

RU-Net: A refining segmentation network for 2D echocardiography

Sarah Leclerc
CREATIS
Universite de Lyon
Lyon, France

Erik Smistad
CIUS
NTNU
Trondheim, Norway

Thomas Grenier
CREATIS
Universite de Lyon
Lyon, France

Carole Lartizien
CREATIS
Universite de Lyon
Lyon, France

Andreas Ostvik
CIUS
NTNU
Trondheim, Norway

Frederic Cervenansky
CREATIS
Universite de Lyon
Lyon, France

Florian Espinosa
Cardiovascular dpt
University hospital of St-Etienne
Saint-Etienne, France

Torvald Espeland
CIUS
St. Olavs Hospital
Trondheim, Norway

Erik Andreas Rye Berg
CIUS
St. Olavs Hospital
Trondheim, Norway

Pierre-Marc Jodoin
Computer Science department
University of Sherbrooke
Sherbrooke, Canada

Lasse Lovstakken
CIUS
NTNU
Trondheim, Norway

Olivier Bernard
CREATIS
Universite de Lyon
Lyon, France

Abstract—In this work, we present a novel attention mechanism to refine the segmentation of the endocardium and epicardium in 2D echocardiography. A combination of two U-Nets is used to derive a region of interest in the image before the segmentation. By relying on parameterised sigmoids to perform thresholding operations, the full pipeline is trainable end-to-end. The Refining U-Net (RU-Net) architecture is evaluated on the CAMUS dataset, comprising 2000 annotated images from the apical 2 and 4 chamber views of 500 patients. Although geometrical scores are only marginally improved, the reduction in outlier predictions (from 20% to 16%) supports the interest of such approach.

Index Terms—2D Echocardiography, Multi-class segmentation, Deep learning, Robustness, Outlier reduction

I. INTRODUCTION

A. Context

The inherent properties of ultrasound imaging (real time, low cost, good soft tissue contrast and high temporal and spatial resolutions) lead it to be the primary

S. Leclerc, T. Grenier, C. Lartizien, F. Cervenansky and O. Bernard are with the University of Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, University of Lyon 1, Villeurbanne, France.

E. Smistad, A. Ostvik and L. Lovstakken are with the Center of Innovative Ultrasound Solutions (CIUS), Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

F. Espinosa is with the Cardiovascular department Centre Hospitalier de Saint-Etienne, Saint-Etienne, France

T. Espeland and E.A. Rye Berg are with the Center of Innovative Ultrasound Solutions (CIUS) and the Clinic of cardiology, St. Olavs Hospital, Trondheim, Norway

P.-M. Jodoin is with the Computer Science Department, University of Sherbrooke, Sherbrooke, Canada.

imaging modality to evaluate the heart function. However, the poor image quality makes the delineation of cardiac structures especially difficult. Recent studies in 2D [1] and 3D echocardiography [2] have demonstrated the potential of fully-automatic solutions in producing accurate reproducible segmentation results. Nevertheless, they also underlined their lack of robustness on unusual cases and situations, and that they were not ready for clinical application.

B. Motivation

Traditionally, the robustness of segmentation algorithms is increased by integrating specifically-designed pre-processing or prior information [2] that regularize the input and/or output spaces of segmentation pipelines. On apical 2- and 4- chamber views of the left ventricle (LV), one strong prior is that the LV is fully present in the image. In this work, we investigate the potential of a segmentation-derived attention mechanism to focus the image processing on a region of interest (ROI) in order to refine the segmentation. The encoder-decoder convolutional neural network (EDN) U-net [3] is used to illustrate the interest of such approach.

C. Related work

In EDN frameworks, refining the segmentation has been addressed by cascaded networks, either trained end-to-end as Stacked Hourglasses (SHG) [4] or trained separately to derive a shape constraint as Anatomically Constrained Neural Networks (ACNN) [5]. Deep supervision as in U-Net ++ [6] is another solution to guide

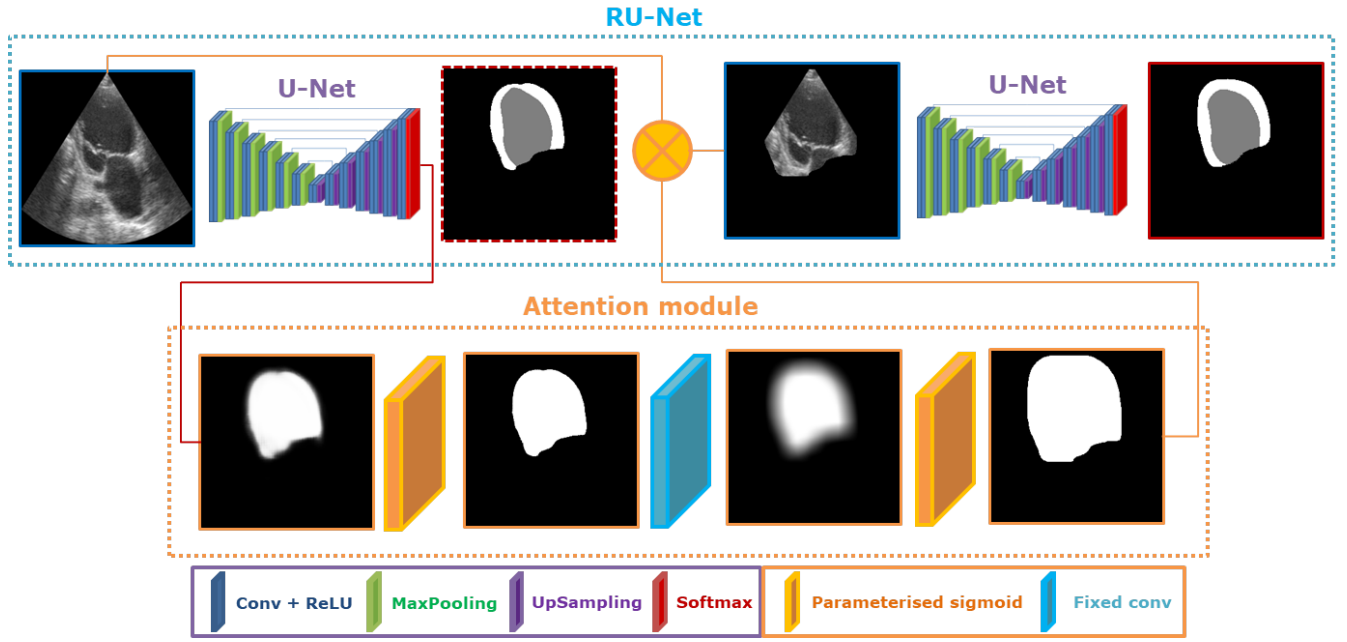


Fig. 1. Illustration of the RU-Net pipeline.

the network towards learning refining steps. All three augmentations have been used in conjunction with the U-Net architecture, but none have improved the overall geometric accuracy nor the number of distorted results on the echocardiographic CAMUS dataset. [1].

Hard attention, which entails extracting a ROI from an image, has the advantage of reducing the task complexity but is often not differentiable [7], and therefore performed by two independently trained networks, which requires a lot of tuning to be efficient. Note there exists ways to easily incorporate “soft” learnable attention, such as the attention layer proposed in [7], whose parameters are trained to reduce the feature response in irrelevant background regions without isolating an ROI.

D. Contributions

We propose a differentiable hard attention mechanism based on the selection and processing of a ROI in a trainable end-to-end pipeline. RU-Net is the combination of two U-Nets, one that pre-processes the image based on a first segmentation, and an other that predicts the final segmentation. By using the segmentation of the first network as a way to simplify the ultrasound image instead of providing a first segmentation as in SHG, the refining network is less dependent on the original segmentation, leading to an improved correction.

II. METHOD

The proposed method consists in two stacked U-Nets. The segmentation of the first network is dilated and multiplied with the image to isolate a region containing the structures of interest. The sigmoid functions $sig(x)$

applied before and after the convolution-based dilation are parametrized sigmoids in order to make the full pipeline trainable end-to-end :

$$sig(x) = \frac{1}{1 + \exp(-slope * (x - shift))} \quad (1)$$

where the slope coefficient is set to 100 to have a quasi-instant transition and the shift left as a hyper-parameter. Please note that since we use a high slope value for the sigmoid, this function can be seen as a binarizing function which justifies the hard attention learning designation of our method

The shift of the second sigmoid is fixed at 0.01 in order to retrieve the energy lost through the convolution (Fig 1). Therefore, RU-Net has 2 additional hyper-parameters:

- 1) the dilation rate, dil ;
- 2) the shift for the first binarization.

Both segmentation outputs are optimized using the same multi-class dice loss function. The full pipeline is illustrated on Fig. 1.

III. EXPERIMENTS

A. Dataset

The three methods are evaluated on the publicly available CAMUS dataset [1], which contains 500 apical 2- and 4- chamber views with expert annotations at end-diastole and end-systole. This study focuses on the segmentation of the LV internal and external borders (the endocardium LV_{endo} and the epicardium LV_{epi}) to observe the impact of ROI information. In the evaluation, we do not differentiate the images either on view, time

TABLE I
GEOMETRICAL PERFORMANCE AND OUTLIERS. RU-NET : $dil = 11, shift = 0.5$
CROSS VALIDATION ON 10 SUBFOLDS OF 200 IMAGES

Model	Prms #	LV-endo					LV-epi					Outliers # %		
		D	MAD	HD	Cx	Sp	D	MAD	HD	Cx	Sp	geo	ana	geo \cap ana
U-net 1	2.0M	0.917 ± 0.055	1.8 ± 1.2	5.8 ± 3.4	0.958 ± 0.022	0.665 ± 0.037	0.946 ± 0.030	2.0 ± 1.1	6.2 ± 3.8	0.976 ± 0.012	0.743 ± 0.743	396 20%	95 4.8%	71 4%
SHG	4.5M	0.917 ± 0.050	1.8 ± 1.1	5.9 ± 3.5	0.960 ± 0.018	0.664 ± 0.036	0.946 ± 0.030	2.0 ± 1.1	6.2 ± 3.9	0.979 ± 0.008	0.742 ± 0.021	425 21%	47 2.4%	29 1.5%
RU-net	3.9M	0.921 ± 0.054	1.8 ± 1.1	5.7 ± 3.6	0.962 ± 0.017	0.669 ± 0.017	0.948 ± 0.033	1.9 ± 1.2	6.1 ± 4.4	0.980 ± 0.006	0.748 ± 0.019	320 16%	23 1.2%	19 1.0%

LV_{endo} : Endocardial contour of the left ventricle; LV_{epi} : Epicardial contour of the left ventricle; ED: End diastole; ES: End systole.
Best values shown in bold.

instant or image quality in order to assess the overall robustness of the algorithms.

B. Evaluation

To thoroughly gauge the performance of the algorithms, three standard geometrical metrics are computed on algorithm predictions: the Dice (D), Mean Absolute Distance (MAD), and Hausdorff Distance (HD). The Dice is a global overlap indice between the ground truth GT and the prediction P , while the MAD and HD respectively give information on the average contour distance and the maximal error committed between the prediction and the ground truth :

The shape correctness is assessed using the anatomical metrics proposed in [8], namely the convexity (Cx) and simplicity (Sp) of shapes:

$$Convexity : Cx(P) = \frac{Area(P)}{Area(ConvHull(P))} \quad (2)$$

$$Simplicity : Sp(P) = \frac{\sqrt{4\pi * Area(P)}}{Perimeter(P)} \quad (3)$$

The robustness of an algorithm may be gauged by its number of outlier predictions. We propose, as in [1] and [8], to represent outlier predictions by setting upper limit values established from the inter-variability for the HD and MAD, and lower limit values for the simplicity and convexity. The corresponding values are summarized in Table II. Geometrical outliers show results above the highest average inter-variability, while anatomical outliers are characterized by a simplicity or convexity lower than the lowest value observed on the 200 images that were annotated by the three experts. Note that the anatomical criteria is totally unsupervised, i.e does not require a ground truth to be evaluated, and can therefore be used as a risk of segmentation failure detector even for unannotated images.

TABLE II
OUTLIER CRITERIA

GEOMETRICAL			ANATOMICAL		
Instant	MAD (mm)	HD (mm)	Structure	Sp	Cx
ED	3.5	4.0	LV	0.741	0.529
ES	8.2	8.8	Myo	0.960	0.694

C. Evaluated methods

Three algorithms are compared in this study, building on the same baseline in order to analyze two different refinement strategies.

- U-Net 1, the baseline, as described in [1]. U-Net 1 shows a similar performance when compared to the original U-Net with a lot less parameters (2M);
- SHG, as described in [1]. SHG corresponds to 3 cascaded light U-Nets 1 (1.5M). The inputs of the second and last U-Nets are the concatenation of the image and the previous segmentation;
- Our proposed RU-Net, defined in the previous section, which is built from U-Net 1.

D. Results

Table I summarizes the results obtained on a 10-fold cross-validation for U-Net 1, SHG and RU-Net. From this table, one can see that RU-Net outperforms the other two networks in terms of geometrical (-4%) and anatomical outliers (down to 1.2%), while maintaining high geometrical accuracy. Such two-step architecture therefore improves the robustness of the segmentation.

On classic geometrical metrics, the SHG architecture does not bring improvement compared to the scores of U-Net 1, however, a decrease of the anatomical outlier rate can be noted. RU-Net in turn does increase the geometrical performance consistently, though marginally, and induces a significant drop in anatomical outliers which only represent 1% of the dataset.

TABLE III
REFINEMENT EFFECT ON RU-NET WITH $dil = 30$ AND $shift = 0.7$
CROSS VALIDATION ON 10 SUBFOLDS OF 200 IMAGES

	LV-endo		LV-epi		Outliers # %	
	MAD	HD	MAD	HD	geo	ana
RU-net- o1	1.8 ± 1.2	5.9 ± 3.6	2.0 ± 1.1	6.5 ± 3.9	403 20%	116 5.8%
RU-net - o2	1.7 ± 1.0	5.6 ± 3.3	1.9 ± 1.1	6.0 ± 4.1	306 15%	33 1.7%

Furthermore, it is interesting to see that most anatomical outliers are also geometrical outliers for all methods, which means that shape irregularities are a strong indicator that the segmentation is at least partially inaccurate.

To observe the refinement at work in RU-Net, and its sensitivity to hyper-parameter changes, Table III shows the comparison between the two outputs of a RU-Net with different dilation rate and shift than the one in Table I. We can first observe that the scores produced by RU-Net are quite stable when changing the two hyper-parameters. Furthermore, the scores of the first output are very close to what is obtained by U-Net 1, which explicitly shows that the refinement occurring along RU-Net is not at the cost of the quality of the first output.

IV. DISCUSSION

The proposed attention mechanism relies on the strong assumption that the first segmentation was already quite successful. Therefore, it cannot correct degenerated results, but instead allows to refine the segmentation results, leading to a significant reduction in the amount of geometrical and anatomical outliers. The comparison with SHG further shows that in cascaded frameworks, using a network to transform the input image is better than transmitting a segmentation mask.

In an attempt to further improve results, we implemented the following alternatives :

- 1) replace the dilation by an inception layer, which allows for three dilation rates for the ROI;
- 2) weigh more strongly the second output than the first;
- 3) use two cascaded RU-Net;
- 4) pre-train the weights of the first network;
- 5) apply RU-Net with a lighter U-Net.

Interestingly, neither replacing the single dilation by an inception layer nor applying a larger weight on the second loss allowed to improve the scores, slightly deteriorating the performance instead. Pre-training the first part of the network or using cascaded RU-Nets lead to similar performances to the baseline RU-Net, with outlier rates of 16% (geometrical) and 2% (anatomical) in both cases. Finally, we noted that with a original network with less parameters the refinement observed between

the two outputs, as in Table III, was more pronounced, showing significant improvement on both outlier rates, and also on geometrical scores.

V. CONCLUSION

The obtained results show the possibility to further improve the robustness of EDN segmentations using a two-steps pipeline. The attention mechanism, as supported by RU-Net scores, appears to be a strong lead to improve results on the 2D echocardiography multi-structure segmentation dataset CAMUS. In the future, the investigation of better ROI selection, possibly completed with the incorporation of shape and temporal priors, should enable EDNs to reach the intra-variability observed on the CAMUS dataset.

ACKNOWLEDGMENT

This work was performed within the framework of the LABEX PRIMES (ANR- 11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). The Centre for Innovative Ultrasound Solutions (CIUS) is funded by the Norwegian Research Council (project code 237887).

REFERENCES

- [1] S. Leclerc, E. Smistad, J. Pedrosa, A. stvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P. Jodoin, T. Grenier, C. Lartizien, J. Dhooze, L. Lovstakken, and O. Bernard, "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, Sep. 2019.
- [2] O. Bernard, J. G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea *et al.*, "Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 967–977, 2016.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
- [4] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 483–499.
- [5] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvaio, T. Dawes, D. P. O'Regan, B. Kainz, B. Glocker, and D. Rueckert, "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384–395, 2018.
- [6] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *in proc. of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11.
- [7] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look for the pancreas," *CoRR*, vol. abs/1804.03999, 2018. [Online]. Available: <http://arxiv.org/abs/1804.03999>
- [8] S. Leclerc, E. Smistad, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, L. Lovstakken, and O. Bernard, "Deep learning segmentation in 2d echocardiography using the camus dataset : Automatic assessment of the anatomical shape validity," in *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track*, London, United Kingdom, 08–10 Jul 2019. [Online]. Available: <https://openreview.net/forum?id=Byx4AM1ntN>