultrasound images segmentation.

II. Method

In this study, we proposed two models based on U-Net, ConvMixer layer, and PDC. In this section, we started by introducing U-Net, our ConvMixer module, and PDC module. Then, we described the overall framework of our networks.

A. U-Net

U-Net is a popular architecture in deep learning for semantic segmentation of images. It consists of a contracting path (encoder) and an expansive path (decoder) with skip connections. In the encoder section, the goal is to receive the ultrasound image as input and achieve a condensed representation of the input image. The decoder section decodes the information extracted from the encoder stage to the size of the input image. The encoder consists of five stages. Each step in this path consists of two 3x3 convolutions. Except in the last step, a 2x2 max pooling operation with a stride of 2 is applied after these two convolutions, serving the purpose of downsampling. The encoder comprises four stages. Each step of the decoder includes a 2x2 up-convolution for upsampling of the feature map and then two 3x3 convolutions. Each convolution followed by a rectified linear unit (ReLU) activation function and batch normalization. Padding and stride are 1 in all 3x3 convolutions. The final layer employs a 1x1 convolution to produce pixel-wise predictions, generating segmentation masks for each class. In U-Net, skip connections are direct connections between corresponding layers in the contracting path (encoder) and the expanding path (decoder), facilitating the seamless flow of high-resolution spatial information to the decoding layers and aiding in the precise localization of features, thereby improving the segmentation accuracy.

B. ConvMixer

As you can see the structure of the ConvMixer layer in the Figure 1, this layer consists of a Depthwise Convolution followed by a Pointwise convolution. Each of the convolutions is followed by an activation function and a batch normalization. The activation function in the original ConvMixer is GELU, but in this work we used ReLU activation function.The computational procedure of ConvMixer is delineated in Equations (1) and (2).

$$z'_l = \text{BN}(\sigma(\text{DepthwiseConv}(z_{l-1}))) + z_{l-1} \tag{1}$$

$$z_l = \text{BN}(\sigma(\text{PointwiseConv}(z'_l))) \tag{2}$$

Where z_l is the output feature map of layer l in the ConvMixer block, σ represents the activation function, and BN denotes batch normalization.

Depthwise convolution involves applying a single convolutional kernel for each input channel and pointwise convolution is employed to form a linear combination of the depthwise convolution output. This approach reduces the number of parameters. Due to the fewer parameters of ConvMixer compared to standard convolution, larger filter sizes can be used in Depthwise convolution. By using filters with a large size, it is possible to combine features in distant spatial locations and extract global context information.

C. Pyramid Dilated Convolution (PDC)

The receptive field of a convolutional layer within a neural network refers to the area of the input image utilized for calculating the output of a specific neuron in the feature map. The larger the receptive field, the more context information is received from the input. Getting more context information brings higher level features. In this regard, Fisher et al. [17] proposed the dilated Convolution which can increase the receptive field without losing the spatial resolution and increasing the number of parameters.

In this work, we used the PDC module to combine context information at different scales. As you can see the PDC Structure in Figure 2, the The input is given in parallel to 4 convolutions with different dilation rates and filter sizes and the outputs of these convolutions are concatenated and then passed through a 1x1 convolution to get different feature maps with combine each other and the number of channels reaches the number of input channels. At the end, batch normalization and ReLU activation function have been applied. The computational procedure of PDC module is delineated in Equation (3).

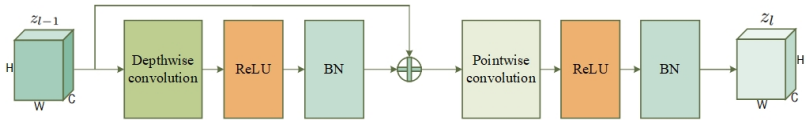

Fig. 1. Pyramid Dilated Convolution (PDC) module structure.

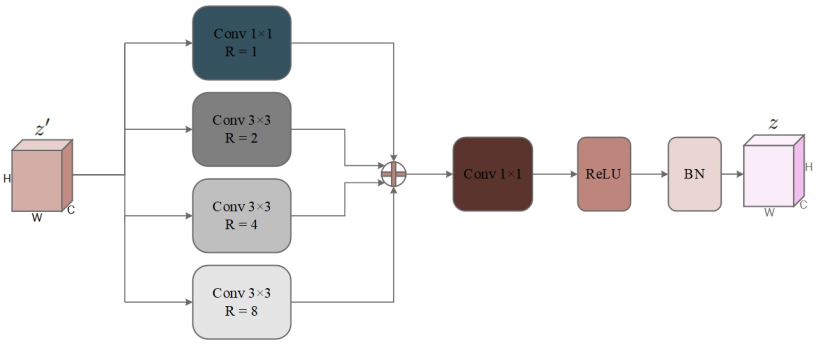

Fig. 2. Pyramid Dilated Convolution (PDC) module structure.

$$z = \text{BN}\left(\sigma\left(\text{PointwiseConv}\left(\text{Concat}\left\{\text{PointwiseConv}(z'),\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\text{DilatedConv2}(z'), \text{DilatedConv4}(z'), \text{DilatedConv8}(z')\right\}\right)\right)\right) \quad (3)$$

Where z represents the output feature map of the PDC block, z' corresponds to the input feature map of the PDC block σ represents the activation function, and BN denotes batch normalization.

D. CU-Net and PCU-Net

As shown in Figure 3, the architecture of the CU-Net model, the encoder includes 5 stages. In each step, the feature maps are first entered into a standard convolution and then into the ConvMixer block. We incorporate max pooling with a window size of 2×2 following each ConvMixer block for down sampling, except in the final step. For each standard convolution, the activation function is ReLU, the kernel size is 3x3, and both padding and stride set to 1. Following each standard convolution, a batch normalization layer is applied. Each ConvMixer block contains 2 ConvMixer layers. The size of the kernels in the depthwise of each layer of ConvMixer is 7 x 7. Padding and stride are both set to 1 in all depthwise and pointwise convolutions, hence the resolution of input and output feature maps of convolution blocks is the same.

The architecture of PCU-Net is the same as Figure 3, except that in the final step of the encoder in CU-Net, the ConvMixer block is replaced by the PDC module with the dilation rates [1, 2, 4, 8]. All convolutions within the PDC module utilize padding and stride values of 1, maintaining uniform resolution for both the input and output feature maps. The decoder path of these two models is the same as the encoder U-Net.

III. Experiments, Results and Discussion

A. DATASET

We used the Multi-Modality Ovarian Tumor Ultrasound (MMOTU) [18] image dataset to evaluate the proposed methods. The images of this dataset were taken from Beijing Shijitan Hospital, Capital Medical University. The MMOTU image dataset consists of 1639 ovarian ultrasound images, collected from 294 patients, which includes two subsets with two modes, OUT_2d and OUT_CEUS, containing 1469 2D ultrasound images and 170 CEUS images, respectively.

In this work we used OUT_2d. As shown in Figure 4, the images in OUT_2d have different scales, so that the width of the images varies from 302 to 1135 pixels and their height from 226 to 794 pixels. Before training, we resized the images to 384 × 384. We have utilized data augmentation techniques such as flipping horizontally and rotating by 90, 180, and 270 degrees. We conduct all our experiments employing a 5-fold cross-validation approach.

B. Experimental setup

In the training phase, a combination of cross-entropy loss (BCE) and dice loss (Dice) is applied as the cost functions:

$$L(t, p) = \text{BCE}(t, p) + \text{Dice}(t, p) \quad (4)$$

where t and p represent the prediction and target, respectively. We employ the Adam optimizer [19] for network optimization. The training process involves using a batch size of 4, and the number of 100 epochs.

C. Results and Comparisons

To quantitatively assess the effectiveness of various segmentation models, we employed different performance evaluation metrics, including the Dice similarity coefficient (DSC), Intersection over Union (IoU), Precision, Recall and Accuracy.

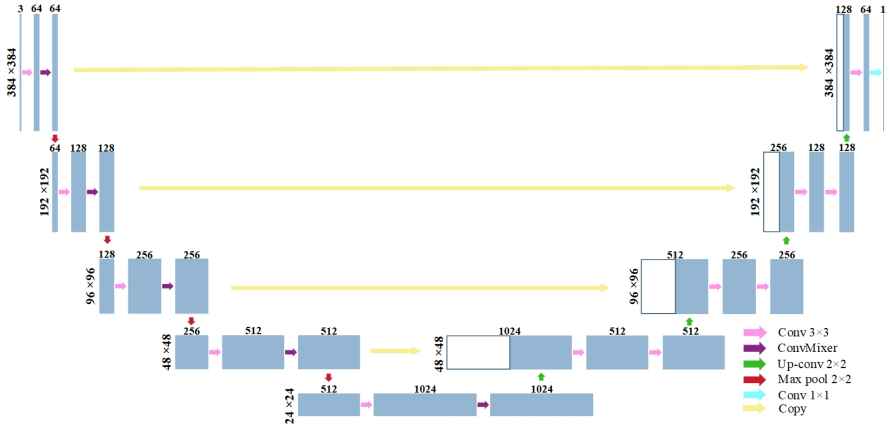

Fig. 3. Overview of the proposed CU-Net architecture.

TABLE I
THE SEGMENTATION RESULTS (MEAN ± STD) ON OTU-2d DATASET (%).

Method	IoU	DSC	Precision	Recall	Accuracy
U-Net	69.82±1.54	76.96±1.07	78.66±3.74	80.58±2.02	96.43±0.13
CU-Net	72.88±0.24	79.20±0.13	81.78±1.21	83.62±2.24	97.06±0.22
PCU-Net	**74.05±0.43**	**79.95±0.94**	**83.79±2.73**	**84.46±0.54**	**97.54±0.05**

TABLE II
THE SEGMENTATION RESULTS (MEAN ± STD) ON AUGMENTED OTU-2d DATASET (%).

Method	IoU	DSC	Precision	Recall	Accuracy
U-Net	78.83±0.10	85.97±0.68	87.48±0.96	85.41±0.93	98.41±0.02
CU-Net	80.81±0.33	87.35±0.19	89.59±0.82	86.82±0.18	98.92±0.01
PCU-Net	**81.65±0.18**	**87.98±0.22**	**90.56±0.51**	**87.31±0.34**	**99.13±0.01**

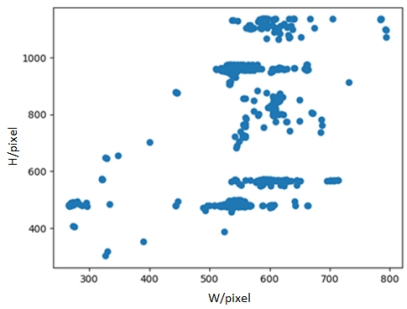

Fig. 4. The scatter plot illustrating the distribution of image scale.

We conducted training for the U-Net, CU-Net, and PCU-Net models using both the original dataset and the augmented dataset. Tables I and II present the quantitative results for these networks trained with the original and augmented datasets, respectively. As evident from Tables I and II, augmenting the number of training samples leads to improved performance in the networks. PCU-Net demonstrates superior segmentation performance compared to the other two methods. Considering the results of the networks on the original data, CU-Net improved 3.06% in terms of IoU and 2.24% in terms of DSC, and PCU-Net improved 4.23% in terms of IoU and 2.99% in terms of DSC compared to the U-Net.

Figure 5 displays a few example outcomes on OTU-2d dataset without augmentation. It is evident from the figures that PCU-Net produces lesion regions and shapes with greater accuracy. Furthermore, Table III presents the parameter counts for each of the models, revealing that CU-Net boasts the smallest parameter count among the three networks. This efficiency is attributed to the incorporation of ConvMixer blocks. Additionally, PCU-Net has fewer parameters compared to U-Net. However, as shown in III, the time required to test a sample in U-Net is less than the rest of the models.

TABLE III
NUMBER OF PARAMETERS AND THE TIME TO TEST A SAMPLE IN EACH MODEL

Method	Number of parameters	Test duration (in seconds)
U-Net	34527041	**0.030**
CU-Net	**24962177**	0.037
PCU-Net	31145089	0.040

Also we compare PCU-Net, U-Net [2], Attention U-Net [3] and TransUnet [4]. Table IV reveals that PCU-Net achieves superior segmentation performance compared to the other three approaches.

IV. CONCLUSION

In this study, we introduced PCU-Net, a rapid and efficient approach for medical ultrasound segmentation. Recognizing

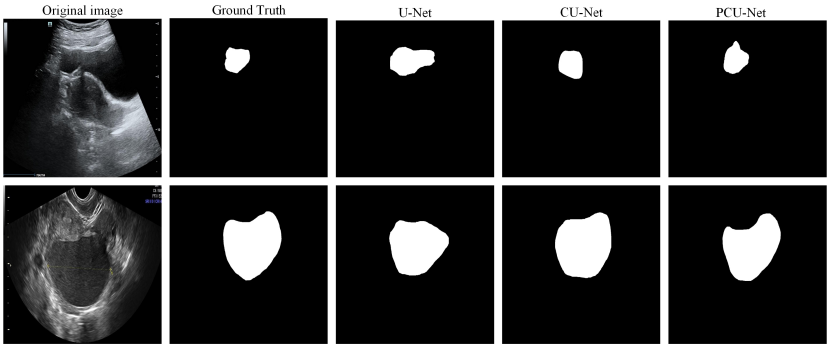

Fig. 5. segmentation results from manual delineation, U-Net, CU-Net, and PCU-Net models. Row 1 - OTU-2d dataset without augmentation, Row 2 – OTU-2d dataset with augmentation.

TABLE IV
THE SEGMENTATION RESULTS (MEAN ± STD) ON OTU-2d DATASET (%).

Method	IoU
U-Net [2]	69.82
Attention U-Net [3]	71.10
TransUnet [4]	69.13
PCU-Net	**74.05**

the significance of capturing global contextual information for achieving favorable segmentation results, we incorporated ConvMixer blocks into a U-Net network. To address challenges posed by variations in tumor size, shape, and tissue pattern, we introduced the PDC module into U-Net, enabling the network to effectively capture multi-scale contextual information.

We validated the performance of PCU-Net on a public ultrasound dataset, and our experimental results demonstrate that our proposed model achieves state-of-the-art performance. Our future plans include testing the model on a diverse range of medical image volumes and applying it to image detection tasks.

REFERENCES

[1] K. Ushijima *et al.*, "Treatment for recurrent ovarian cancer—at first relapse," *Journal of oncology*, vol. 2010, 2010.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[3] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[4] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[5] J. M. J. Valanarasu and V. M. Patel, "Unext: Mlp-based rapid medical image segmentation network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 23–33.

[6] Y. Zhang, H. Liu, and Q. Hu, "Transfuse: Fusing transformers and cnns for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, 2021, pp. 14–24.

[7] A. Mansouri, M. Noei, and M. Saniee Abadeh, "A hybrid machine learning approach for early mortality prediction of icu patients," *Progress in Artificial Intelligence*, vol. 11, no. 4, pp. 333–347, 2022.

[8] A. Mansouri, M. Noei, and M. S. Abadeh, "Predicting hospital length of stay of neonates admitted to the nicu using data mining techniques," in *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, 2020, pp. 629–635.

[9] M. Noei and M. S. Abadeh, "A genetic asexual reproduction optimization algorithm for imputing missing values," in *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*. IEEE, 2019, pp. 214–218.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[11] M. Noei, M. Parvizimosaed, A. S. Bigdeli, and M. Yalpanian, "A secure hybrid permissioned blockchain and deep learning platform for ct image classification," in *2022 International Conference on Machine Vision and Image Processing (MVIP)*. IEEE, 2022, pp. 1–5.

[12] 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: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, 2018, pp. 3–11.

[13] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2020, pp. 1055–1059.

[14] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, and M. Andre, "Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network," *Biomedical Signal Processing and Control*, vol. 61, p. 102027, 2020.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[16] A. Trockman and J. Z. Kolter, "Patches are all you need?" *arXiv preprint arXiv:2201.09792*, 2022.

[17] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[18] Q. Zhao, S. Lyu, W. Bai, L. Cai, B. Liu, M. Wu, X. Sang, M. Yang, and L. Chen, "A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation," *arXiv preprint arXiv:2207.06799*, 2022.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.