2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340931

Dual consistency loss for contour-aware segmentation in medical images

Helena R. Torres, Bruno Oliveira, Jaime C. Fonseca, Pedro Morais, João L. Vilaça

***Abstract*— Medical image segmentation is a paramount task for several clinical applications, namely for the diagnosis of pathologies, for treatment planning, and for aiding image-guided surgeries. With the development of deep learning, Convolutional Neural Networks (CNN) have become the state-of-the-art for medical image segmentation. However, issues are still raised concerning the precise object boundary delineation, since traditional CNNs can produce non-smooth segmentations with boundary discontinuities. In this work, a U-shaped CNN architecture is proposed to generate both pixel-wise segmentation and probabilistic contour maps of the object to segment, in order to generate reliable segmentations at the object's boundaries. Moreover, since the segmentation and contour maps must be inherently related to each other, a dual consistency loss that relates the two outputs of the network is proposed. Thus, the network is enforced to consistently learn the segmentation and contour delineation tasks during the training. The proposed method was applied and validated on a public dataset of cardiac 3D ultrasound images of the left ventricle. The results obtained showed the good performance of the method and its applicability for the cardiac dataset, showing its potential to be used in clinical practice for medical image segmentation.**

***Clinical Relevance*— The proposed network with dual consistency loss scheme can improve the performance of state-of-the-art CNNs for medical image segmentation, proving its value to be applied for computer-aided diagnosis.**

I. INTRODUCTION

Image segmentation is a fundamental task in medical image analysis, and it is essential for several clinical tasks, such as clinical diagnoses, image-guided interventions, surgical planning, and treatment follow-ups [1]. However, medical image segmentation is a challenging task since it requires strong expertise in image analysis. In fact, in clinical practice, this task is often performed manually, being time-consuming, labor-intensive, and prone to inter and intra-observer variability. Thus, automatic methods can be helpful to design an effective segmentation approach, allowing to avoid the drawbacks of the traditional manual segmentation [2], [3]. Currently, the development of such automatic methods was boosted due to the advances of deep learning (DL). Here, Convolutional Neural Networks (CNNs) have superseded the image processing field due to its capability to generate highly accurate segmentations [4]. However, challenges in the current DL-based segmentation approaches can be still identified. One of these challenges concerns the accuracy of the delineation near the object's boundaries, since unsatisfactory segmentation results may occur near these regions [5]. This can be explained by two main reasons. Firstly, the anatomical structures in medical images may have ambiguous or undefined boundaries, hampering the segmentation task. Secondly, CNNs are trained to capture edge, texture, and intensity features simultaneously, which may divert their focus and ability to capture boundaries' details [6]. Additionally, most of CNN architectures contain down-sample processes that results in loss of spatial information, naturally impacting the object's boundaries recognition. To overcome these issues, several works were proposed in the literature, mostly focused on the development of contour-aware networks to account for the object boundaries [7]–[11]. Yet, the solutions proposed in the state-of-the-art did not properly explore the inherent relationship between the segmentation and contour predictions.

Motivated by these observations, a contour-aware CNN is proposed in this work for the task of medical image segmentation. Specifically, a U-shaped network architecture is applied to predict segmentation and probabilistic contour maps simultaneously. Moreover, to ensure the consistent learning between both predictions, a dual consistency loss is proposed to explore the intrinsic relation between them. The proposed method was applied for the task of left ventricle segmentation in 3D ultrasound (US) images. As such, the main contributions of this work can be described as follows:

- A contour-aware network architecture that generates both segmentation and contour maps is proposed;
- A dual consistency loss scheme is applied during training to explore the inherent relation between segmentation and contour maps;

This work was funded by projects NORTE-01–0145-FEDER000059 and NORTE-01–0145-FEDER-000045, supported by Northern Portugal Regional Operational Programme Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). It was also funded by national funds, through the FCT–Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020 and LASI-LA/P/0104/2020 and by FCT and FCT/MCTES in the scope of the projects UIDB/05549/2020, UIDP/05549/2020, and CEECINST/00039/2021. This project was also funded by the Innovation Pact HfFP – Health From Portugal, co-funded from the "Mobilizing Agendas for Business Innovation" of the "Next Generation EU" program of Component 5 of the Recovery and Resilience Plan (RRP), concerning "Capitalization and Business Innovation", under the Regulation of the Incentive System "Agendas for Business Innovation". The authors also acknowledge support from FCT and the European Social Found, through Programa Operacional Capital Humano (POCH), in the scope of the PhD grant SFRH/BD/136670/2018, SFRH/BD/136721/2018 and COVID/BD/154328/2023.

Helena R. Torres and Bruno Oliveira are with 2Ai – School of Technology, IPCA, Barcelos, Portugal, with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal, with Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, with ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal (email: htorres@ipca.pt, boliveira@ipca.pt), and with LASI – Associate Laboratory of Intelligent Systems, Guimarães, Portugal. Jaime Fonseca is with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal (email: jaime@dei.uminho.pt). Pedro Morais and João Vilaça are with 2Ai – Polytechnic Institute of Cávado and Ave, Barcelos, Portugal (email: pmorais@ipca.pt, jvilaca@ipca.pt) and with LASI – Associate Laboratory of Intelligent Systems, Guimarães, Portugal.

- The proposed method is validated on a clinical dataset of 3D US for the challenging task of left ventricle segmentation and compared against state-of-the-art methods.

The remainder of the paper is organized as follows. In Section II, the proposed method is described. The experiments are described in Section III, with the results being presented in Section IV. In Section V, the performance of the method is discussed, and the main conclusions of this paper are given in Section VI.

II. Methods

In this section, we present the formulation of the proposed method in detail. Firstly, the contour-aware CNN is detailed. Secondly, the proposed consistency learning scheme is described. An overview of the proposed method is presented in Figure 1.

A. Segmentation and contour learning network

A U-shaped encoder-decoder network based on the well-known U-Net [12] is adapted is this work to predict pixel-wise segmentation and probabilistic contour maps simultaneously. The network architecture is composed of two paths: (1) an encoding network, also called contraction path, and (2) a decoding network, also called expanding path. The encoding path is composed of downsampling blocks used to capture context from the image and to learn scale invariance. Each encoding block is composed of two convolutional layers followed by batch normalization and a leaky rectified linear unit (ReLU). The downsampling is implemented using a stride convolution on the first layer of each encoding block. The initial number of feature maps is defined to be 32, which is doubled in each downsampling stride convolution operation. The decoding path corresponds to a symmetric expanding path that has the goal of compensating the spatial information loss caused by the down-sampling process. To perform the up-sampling process during the expanding process, transposed convolution is applied between each decoding block of this path. To ensure information sharing between encoder and decoder, skip connections are used as proposed in [12].

To enhance the network learning at the object's boundaries, two outputs are extracted from the network, namely: 1) a pixel-wise segmentation map; and 2) a contour probability map (Figure 1). This contour probability map is represented by a gaussian-like curve defined around the contour of the object to be segmented. Both network outputs are extracted after application of a Sigmoid activation layer at the end of the network. To guide the network to predict the maps, two different loss functions are used. For the pixel-wise segmentation prediction, the Dice loss function was used to compare the ground-truth segmentation Y_s with the predicted segmentation map $\hat{Y}_s$, such as:

$$\mathcal{L}_s = \frac{2Y_s.\hat{Y}_s}{{Y_s}^2 + {\hat{Y}_s}^2}. \tag{1}$$

For contour prediction, the mean squared error (MSE) loss was applied to accomplish a regression-based task. This loss is defined as:

$$\mathcal{L}_c = ||Y_c - \hat{Y}_c||^2, \tag{2}$$

where Y_c is the ground-truth for the gaussian-like contour map and $\hat{Y}_c$ is the predicted probability contour map.

This procedure allows to accomplish a contour-aware network, since the learning process is not only guided by segmentation predictions but also by contour probability estimation.

B. Consistent learning

Despite the proposed network being able to generate both segmentation and contour maps, enhancing the learning process at the object's boundaries, inconsistent results may appear between both outputs of the network. Since the pixel-wise segmentations and the contour maps are explicitly related, a consistency learning between the two outputs of the network must be ensured. To achieve this goal, a dual consistency learning scheme is applied in this work, where the duality is achieved by establishing segmentation-to-contour (S2C) and contour-to-segmentation (C2S) consistencies.

The S2C consistency is achieved by forcing the segmentation output to have a probabilistic contour map similar to the one generated by the network. The process for computing this consistency is relatively straightforward. Firstly, a dilation operation is performed to the segmentation output and a subtraction process is applied to only remain the contour of the segmentation. To ensure the differentiability needed during training, the dilation process is applied using a one-strided max-pooling layer, mimicking the morphological dilation. After this process, the contour map is smoothed by a convolutional layer with a gaussian kernel to achieve a gaussian-like curve of the segmentation contour. To compute the consistency loss for the S2C, the MSE was applied:

$$\mathcal{L}_{S2C} = ||\hat{Y}_c - \hat{Y}_{S2C}||^2, \tag{3}$$

where $\hat{Y}_{S2C}$ is the contour map of the segmentation output.

To promote the C2S consistency, the region enclosed by the contour map must be coherent with the segmentation output, *i.e.,* must correspond to the same region. Thus, to compute the C2S consistency, a segmentation mask must be retrieved from the predicted contour map. To do that, maps corresponding to the maximum cumulative intensities at each direction in each image dimension are retrieved from the contour map output. These maps are then multiplied to only maintain the interior of the contour, which corresponds to a segmentation-like map. The corresponding consistency loss is then given by the Dice loss between the segmentation-like map $\hat{Y}_{C2S}$ and the output segmentation map $\hat{Y}_s$:

$$\mathcal{L}_{C2S} = \frac{2\hat{Y}_{C2S}.\hat{Y}_s}{{\hat{Y}_{C2S}}^2 + {\hat{Y}_s}^2}. \tag{4}$$

The overall loss function is given by the sum of the four individual losses, being defined as:

$$\mathcal{L} = \alpha_1\mathcal{L}_C + \alpha_2\mathcal{L}_C + \alpha_3\mathcal{L}_{C2S} + \alpha_4\mathcal{L}_{S2C}. \tag{5}$$

where α is a weight factor applied to each term, allowing to control the influence of each loss during the training.

III. Experiments

A. Dataset

The proposed method was evaluated for the task of left ventricle segmentation in 3D US images. For that, the public dataset CETUS was used [13]. This dataset is composed of 3D image sequences of the left ventricle from 45 subjects,

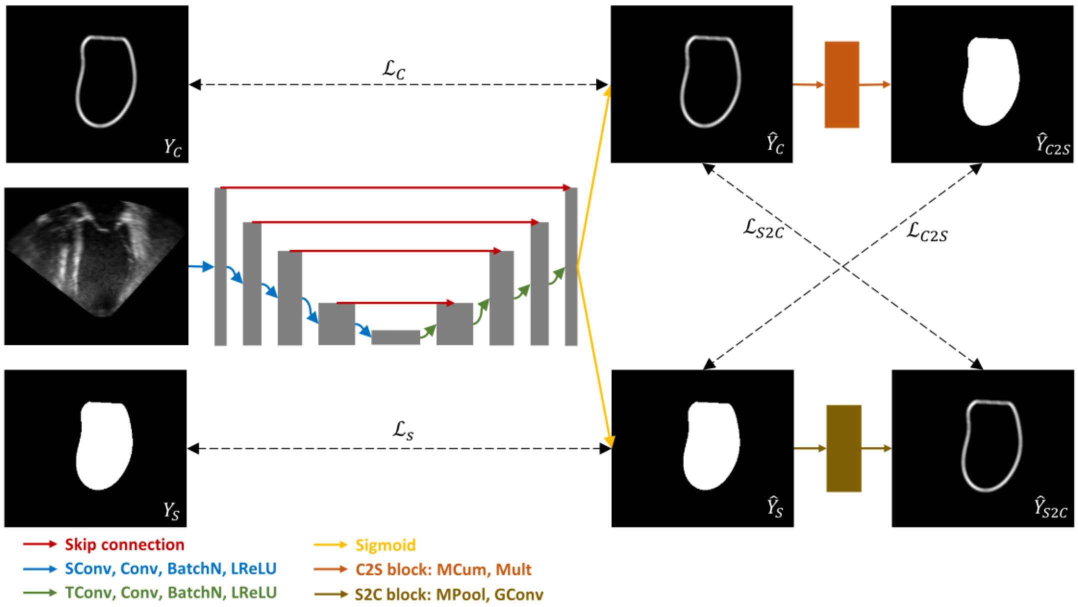

Figure 1 – Overview of the proposed contour-aware network with dual consistency. SConv – strided convolution; Conv – convolution; BatchN – batch normalization; LReLu - leaky rectified linear unit; TConv – transpose convolution; MCum – Maximum cumulative; Mult – Multiplication; MPool – Max pooling; GConv – Gaussian convolution.

from which 15 subjects are healthy and 30 are patients diagnosed with myocardial infarction or dilated cardiomyopathy. For each subject, two image frames are provided corresponding to end-diastole (ED) and end-systole (ES) phases. The segmentation ground-truths were performed by expert cardiologists and are available in the dataset. The dataset is divided into training set, corresponding to 15 subjects, and testing set, corresponding to 30 subjects. In this work, two subjects were removed from the original training set to create a validation set, used to analyse the network performance and loss evolution during training, and to choose the best epoch for early stop of the training.

B. Implementation details

The network was trained with a mini-batch size of 2 and using the Adam optimizer with an initial learning rate of 0.0001. At the end of each epoch, the learning rate was updated using a polynomial learning rate decay policy. To overcome overfitting problems during training, low probability dropout was used in the convolutional layers. Moreover, on-the-fly data augmentation techniques were applied during training, consisting of both intensity-based transformations and spatial transformations.

To decrease the difficulty in the optimization of network with the proposed consistency-based loss function and to avoid the vanishing gradients phenomenon, the network was initially trained with α_3 and α_4 set to zero. After few epochs, these α were set to one to allow the dual consistency scheme. Finally, the training and testing were performed using the PyTorch python library.

C. Evaluation metrics

The segmentation performance of the proposed method was assessed in terms of Dice coefficient, average symmetric surface distance (ASD), and Hausdorff distance (HD). These evaluation metrics are retrieved between the final segmentations results and respective ground-truth maps. To compute the final segmentation results, the segmentation maps $\hat{Y}_{C2S}$ and $\hat{Y}_S$ are averaged and thresholded.

IV. Results

Table 1 summarizes the performance of proposed method for left ventricle segmentation, assessed in terms of Dice, ASD, and HD. The results are presented for ES and ED phases individually. Overall, an average Dice of 90.8 ± 4.1 %, an average ASD of 1.6 ± 0.5 mm, and an average HD of 7.2 ± 2.8 mm were obtained, suggesting the good performance of the method. Moreover, Table 1 also presents the comparison of the proposed method against literature approaches that also used the CETUS dataset [14]–[16]. Here, we can conclude that the proposed method presents results competitive with the state-of-the-art. To visually demonstrate the method's performance, Figure 2 is presented, with one example of a bad and a good segmentation for each phase.

V. Discussion

In this work, a new method to segment the left ventricle in 3D US images was proposed. The proposed method consisted of a CNN network that simultaneously and consistently predicts segmentation and contour maps. Analyzing Table 1, it is possible to conclude that the proposed method generated accurate results for both image sequence corresponding to the ED and ES phases. The proposed network was designed to not only optimize segmentation maps but also contour maps, allowing to recover the boundary of the object and produce smooth segmentations, since the network is reinforced to learn boundary features during training. Moreover, the proposed dual consistency training scheme allowed to achieve accurate segmentations since inconsistent predictions between segmentation and contour maps are avoided.

The results obtained were also shown to be competitive with state-of-the-art approaches. These good results are corroborated by Figure 2, where good segmentation results

Table 1 – Method's performance in the CETUS dataset and comparison with state-of-the-art methods. Best results are presented in bold.

	End Diastole (ED)			End Systole (ES)		
	Dice (%)	ASD (mm)	HD (mm)	Dice (%)	ASD (mm)	HD (mm)
BEAS [14]	89.4 ± 4.1	2.26 ± 0.73	8.10 ± 2.66	85.6 ± 5.7	2.43 ± 0.91	8.13 ± 3.08
FCN [15]	90.6 ± 2.6	1.98 ± 1.03	11.94 ± 9.46	87.2 ± 5.0	2.83 ± 1.89	12.45 ± 10.69
ACNN [16]	91.2 ± 2.3	1.89 ± 0.51	**6.96 ± 1.75**	87.3 ± 5.1	2.09 ± 0.77	7.75 ± 2.65
Proposed	**92.3 ± 2.3**	**1.49 ± 0.41**	6.98 ± 2.79	**89.3 ± 4.8**	**1.68 ± 0.64**	**7.39 ± 2.9**

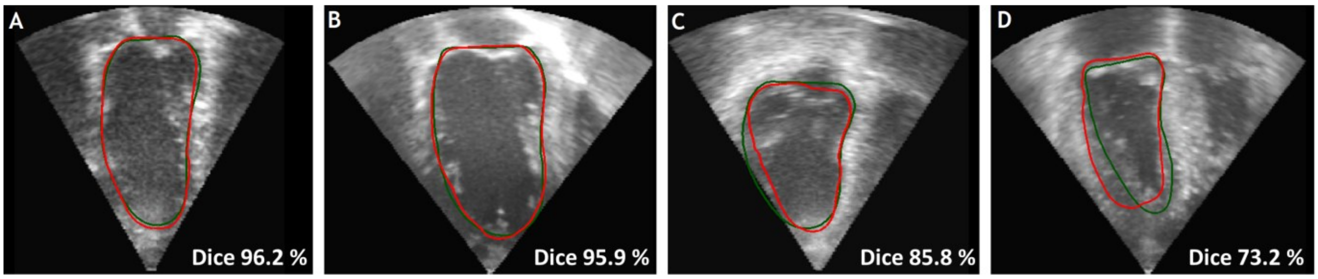

Figure 2 – Example of segmentation results (red) and respective ground-truth (green). A) good result for the ED phase; B) bad result for the ED phase; C) good result for the ES phase; D) bad result for the ES phase;

can be visualized. However, less accurate segmentation results can be also seen, potentially corresponding to pathological images with particular features. Nevertheless, the overall performance of the method showed its potential to be used for reliable medical image segmentation.

In future work, the proposed method will be applied and evaluated for different anatomical structures segmentation in different imaging modalities, since the proposed approach can be applied to a variety of medical segmentation tasks. Moreover, the addition of the contour information generated by the network will be used jointly with the predicted segmentation masks to achieve the final segmentation, potentially improving the results.

VI. Conclusion

A contour-aware CNN with dual consistency learning was applied for the task of left ventricle segmentation in 3D US images. The proposed method was shown to produce accurate segmentations, having potential to be used in clinical practice for medical image analysis.

References

[1] K.-P. Wong, "Medical image segmentation: methods and applications in functional imaging," in *Handbook of biomedical image analysis*, Springer, Boston, 2005, pp. 111–182.

[2] H. R. Torres, S. Queirós, P. Morais, B. Oliveira, J. C. Fonseca, and J. L. Vilaça, "Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review," *Comput. Methods Programs Biomed.*, vol. 157, pp. 49–67, Apr. 2018.

[3] H. R. Torres *et al.*, "A review of image processing methods for fetal head and brain analysis in ultrasound images," *Comput. Methods Programs Biomed.*, vol. 215, p. 106629, 2022.

[4] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *J. Digit. Imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[5] M. Zhang, B. Dong, and Q. Li, "Deep Active Contour Network for Medical Image Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 2020, pp. 321–331.

[6] R. Geirhos, C. Michaelis, F. A. Wichmann, P. Rubisch, M. Bethge, and W. Brendel, "Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," *7th Int. Conf. Learn. Represent. ICLR 2019*, no. c, pp. 1–22, 2019.

[7] H. R. Torres *et al.*, "Fetal head circumference delineation using convolutional neural networks with registration-based ellipse fitting," in *SPIE Medical Imaging*, 2022, no. April 2022, p. 126.

[8] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P. A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image Anal.*, vol. 36, pp. 135–146, 2017.

[9] C. Tan, L. Zhao, Z. Yan, K. Li, D. Metaxas, and Y. Zhan, "Deep multi-task and task-specific feature learning network for robust shape preserved organ segmentation," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 1221–1224, 2018.

[10] M. I. Razzak, "Microscopic Blood Smear Segmentation and Classification using Deep Contour Aware CNN and Extreme Machine Learning," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017.

[11] B. Murugesan, K. Sarveswaran, S. M. Shankaranarayana, K. Ram, J. Joseph, and M. Sivaprakasam, "Psi-Net: Shape and boundary aware joint multi-task deep network for medical image segmentation," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 7223–7226, 2019.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[13] O. Bernard *et al.*, "Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography," *IEEE Trans. Med. Imaging*, vol. 35, no. 4, pp. 967–977, 2016.

[14] D. Barbosa *et al.*, "Fast and Fully Automatic 3-D Echocardiographic Segmentation Using B-Spline Explicit Active Surfaces: Feasibility Study and Validation in a Clinical Setting," *Ultrasound Med. Biol.*, vol. 39, no. 1, pp. 89–101, 2013.

[15] H. Chen, Y. Zheng, J. H. Park, P. A. Heng, and S. K. Zhou, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 487–495, 2016.

[16] O. Oktay *et al.*, "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 384–395, 2018.