An Efficient Capsule-based Network for 2D Left Ventricle Segmentation in Echocardiography Images

R. Naghne, A. Kazemi, H. Moghaddasi, M. Rahmani, P. Farnia, A. Ahmadian, J. Alirezaie*

***Abstract*— The segmentation of cardiac chambers is essential for the clinical diagnosis and treatment of cardiovascular diseases. It is demonstrated that in cardiac disease, the left ventricle (LV) is extensively involved. Therefore, segmentation of the LV in echocardiographic images is critical for the precise evaluation of factors that influence cardiac function such as LV volume, ejection fraction, and LV mass. Although these measurements could be obtained by manual segmentation of the LV, it would be time-consuming and inaccurate because of the poor quality and low contrast of these images. Convolutional neural networks, commonly referred to as CNNs, have emerged as a highly favored deep learning technique for medical image segmentation. Despite their popularity, the pooling layers in CNNs ignore the spatial information and do not consider the part-whole hierarchy relationships. Furthermore, they require a large training dataset and a large number of parameters. Therefore, Capsule Networks are proposed to address the CNNs limitations. In this study, for the first time, an optimized capsule-based network for object segmentation called SegCaps is proposed to achieve accurate LV segmentation on echocardiography images applied to the CAMUS dataset. The result was compared against the standard 2D-UNet. The modified SegCaps and 2D-UNet achieved an average Dice similarity coefficient (DSC) of 84.48% and 83.28% on LV segmentation, respectively. The capabilities of the CapsNet led to an improvement of 1.44% in DSC with 92.77% fewer parameters than the U-Net. The results indicate that the proposed method leads to accurate and efficient LV segmentation.**

***Clinical Relevance*— From a clinical point of view, our findings lead to more precise evaluations of critical cardiac parameters, including ejection fraction as well as left ventricle volume at end-diastole and end-systole.**

I. INTRODUCTION

Cardiovascular diseases are among the top causes of death worldwide, which jumped from third place in 2009 to first place in 2019 [1]. The first step towards an adequate clinical diagnosis of cardiovascular diseases is using various current imaging techniques, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound. These imaging modalities visualize various cardiac structures in detail non-invasively for diagnosing cardiovascular diseases [2]. Among the different imaging modalities, cardiac ultrasound (echocardiography) is a commonly used method for cardiovascular disease diagnosis because of its real-time visualization and non-ionizing radiation. Different cardiovascular diseases are evident in the left ventricle (LV), which can be detected in echocardiography images. According to studies, most cardiovascular diseases could be identified by an inspection of the LV in terms of its dimensions, shape, wall thickness, and cavity size [3]. Accurate segmentation of the LV is critical for the precise evaluation of cardiac function factors such as LV volume, ejection fraction, and LV mass. Although the manual segmentation of LV by physicians is considered clinical ground truth, it would be time-consuming and tedious. Therefore, automated methods are needed to achieve more efficient segmentation results [4]. In this regard, LV segmentation strategies based on conventional approaches, including shape-based [5], appearance-based [6], and active contour approaches [7] have been proposed. Recently, Gomez et al. [8] proposed a method to segment the LV from echocardiography images using a full multiresolution active shape model (ASM). Recently, in 2021, a novel echocardiographic image segmentation approach, which combines the ASM, Nakagami distribution, and mean squared eigenvalue error was proposed [9]. Despite the wide use and popularity of the explained methods, accurate detection of the endocardial border in LV segmentation remains a challenging task.

Compared to conventional methods, deep learning approaches have been widely applied in medical image segmentation and have allowed us to achieve impressive accuracy [10-12]. Among various deep learning techniques, U-Net has been the standard convolutional neural network (CNN) framework for medical image segmentation [13]. In this regard, in 2017, Smistad et al. decrease the need for manual segmentation by pre-training a CNN using via a Kalman filter (KF)-based segmentation method. Their findings showed that a CNN trained only on generated data could even achieve the same accuracy as the conventional methods [14]. In 2021, Zhao et al. used bidirectional feature fusion (BFF-Net) and wavelet methods to create integrated wavelet networks in order to segment the left atrium and LV of the echocardiogram. They reduced the effect of noise, optimized the loss function, and improved the accuracy of segmentation [15]. In 2022, multiple designs are developed by Belfilali et al. [16], which build a set of pre-trained CNNs in order to use them in five segmentation topologies, including LinkNet, TransUNet, Attention UNet, UNet1, and UNet2. The results of the experiments showed that the suggested framework using the UNet1VGG19 outperformed other networks with a Dice similarity coefficient (DSC) of 93.30% and a Hausdorff distance of 4.01 mm. Recently, in 2023, Sfakianakis et al. [17] used a UNet-based network to extract the myocardium, left atrium, and LV from

This work was supported by the Faculty of Medicine, Tehran University of Medical Sciences under grant number 58684.

R. Naghne, A. Kazemi, H. Moghaddasi, M. Rahmani, P. Farnia, and A. Ahmadian* are with the Image-Guided Surgery Group, Research Center of Biomedical Technology and Robotics (RCBTR), Advanced Medical Technologies and Equipment Institute (AMTEI), Tehran University of Medical Sciences, Tehran, Iran and with the Medical Physics and Biomedical Engineering Department, Faculty of Medicine, Tehran University of Medical Sciences (TUMS).

J. Alirezaie is with the Department of Electrical and Computer Engineering, Toronto Metropolitan University, Toronto, Ontario, Canada.

*Corresponding author: ahmadian@sina.tums.ac.ir

a cardiac ultrasound image dataset. They've introduced a substantial data augmentation strategy inspired by medical images, which is used to significantly improve the training procedure. They achieved DSC of 94.6% at the end of the diastole (ED) phase and 92.9% at the end of the systole (ES) phase.

Despite the popularity of CNNs in medical image segmentation, inaccuracies in the output are noticeable. Pooling layers in CNNs ignore the spatial information and do not consider the part-whole hierarchical relationships, which also require a large training dataset. Therefore, in 2017, capsule networks (CapsNet) were introduced by Sabour et al. as an alternative to CNNs [18]. Their ability to extract inherent spatial relationships, understand various image poses such as position, size, and orientation, and train on small datasets takes them into consideration [19, 20]. CapsNets use a "routing by agreement" algorithm between different layers, which replaces pooling in CNNs. CapsNets have the potential to achieve high performance [21] for segmentation tasks in medical image analysis, LaLonde et al. used capsule networks for the object segmentation task [22]. They expanded the idea of capsule networks and introduced the concept of deconvolutional capsules, which led to the development of a novel convolutional-deconvolutional capsule architecture called SegCaps. This architecture produces impressive outcomes for object segmentation. To evaluate its performance, they applied the proposed SegCaps for the segmentation of pathological lung CT images and compared its accuracy with other UNet-based architectures. As opposed to main capsules, images of large size can be fed into SegCaps. SegCaps architecture can provide a noticeable decrease in the number of UNet architecture parameters by 95.4% while obtaining more accurate segmentation results [23]. They also have an improved representation of spatial information and greater segmentation accuracy while using fewer parameters than the current state-of-the-art models.

Therefore, in this study, we proposed a novel work that promotes the optimization of capsule network architecture for LV segmentation in echocardiography images.

II. Method and Material

A. Data Description

In this study we have used Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) dataset published by Leclerc et al. [24]. The CAMUS dataset was acquired from GE Vivid E95 ultrasound scanners, with a GE M5S probe. In order to estimate the LV volumes with high resolution, two sequences were selected for each patient, namely 2D four-chamber and two-chamber views. The resolution along the x-axis (parallel to the probe) was set at λ/2 = 0.3 mm, while the resolution along the z-axis (perpendicular to the probe) was set at λ/4 = 0.15 mm, where λ represents the wavelength of the ultrasound probe with ground truth of cardiac structures at ED and ES. This dataset includes different quality B-mode images of 500 patients (10% were categorized as good or average, while 90% were of low quality) and with varying degrees of ejection fraction (EF) (40% of patients with an EF lower than 45% and 60% of patients with an EF higher than 55%).

B. Capsule Network

The CapsNets' architecture has two significant advantages over CNNs; Namely, the capsule concept and route by agreement algorithm. In a capsule network architecture, each layer comprises a set of capsules (groups of neurons) that extract features from different areas of feature space. Each capsule layer represents the parameters of an object, like its orientation and pose, by using dynamic routing and route by agreement algorithms. This property provides part-to-whole relationships that cannot be achieved using CNNs. The output of each capsule is a vector that describes the retrieved pose information, such as orientation. The vector's length indicates the entity's probability of existence and its orientation indicates the pose information [22].

C. SegCaps Architecture

SegCaps expands the use of CapsNet to the task of object segmentation [23]. In this study, SegCaps is optimized to accurately segment the LV of the heart from echocardiography images, which is implemented in Fig. 1.

At first, the input is fed to SegCaps network as a 224×224 pixel image. Then it passed through a 2D convolutional layer that produces 16 feature maps with the same spatial dimensions. The obtained feature map forms the first set of capsules, with a 16-dimensional vector and 224×224 grid size, and attempts to reconstruct the same image from scratch. In the following, the process is generalized for any given layer l in the network.

In convolutional capsules, every capsule receives a set of prediction vectors, $\hat{u}_{(x,y),k,l}$ which are produced in coordinate (x, y) and layer l. The prediction vectors are calculated by multiplying a locally learned transformation matrix for each capsule type by a user-defined window of outputs from the preceding layer, $U_{(x,y),k,l}$. To determine the final input to each capsule, routing coefficients are determined by the dynamic routing algorithm. The output of the iterative routing procedure $r_{(x,y),k,l}$ is the coefficients used by prediction vectors to create the input for each capsule in the following layer (1).

$$P_{(x,y),k,l+1} = \sum r_{(x,y),k,l} \cdot \hat{u}_{(x,y),k,l} \quad (1)$$

In the subsequent phase, a squash function is applied to each capsule layer in order to obtain the capsule output vector (2).

$$V_{(x,y)} = \frac{\left\|P_{(x,y)}\right\|^2}{1+\left\|P_{(x,y)}\right\|^2} \cdot \frac{P_{(x,y)}}{\left\|P_{(x,y)}\right\|} \quad (2)$$

$P_{(x,y)}$ is the final output, and $V_{(x,y)}$ is the vector output, which maintain its directional independence and spatial location (x,y). The last capsule layer calculates the dot product of the input vector with the prediction vector as follows (3):

$$a_{(x,y),k,l} = V_{(x,y)} \cdot \hat{u}_{(x,y),k,l} \quad (3)$$

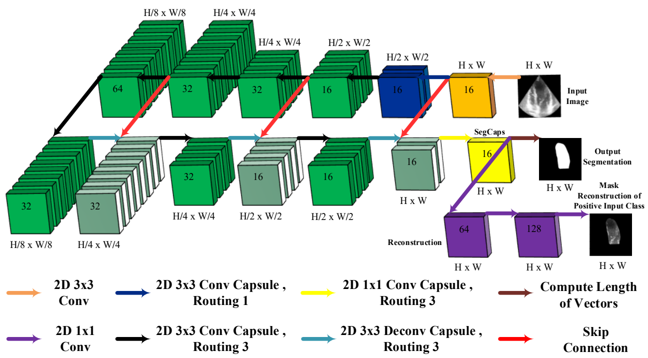

Figure 1. Illustration of the proposed SegCaps network architecture

D. 2D-UNet

The proposed network is compared against a network that is conceptually identical to the standard 2D-UNet (Fig. 2). In the first path of UNet, called the encoder, the number of features is gradually increased by going through multiple layers of convolution and downsampling. When it comes to the second path, the decoder, multiple steps of upsampling are used to get back to the original image resolution.

III. Experiments and Results

A. Experimental Configuration

All experiments are executed using the TensorFlow library in Python and an NVIDIA GeForce GTX 1050 Ti graphics processing unit (GPU). In this study, we employed the LV ground truth from the CAMUS dataset, which consists of 500 cases in total, and its detailed information is provided in Table I.

B. Analysis of Experiments

To quantify the experiment's result, DSC is defined as a statistical validation metric, using Equation 4. The final DSC results and the number of network parameters for the proposed method and 2D-UNet for 140 epochs of training are shown in Table II.

$$\mathrm{DSC} = \frac{2|A_G \cap A_S|}{|A_G| + |A_S|} \quad (4)$$

$$\textit{Total Disagreement} = 0 \leq DSC \leq 1 = \textit{Total Agreement}$$

The DSC is a spatial overlap index and measures the overlap between the segmentation results A_S and the ground truth A_G.

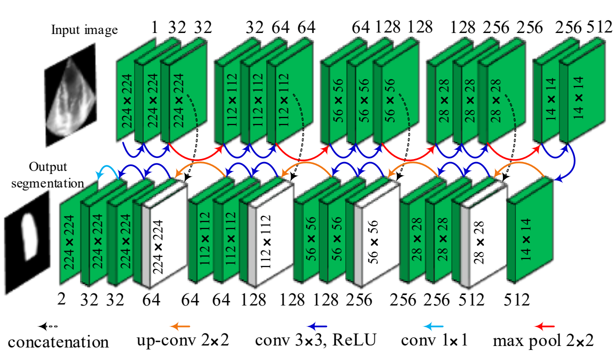

Figure 2. A diagrammatic representation of the 2D-UNet.

TABLE I. DETAILED DATASET SPLITTING INFORMATION

500 Case							
450 Case				50 Case			
Quality		*Ejection Fraction*		*Quality*		*Ejection Fraction*	
Low	*Good or Average*	*Lower than 45%*	*Higher than 55%*	*Low*	*Good or Average*	*Lower than 45%*	*Higher than 55%*
84 case	366 case	222 case	87 case	10 case	40 case	24 case	10 case
1800 Image				200 Image			
Train		*Validation*		*Test*			
70%		30%		100%			

TABLE II. NUMBER OF PARAMETERS AND DSC RESULTS OF 2D-UNET AND PROPOSED METHOD ON CAMUS DATASET.

Approach	DSC score (%)	Parameters
SegCaps	***84.48***	***0.6M***
2D-UNet	*83.28*	*7.7M*

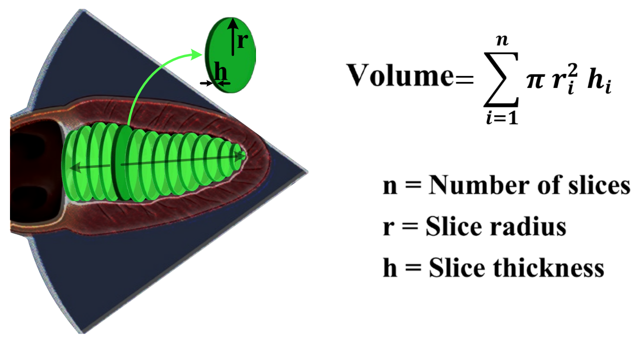

Figure 3. Simpson's biplane method

In addition, the accurate segmentation results lead to determining the LV volume of ED (LV_{ED}) and ES (LV_{ES}) frames using Simpson's biplane method of discs (Fig. 3) [25].

Consequently, the LV Ejection Fraction (LV_{EF}) was calculated based on the difference between the LV_{ED} and LV_{ES} according to Equation 5.

$$LV_{EF} = \frac{LV_{ED} - LV_{ES}}{LV_{ED}} \times 100 \quad (5)$$

To evaluate the segmentation performance of SegCaps and 2D-UNet on the CAMUS dataset qualitatively, the segmentation results for case 2 of the dataset at the ES and ED for two-chamber and four-chamber views are shown in Fig. 4.

IV. Conclusion

Accurate segmentation of LV enables physicians to evaluate cardiac function and diseases. Motivated by the prominence of capsule network architecture, this study aimed to apply the SegCaps network for LV segmentation in echocardiography images. Our findings proved that SegCaps outperformed other methods in improving the accuracy and efficiency of automatic LV segmentation in 2D echocardiographic images. In this study, the CAMUS dataset was used to evaluate the modified SegCaps. This network is able to achieve a mean DSC of 84.48% with 92.77% fewer parameters than 2D-UNet, as shown in Table II.

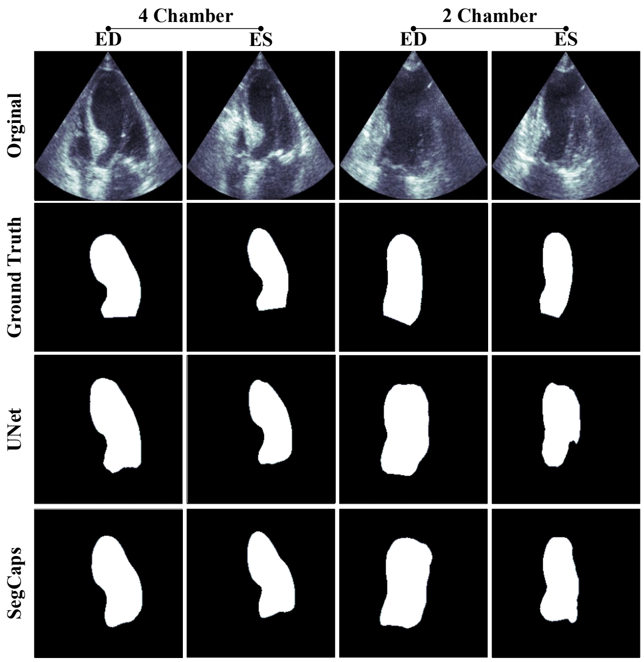

Figure 4. Segmentation results of case 2 for four-chamber and two-chamber views in ED and ES by using the third row: 2D-UNet, and fourth row: SegCaps methods. Original and ground truth images are shown in the first and second rows, respectively.

The findings of our modified SegCaps have revealed that, in addition to increasing accuracy, the computational cost has decreased. The modified SegCaps have segmented the LV endocardium for diverse views of images, including diastole and systole efficiently, as illustrated in Fig. 4. In this study, we have successfully managed to overcome the limitations of CNN by taking advantage of capsule networks. This work provides a promising approach for further implementation of advanced segmentation methods in echocardiography imaging systems.

REFERENCES

[1] G. Compare. Institute for Health Metrics and Evaluation. https://vizhub.healthdata.org/gbd-compare/ (accessed 2023-01-16.

[2] V. Panis and E. Donal, "Imaging techniques for cardiac function," *Applied Sciences,* vol. 11, no. 22, p. 10549, 2021.

[3] C. Chen *et al.*, "Deep learning for cardiac image segmentation: a review," *Frontiers in Cardiovascular Medicine,* vol. 7, p. 25, 2020.

[4] M. N. Berman, C. Tupper, and A. Bhardwaj, "Physiology, Left Ventricular Function," 2019.

[5] N. Paragios, "A level set approach for shape-driven segmentation and tracking of the left ventricle," *IEEE transactions on medical imaging,* vol. 22, no. 6, pp. 773-776, 2003.

[6] J. G. Bosch *et al.*, "Automatic segmentation of echocardiographic sequences by active appearance motion models," *IEEE transactions on medical imaging,* vol. 21, no. 11, pp. 1374-1383, 2002.

[7] J. Yan and T. Zhuang, "Applying improved fast marching method to endocardial boundary detection in echocardiographic images," *Pattern Recognition Letters,* vol. 24, no. 15, pp. 2777-2784, 2003.

[8] E. G. Gomez, F. T. Robles, B. E. Ramirez, J. Olveres, and F. A. Cosío, "Full multi resolution active shape model for left ventricle segmentation," in *17th International Symposium on Medical Information Processing and Analysis*, 2021, vol. 12088: SPIE, pp. 422-432.

[9] Y. Ali, S. Beheshti, and F. Janabi-Sharifi, "Echocardiogram segmentation using active shape model and mean squared eigenvalue error," *Biomedical Signal Processing and Control,* vol. 69, p. 102807, 2021.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.

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

[12] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence,* 2021.

[13] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence,* vol. 9, no. 2, pp. 85-112, 2020.

[14] E. Smistad and A. Østvik, "2D left ventricle segmentation using deep learning," in *2017 IEEE international ultrasonics symposium (IUS)*, 2017: IEEE, pp. 1-4.

[15] C. Zhao *et al.*, "Multi-scale wavelet network algorithm for pediatric echocardiographic segmentation via hierarchical feature guided fusion," *Applied Soft Computing,* vol. 107, p. 107386, 2021.

[16] H. Belfilali, F. Bousefsaf, and M. Messadi, "Left ventricle analysis in echocardiographic images using transfer learning," *Physical and Engineering Sciences in Medicine,* vol. 45, no. 4, pp. 1123-1138, 2022.

[17] C. Sfakianakis, G. Simantiris, and G. Tziritas, "GUDU: Geometrically-constrained Ultrasound Data augmentation in U-Net for echocardiography semantic segmentation," *Biomedical Signal Processing and Control,* vol. 82, p. 104557, 2023.

[18] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Advances in neural information processing systems,* vol. 30, 2017.

[19] M. K. Patrick, A. F. Adekoya, A. A. Mighty, and B. Y. Edward, "Capsule networks–a survey," *Journal of King Saud University-computer and information sciences,* vol. 34, no. 1, pp. 1295-1310, 2022.

[20] H. Moghaddasi *et al.*, "A Hybrid Capsule Network for Automatic 3D Mandible Segmentation applied in Virtual Surgical Planning," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022: IEEE, pp. 3768-3771.

[21] A. Jiménez-Sánchez, S. Albarqouni, and D. Mateus, "Capsule networks against medical imaging data challenges," in *Intravascular imaging and computer assisted stenting and large-scale annotation of biomedical data and expert label synthesis*: Springer, 2018, pp. 150-160.

[22] R. LaLonde and U. Bagci, "Capsules for object segmentation," *arXiv preprint arXiv:1804.04241,* 2018.

[23] R. LaLonde, Z. Xu, I. Irmakci, S. Jain, and U. Bagci, "Capsules for biomedical image segmentation," *Medical image analysis,* vol. 68, p. 101889, 2021.

[24] S. Leclerc *et al.*, "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE transactions on medical imaging,* vol. 38, no. 9, pp. 2198-2210, 2019.

[25] E. Folland, A. Parisi, P. Moynihan, D. R. Jones, C. L. Feldman, and D. Tow, "Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. A comparison of cineangiographic and radionuclide techniques," *Circulation,* vol. 60, no. 4, pp. 760-766, 1979.