Segmenting the carotid-artery wall in ultrasound image sequences with a dual-resolution U-net

Nolann Lainé, Guillaume Zahnd, Herve Liebgott, Maciej Orkisz

Segmenting the carotid-artery wall in ultrasound image sequences with a dual-resolution U-net

1st Nolann Lainé* 2nd Guillaume Zahnd† 3rd Hervé Liebgott* 4th Maciej Orkisz*

*Univ Lyon, Université Lyon 1, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France
†Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg, Germany

***Abstract*—Thickening of intima-media complex in the common carotid artery is a biomarker of atherosclerosis. To automatically measure this thickness, we propose a region-based segmentation method, involving a supervised deep-learning approach based on the dilated U-net architecture, named caroSegDeep. It was trained and evaluated using 5-fold cross-validation on two open-access databases containing a total of 2676 annotated images. Compared with the methods already evaluated on these databases, caroSegDeep established a new benchmark and achieved a mean absolute error twice smaller than the inter-observer variability.**

***Index Terms*—Atherosclerosis, carotid artery, segmentation, ultrasound, deep learning.**

I. Introduction

World Health Organization reports the cardiovascular diseases, particularly atherosclerosis, as the leading cause of death worldwide [2]. Their prevention [4] requires screening by means of a non-ionising and inexpensive imaging modality. Ultrasound (*US*) imaging has these characteristics and is routinely used to explore the common carotid artery (*CCA*), which is often considered as the sentinel of atherosclerosis [15]. An early sign of this disease onset is arterial wall thickening. To measure the thickness of interest, the contours of the intima-media complex (*IMC*), namely, lumen-intima (*LI*) and media-adventitia (*MA*) interfaces, need to be identified (Fig. 1).

This task is most frequently addressed by contour-based approaches [1], [3], [11], [16], which exploit intensity changes caused by the echoes at acoustic interfaces. Region-based segmentation methods based on texture, adaptive thresholding [9], [14] or clustering have also been proposed. Recently, deep-learning (*DL*) has been successfully used in vascular US-image segmentation to enhance the structures of interest prior to the actual delineation by more conventional contour-based methods [7], [10], [13]. The drawback of these approaches is the necessity to combine a learnable pre-processing operation with an analytic segmentation task.

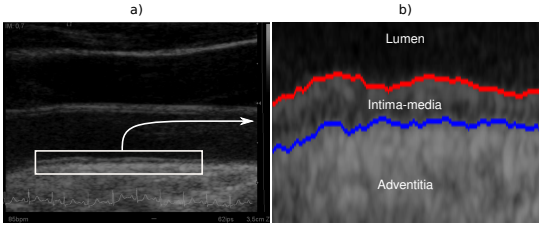

Fig. 1. Example longitudinal B-mode US image of a *CCA*. *a*) Far wall encompassed by a rectangle. *b*) Enlarged region detailing the *IMC* with its interfaces segmented by caroSegDeep: *LI* (red) and *MA* (blue).

The main contribution of this work is a region-based segmentation method using supervised deep learning—dubbed "caroSegDeep"— designed to extract the *IMC* in B-mode *US* images. Using a collection of overlapping patches, caroSegDeep successfully segments the entire exploitable part of the *IMC* with a simple network architecture. All training and evaluation processes were carried out on the open-access multi-center databases CUBS 1 [6][1] and CUBS 2 [5][2]. Each image was annotated by at least two experts (A1 and A2) and one of them performed a second annotation round (A1') to assess inter- and intra-expert variability.

II. Method

The proposed solution builds on the U-net [12] architecture, with dilated convolutions on the bottleneck to increase the receptive field [8]. It works in two steps and uses two identical networks with different weights (Fig. 2). First, the user's two mouse clicks define the region of interest (*ROI*) to be segmented. Next the far wall is detected. Eventually the *IMC* segmentation is done in the vicinity of the detected location. The code is available at https://github.com/nl3769/caroSegDeep.

A. Detection of the far wall

Similarly to many state-of-the-art methods [7], [10], [14], [16], the far wall is first detected to drive the subsequent actual *IMC* segmentation. The corresponding U-net, referred to as Θ_{FW}, uses patches of full image height and 128-pixel width. The next paragraph describes the data preparation for the training phase of Θ_{FW}, while the subsequent one specifies how the patch-wise predictions inferred using Θ_{FW} are post-processed to obtain the curve approximately localizing the far wall on the entire *ROI*.

Pre-processing and training: Each image was resampled to a 512-pixel height. For training, the *IMC* median axis was defined as the line halfway between *LI* and *MA* annotations, extending across the entire width of the *ROI*. Pixels below the median axis were set to 1 and the others to 0, to generate a

[1]http://dx.doi.org/10.17632/fpv535fss7.1
[2]http://dx.doi.org/10.17632/m7ndn58sv6.1

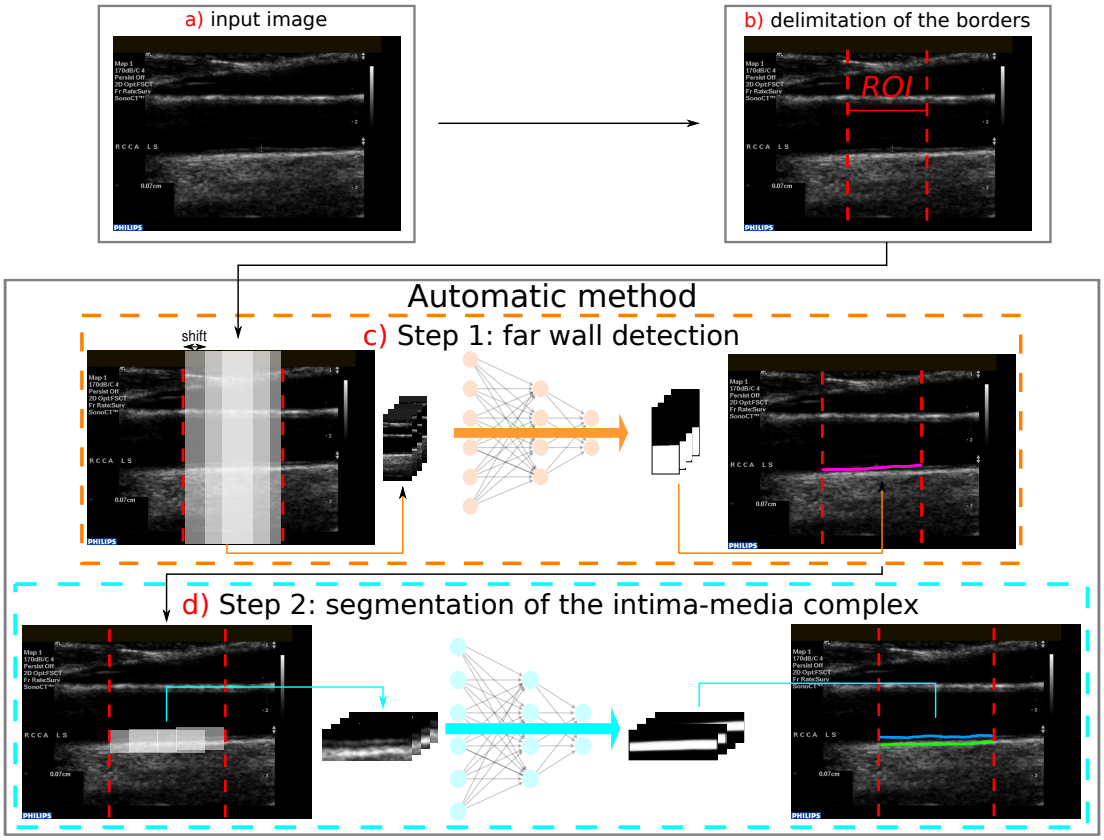

Fig. 2. Outline of the proposed method. *a*) Input image. *b*) Interactive delimitation of the left and right borders of the *ROI* (two mouse clicks). *c*) Far wall detection: median axis (magenta) deduced from overlapping masks predicted within full image-height patches. *d*) *IMC* segmentation: *LI* (cyan) and *MA* (green) interfaces deduced from predicted overlapping masks within high-resolution patches distributed along the median axis.

reference mask (M_{ROI}). Image *ROI* and M_{ROI} were identically cut into patches with a 100-pixel overlap aiming at data augmentation. These gray-level and binary patches (Fig. 3) were the inputs of the training process.

Inference and post-processing: The same resampling and cutting of the *ROI* into overlapping 128×512-pixel patches is performed to prepare each image for inference. Then, each patch is segmented using Θ_{FW}. Two maps are created:

- ***overlay map:*** containing, for each pixel, the number of overlapping patches covering this pixel,
- ***prediction map:*** containing, for each pixel, the sum of values predicted by Θ_{FW} within these patches.

Subsequently, an average value in the range $[0, 1]$ is calculated for each pixel by dividing the prediction map by the overlay map. The averaged prediction map is binarized using a threshold of 0.5 and then cleaned by retaining the largest connected component. The median axis we seek is the upper boundary of the region thus obtained.

B. Segmentation of the IMC

The actual segmentation of the *IMC* uses the same network architecture (the dilated U-net used here is referred to as Θ_{IMC}) and several concepts from Section II-A: overlapping 128×512-pixel patches, overlay and prediction maps, as well as a similar post-processing except that two contours are extracted (the *LI* and *MA* interfaces). The following paragraphs focus on the specific choices made for this step.

Pre-processing and training: Here, the height of the patches just needs to encompass the *IMC*, which is 0.8-millimeter thick, on average. Therefore, each image is vertically resampled to $5\,\mu$m/pixel, so that the 512-pixel patch height roughly corresponds to 2.6 mm. The ground truth for training was deduced from the images thus interpolated (Fig. 4): pixels located between the annotated *LI* and *MA* interfaces were set to 1, and the others to 0. The patches were picked along the median axis: at each abscissa x_i, three patches were extracted, centered at y_i and $y_i \pm 128$, where y_i is the mean ordinate of the median axis on the patch width. This choice aimed at data augmentation coping with possibly inaccurate far-wall approximation, as well as with tilted arteries.

Inference and post-processing: The patches are also extracted along the far-wall approximation, which results here from the first step (Section II-A), and the number of vertically overlapping patches extracted at each abscissa x_i may be three or more, depending on the tilt of the median axis. After combining the predictions made by Θ_{IMC} in all patches into a prediction map, the segmentation map is derived thereof, as described before. The *LI* and *MA* interfaces are eventually

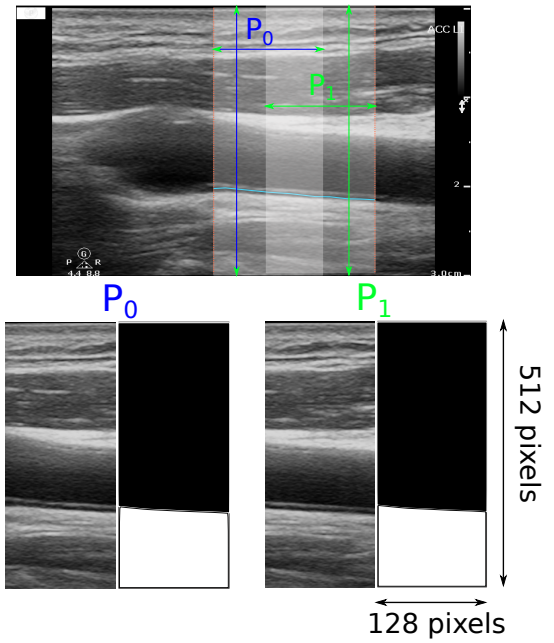

Fig. 3. Schematic representation of image patches and corresponding masks used during the training phase of the far-wall detection network.

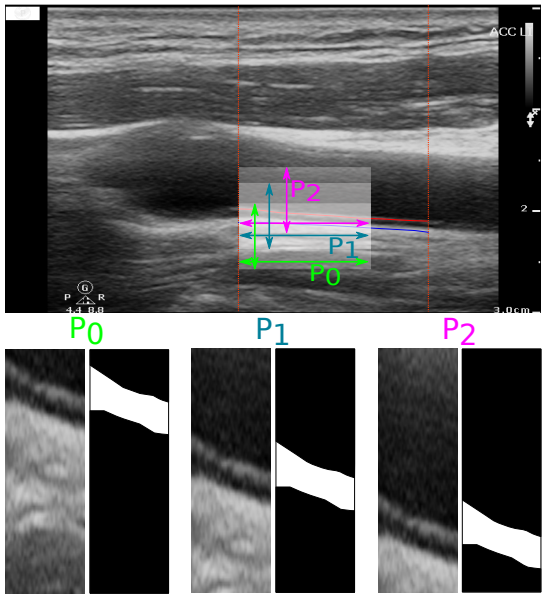

Fig. 4. Schematic representation (differing from the actual patch size) of data used during the training phase of the *IMC*-segmentation network. Image patches with associated masks located at (x_i, y_i), and $(x_i, y_i \pm 128)$.

deduced as the respective upper and lower boundaries of the region thus segmented.

III. Results

To assess each network on data not seen during its training, a 5-fold cross-validation was performed: the database was divided into five subsets of equal size, and five pairs of Θ_{FW} and Θ_{IMC} networks were trained and tested independently using in turn three subsets for training (60%), one subset for validation (20%) and one subset for testing (20%). The results obtained on the test subsets of these five pairs were merged to assess the performance of the method on the entire database.

The second step (actual *IMC* segmentation) of the proposed cascade approach assumes that the median axis resulting from the first step (far wall detection) is approximately correct: if the far wall is wrongly detected, the *ICM* segmentation is likely to fail. Therefore, we first quantified the success rate of the first step alone, then incorrect median axes were manually redrawn using a home-made graphical interface before assessing the accuracy of the second step.

Robustness of the far wall detection: Images where the distance between the predicted and reference median axes was greater than 0.8 mm were selected for visual inspection, which confirmed that 36 of them (1.3% of the database) were failures, *i.e.* curves unusable to initialize the *IMC* segmentation step. Hence, the success rate was of 98.7%. In the 36 images with failures, the median axis was interactively reinitialized, in preparation for the subsequent segmentation step.

Accuracy of the *IMC* segmentation: The *IMC* segmentation inaccuracies were calculated with respect to the annotations performed by the expert *A1*. We assessed the median absolute difference (*MAD*) on *IMT*, which is the target measure, as well as the Hausdorff distance on the *LI* and *MA* contours. These errors were compared with the inter- and intra-observer variabilities, A2 vs. A1 and A1' vs. A1, respectively. Table I summarizes the segmentation errors separately calculated for images from the CUBS 1 and CUBS 2 databases, which allows a comparison with other methods evaluated on these databases in previous studies [5], [6]. Thus, the previous benchmark on the CUBS 1 database was a *MAD* value on *IMT* equal to $114 \pm 117\,\mu$m, while on the CUBS 2 database the best conventional method achieved $139 \pm 119\,\mu$m and another U-net-based method obtained $178 \pm 120\,\mu$m.

TABLE I
SEGMENTATION INACCURACIES (MEAN ± STD) FOR CUBS 1 AND CUBS 2 DATASETS: MEAN ABSOLUTE DIFFERENCE (MAD) FOR THICKNESS QUANTIFICATION (IMT) AND HAUSDORFF DISTANCE (HD) FOR CONTOUR LOCATIONS (LI AND MA). THE ERRORS OF THE PROPOSED CAROSEGDEEP METHOD WITH RESPECT TO REFERENCE ANNOTATIONS ARE COMPARED TO THE INTER- AND INTRA-OBSERVER VARIABILITIES.

Measure	Dataset / Compare	CUBS 1	CUBS 2
IMT: MAD (μm)	caroSegDeep vs. A1	99 ± 89	106 ± 89
	A2 vs. A1	206 ± 168	192 ± 166
	A1' vs. A1	144 ± 123	160 ± 140
LI: HD (μm)	caroSegDeep vs. A1	320 ± 193	305 ± 197
	A2 vs. A1	380 ± 207	327 ± 138
	A1' vs. A1	357 ± 204	352 ± 140
MA: HD (μm)	caroSegDeep vs. A1	287 ± 153	289 ± 147
	A2 vs. A1	351 ± 161	338 ± 184
	A1' vs. A1	319 ± 155	346 ± 185

IV. Discussion

We proposed and fairly evaluated on open-access databases an almost-automatic (two user mouse-clicks) deep-learning method devised to extract the contours of the intima-media

complex in longitudinal B-mode ultrasound images of the carotid artery. The proposed approach is patch-based, which allows for segmenting variable-width *ROIs* of without resizing the images.

The far-wall localization step was successful in all but 1.3% of the images. This robustness is a prerequisite for overall correct segmentation. The latter achieved errors smaller than the inter-observer variability, both in terms of estimated thickness and contour location. Compared to the methods already evaluated on the same data [5], [6], caroSegDeep established a new benchmark. As the proposed method is based on supervised learning, it has the potential to increase its performance by using larger and more diverse data for training.

In particular, the largest errors occurred in the presence of calcified plaques, which were present in few images. This work was oriented towards asymptomatic plaque-free subjects, images with plaques were not expected. Nevertheless, we expect that results might be improved by re-training the networks on a database enriched with such images. This avenue deserves investigation. Another direction might exploit multiple annotations available for each image to increase the robustness by learning how to account for such uncertainties.

V. Acknowledgments

This work was partly supported, via NL's doctoral grant, by 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 authors have no relevant financial or non-financial interests to disclose.

VI. Compliance with ethical standards information

The data from human subjects used in this work were obtained and treated in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committees of the institutions involved in creating the multicentric database, from which these data were accessed.

References

[1] S. Delsanto, F. Molinari, P. Giustetto, W. Liboni, S. Badalamenti, and J. Suri, "Characterization of a completely user-independent algorithm for carotid artery segmentation in 2-D ultrasound images," *IEEE Trans Instrum Meas*, vol. 56, no. 4, pp. 1265–1274, 2007.

[2] S. Kaptoge, L. Pennells, D. De Bacquer, M. Cooney, M. Kavousi, G. Stevens *et al.*, "World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," *Lancet Glob Health*, vol. 7, no. 10, pp. e1332–e1345, 2019.

[3] C. Loizou, C. Pattichis, M. Pantziaris, T. Tyllis, and A. Nicolaides, "Snakes based segmentation of the common carotid artery intima media," *Med Biol Eng Comput*, vol. 45, no. 1, pp. 35–49, 2007.

[4] H. McGill Jr, C. McMahan, and S. Gidding, "Preventing heart disease in the 21st century: implications of the pathobiological determinants of atherosclerosis in youth (PDAY) study," *Circulation*, vol. 117, no. 9, pp. 1216–1227, 2008.

[5] K. Meiburger, F. Marzola, G. Zahnd, F. Faita, C. P. Loizou, N. Lainé, C. Carvalho, D. A. Steinman, L. Gibello, R. M. Bruno *et al.*, "Carotid ultrasound boundary study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement," *Computers Biol Med*, 2022.

[6] K. Meiburger, G. Zahnd, F. Faita, C. P. Loizou, C. Carvalho, D. A. Steinman, L. Gibello, R. M. B, F. Marzola, R. Clarenbach *et al.*, "Carotid ultrasound boundary study (CUBS): An open multicenter analysis of computerized intima–media thickness measurement systems and their clinical impact," *Ultrasound Med Biol*, vol. 47, no. 8, pp. 2442–2445, 2021.

[7] R.-M. Menchón-Lara, J.-L. Sancho-Gómez, and A. Bueno-Crespo, "Early-stage atherosclerosis detection using deep learning over carotid ultrasound images," *Appl Soft Comput*, vol. 49, pp. 616–628, 2016.

[8] N. Meshram, C. Mitchell, S. Wilbrand, R. Dempsey, and T. Varghese, "Deep learning for carotid plaque segmentation using a dilated U-net architecture," *Ultrason Imaging*, vol. 42, no. 4-5, pp. 221–230, 2020.

[9] Y. Nagaraj, P. Madipalli, J. Rajan, P. Kumar, and A. Narasimhadhan, "Segmentation of intima media complex from carotid ultrasound images using wind driven optimization technique," *Biomed Signal Process Control*, vol. 40, pp. 462–472, 2018.

[10] C. Qian, E. Su, and X. Yang, "Segmentation of the common carotid intima-media complex in ultrasound images using 2-D continuous max-flow and stacked sparse auto-encoder," *Ultrasound Med Biol*, vol. 46, no. 11, pp. 3104–3124, 2020.

[11] K. Raj, J. Joseph, P. Nabeel, and M. Sivaprakasam, "Automated measurement of compression-decompression in arterial diameter and wall thickness by image-free ultrasound," *Comput Methods Programs Biomed*, vol. 194, p. 105557, 2020.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Int Conf Med Image Computing and Computer-assisted Intervention – MICCAI*, 2015, pp. 234–241.

[13] J. Shin, N. Tajbakhsh, R. Hurst, C. Kendall, and J. Liang, "Automating carotid intima-media thickness video interpretation with convolutional neural networks," in *Proc IEEE Conf Computer Vision Pattern Recogn*, 2016, pp. 2526–2535.

[14] K. Wang, Y. Pu, Y. Zhang, and P. Wang, "Fully automatic measurement of intima-media thickness in ultrasound images of the common carotid artery based on improved Otsu's method and adaptive wind driven optimization," *Ultrason Imaging*, vol. 42, no. 6, pp. 245–260, 2020.

[15] F. Yousefi Rizi, J. Au, H. Yli-Ollila, S. Golemati, M. Makūnaitė, M. Orkisz *et al.*, "Carotid wall longitudinal motion in ultrasound imaging: An expert consensus review," *Ultrasound Med Biol*, vol. 46, no. 10, pp. 2605–2624, October 2020.

[16] G. Zahnd, K. Kapellas, M. Van Hattem, A. Van Dijk, A. Sérusclat, P. Moulin *et al.*, "A fully-automatic method to segment the carotid artery layers in ultrasound imaging: Application to quantify the compression-decompression pattern of the intima-media complex during the cardiac cycle," *Ultrasound Med Biol*, vol. 43, no. 1, pp. 239–257, 2017.