IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 November 2024, accepted 14 December 2024, date of publication 24 December 2024,
date of current version 31 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3522022

RESEARCH ARTICLE

Performance Analysis of Dilated One-to-Many U-Net Model for Medical Image Segmentation

VAHID ASHKANI CHENARLOGH[1,2], ARMAN HASSANPOUR[2,3], KATARINA GROLINGER[1], (Senior Member, IEEE), AND VIJAY PARSA[1,2,3,4]

[1]Department of Electrical and Computer Engineering, Western University, London, ON N6A 3K7, Canada
[2]National Centre for Audiology, Western University, London, ON N6A 3K7, Canada
[3]Health and Rehabilitation Sciences Program, Faculty of Health Sciences, Western University, London, ON N6A 3K7, Canada
[4]School of Communication Sciences and Disorders, Western University, London, ON N6A 3K7, Canada

Corresponding author: Vahid Ashkani Chenarlogh (vashkani@uwo.ca)

The work of Vahid Ashkani Chenarlogh, Arman Hassanpour, and Vijay Parsa was supported by the Ph.D. Funding Support to the Natural Sciences and Engineering Research Council (NSERC) Canada, through the Discovery Grant.

ABSTRACT Medical image processing applications typically demand highly accurate image segmentation. However, existing segmentation approaches exhibit performance degradation when faced with diverse medical imaging modalities and varied segmentation target sizes. In this paper, we propose and evaluate a dilated One-to-Many U-Net deep learning model that addresses these challenges. The proposed model comprises of four rows of encoder-decoder modules, with each module consisting of three trainable blocks with different layers. The last three rows of the U-Net are extended versions of the three blocks in the first row, with the encoder-decoder blocks connected through the skip connections to the previous rows. The outputs of the last blocks from the last three rows in the decoder are concatenated, and finally, a dilation network is employed to improve the small target segmentation in different medical images. Two datasets have been used for the evaluation: the HC18 grand challenge ultrasound dataset for fetal head segmentation and the Multi-site MRI dataset, including the BIDMC and HK sites, for prostate segmentation in MRI images. The proposed approach achieved Dice and Jaccard coefficients of 96.54% and 93.93%, respectively, for the HC18 grand challenge dataset, 96.76% and 93.97% for the BIDMC site dataset, and 92.58% and 86.96% for the HK site dataset. Statistical analyses showed that the proposed model outperformed several other U-Net-based models.

INDEX TERMS U-Net, medical image segmentation, deep learning, dilation network, neural network.

I. INTRODUCTION

Automated medical image segmentation has been widely investigated in the image analysis community. Precise and reliable solutions are yearned to enhance clinical workflow effectiveness and support decision-making through fast and automatic extraction of quantitative measurements. A critical step in diagnosing, treating, and monitoring many diseases is the semantic segmentation of medical images. Although the automation of this task has been widely studied in the past, manual annotations are commonly utilized in clinical practice [1], which are time-consuming and susceptible to inter and intra-observer variability. Therefore, there is a high interest in accurate and reliable automatic segmentation approaches that enhance workflow productivity in clinical scenarios and lighten the workload of radiologists and other medical experts.

Over the decades, many automated medical image segmentation techniques have been proposed, principally concentrated on images of specific modalities. In the early days, simple rule-based approaches were utilized, but those methods failed to preserve robustness when tested on a massive variety of data [2]. Prior methods to medical image segmentation are usually based on edge detection and template matching [3]. Furthermore, different adaptive algorithms were presented, relying on geometric shape priors

The associate editor coordinating the review of this manuscript and approving it for publication was Hasan S. Mir.

with soft computing tools [4] and fuzzy algorithms [5]. The disadvantages of these methods lie in using hand-crafted features to perform the segmentation.

Recent advancements in deep learning [6] have shown great promise in solving such issues. In this regard, Convolutional Neural Networks (CNNs) [7] have been more successful than other architectures, and have in particular revolutionized semantic segmentation tasks. Therefore, in recent years, the CNNs have attracted significant attention in the medical image segmentation community [8], with the U-Net architecture providing improved segmentation performance [9]. A U–Net model is comprised of encoding-decoding modules. In the encoding module, numerous feature maps with reduced dimensionality are elicited from the input data. From those feature maps, the decoding module produces segmentation maps of the same size as the input by performing transposed convolutions. While the basic U-Net model has been shown to detect targets with predefined shapes or at expected locations, it has limitations in extracting the necessary complex features when the target object has a non-standard shape and random location [10]. For this reason, many extensions of the U-Net architecture have recently been proposed for specific applications, each with its own pros and cons. Many of these models exhibit suboptimal performance when processing a diverse array of image types.

This paper presents a novel U-Net architecture, named ''dilated One-to-Many U-Net model'', which exhibits enhanced performance compared to a number of state-of-the-art U-Net models when applied to challenging image segmentation problems. First, by proposing the one-to-many model, we are increasing how much information is transferred from the encoder to the decoder module. Then, we take advantage of a sequential multi-stream dilated network to enable the model to precisely segment the small targets. Experiments on different datasets show that the proposed model outperforms many other state-of-the-art U-Net models in segmenting targets with different sizes and variable shapes.

The rest of the manuscript is organized as follows: Section II reviews the relevant literature surrounding U-Nets. Section III details the proposed U-Net architecture. Section IV describes the image segmentation datasets and experiments setup. Section V, reports the performance of the proposed U-Net architecture. Finally, the manuscript concludes with the summary of our investigations in Section VI.

II. RELATED WORKS

Lately, deep learning-based models with the encoder-decoder structure have been widely applied in many different domains due to their capability to achieve accurate performance across diverse real-world applications. For example, Khan et al. [11] employed an encoder-decoder-based structure for footprints extraction from aerial images where the encoder module employs a dense convolutional architecture to extract global multi-scale features, and the decoder module uses consecutive deconvolution layers to create a dense segmentation map in the output. In another work, Khan et al. [12] proposed a pixel-wise classification using multi-scale extracted features through the combination of the DenseNet and U-Net models.

Owing to the promising results obtained from the U-Net structure, this architecture has also been used to analyze various medical images collected through the ultrasound, MRI, and CT modalities. In many segmentation tasks, the U-Net architecture has attracted more attention than other deep learning-based solutions for medical image segmentation [13], [14], [15], [16], [17], [18], [19]. Furthermore, many extensions of this architecture have also been proposed recently [13], [20].

Sevastopolsky et al. [21] applied U-Net to directly segment the optic disc and optic cup in retinal fundus images for glaucoma diagnosis. Mubashar et al. [14] used a U-Net-based network, a multi-scale recurrent, residual model with dense skip connections for medical image segmentation in CT images. Zhou et al. [15] introduced the modified version of the main U-Net [9] architecture called UNet++. Additionally, the authors enhanced the UNet++ architecture by reconfiguring the skip connections between the encoder and decoder modules, enabling the integration of features extracted at various scales.

Owais et al. [16] used a modified U-Net model to carry out the medical image segmentation task throughout the edge devices. The SE-U-Net proposed by Guo and Matuszewski [19] is a U-Net network augmented by the dilation kernel to segment the polyp in colonoscopy images. A modified encoder-decoder with several integrated sequential depths dilated inception blocks based on deep learning has also been proposed by Mahmud et al. [22] to overcome the limitations of traditional approaches by aggregating features from different receptive areas of dilated convolutions.

Moradi et al. [23] proposed a Multi-Feature Pyramid U-Net (MFP U-Net) model for left ventricle segmentation. They equalized the depth of all feature maps within the decoder module in order to increase the segmentation accuracy. Automated concentration on different regions of interest and/or targets through the use of Attention Gates (AGs), known as the Attention U-Net model, has been proposed by Oktay et al. [13]. Generating different scales of context information with minimal loss is one of the dilated U-Net advantages presented in [24]. In another work, Chalana and Kim [25] studied a method for evaluating fetal head segmentation tasks in ultrasound images. They introduced a protocol for assessing algorithms for medical image segmentation tasks using available boundaries extracted by multiple expert observers. The authors have employed statistics to find the common points of the boundaries between the computer-generated and the expert annotated hand-outlined boundaries.

Heuvel et al. [26] employed a random forest classifier to extract the Haar-like features from the ultrasound images of

the fetal skull and investigated Hough transforms for Head Circumference (HC) extraction. Lu et al. [27] addressed the challenge of filtering and detecting incomplete curves of the extracted fetal head border using the direct inverse randomized Hough transform method. In their study [27], the borders are highlighted and detected by iteratively applying the mentioned method to the ultrasound images to extract the fetal head area. Li et al. [28] localized the fetal head using a random forest classifier trained with prior knowledge about the gestational age and depth of the ultrasound scanning system. Moreover, they have used a phase symmetry fast ellipse (Elli-Fit) to fit the HC ellipse for head measurement. In another work, Avalokita et al. [29] represented optimum pixels of the ultrasound images for fetal head measurement using the ellipse fitting method throughout the pre-localized region of interest as the fetal head area.

Sobhaninia et al. [30] employed a multi-task deep CNN model for HC segmentation and estimation using ultrasound 2D images. Fiorentino et al. [31] proposed an approach with two CNN models. The first CNN model is based on transfer learning for head localization and centring, and the second regression CNN is based on distance fields to delineate the HC. Amini [32] proposed a deep learning model using multi-scale ultrasound images to extract fetal head circumference. They introduced a DeepLinkNet loss function to weigh the border pixels of the fetal head that can improve the segmentation performance of the model and reduce the number of training parameters. Moreover, HC extraction and measurement using different base approaches have been studied in [33], [34], and [35].

Prostate segmentation using MRI data has been studied by Zhang et al. [36]; they proposed a deep-stacked transformation approach using transfer learning for domain generalization in segmentation tasks. Liu et al. [37] improved model generalization through prostate MRI segmentation using the shape-aware meta-learning strategy. Ashkani et al. [38] proposed a U-Net-based model along with the attention gates and residual blocks for the segmentation of clinical targets in ultrasound and MRI data. They offered a segmentation architecture that comprises two consecutive U-Net models. In each model, they employed residual blocks between the encoder-decoder modules and multi-scale attention gates to produce richer contextual information within the skip connections that help the networks segment various targets.

Table 1 summarizes the discussed U-Net-based models. The existing U-Net-based architectures are limited in their ability to segment challenging targets such as small-size targets and those with shape variability across different types of medical images [10]. This causes inconsistencies in segmentation accuracy when different target sizes with various shapes are involved in the different medical images. This inconsistency hampers precise segmentation efforts in medical imaging, highlighting the need for a stable, accurate model to improve segmentation performance in real-world applications.

In our research, we improved the segmentation of medical targets when it comes to generalized segmentation across different imaging modalities and with different target sizes. We have proposed a different U-Net-based segmentation architecture that is able to maintain performance consistency when faced with different image types and target sizes by exploiting multi-level extended features from previous layers in both encoder and decoder modules.

III. THE PROPOSED MODEL

This paper introduces a novel approach through dilated multi-level feature modeling to enhance segmentation stability for small targets and variable shapes within various medical imaging systems. The proposed architecture, building upon the innovative U-Net-based model, facilitates the transfer of a broad spectrum of information levels to the output. It utilizes features from diverse levels to focus on different target shapes and sizes, thereby achieving more accurate predictions.

The proposed model is comprised of two sequentially connected networks, including a One-to-Many network and a dilation network. We named the first part of our model *one-to-many* because each block in both the encoder and decoder in the first row is extended to many other blocks through the next rows. In other words, one-to-many refers to a multi-level U-Net architecture that is extended from previous layers in both encoder and decoder modules. This helps the network to perceive the extracted features well by transferring more detailed encoded information toward the decoder module through the many skip connections. Furthermore, transferring more information from the encoder module to the decoder module improves the robustness of the segmentation task, which is the philosophy of the proposed architecture. Next, a multi-stream dilated network has been added for segmenting small targets within the input data. In the proposed multi-stream network, each stream running parallelly using a distinct dilation rate provides different features and enables the network to concentrate on small targets.

Fig. 1 illustrates the proposed dilated One-to-Many model. Encoder and decoder modules are comprised of several blocks and connected through the bottleneck layer. The encoder blocks, the bottleneck layer and the decoder blocks are described as follows:

The Encoder Blocks: In the proposed model (Fig. 1), there are four rows in the encoder module, and each row consists of three encoder blocks. Hyperparameters (e.g., filter size, number of filters) for all blocks are determined empirically through the grid search strategy. The last three rows are an extended version of the three blocks from the first row: the connections to extended rows are depicted with the dotted orange line in Fig. 1. Each block in the encoder module, as seen in Fig. 1, is comprised of consecutive 2D-convolution, a Rectified Linear Unit (ReLU) [38] as an activation layer, and a batch normalization [35] layer that improves convergence [32]. We have used a convolutional

TABLE 1. Summarized and clarified the most similar works and their contributions.

Most similar works	Main contribution
Attention U-Net	Attention gates are used to highlight informative features and suppress less informative ones. These gates caused concentration on informative regions of interest or targets.
Dilated U-Net	Generated different scales of context information with minimal loss.
MFP U-Net	Equalized the depth of all feature maps within the decoder module in order to increase segmentation accuracy.
R2U-Net	Employed embedding recurrent residual convolution blocks to feature accumulation through the deeper network to be ensured regarding better feature representation.
U-Net	The basic U-Net model with the encode-decoder structure suffers from a limited capabilities problem.

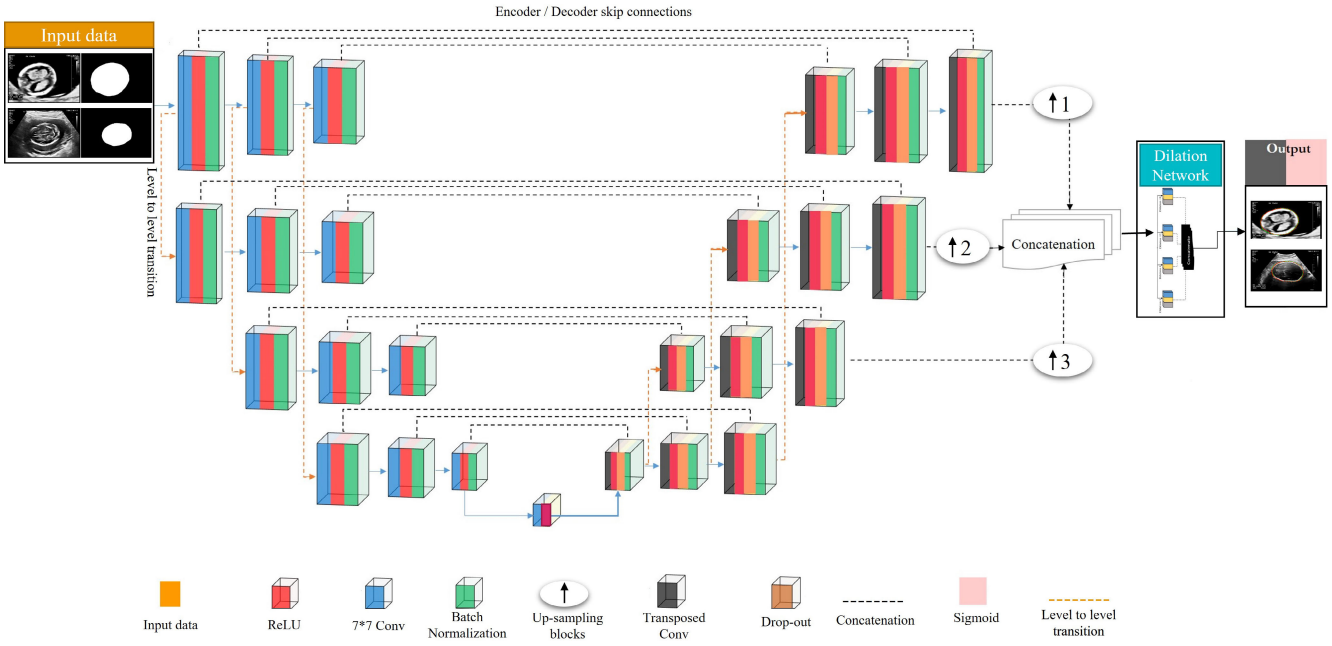

FIGURE 1. Overview of the proposed dilated One-to-Many U-Net based model.

kernel of 7 × 7, which is swept by 2 × 2 strides, and initialized these convolutional layers randomly by a normal distribution with a standard deviation of 0.02. The image is padded, and in order to extract various feature maps in different blocks, 20, 40, and 80 convolutional filters were used for layers in the first encoder row, 40, 80, and 160 filters in the second row, 80, 160, and 320 filters in the third row, and finally, 160, 320, and 640 filters for the three layers in the fourth row of the encoder module, respectively.

Bottleneck Layer: This is a 2D convolutional layer with 640 convolutional filters and a 7 × 7 kernel, which sweeps with 2 × 2 strides. This convolutional layer is initialized with a standard deviation of 0.02 and a ReLU layer was used as an activation function after this layer.

The Decoder Blocks: In the proposed architecture, each block of the decoder module is comprised of 2D transposed convolution layers that are initialized randomly using a normal distribution with a standard deviation of 0.02. A 7 × 7 kernel with 2 × 2 strides swept over the inputs in all decoder layers. The number of filters in the decoder blocks is the reverse of that in the encoder blocks. Thus, there are 640, 320, and 160 filters in the first row of the decoder module, 320, 160, and 80 filters in the second row, 160, 80, and 40 filters in the third row, and finally, 80, 40, and 20 filters for the three layers in the fourth row of the decoder module, respectively.

A ReLU activation function followed each deconvolution layer. A dropout layer with a probability of 50% has been employed after each activation function to avoid overfitting during the training. Finally, the batch normalization layer is added after concatenating the output of the dropout layer with the appropriate skip connection feature maps from the encoder in order to improve convergence.

The decoder module adopts a similar extension strategy as the previously described encoder, utilizing reverse skip connections that extend from the lower to the upper rows. The output of the One-to-Many model is obtained by concatenating the last blocks in the last three rows of the decoder module, where the features are processed using the up-sampling block. This up-sampling block is comprised of a 2D transposed convolutional layer and a ReLU activation function. Each 2D transposed convolutional layer includes 20 convolutional filters that are initialized randomly using a normal distribution with a standard deviation of 0.02. A 7×7 kernel with a 2×2 stride swept over the inputs in all decoder layers. In the decoder module, as seen in Fig. 1, the up-sampling ratio is 3, 2, and 1 for the second, third, and last rows, respectively.

Up-sampling was accomplished through consecutive 2D transposed convolutional layers, the number of which depended on the up-sampling ratio. In other words, for an up-sampling with a ratio of 3, there are three consecutive 2D transposed convolutional layers and so on. Then, the outputs of these blocks were concatenated to extract feature maps from all levels of the expansion module to be included in the segmentation process in the output.

Fig. 2 represents the proposed dilation network which has been applied to the output of the One-to-Many network to render the proposed model more robust in the face of different sizes of segmentation targets in medical images. In this step, we take advantage of dilated convolutions for segmenting small segmentation targets [40] of the input data. Different dilation rates are used throughout the multi-stream dilation network architecture and concatenated. Thus, the output is composed of features with different dilation ratios.

Specifically, the proposed dilation network is a multi-stream model where the input has been applied to four different dilation blocks. Each block comprises of a 2D convolutional layer containing a dilated 7×7 kernel. After each dilated convolutional layer, the ReLU activation layer and the batch normalization layer have been employed. The dilation rate equals to 2, 4, 6, and 8 in the four blocks, respectively. Finally, the outputs of the four streams are concatenated. The number of filters through the four paralleled dilated layers are 20, 40, 80, and 160.

Finally, we applied a 2D transposed convolutional layer with a 7×7 kernel size which were swept by 1×1 strides in the output layer. A sigmoid function followed this layer to generate the mask of corresponding inputs. The loss function we have used to train the model is the Dice loss function. Dice loss is particularly effective for medical image segmentation due to its ability to handle class imbalance, be robust to variation, and facilitate convergence in imbalanced datasets. The Adam optimizer with a learning rate equal to 0.0001 is used to train the complete model.

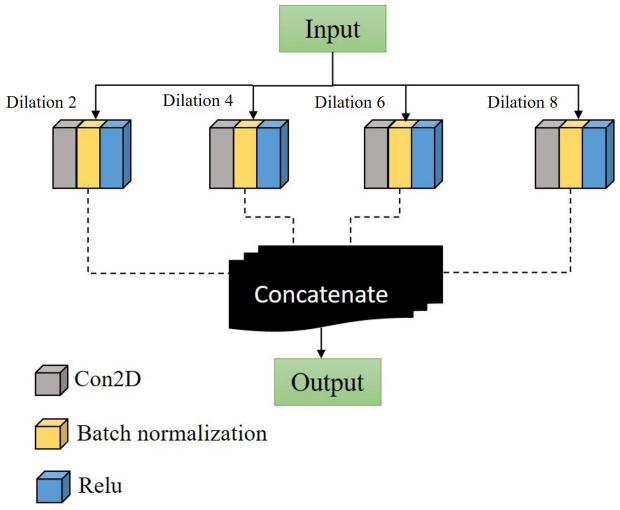

FIGURE 2. The proposed dilation network architecture.

It is pertinent to note that due to deep learning-based structure, hyper-parameters must be tuned to achieve high accuracy. Specifically, the grid search strategy is employed and the hyper-parameters are empirically determined through many experiments using the grid search strategy.

IV. EXPERIMENTS

A. DATASETS

We conducted experiments on two datasets with different image types, including the HC18 Grand challenge dataset [26] and the Multi-site MRI dataset [41]. A few sample images, as well as their corresponding masks, are shown in Fig. 3.

HC18-Grand Challenge Dataset: This public dataset comprises 1334 two-dimensional ultrasound images to measure the fetal HC parameter. In this dataset, there are 999 images manually annotated by an expert and 335 unannotated images. The resolution of all ultrasound images is 800 by 540 pixels, with a pixel size ranging from 0.052 to 0.326 mm.

Multi-Site MRI Dataset: The T2-weighted MRI dataset was collected for prostate identification purposes. This dataset is comprised of two public sources - sites E and F. The detailed information of each source E and F containing the number of samples, resolution of images, and imaging protocols is summarized in Table 2. These sites are more commonly used than others as MRI data for prostate segmentation in research communities [37], [38]. Thus, we evaluated the datasets from both the sites E and F to compare our approach with state-of-the-art approaches.

TABLE 2. Summary information for the two selected sites in the multi-site MRI dataset.

Dataset	Institution	Case num	Field strength (T)	Resolution (mm)	Endorectal Coil	Manufacturer
Site E	BIDMC	12	3	0.25/2.2-3	Endorectal	GE
Site F	HK	12	1.5	0.625/3.6	Endorectal	Siemens

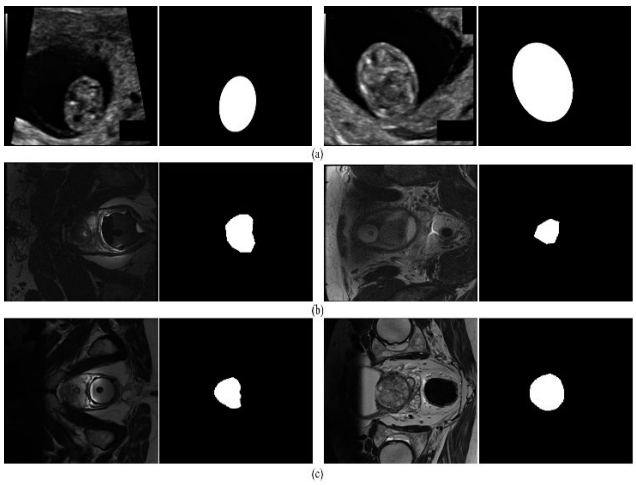

FIGURE 3. Samples from the HC18 Grand challenge, and Multi-site MRI datasets with the corresponding annotations of the target structures. (a): HC18, (b): Site E or BIDMC, (c): Site F or HK.

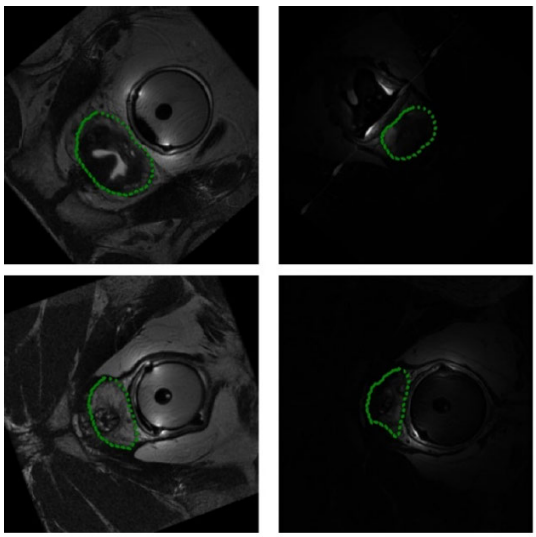

FIGURE 4. Some samples of prostate datasets that R2U-Net and U-Net models could not predicted the output mask to calculating the mean of HD. Green dotted contours represent the ground truth mask.

B. PRE-PROCESSING

The proposed model was independently trained using the datasets mentioned earlier. In total, in our experiments, we have used all 999 labeled samples from the HC18 Grand challenge dataset, 1215 samples for site E, and 605 samples for site F in the multi-site MRI dataset. From these datasets, 20% of data was randomly selected to evaluate the model in the test phase and the remaining 80% of each dataset was used to train the model independently.

All data were resized to 256 by 256 pixels resolution. Data augmentation was performed on the Multi-site MRI dataset using rotation with two angles (15 and 30 degrees), and horizontal flip of rotated images to generate five images from each sample. The entire set of input images was converted to grayscale and normalized by their standard deviation before the model training. The normalization was carried out as shown in Eq. 1 where x_i and z_i are the sample and normalized sample respectively and *sigma* is the standard deviation of x.

$$z_i = \frac{x_i - mean(x)}{\sigma} \tag{1}$$

C. EVALUATION METRICS

First, to compare the estimated volumes of the target structures, we used the Dice Similarity Coefficient (DSC). Further, we evaluated the segmentation performance by comparing the ground truth contours with the predicted ones using the Jaccard Similarity Coefficient (JSC) and Hausdorff Distance (HD). The DSC indices were calculated using Eq. 2, where B indicates ground truth contours and A represents the model's predicted contours. Finally, JSCs were calculated using equation Eq. 3.

$$DSC = \frac{2\,|A \cap B|}{|A| + |B|} \tag{2}$$

$$JSC = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{3}$$

In order to measure the maximum distance of the predicted contour to the nearest point in the reference contours, we have calculated HD (see Eq. 4), where A and B denote the two contours, and $d(a, b)$ indicates Euclidean distance.

$$HD = max\left(\left(\max_{a\in A}\left(\min_{b\in B} d\,(a,b)\right)\right), \left(\max_{b\in B}\left(\min_{a\in A} d\,(a,b)\right)\right)\right) \tag{4}$$

We further computed an expanded set of metrics, including the Intersection over Union (IoU), Precision, Recall, and F1 Score, for a more comprehensive quantitative evaluation of the model performance. These metrics are defined within equations Eqs. 5–8. The IoU metric is defined by the number of common ground truth pixels and predicted output masks divided by the whole pixels across both ground truth and predicted masks. Precision is described using the proportion of positive detections (true positive and false positive per

TABLE 3. The maximum (i.e. highest) Dice and Jaccard coefficients across all 150 epochs for different U-Net models and different datasets. Bolded values indicate the best model under each column.

Models	HC18		BIDMC		HK		Number of training parameters
	Dice	Jaccard	Dice	Jaccard	Dice	Jaccard	
Proposed Model	0.9654	0.9393	**0.9676**	**0.9397**	0.9258	0.8696	21,798,641
Attention U-Net	0.9686	0.9452	0.8697	0.7808	0.9225	0.8599	1,502,954
Dilated U-Net	**0.9714**	**0.9491**	0.9475	0.9030	0.9351	0.8807	1,486,033
MFP U-Net	0.9691	0.9454	0.9635	0.9324	0.9167	0.8515	1,599,066
R2U-Net	0.9448	0.9124	0.5615	0.4809	0.8195	0.7479	1,590,673
U-Net	0.9704	0.9481	0.7955	0.6851	**0.9414**	**0.8920**	1,486,033

TABLE 4. Experiment results on HC18 dataset using the proposed and other U-Net based models.

Models	HD(mm)	Precision%	Recall%	F1 score%	IoU%
Proposed model	6.08	97.23	97.21	**97.14**	**94.94**
Attention U-Net	**5.82**	96.67	97.27	96.85	94.51
Dilated U-Net	6.90	96.77	96.50	96.51	93.89
MFP U-Net	6.0	96.76	**97.29**	96.90	94.54
R2U-Net	12.33	**97.56**	92.66	94.05	90.83
U-Net	6.15	97.01	97.27	97.04	94.79

TABLE 5. Experiment results on BIDMC dataset using the proposed and other U-Net based models.

Models	HD(mm)	Precision%	Recall%	F1 score%	IoU%
Proposed model	4.24	**97.96**	95.65	**96.67**	98.14
Attention U-Net	19.43	68.68	85.18	72.06	72.09
Dilated U-Net	4.47	94.72	95.11	94.75	92.08
MFP U-Net	**4.08**	96.93	**95.94**	96.34	**98.22**
R2U-Net	-	62.56	42.81	46.94	88.33
U-Net	-	68.77	77.95	70.11	68.50

TABLE 6. Experiment results on HK dataset using the proposed and other U-Net based models.

Models	HD(mm)	Precision%	Recall%	F1 score%	IoU%
Proposed model	5.51	**94.67**	90.37	91.91	89.75
Attention U-Net	5.54	91.52	**93.54**	**92.22**	89.73
Dilated U-Net	5.67	91.01	91.37	91.51	91.22
MFP U-Net	**5.42**	92.27	91.94	91.67	90.45
R2U-Net	-	88.03	76.88	80.39	89.56
U-Net	5.63	92.36	92.23	92.14	**92.91**

TABLE 7. Comparison of the proposed model with the results of the state-of-the-art models reported in literature in terms of the Dice coefficient.

Algorithm	HC18	BIDMC	HK
Proposed method	96.54	**96.76**	**92.58**
Heuvel *et al.* [26]	97.00	—	—
Sobhaninia *et al.* [30]	96.84	—	—
Fiorentino *et al.* [31]	**97.76**	—	—
Ciurte *et al.* [33]	94.45	—	—
Sun *et al.* [34]	96.97	—	—
Rong *et al.* [35]	95.49	—	—
Srivastava *et al.* [44]	92.0	—	—
Zhang *et al.* [36]	—	81.20	88.96
Liu *et al.* [37]	—	87.37	88.34
Ashkani *et al.* [38]	—	90.85	90.75
Wei *et al.* [45]	—	82.91	81.48

common pixel) relative to the ground truth mask. This metric determines the number of matching ground truth annotations per common pixel through the predicted object in a given image. Recall metric describes the completeness of the positive ground truth mask and determines the number of positive predictions of the objects annotated in the ground truth mask. Lastly, the F1 Score metric is defined as a harmonic mean of the precision and recall metrics.

$$IoU = \frac{Target \cap Prediction}{Target \cup Prediction} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1\ Score = \frac{2\ TP}{(2\ TP + FP + FN)} \tag{8}$$

where TP, FP, and FN are true positives, false positives, and false negatives respectively.

V. RESULTS AND THEIR ANALYSES

A. DESCRIPTIVE RESULTS

In this section, we compare the results of the proposed architecture with the state-of-the-art U-Net-based architectures

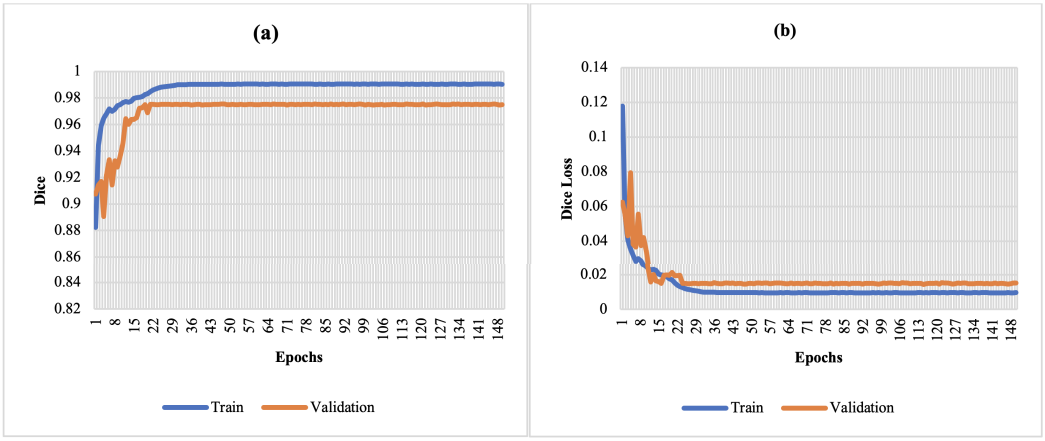

FIGURE 5. (a) Profile of the Dice coefficient variation during the training and validation processes with the HC18 dataset using the proposed model. (b) rofile of the Dice loss variation during the training and validation processes with the HC18 dataset using the proposed model.

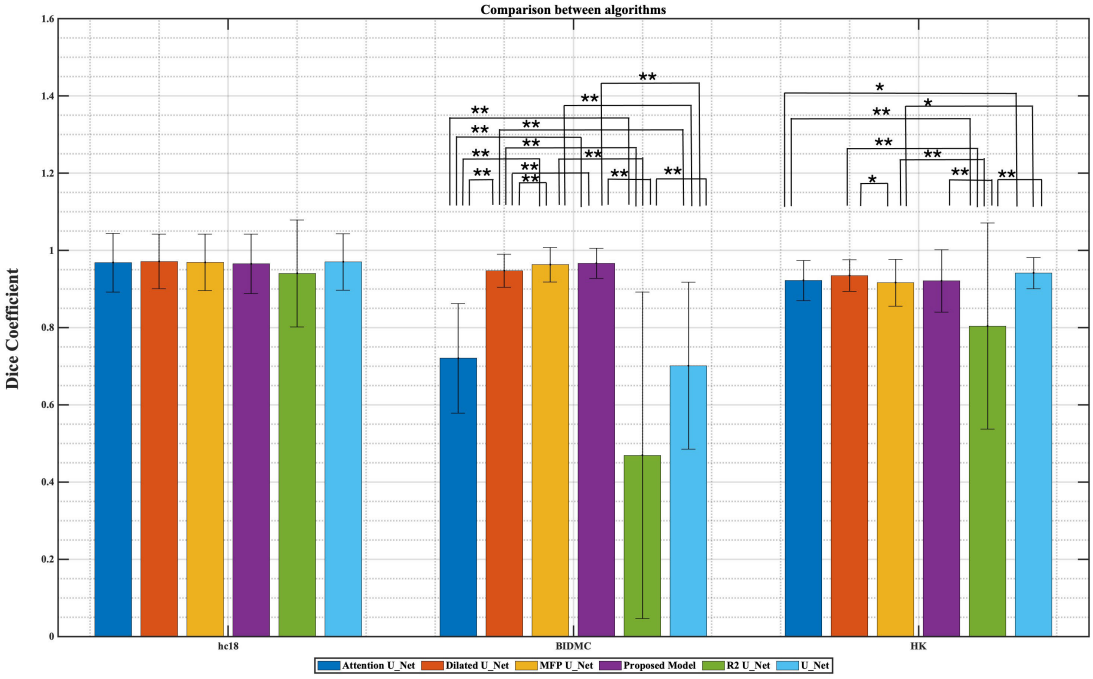

FIGURE 6. Comparison between AttentionU-Net, DilatedU-Net, MFP U-Net, R2U-Net, U-Net, and the proposed model with Dice coefficient using HC18 Grand challenge, BIDMC, and HK datasets. The error bars represent the 95% confidence interval. The "$*$" symbols on the bridge indicate the level of significance measured by p-value where one "$*$" indicates significant difference with $p < 0.05$ and two "$**$" indicates a highly significant difference with $p < 0.001$. The insignificant conditions are not depicted on the bridge.

on the HC18 Grand challenge and multi-site MRI datasets. Specifically, we compared the proposed model with five cutting-edge deep-learning algorithms in medical image segmentation tasks:

1) U-Net [9]
2) Dilated U-Net [24]
3) Attention U-Net [13]
4) R2U-Net [20]
5) MFP-U-Net [23]

The experiments were performed with 150 epochs on a system with 16 GB RAM, a GPU-based graphic card with 2176 CUDA cores (GeForce RTX 2060-A8G), and the Intel Xeon CPU. The Tensorflow backend Keras was employed to implement the deep learning models.

The highest Dice and Jaccard coefficients across all epochs from each of the six algorithms with different datasets are shown in Table 3. From this table, we can see that the proposed approach overall performs better than other

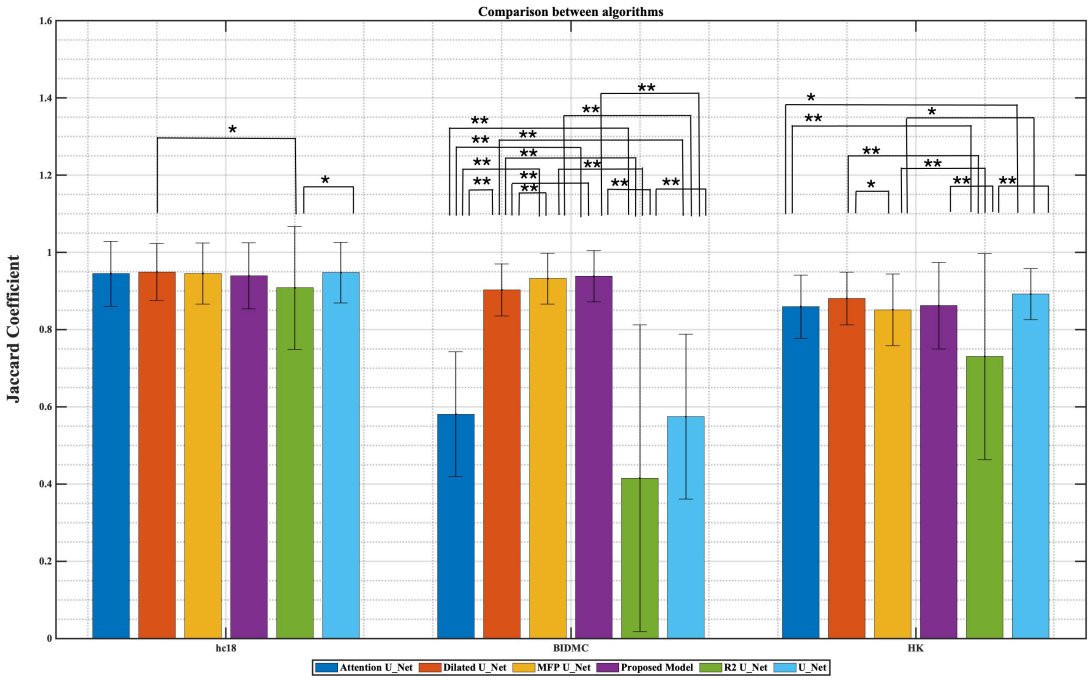

FIGURE 7. Comparison between AttentionU-Net, DilatedU-Net, MFP U-Net, R2U-Net, U-Net, and the proposed model with Jaccard coefficient using HC18 Grand challenge, BIDMC, and HK datasets. The error bars represent the 95% confidence interval. The "∗" symbols on the bridge indicate the level of significance measured by p-value where one "∗" indicates significant difference with $p < 0.05$ and two "∗∗" indicate a highly significant difference with $p < 0.001$. The insignificant conditions were not depicted on the bridge.

state-of-the-art techniques considered in this paper. The proposed model has the highest value of the Dice and Jaccard coefficients for the BIDMC datasets. For the HK dataset, the U-Net and Dilated U-Net achieved slightly better values. The lower Dice and Jaccard coefficients associated with the proposed model for the HK dataset are plausibly due to the lower number of training samples in the HK dataset (484 samples) as well as the low contrast found in the HK dataset images. Nevertheless, across the three datasets, the proposed approach performs better overall.

On the HC18 dataset, all models showed competitive results with the proposed model achieving 96.54% and 93.93% in terms of the Dice and Jaccard coefficients, respectively. Furthermore, to demonstrate the efficacy of the proposed architecture, an evaluation was conducted on the multi-site MRI dataset, encompassing sites E and F, relative to other U-Net-based architectures. It is important to acknowledge that the inherent low contrast of structures within this dataset may contribute to the increased difficulty in accurately identifying the prostate.

Experiments on BIDMC resulted in 96.76% and 93.97% for Dice and Jaccard coefficients, respectively, using the proposed model. Results indicated that the performance remarkably decreased using the Attention U-Net, R2U-Net, and U-Net. Thus, these models have instability in the face of more challenging datasets.

Our approach achieved 92.58% for Dice and 86.96% for Jaccard coefficients using the HK dataset, where the R2U-Net model resulted in the lowest performance of 81.95% and 74.79% in terms of Dice and Jaccard coefficients, respectively. It can also be noted from Table 3, that the Jaccard coefficient reported by the proposed model was better than the average across all other models by 0.94%, 18.33% and 2.32% in the HC18, BIDMC, and HK datasets, respectively. In a similar manner, focusing on the Dice coefficient results, the proposed model was better than the average across all other models by 1.05%, 31.14% and 1.88% in the HC18, BIDMC, and HK datasets, respectively.

It is worth noting that the last column of Table 3 indicates the number of training parameters for each model. While the state-of-the-art U-Net models achieve great accuracy in many scenarios, as already discussed, these architectures struggle with segmenting challenging medical datasets across different modalities with variability in the shape and size of the region of interest in different types of medical images. In comparison, the proposed model is of higher complexity, but results in higher stability when faced with such challenging targets.

To further examine the behaviour of our model, we calculated the HD parameter, IoU, precision, Recall, and F1 Score for all the test sets. Tables 4–6 summarize the results of these metrics. Lower values of HD (reported in millimetres (mm)) indicate that the two contours match closely. The proposed

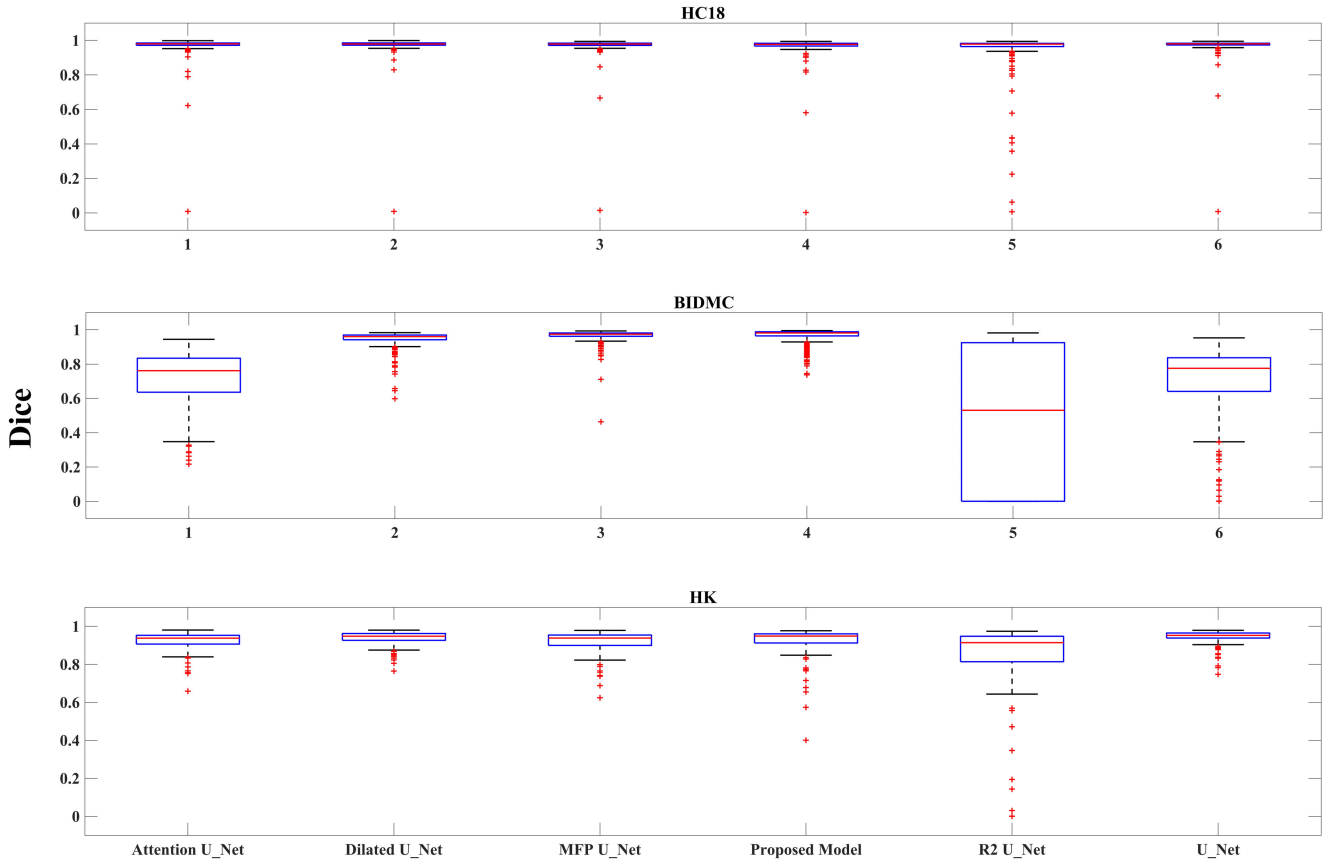

FIGURE 8. The boxplots of Dice metric indicate the absolute error values of the Attention U-Net, Dilated U-Net, MFP U-Net, the proposed model, R2U-Net, and U-Net for each dataset, HC18 Grand challenge, BIDMC, and HK.

model achieved 6.08 mm, 4.24 mm, and 5.51 mm as HD values with the HC18, BIDMC, and HK test sets, which were better than those reported by most of the other U-Net-based models. In these tables, the dash "-" means that models did not produce a predicted mask in the output for some samples, and thus the HD could not be computed for those models. For example, Fig. 4 demonstrates some example images where these models could not predict the output mask; these images only depict the ground truth using a green contour.

Also, in Table 7, we compare the proposed model with the additional state-of-the-art approaches using the results they reported for the same datasets. From the results in Table 7, we concluded that the proposed method achieved better results overall than the other approaches. On the BIDMC and HK datasets, our approach demonstrated better performance in terms of the Dice accuracy than the other considered state-of-the-art methods, with an improvement over the next best performing model of approximately 6% and 2% for the BIDMC and HK datasets, respectively.

Fig. 5 illustrates the learning curves for the proposed model using the HC18 dataset: specifically, Fig. 5(a) and Fig. 5(b) display the variations in the Dice coefficient and Dice loss across epochs for the training and validation sets. It can be observed that the network converges around the 29th epoch with the Dice coefficient reaching 99% on the training set and 96.54% on the validation set. Similar learning curve patterns were observed for the MRI dataset as well.

B. STATISTICAL ANALYSES AND VISUALIZATION

The statistical significance of the differences in the Dice and Jaccard coefficients obtained from each dataset was analyzed as well. First, to investigate the normality of the data, the Kolmogorov-Smirnov test was conducted, and the results showed that the data were not normally distributed. Therefore, a Kruskal-Wallis test with the Bonferroni correction was employed to perform the between-group comparisons for the non-normally distributed data. The IBM Statistical Package Social Sciences (SPSS Version 27) software was utilized to execute this analysis, where the confidence interval value was set to 95%, and an 80% power was assumed.

Figs. 6 and 7 show the evaluation results in terms of the Dice and Jaccard coefficients for the test sets using the algorithms mentioned earlier across the three considered datasets, along with the statistical significance. In these figures, bar graphs illustrate the mean of the Dice and Jaccard coefficients across all epochs. Note that the error bars shown for the results indicate each algorithm's standard deviation (SD) of the experimental data. The Dice and

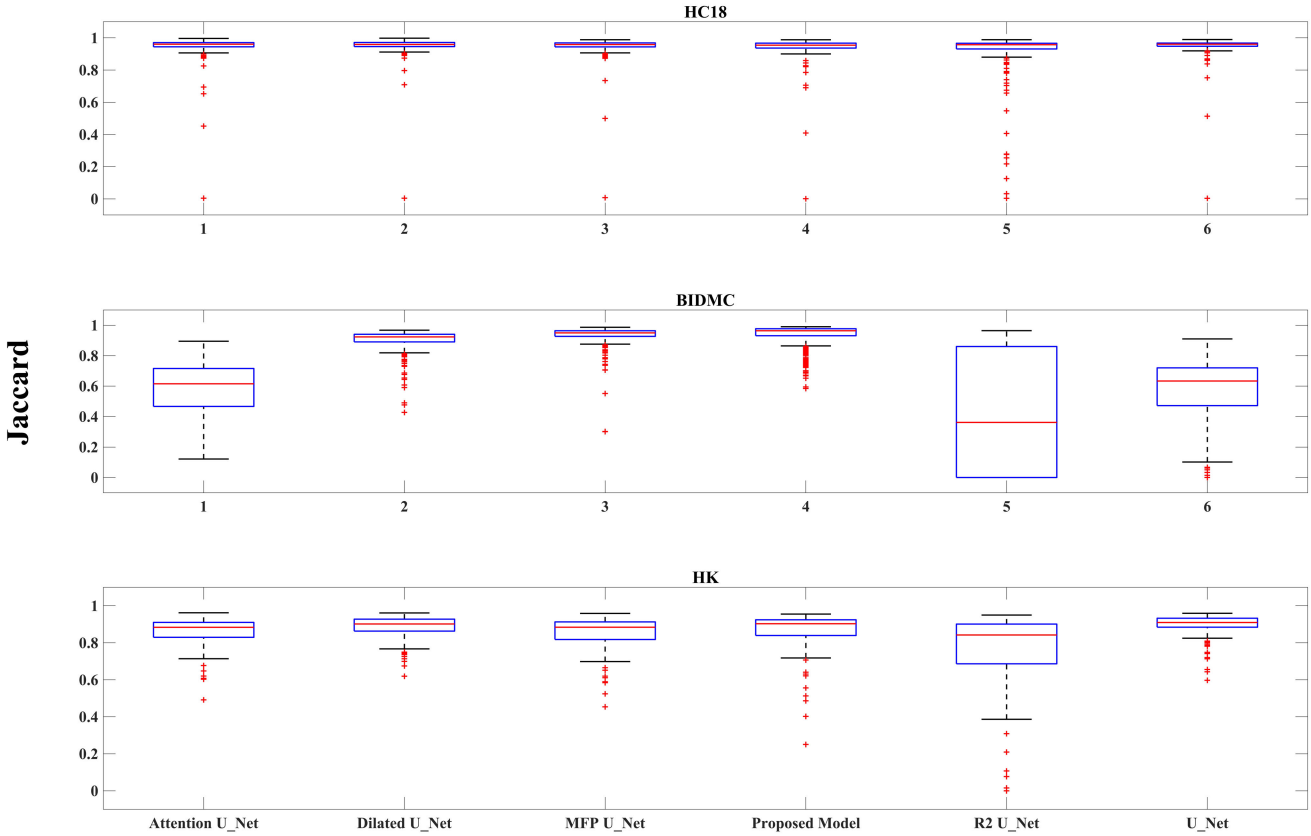

FIGURE 9. The boxplots of the Jaccard metric indicate the absolute error values of the Attention U-Net, Dilated U-Net, MFP U-Net, the proposed model, R2U-Net, and U-Net for each dataset, such as HC18 Grand challenge, BIDMC, and HK.

Jaccard coefficients from the six models were statistically compared, and the analysis results are shown in Figs. 6 and 7. In these figures, "*" indicates a significant difference with $p < 0.05$ and "**" indicates a highly significant difference with $p < 0.001$. It is worth mentioning that the insignificant statistical comparisons are not shown in these figures.

The results of the multi-experiment analysis from Fig. 6 show that the proposed model using the Dice coefficients regardless of the datasets is highly significantly different ($p < 0.001$) except for the HC18 dataset. For the HK dataset, our approach resulted in a significant performance improvement relative to the R2U-Net model.

For the Jaccard coefficients (Fig.7), the multivariate analysis of variance of the proposed model is highly significantly different ($p < 0.001$) from the others with the BIDMC dataset and highly significantly different from the R2U-Net algorithm in the HK dataset. The proposed algorithm is not significant for the HC18 dataset compared to the other models. Results have demonstrated that R2U-Net has the worst performance in all sets. Still, the most significant conclusion from these results is about the consistent performance of our model when faced with dataset variations. Moreover, results demonstrated that the proposed model is less sensitive to the outliers, indicating higher stability and reliability in its predictions.

The results are also shown in the boxplots for further visualization: Figs. 8 and 9 compare the six different architectures using the Dice and Jaccard coefficients, respectively. The upper, middle, and lower subplots of Figs. 8 and 9 correspond to the HC18, BIDMC, and HK datasets. The outliers are red plus signs while the red line shows the median of Dice and Jaccard coefficients. For our model, all plus signs are densely situated near the median line which indicates better performance than other models. Moreover, with a smaller number of plus signs and sparse dispersal of plus signs close to the median line, the proposed model demonstrates better segmentation accuracy than all other models across the HC18, BIDMC, and HK datasets.

Fig. 10 shows a visual comparison between our model and others on a few examples from the HC18 dataset. In the figure, green and red dotted contours illustrate ground truth and predicted mask, respectively. The visual analysis shows that the proposed model is robust with challenging data (small targets). Rows 6 and 7 in Fig. 10 show small HC targets challenging for segmentation; in such conditions, the proposed model resulted in good performance compared to other models. From this figure, in rows six and seven, it can

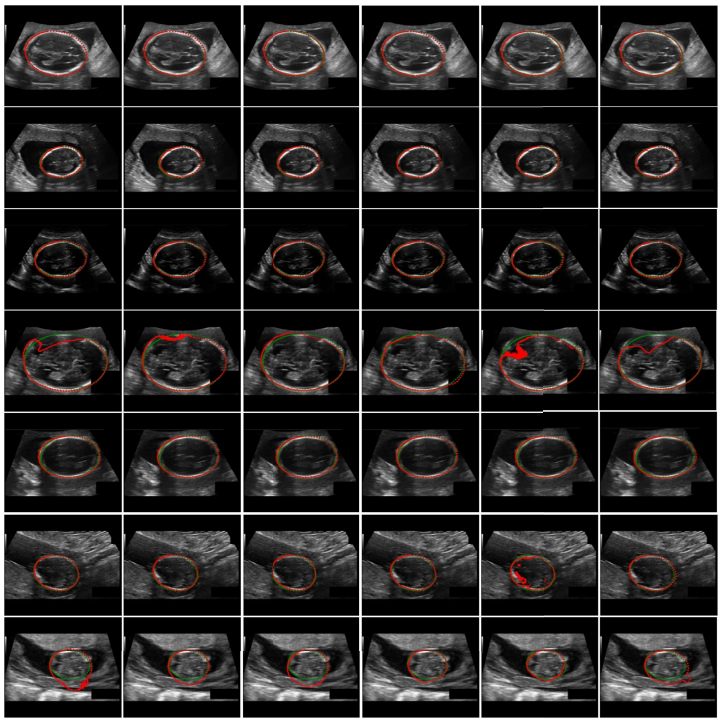

FIGURE 10. Examples of results from the HC18 dataset obtained with the proposed model in comparison with other state-of-the-art U-Net-based models. The green and red lines show the ground truth and predicted outputs. (From right to left: first column: Attention U-Net, second column: Dilated U-Net, third column: MFP U-Net, fourth column: Proposed Model, fifth column: R2U-Net, and sixth column: U-Net).

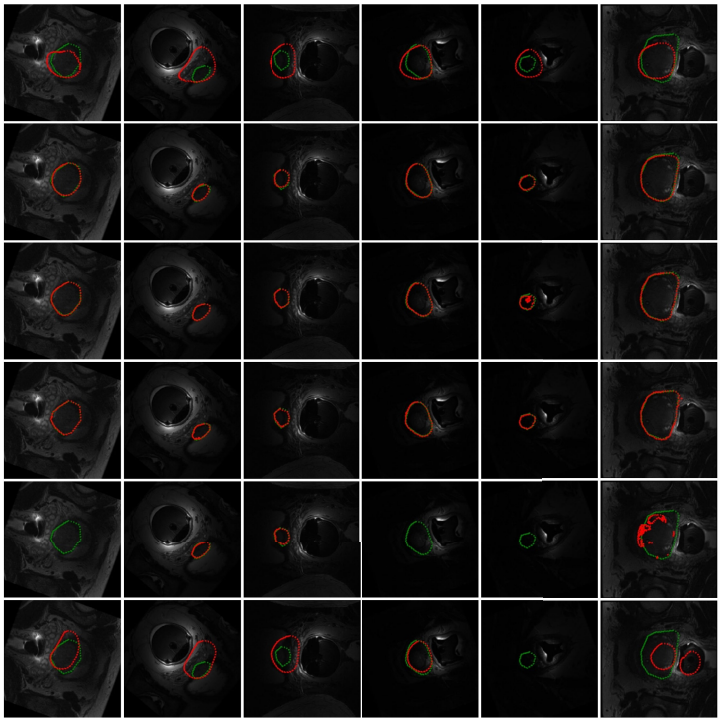

FIGURE 11. Examples of results from the BIDMC dataset obtained with the proposed model in comparison with other state-of-the-arts U-Net-based models. The green and red lines show the ground truth and predicted outputs. (First row: Attention U-Net, second row: Dilated U-Net, third row: MFP U-Net, fourth row: Proposed Model, fifth row: R2U-Net, and sixth row: U-Net).

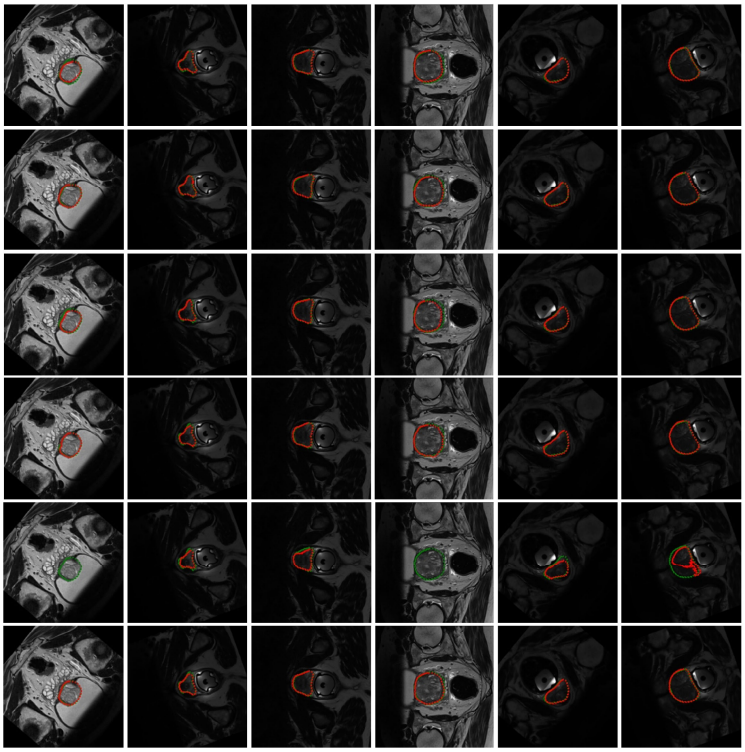

FIGURE 12. Examples of results from the HK dataset obtained with the proposed model in comparison with other state-of-the-art U-Net based models. The green and red lines show the ground truth and predicted outputs. (First row: Attention U-Net, second row: Dilated U-Net, third row: MFP U-Net, fourth row: Proposed Model, fifth row: R2U-Net, and sixth row: U-Net).

be observed that the Attention U-Net, Dilated U-Net, MFP U-Net, R2U-Net, and U-Net may not accurately identify the small clinical targets. At the same time, the proposed model performed much better in such challenging target segmentation. Moreover, the proposed model demonstrated better performance with images where the target region is flat and similar to other image surfaces as seen in the fourth row.

Figs. 11 and 12 depict the segmentation results obtained from different models on BIDMC and HK datasets, respectively. The overall observation was that the proposed model performed better for challenging images with flat and small prostates, such as those in the fourth, fifth and sixth columns in Fig 11. From this figure, we also observed that the R2U-Net and U-Net models have the worst performance in challenging (small and low contrast) target segmentation. From the visual analysis of results on the HK samples in Fig. 12, we concluded that this dataset, due to better resolution, is less challenging than the BIDMC dataset for prostate segmentation purposes. Thus, most of the models performed well in the target identification through this dataset, except the R2U-Net model, which has a problem segmenting the targets accurately. For example, in the fifth row, we can observe green contours as the ground truth masks in the first and fourth columns without any predicted contours (red line) in the fourth column corresponding to the R2U-Net model.

Given that the accuracy on test samples does not reach 100%, there are instances where the proposed method fails to predict outcomes precisely. Notably, the incidence of such inaccuracies is higher within the HK and BIDMC datasets compared to the HC18 dataset. Fig. 13 illustrates cases where the proposed model encountered difficulties in accurately identifying the target. This is particularly evident in some samples where critical details at the edges of the region of interest are overlooked, compounded by a significant resemblance between the target region and surrounding areas in the image.

In conclusion, the proposed model demonstrates enhanced segmentation precision consistently across diverse datasets. It is important to address the complexity of our model in comparison with others. The distinctive features of our model, namely, the encoder-decoder skip connections, the implementation of multi-level feature transitions, and the incorporation of a multi-stream dilated network, are key contributors to the observed improvements in performance.

In the level-to-level transition procedure, different levels of information are extracted through layers. Features extracted from the first layer are transmitted to the next lower layer to extract different features, and so on. Thus, from the information entered into the first blocks in the first layer, multi-level information is extracted through the layer

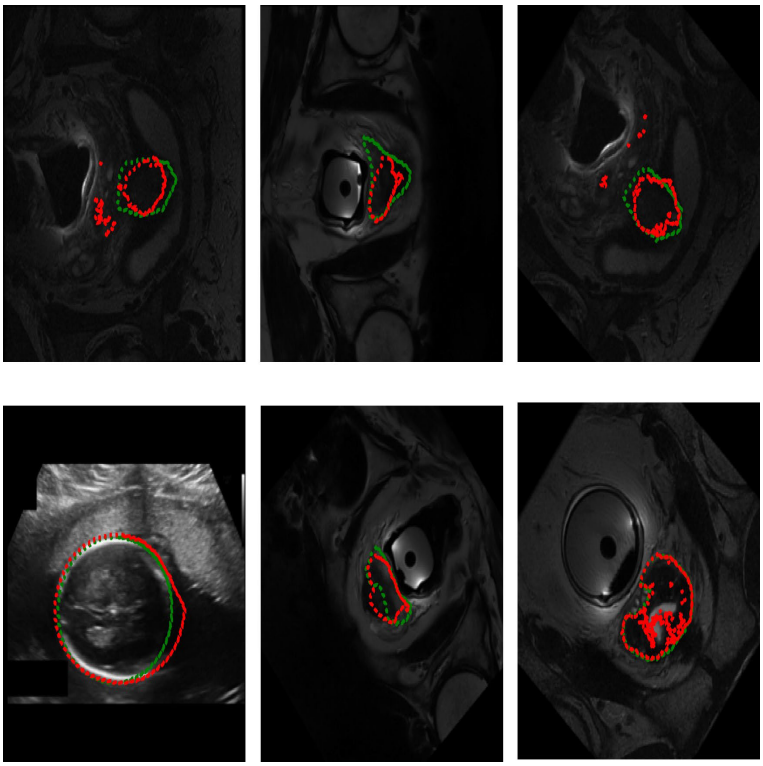

FIGURE 13. Some failure cases of the accurate prediction using the proposed model.

hierarchy. However, this led to a significant increase in the feature dimensionality and hence the complexity of our model. Nevertheless, as demonstrated, this increased complexity was necessary to achieve robust and consistent performance across varied datasets and segmentation target sizes.

VI. CONCLUSION

This study proposed a novel U-Net-based model named the One-to-Many U-Net model for segmenting different clinical targets in various medical images such as fetal head segmentation for head circumference measurements purpose in ultrasound imaging systems and prostate segmentation in MRI images. The proposed model comprises the multi-level feature extraction strategy that extends from the previous blocks in both encoder and decoder modules and provides multiple connections between encoder and decoder blocks that cause the improvement of segmentation tasks for different target sizes and increased performance consistency through the various imaging systems. This model contains four rows of encoder-decoder modules with three feature extraction blocks in each row. After concatenating the outputs of the decoder modules, a multi-stream dilation network is applied to the output of the One-to-Many network to accurately identify the small segmentation targets.

To evaluate the proposed approach, two different types of medical images were used: fetal head segmentation from ultrasound images and prostate segmentation from the MRI imaging system. The quantitative and statistical analysis showed the superior performance of the proposed model compared to most of the state-of-the-art U-Net-based models as well as several other approaches from the literature. Moreover, the proposed architecture demonstrated significant improvement in accuracy and stability in performance for prostate segmentation compared to other U-Net-based models. It achieved 96.54%, 96.76%, and 92.58% in Dice coefficients for the HC18, BIDMC, and HK datasets. In addition, the proposed model resulted in Jaccard coefficients of 93.93%, 93.97%, and 86.96% for the HC18, BIDMC, and HK datasets, respectively. In the visual analysis, we observed that the proposed model has good performance when facing challenging data (small targets), in contrast to other models which did not perform well with such data.

ACKNOWLEDGMENT

The authors would like to express their appreciation to the managers at Med Fanavaran Plus (MFP), Iran, for their cooperation to use their processors in the earlier phases of this research.

REFERENCES

[1] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps: Automation of Decision Making*, 2018, pp. 323–350.

[2] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, no. 1, pp. 315–337, Aug. 2000, doi: 10.1146/annurev.bioeng.2.1.315.

[3] Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 595–604, Jul. 2001, doi: 10.1109/42.932744.

[4] P. Mesejo, A. Valsecchi, L. Marrakchi-Kacem, S. Cagnoni, and S. Damas, "Biomedical image segmentation using geometric deformable models and metaheuristics," *Computerized Med. Imag. Graph.*, vol. 43, pp. 167–178, Jan. 2014, doi: 10.1016/j.compmedimag.2013.12.005.

[5] Y. Zheng, B. Jeon, D. Xu, Q. M. J. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *J. Intell. Fuzzy Syst.*, vol. 28, no. 2, pp. 961–973, 2015, doi: 10.3233/ifs-141378.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[8] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: A review," *J. Med. Syst.*, vol. 42, no. 11, pp. 1–13, Nov. 2018, doi: 10.1007/s10916-018-1088-1.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, Jan. 2015, pp. 234–241.

[10] M. A. Shahedi, "Study on U-Net limitations in object localization and image segmentation," in *Proc. Soc. Imag. Informat. Med. (SIIM), Virtual Meeting*, 2020.

[11] S. D. Khan, L. Alarabi, and S. Basalamah, "An encoder–decoder deep learning framework for building footprints extraction from aerial imagery," *Arabian J. Sci. Eng.*, vol. 48, no. 2, pp. 1273–1284, Feb. 2023, doi: 10.1007/s13369-022-06768-8.

[12] S. D. Khan, L. Alarabi, and S. Basalamah, "Deep hybrid network for land cover semantic segmentation in high-spatial resolution satellite images," *Information*, vol. 12, no. 6, p. 230, May 2021, doi: 10.3390/info12060230.

[13] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," pp. 1–10, 2018, *arXiv:1804.03999*.

[14] M. Mubashar, H. Ali, C. Grönlund, and S. Azmat, "R2U++: A multiscale recurrent residual U-Net with dense skip connections for medical image segmentation," *Neural Comput. Appl.*, vol. 34, no. 20, pp. 17723–17739, Oct. 2022, doi: 10.1007/s00521-022-07419-7.

[15] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *Proc. IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Dec. 2019, doi: 10.1109/TMI.2019.2959609.

[16] O. Ali, H. Ali, S. A. A. Shah, and A. Shahzad, "Implementation of a modified U-Net for medical image segmentation on edge devices," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 11, pp. 4593–4597, Nov. 2022, doi: 10.1109/TCSII.2022.3181132.

[17] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[18] B. Ait Skourt, A. El Hassani, and A. Majda, "Lung CT image segmentation using deep neural networks," *Proc. Comput. Sci.*, vol. 127, pp. 109–113, Jan. 2018.

[19] Y. Guo and B. Matuszewski, "GIANA polyp segmentation with fully convolutional dilation neural networks," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, Prague, Czech Republic, 2019, pp. 632–641.

[20] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, Mar. 2019, Art. no. 014006, doi: 10.1117/1.jmi.6.1.014006.

[21] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognit. Image Anal.*, vol. 27, no. 3, pp. 618–624, Jul. 2017, doi: 10.1134/s1054661817030269.

[22] T. Mahmud, B. Paul, and S. A. Fattah, "PolypSegNet: A modified encoder–decoder architecture for automated polyp segmentation from colonoscopy images," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104119, doi: 10.1016/j.compbiomed.2020.104119.

[23] S. Moradi, M. G. Oghli, A. Alizadehasl, I. Shiri, N. Oveisi, M. Oveisi, M. Maleki, and J. Dhooge, "MFP-unet: A novel deep learning based approach for left ventricle segmentation in echocardiography," *Phys. Medica*, vol. 67, pp. 58–69, Nov. 2019, doi: 10.1016/j.ejmp.2019.10.001.

[24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," pp. 1–13, 2015, *arXiv:151107122*.

[25] V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Trans. Med. Imag.*, vol. 16, no. 5, pp. 642–652, Oct. 1997, doi: 10.1109/42.640755.

[26] T. L. A. van den Heuvel, D. de Bruijn, C. L. de Korte, and B. V. Ginneken, "Automated measurement of fetal head circumference using 2D ultrasound images," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0200412, doi: 10.1371/journal.pone.0200412.

[27] W. Lu, J. Tan, and R. Floyd, "Fetal head detection and measurement in ultrasound images by a direct inverse randomized Hough transform," *Proc. SPIE*, vol. 5747, pp. 715–722, Apr. 2005, doi: 10.1117/12.594813.

[28] J. Li, Y. Wang, B. Lei, J.-Z. Cheng, J. Qin, T. Wang, S. Li, and D. Ni, "Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 215–223, Jan. 2018.

[29] D. T. Avalokita, T. Rismonita, A. Handayani, and A. W. Setiawan, "Automatic fetal head circumference measurement in 2D ultrasound images based on optimized fast ellipse fitting," in *Proc. IEEE REGION 10 Conf.*, Nov. 2020, pp. 37–42, doi: 10.1109/TENCON50793.2020.9293786.

[30] Z. Sobhaninia, S. Rafiei, A. Emami, N. Karimi, K. Najarian, S. Samavi, and S. M. R. Soroushmehr, "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 6545–6548, doi: 10.1109/EMBC.2019.8856981.

[31] M. C. Fiorentino, S. Moccia, M. Capparuccini, S. Giamberini, and E. Frontoni, "A regression framework to head-circumference delineation from US fetal images," *Comput. Methods Programs Biomed.*, vol. 198, Jan. 2021, Art. no. 105771, doi: 10.1016/j.cmpb.2020.105771.

[32] S. M. Amini, "Head circumference measurement with deep learning approach based on multi-scale ultrasound images," *Multimedia Tools Appl.*, vol. 81, no. 23, pp. 32981–32993, Sep. 2022, doi: 10.1007/s11042-022-13107-4.

[33] A. Ciurte, X. Bresson, and M. B. Cuadra, "A semi-supervised patch-based approach for segmentation of fetal ultrasound imaging," in *Proc. Challenge US, Biometric Meas. Fetal Ultrasound Images*, 2012, pp. 5–7.

[34] C. Sun, "Automatic fetal head measurements from ultrasound images using circular shortest paths," in *Proc. Challenge US, Biometric Meas. Fetal Ultrasound Images*, 2012, pp. 13–15.

[35] Y. Rong, D. Xiang, W. Zhu, F. Shi, E. Gao, Z. Fan, and X. Chen, "Deriving external forces via convolutional neural networks for biomedical image segmentation," *Biomed. Opt. Exp.*, vol. 10, no. 8, pp. 3800–3814, 2019, doi: 10.1364/boe.10.003800.

[36] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, and Z. Xu, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2531–2540, Jul. 2020, doi: 10.1109/TMI.2020.2973595.

[37] Q. Liu, Q. Dou, and P. Heng, "Shape-aware meta-learning for generalizing prostate MRI segmentation to unseen domains," in *Proc. 23rd Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, Lima, Peru, Jan. 2020, pp. 475–485.

[38] V. A. Chenarlogh, A. Shabanzadeh, M. G. Oghli, N. Sirjani, S. F. Moghadam, A. Akhavan, H. Arabi, I. Shiri, Z. Shabanzadeh, M. S. Taheri, and M. K. Tarzamni, "Clinical target segmentation using a novel deep neural network: Double attention Res-U-Net," *Sci. Rep.*, vol. 12, no. 1, pp. 1–17, Apr. 2022, doi: 10.1038/s41598-022-10429-z.

[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, Jan. 2015, pp. 448–456.

[40] V. A. Chenarlogh and F. Razzazi, "Multi-stream 3D CNN structure for human action recognition trained by limited data," *IET Comput. Vis.*, vol. 13, no. 3, pp. 338–344, Apr. 2019.

[41] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "MS-net: Multi-site network for improving prostate segmentation with heterogeneous MRI data," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2713–2724, Sep. 2020, doi: 10.1109/TMI.2020.2974574.

[42] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006, doi: 10.1109/TMI.2006.880587.

[43] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Stanford, CA, USA, Oct. 2016, pp. 565–571.
[44] S. Srivastava, A. Vidyarthi, and S. Jain, "Analytical study of the encoder–decoder models for ultrasound image segmentation," *Service Oriented Comput. Appl.*, vol. 18, no. 1, pp. 81–100, Mar. 2024.
[45] Z. Wei, W. Dong, P. Zhou, Y. Gu, Z. Zhao, and Y. Xu, "Prompting segment anything model with domain-adaptive prototype for generalizable medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Jan. 2024, pp. 533–543.

VAHID ASHKANI CHENARLOGH received the B.Sc. degree in electrical and electronic engineering from Qazvin Islamic Azad University, Iran, in 2015, the M.Sc. degree in electrical and telecommunication engineering from Islamic Azad University Science and Research Branch, Iran, in 2018. He is currently pursuing the Ph.D. degree with Western University, Canada. He is involved in research works with the National Centre for Audiology (NCA) and the dspfactory Digital Signal Processing Research Laboratory. His current research interests include digital signal processing, speech intelligibility and enhancement, machine learning, and deep learning.

ARMAN HASSANPOUR received the B.Sc. degree in computer engineering-hardware and the M.Sc. degree in mechatronics engineering from Qazvin Islamic Azad University, in 2008 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Health and Rehabilitation Sciences Graduate Program (Hearing Science field), Faculty of Health Sciences, Western University. His research interests include digital signal processing, speech intelligibility, electroacoustic assessment of hearing aids, and machine learning in hearing applications. He is an Active Member of the National Centre for Audiology (NCA) and the dspfactory Digital Signal Processing Research Laboratory.

KATARINA GROLINGER (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in mechanical engineering from the University of Zagreb, Croatia, and the M.Eng. and Ph.D. degrees in software engineering from Western University. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, Western University. She is also Canada Research Chair (Tier 2) in engineering applications of machine learning and the Vector Institute Faculty Affiliate. Before joining Western University as an Assistant Professor, she was a Postdoctoral Fellow with Western University, where she was working on developing data analytics solutions for energy domain, clickstream data analytics, and big data processing. Her work is largely carried out in collaboration with industry where she develops research to solve industry problems. Also, she has more than ten years of industry experience in various roles, including a Software Engineer, an Oracle Certified Database Administrator, and a Technical Team Leader.

VIJAY PARSA received the Ph.D. degree in biomedical engineering from the University of New Brunswick, Canada, in 1996. He then joined the Hearing Health Care Research Unit, University of Western Ontario, where he worked on developing speech processing algorithms for audiology and speech language pathology applications. From 2002 to 2007, he was the Oticon Foundation Chair in acoustic signal processing. He is currently an Associate Professor jointly appointed across the Faculties of Health Sciences and Engineering. His research interests include speech signal processing with applications to hearing aids, assistive listening devices, and augmentative communication devices.