Semantic Segmentation with DenseNets for Carotid Artery Ultrasound Plaque Segmentation and CIMT estimation

Maria del Mar Vila[a,b,c], Beatriz Remeseiro[d], Maria Grau[b,a,e], Roberto Elosua[b,a,f], Àngels Betriu[g], Elvira Fernandez-Giraldez[g], Laura Igual[c]

[a] *CIBER Enfermedades Cardiovasculares, Instituto de Salud Carlos III Monforte de Lemos 3-5, Pabellón 11, 28029 Madrid, Spain*
[b] *IMIM, Institut Hospital del Mar d'Investigacions Mèdiques Dr. Aiguader 88, 08003 Barcelona, Spain*
[c] *Departament de Matemàtiques i Informàtica, Universitat de Barcelona Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain*
[d] *Department of Computer Science, Universidad de Oviedo Campus de Gijón s/n, 33203 Gijón, Spain*
[e] *Departament de Medicina, Universitat de Barcelona Carrer Casanova, 143, 08036 Barcelona, Spain*
[f] *Facultat de Medicina, Universitat de Vic - Universitat Central de Catalunya Ctra. Roda, 70, 08500 Vic, Spain*
[g] *UDETMA i Servei Nefrologia, Hospital Universitari Arnau de Vilanova, IRB (Institut de Recerca Biomèdica de Lleida) Av. Rovira Roure, 80, 25198 Lleida, Spain*

Abstract

Background and Objective: The measurement of Carotid Intima Media Thickness (CIMT) in ultrasound images can be used to detect the presence of atherosclerotic plaques. Usually, the CIMT estimation strategy is semi-automatic, since it requires: 1) a manual examination of the ultrasound image for the localization of a Region Of Interest (ROI), a fast and useful operation when only a small number of images need to be measured; and 2) an automatic delineation of the CIM region within the ROI. The existing efforts for automating the process have replicated the same two-step structure, resulting in two consecutive independent approaches. In this work, we propose a fully automatic single-step approach based on semantic segmentation that allows us to segment the plaque and to estimate the CIMT in a fast and useful manner for large data sets of images.

Methods: Our single-step approach is based on Densely Connected Convolu-

*Principal corresponding author: Laura Igual (ligual@ub.edu)

tional Neural Networks (DenseNets) for semantic segmentation of the whole image. It has two remarkable characteristics: (1) it avoids ROI definition, and (2) it captures multi-scale contextual information in the complete image interpretation, due to the concatenation of feature maps carried out in DenseNets. Once the input image is segmented, a straightforward method for CIMT estimation and plaque detection is applied.

Results: The proposed method has been validated with a large data set (REGICOR) of more than 8,000 images, corresponding to two territories of the Carotid Artery: Common Carotid Artery (CCA) and Bulb. Among them, a subset of 331 images has been used to evaluate the performance of semantic segmentation ($\approx 90\%$ for train, $\approx 10\%$ for test). The experimental results demonstrated that our method outperforms other deep models and shallow approaches found in the literature. In particular, our CIMT estimation reaches a correlation coefficient of 0.81, and a CIMT mean error of 0.02 mm and 0.06 mm in CCA and Bulb images, respectively. Furthermore, the accuracy for plaque detection is 96.45% and 78.09% in CCA and Bulb, respectively. To test the generalization power, the method has also been tested with another data set (NEFRONA) that includes images acquired with different equipment.

Conclusions: The validation carried out demonstrates that the proposed method is accurate and objective for both plaque detection and CIMT measurement. Moreover, the robustness and generalization capacity of the method have been proven with two different data sets.

Keywords: Semantic Segmentation of Carotid Artery, Intima Media Thickness, Ultrasound Images, Atherosclerotic Plaque Detection, Fully Convolutional Neural Networks

1. Introduction

Cardiovascular diseases are the leading cause of death in developed countries. Most of at-risk individuals of cardiovascular events suffer atherosclerosis, a chronic inflammatory process characterized morphologically by an asymmetric focal thickening of the innermost layer of the artery. Thus, monitoring the detection of the atherosclerotic plaque as well as its characteristics or changes may have significant clinical relevance for the assessment of future cardiovascular events. The Ultrasound (US) Carotid Artery (CA) images are used to detect the burden of atherosclerosis, since they provide the possibility to measure the Carotid Intima Media Thickness (CIMT) of the artery and identify the presence of atherosclerotic plaques. The CIM region is defined by the Lumen-Intima (LI) and Media-Adventitia (MA) interfaces (see Figure 1), and the CIMT is commonly estimated in the far wall (interfaces at the bottom of the image) of the CA. To simplify, we use the term CIM region to refer to the region located at the far wall of the CA. The Mannheim Consensus [1] defines a sufficient criterion for plaque detection: plaques are structures inside the arterial lumen showing CIMT $\geq$ 1.5mm.

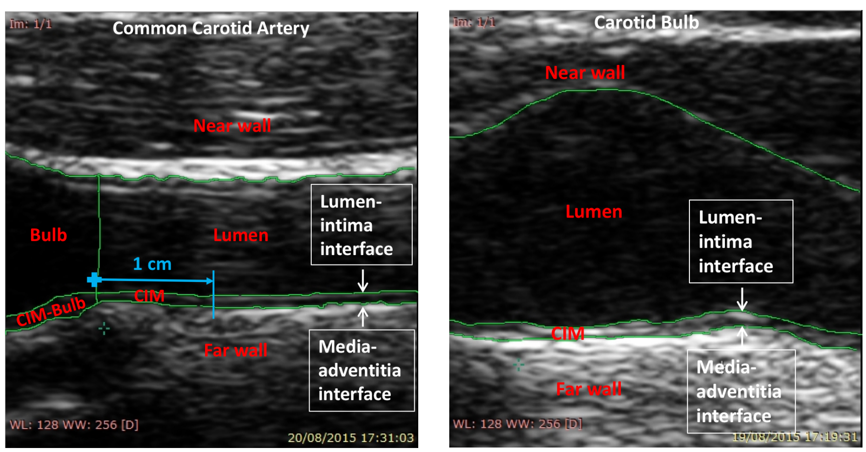


Figure 1: Common Carotid Artery (CCA) (left) and Bulb (right) US longitudinal images. The different parts of the CA are delimited with green lines. In both cases the CIMT is estimated in the CIM region from the far wall. The CIMT in the CCA is measured approximately 1 cm distal from the carotid Bulb.

1.1. Related Work

Basic techniques for CIM region delineation and plaque segmentation presented in the literature include, among others, Hough transform, edge detection [2, 3], active contours [4], snakes [5, 6], and other solutions such as integrated approaches that combine several basic Machine Learning (ML) methods [7, 8]. Interest readers can refer to the review studies [9, 10] for more references.

Following [10], the methods can be broadly classified into two categories. The first category includes techniques that are fully automatic, whereas the second one includes those that require user interaction, i.e. semi-automatic. Semi-automatic approaches [3, 6] require user interaction for manual initialization in order to select a Region of Interest (ROI), and/or to correct wrong results during examination. In general, the manual ROI selection together with these type of interactions result in better performance. The best semi-automatic methods found in the literature for clinical practice are the ones that offer visual feedback during image acquisition instead of analyzing stored images [3].

In contrast, fully automatic methods [4, 2, 7, 8, 11, 12] run without any initial setting, or user interaction. The main advantage of these techniques is that they are able to process large amounts of data. Furthermore, they allow the reproducibility of results, and save time and resources.

Preliminary efforts using ML and Deep Learning (DL) in fully automatic CIMT evaluation have been presented [11, 8, 12, 7]. In [11], a standard multi-layer perceptron with an auto-encoder is proposed for CA image interpretation, but it does not outperform the snake-based method in [4]. In [8], Zhang et al. proposes a two-step segmentation method of the CIM region based on patch-based classification and Stacked Sequential Learning. More recently, in [12], patch-based Convolutional Neural Networks (CNNs) are used in the different steps for CIMT estimation. This work uses US video instead of unique frame (as used in many works and in our paper, see Section 3.1), and thus adds an extra first step for selecting three end-diastolic ultrasound frames. The most recent work in the literature to automatically segment plaque is presented in [7], an approach that uses several ML methods and combines them in an iterative algorithm.

Table 1 summarizes the most relevant methods presented in the literature regarding the CIMT error and compares several characteristics that are explained in the next subsection.

Author	Year	Segmentation Method	Method SA/FA	Proc. Time per Frame	Data Type	Artery Territory	Presence of Plaque	N	Different Acquisition Devices	Mean CIMT Error (mm)
Faita et al.[3]	2008	Edge Detection	SA(2)	*NS*	V	CCA	No	150	No	0.001
Molinari et al.[2]	2012	Edge Detection	FA(2)	<15s	UF	CCA	*NS*	365	Yes	0.078
Loizou et al.[5]	2013	Snakes	SA(2)	28s	UF	CCA	Yes	20	No	0.065
Menchón-Lara et al.[11]	2015	NN Auto-Encoders	FA(2)	1.4s	UF	CCA	*NS*	55	No	0.018
Bastida-Jumilla et al.[4]	2015	Frequency-Domain Snakes	FA(2)	12.2s	UF	CCA	*NS*	46	No	0.014
Zhang et al.[8]	2015	Patch-based Basic ML techniques	FA(2)	*NS*	UF	CCA	Yes	100	Yes	1.37% (point-to-point relative error)
Shin et al.[12]	2016	CNN Patch-based	FA(2)	*NS*	V	CCA	*NS*	92	No	0.023 per interface (LI and MA)
Zhao et al.[6]	2017	Snakes	SA(2)	0.24s	V	*NS*	Yes	*NS*	Yes	0.053
Qian et al.[7]	2018	Patch-based Basic ML techniques	FA(2)	6min	UF	CCA	Yes	29	No	0.34 (average point-to-point distance)
Our proposal	2019	Fully CNN Semantic Segmentation	FA(1)	0.79s	UF	CCA & Bulb	Yes	4,751 (CCA) 3,733 (Bulb)	Generalization Test	0.022 (CCA) 0.06 (Bulb)

Table 1: Most recent/relevant techniques for CA segmentation and CIMT estimation together with their main characteristics: Author and reference, year of publication, the used segmentation method, if the method is Semi-Automatic (SA) or Fully-Automatic (FA) (“(1)”: one-step and “(2)”: two-step), the processing time per frame, the type of data as a Unique Frame (UF) or Video (V), the artery territory, the presence of plaque in the images of the data set, the number of subjects of the data set (N) and the mean CIMT error in mm. Finally, *NS* means “Not Specified” in the referenced paper.

1.2. Contributions

To the best of our knowledge, all the aforementioned DL segmentation techniques are two-step approaches that define separate methods to, first, localize the ROI (made manually in case of semi-automatic methods); and second, delineate the CIM region within the ROI.

In our paper, we propose a novel single-step (see column named "Method SA/FA" Table 1) DL approach for automatic CA image interpretation. This approach is based on Semantic Segmentation (SS) using Densely Connected Convolutional Networks (DenseNets) [13], which were designed to facilitate the training of very deep networks due to a reduction in the number of parameters used and the reuse of feature maps. Our proposal represents the first attempt in the literature to accurately localize and interpret the different anatomical components of the CA (lumen, far wall, near wall, bulb, CIM region and CIM-bulb region, see Figure 1), which can be helpful in the proper estimation of the CIMT. Using the segmented region, we define a straightforward approach for CIMT estimation and plaque detection.

Moreover, the majority of the proposed techniques in the literature restrict their application to five particular conditions of the CA images and data sets, which are summarized in the columns 7-11 in Table 1 and are explained below.

1) Most of the presented works and reference values from the guidelines focus only on Common Carotid Artery (CCA) images. The image quality of other territories, such as Bulb, is worse than CCA (poorer contrast and more affected by noise). Also, successful imaging depends on the subjects anatomy. These facts make the segmentation of the CIM region in Bulb difficult. None of the revised methods deal with Bulb images (see column named "Artery Territory" in Table 1). However, we demonstrate that the method proposed in our paper is easily extensible to this different CA territory, after being successfully trained for both CCA and Bulb.

2) In the non-plaque images (i.e. images in which the plaque does not appear), the CIM region is observed as a straight thin shape, whereas the presence of plaque leads to a focal thickening of the CIM region, resulting in an irreg-

ular shape (see Figure 2). The shape variability of the CIM region makes the definition of a robust segmentation method more difficult. As a consequence, most of the previous works only measure the CIMT within plaque free regions and discard images with the presence of plaque (see column named "Presence of Plaque" in Table 1). Unlike most previous works, we broaden the target and build a more general method able to accurately estimate the CIMT, even in the presence of plaque. This feature makes our method useful for data sets of population studies, such as the one considered in this paper. Moreover, the presence of plaque in the data set allows us to evaluate the plaque detection of our method.

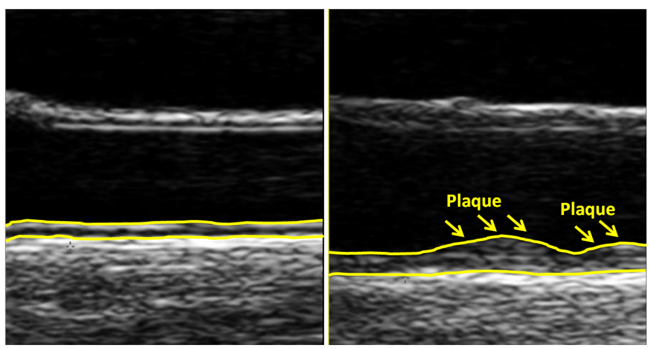


Figure 2: US images from CCA without plaque (left) and with atherosclerotic plaque (right).

3) In terms of the number of images processed, the size of the considered data sets in the previous CIMT estimation studies is quite small (see column named "N" in Table 1). Although these sample sizes guarantee an adequate level of study power, a large-scale study —such as the one presented in our paper— is required to carefully assess the effect of variability on segmentation performance, and also to evaluate the systems before their application in the real praxis. In particular, we show an extensive evaluation of the CIMT measurement and plaque detection in a large data set (REGICOR), which contains 8,484 images.

4) The different devices and settings used for image acquisition provide data sets with different image characteristics. These differences imply difficulties for the robust segmentation of CA components and CIMT estimation. For

this reason, most of the methods in the literature use data sets provided by a single device (see column named “Different Acquisition Devices” in Table 1). In contrast, we validate the robustness and generalization power of our method by applying it to the NEFRONA data set, which contains images provided by different equipment (see Section 3.1).

5) Regarding the validation procedure, we extensively evaluate our proposals. We compare the obtained CIMT estimation with other state-of-the-art approaches to demonstrate the outperformance of the proposed method (see column “Mean CIMT Error (mm)” in Table 1). Moreover, we compare the CIM segmentation results with other approaches and we measure the Inter-Observer Variability (IOV) of the manual segmentation showing the degree of difficulty of the problem at hand, especially in the case of Bulb images (see Section 3). Lastly we evaluate plaque detection in the large data set, REGICOR, for which we obtain very promising.

This paper is structured as follows: the current section introduces the problem, exposes the related work and details the contributions of the paper. In Section 2, we present the proposed CIM region segmentation method, the CIMT estimation approach and the plaque detection method. The used data sets and the results obtained are presented in Section 3. Finally, Section 4 closes the paper with conclusions and future challenges.

2. Methodology

This work proposes a method for automatic CA image interpretation that integrates semantic segmentation with other image analysis techniques for CIMT estimation. Figure 3 depicts the workflow of our approach, subsequently explained in depth.

2.1. Semantic Segmentation

In our research, CA segmentation is about solving the problem of separating the different anatomical components of the CA (i.e. lumen, far wall, near wall,

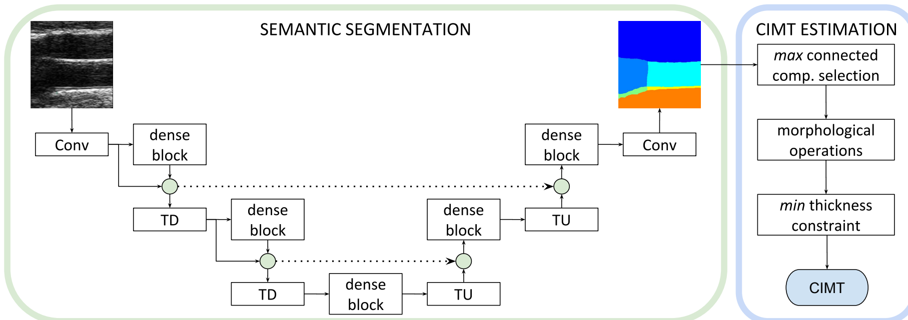
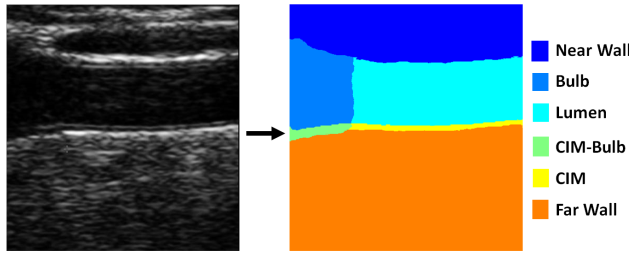


Figure 3: Workflow of the proposed method for semantic carotid artery segmentation and CIMT estimation. The SS model is composed of a down-sampling path with Transition Down (TD) blocks, and an up-sampling path with Transition Up (TU) blocks, both including dense blocks that create the feature maps. A Convolution (Conv) is applied at the input of the network as well as at the end, to generate the final segmentation. The small circles represent concatenations, and the dotted arrows are the skip connections.

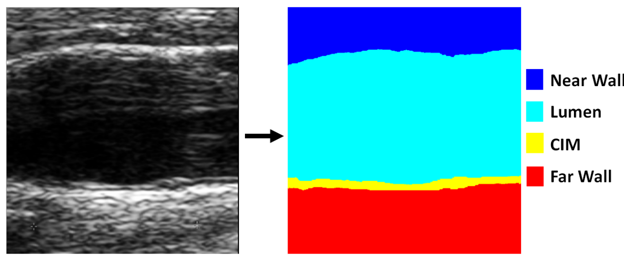
bulb, CIM region, and CIM-bulb region, see Section 1), thus obtaining a mask with six or four different labels, depending on whether CCA or Bulb images are being analyzed, respectively. For this purpose, we propose the use of semantic segmentation (SS) algorithms that work in a supervised learning framework, instead of using image features such as shapes or pixel-based features.

Fully Convolutional Networks (FCN) [14], commonly used in SS problems, are a particular case of Convolutional Neural Networks (CNN) that do not use fully-connected layers. They take an image of any size as input data and transform it to obtain a segmented image, with the same spatial resolution, by means of an inference, learning process. Figure 4 shows an example of two CA images (inputs to the SS model) and their corresponding segmented images (expected outputs of the SS model).

Any CNN model can be extended to be used as FCNs and so applied to a SS problem. From the state-of-the-art architectures, we have selected Densely Connected Convolutional Networks (DenseNets) [13], an extension of the well-known Residual Networks (ResNets) [15]. DenseNets has been designed to ease the training of very deep networks, and present some characteristics that make them very appropriate for SS: parameter efficiency, implicit deep supervision, and feature reuse.



(a) CCA



(b) Bulb

Figure 4: Example of the input (left) and expected output (right) of the SS model for images of both territories: (a) CCA and (b) Bulb. The legend at right details the segmentation labels.

As a result of all of these reasons, we have considered the so-called Tiramisu [16], an extension of DenseNets such as FCNs, to solve the CA segmentation problem. The Tiramisu architecture (see Figure 3, left) is composed of a down-sampling path with transition down (TD) blocks to extract coarse semantic features, and an up-sampling path with transition up (TU) blocks to recover the input image resolution at the output level. Both paths are connected by means of skip connections that allow the recovery of fine-grained information, and they are defined by a sequence of dense blocks that contain a set of concatenated layers, as proposed in DenseNets. The three types of blocks used in the Tiramisu model are defined as follows:

- Dense blocks are composed of concatenated layers that include Batch Normalization [17], Rectified Linear Unit [18], 3×3 convolution, and Dropout [19] (probability 0.2).

- TD blocks are composed of Batch Normalization, Rectified Linear Unit,

1×1 convolution, Dropout (probability 0.2) and 2×2 max-pooling (stride 2).

- TU blocks are composed of 3×3 transposed convolution (stride 2).

Our implementation of the semantic segmentation model is in Keras[1], with Theano as backend, and is publicly available for download[2].

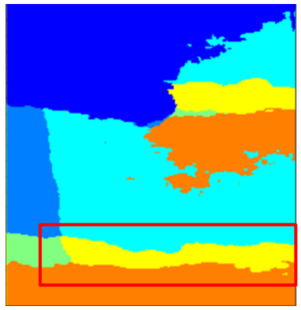
2.2. CIMT estimation and Plaque Detection

The output of the semantic segmentation process is a mask divided in different regions (see Figure 4: six for CCA images, and four for Bulb images). The information provided by the different regions identified in the mask are used to estimate the CIMT, following the next procedure (partially illustrated in Figure 5):

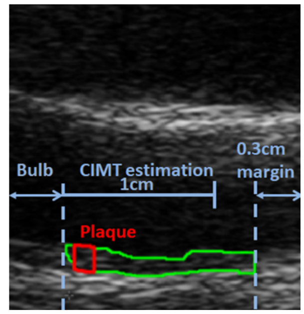
1. The biggest connected component, corresponding to the CIM label, is identified (Figure 5(a)). In the case that the two biggest connected components have a similar size, we select the largest one that is more similar to the rectangular shape of the CIM region.
2. The borders of the CIM region are smoothed with basic morphological operations. In particular, these operations are *opening*, to remove small objects; and *closing*, to avoid small holes. Rectangles are used as structuring elements for these operations, with dimensions 4×8 for closing and 2×25 for opening.
3. According to the experience of technicians, image quality is not good at the ends of the image (approximately 0.3 cm in each side). For this reason, we define a margin of 0.3 cm in the right part of the CCA images (see Figure 5(b)), and two margins of 0.3 cm in the right and left parts of the Bulb images (see Figure 5(d)). Moreover, the mean values from CIMT in

[1] https://keras.io/

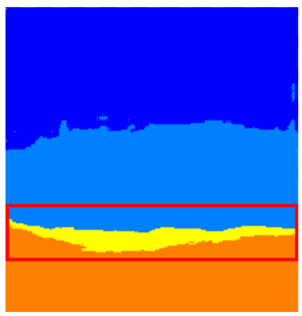
[2] https://github.com/beareme/keras_semantic_segmentation



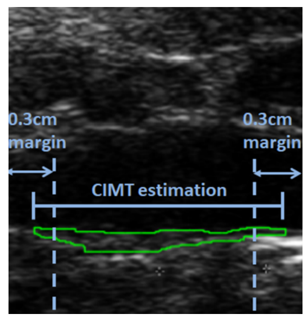
(a) Mask from a CCA image



(b) CIMT estimation in CCA



(c) Mask from a Bulb image



(d) CIMT estimation in Bulb

Figure 5: Representative example of the CIMT estimation procedure for CCA (top) and Bulb images (bottom). At left, masks obtained from the semantic segmentation model (yellow pixels correspond to CIM region label); and the biggest, largest connected component selected as CIM region (red rectangle). At right, the CIM region obtained from the semantic segmentation result (in green). (b) Left margin used to discard pixels with CIMT value greater than 1.5 mm, CIMT estimation area (1 cm after the bulb), and plaque region (in red). (d) Right and left margins are used to discard pixels with the CIMT value lower than 0.4 mm, and the CIMT estimation area.

CCA are, in general, between 0.4 mm and 1.5 mm [20]. Based on this, we discard the pixels of the CIM region that are within the lateral margins, and have a CIMT value outside the range [0.4, 1.5] mm.

4. Once the CIM region is obtained, we divide the CIM region in vertical lines (each line corresponds to one pixel). For each vertical line, the absolute distance between the two borders is considered. Finally, we compute the CIMT measurement as the mean from all these values.

For CCA images the CIMT is estimated 1 cm distal from the Bulb, justified from a clinical standpoint [1] (see Figure 5(b)).

5. Afterwards, each image is classified as containing plaque or non-plaque, using the CIMT measurement and following the Mannheim Consensus (see Section 1).

3. Experiments

3.1. Data Set

In this work, we consider two different data sets: REGICOR and NEFRONA. REGICOR[3] consists of a sample of 2,379 subjects from *Girona's Heart Registry* [20]. The images were collected from 2007 to 2010, and the subjects represent general population aged between 35 and 84. Two trained sonographers performed the CA US scans with an Acuson XP128 US system equipped with L75-10 MHz transducer and a computer program extended frequency (Siemens-Acuson). US longitudinal images were obtained in B-mode with resolution 23.5 pixels/mm. The original images were saved in DICOM format and then converted to PNG. The set of images collected for each patient were obtained from left and right CA in two different territories (CCA and Bulb), resulting in a total of 8,484 images (4,751 CCA images, and 3,733 Bulb images). The CIMT reference values, given by the Amsterdam Medical Center[4] (AMC), were used as the Ground-Truth (GT) for the CIMT estimation. Note that all the images were analyzed by an AMC expert using the semi-automatic validated software e-track [21]. Regarding the GT for plaque detection, it was obtained using the provided CIMT reference values and applying the Mannheim consensus. Furthermore, the images containing plaque were finally supervised by an expert.

Besides the GT for CIMT estimation and plaque detection, a segmentation GT was defined for a subset of the REGICOR images. In order to obtain it, an expert (Expert1) manually delineated and labeled the different regions of the original images, using six labels for CCA and four for Bulb (written in red in

[3]https://www.regicor.org/en_index.html

[4]https://www.abc.uva.nl/research/institutes/institute-articles/academic-medical-center-amc.html

Figure 1). Since this manual task is difficult and very time-consuming, only a representative subset of REGICOR images was labeled, including 159 CCA images (51 with plaque and 108 without plaque), and 172 Bulb images (68 with plaque and 104 without). The training set contains 141 images for the CCA and 155 images for the Bulb, whilst the rest of them were used for testing. The test images were used for the comparison of the segmentation approaches presented in Section 3.3. Additionally, the test images were manually segmented by a second expert (Expert2) to measure the IOV.

The second data set, NEFRONA[5], from Atherotrombotic Diseases Unit Detection Hospital Arnau de Vilanova, consists of a collection of B-Mode US of the CA obtained by a Vivid BT09 device (from General Electric), with a 6-13 MHz band. For each subject of the study CCA images were captured. This data set is formed by 27 images with the corresponding CIM regions and their CIMT values (NEFRONA GT), provided by the General Electric device.

Note that data from both data sets, REGICOR and NEFRONA, can be requested to the corresponding contacts.

3.2. Validation Setup

This section includes the different experiments carried out to validate our approach results, which are summarized in Table 2 and following described in depth.

Experiment 1: Segmentation. In order to validate the proposed segmentation method, we compared six different approaches applied to a subset of the REGICOR dataset: four DenseNets models based on Tiramisu, the U-Net method [22], and a two-step approach based on the shallow method Random Forest (RF). Regarding the Tiramisu model, we have considered two different configurations varying the depth of the network: Tiramisu56 (a total of 56 layers, 4 per dense block) and Tiramisu103 (a total of 103 layers, from 4 to 12 per block). In order to show if the SS of several anatomical components helps in the CIM

[5]http://www.nefrona.es/

Experiment 1: Segmentation
Purpose: comparison of different segmentation approaches
Data set: subset of REGICOR. GT: manually segmented images
images: 159 (CCA), 172 (Bulb). Train/test split: $\approx 90\% - 10\%$
Performance measures: accuracy, specificity, sensitivity, precision, Dice coefficient
Experiment 2: CIMT estimation
Purpose: comparison of different methods for CIMT estimation
Data set: REGICOR. GT: CIMT values
images: 8,484 (all of them used for validation)
Error measurement: correlation coefficient and Bland-Altman analysis
Experiment 3: Plaque detection
Purpose: comparison of different methods for plaque detection
Data set: REGICOR. GT: plaque detection (yes/no)
images: 8,484 (all of them used for validation)
Performance measures: accuracy, specificity, sensitivity
Experiment 4: Generalization power
Purpose: assessment of the generalization power of the proposed method
Data set: NEFRONA. GT: CIMT values
images: 27 (all of them used for validation)
Error measurement: correlation coefficient and Bland-Altman analysis

Table 2: Summary of the different experiments carried out for validation purposes.

region segmentation, we also compared the results provided by the two Tiramisu models (Tiramisu56 and Tiramisu103), but using only two labels (CIM region and *background*). We called this second approach Binary Segmentation (BS), whilst the one with all the labels is referred as Semantic Segmentation (SS). Notice that both approaches, BS and SS, were compared by considering two labels in the evaluation measure. In order to demonstrate the adequacy of using DenseNets, the U-Net was also considered in the experimentation. In this sense, it is worthy to point out that the main difference between U-Net and Tiramisu is that U-Net uses standard convolutions instead of the dense blocks proposed in the DenseNet architecture. Finally, in order to compare the NNs with classical methods, we have also considered a two-step approach based on RF. Particularly, we refer as RF2 to the two-step approach in which a ROI is first automatically extracted (pre-processing) and then a patch-based RF (multi-class) is used for pixel-wise classification. In this case, a post-processing specifically designed for this method [8] can be applied, which is referred as RF2-PP.

All the NN models were trained using a GeForce Titan X (Pascal) 12GB GPU from NVIDIA. The models' weights were initialized using the HeUniform initialization [23], and the RMSprop algorithm [24] was used as optimizer. The training process was carried out in two steps, as in [16]: 1) pre-training with random cropping for data augmentation (crop dimension: 224×224 px), learning rate $1e-3$, and batch size 3; 2) fine-tuning with full size images (image dimension: 470×445 px), learning rate $1e-4$ and batch size 1. The outputs were monitored using the pixel-wise accuracy and the dice coefficient, with a patience of 100 during pre-training and 50 during fine-tuning.

A complete set of measures was used to evaluate the performance of the different segmentation models. All of them are defined as follows, considering CIM region (positive) and Background (negative), and using the terms true positive (TP), true negative (TN) false positive (FP), and false negative (FN).

- The pixel-wise accuracy, i.e. the percentage of pixels correctly classified.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

- Specificity, i.e. the proportion of negatives correctly classified.

$$Spec = \frac{TN}{TN + FP}$$

- Sensitivity, i.e. the proportion of positives correctly classified.

$$Sens = \frac{TP}{TP + FN}$$

- Precision, i.e. the proportion of true positives against all the positives.

$$Prec = \frac{TP}{TP + FP}$$

- Dice coefficient, i.e. the similarity over classes.

$$DC = \frac{2TP}{2TP + FP + FN}$$

Experiment 2: CIMT estimation. With the aim of evaluating our method in terms of CIMT estimation over the REGICOR dataset, we have considered the correlation coefficient (cc) between the GT and the predicted CIMT values as well as the Bland-Altman analysis. For a deep comparison, we have considered not only the methods used in the *Experiment 1* (Tiramisu56, Tiramisu103 and RF2-PP), but also other approaches found in the literature (see Section 1.1).

Experiment 3: Plaque detection. The target is to evaluate our method in terms of plaque detection over the REGICOR dataset, including a comparison with the two-step approaches (RF2 and RF2-PP). For this purpose, we have used the following metrics, previously defined, considering the presence of plaque as positive and the absence of plaque as negative: Accuracy (Acc), Specificity (Spec), and sensitivity (Sens).

Experiment 4: Generalization power. To validate the generalization power of our method, we trained it with the subset of REGICOR used in the *Experiment 1* and evaluate its performance in terms of CIMT estimation over the NEFRONA dataset. Images from the two data sets were acquired by different devices, thus, they have different resolutions and image intensity distributions. Hence, we process the data to equate the intensity distribution of all the images and adapt the resolution. In first place, we modify the image gray levels to saturate the bottom 1% and the top 1% of all the image pixels in the two data sets. Next, we transform NEFRONA images so that they have the same resolution than REGICOR images; more precisely, from a resolution of 10.4 pixels/mm (NEFRONA) to 23.5 pixels/mm (REGICOR). In order to do that, we apply a bilinear interpolation, in which the output pixel value is a weighted average of pixels in the nearest 2-by-2 neighborhood. The CIM region of NEFRONA GT was delineated only in a small part of the image and following a different criterion than in REGICOR. For this reason, the validation of the segmentation can only be qualitative. Regarding the validation of the CIMT estimation, we consider the correlation coefficient (cc) between CIMT value from NEFRONA data set GT and the estimated CIMT, and also show the Bland-Altman analysis.

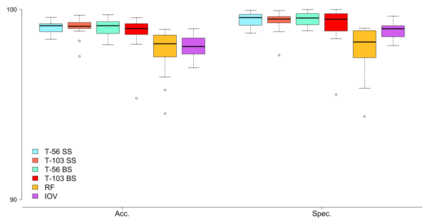
3.3. Results

In this section we report the results obtained in the four experiments previously described, summarized in Table 2.

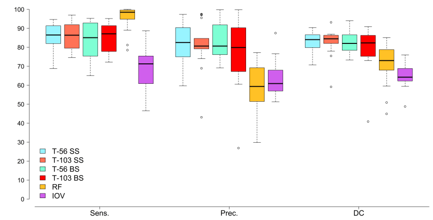
Experiment 1: Segmentation. Figure 6 depicts the comparison between the different segmentation approaches in CCA (left) and Bulb (right) test images. It can be seen that the different Tiramisu architectures clearly improve the RF2 results (mainly note improvement in DC). Moreover, making the Tiramisu model deeper by increasing the number of parameters (from 56 to 103) does not improve the results, probably due to the size of the training set. Although the BS is equivalent to SS in CCA images, the semantic information is crucial for the CIMT estimation step in these images (see Section 2). Note that the improvement using SS is more evident in Bulb images. Regarding U-Net, its results are slightly worse than Tiramisu103 BS and are not included in the graphic. Finally, the IOV results (considering Expert1 as GT, versus Expert2) are low compared with the automatic methods results, specially in Sensitivity and DC, in both CCA and Bulb images. These results and the high standard deviations show the difficulty of reproducing the CA results in clinical trials. It is worth noting that all the measures have been computed using the Expert1's labels as GT, but the values are equivalent for the labels of Expert2.

Figure 7 shows qualitative examples of the CIM segmentation results regarding three methods: a shallow method (RF) and two methods based on CNN (U-Net method and Tiramisu56 method). As can be seen, U-Net does not give an accurate result of the different areas of the image and RF oversegments the CIM region.

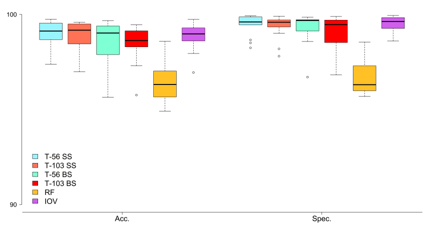
Experiment 2: CIMT estimation. Figure 8(a) shows the correlation between the CIMT values (GT and predicted) in CCA images for the best method, i.e. "Tiramisu56 SS+CIMT estimation", which reaches a high cc of 0.81 (cc=0.77 when applying only Tiramisu56 SS). The result is very similar to Tiramisu103 (cc=0.80), in contrast to RF2-PP (cc=0.72). Regarding Bulb images (see Figure 8(c)), "Tiramisu56 SS+CIMT estimation" achieves a lower cc of 0.43 (cc=0.34



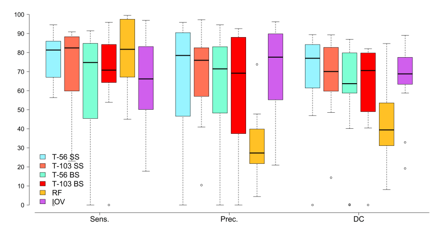
(a) Accuracy and Specificity measurements for CCA images



(b) Sensitivity, Precision and Dice Coefficient measurements for CCA images



(c) Accuracy and Specificity measurements for Bulb images



(d) Sensitivity, Precision and Dice Coefficient measurements for Bulb images

Figure 6: Box-plot of metrics results for the different segmentation methods and IOV. Note that the overlap measurements are split up for visualization purposes, using different scales in the abscissa axis.

when applying only Tiramisu56 SS), probably due to the worse quality of the images in Bulb, which makes the task more difficult in this territory. However, our proposal still outperforms RF2-PP, which only reaches a cc of 0.41.

In Figure 8(b), the Bland-Altman plot depicts the difference, in CCA images, between the CIMT of the corresponding two values (estimated and GT) against the average of both values. This plot shows a high degree of agreement between the two measures, especially in the cases where the CIMT is small ($<$0.5mm), which correspond to healthy population (i.e. without plaque) [20]. Furthermore, this plot shows that the predicted CIMT is, on average, slightly underestimated (mean -0.02). The confidence intervals for the "mean of the differences line" (shown in red in Figure 8) shows that this bias is statistically significant. Therefore, in order to achieve the interchangeability of the techniques this bias cannot be avoided.

(a) Tiramisu56 (b) U-Net (c) RF

Figure 7: Qualitative results of the semantic segmentation procedure using three different methods.

The results are similar for Bland-Altman analysis in Bulb images (see Figure 8(d)) and, in this case, the average slightly overestimates the CIMT measure (the mean of the differences is 0.06 and this bias is also statistically significant). Column named "Mean CIMT Error (mm)" in Table 1 compares the mean CIMT error for our method and several methods in the literature. It should be highlighted that our CIMT error is low compared with other fully automatic methods reviewed in the Table. In particular, only the two-step methods [4, 11] reach a CIMT error lower than our method, but in a much smaller data set and only in one territory (CCA). In fact, the size of our data set is much larger than the ones considered in all the rest of papers (our data set: 2,379 subj. vs revised data sets: [36-365] subj.). Note that, as can be seen in this column of the Table, the CIMT error is not always presented as the mean of the CIMT error, in some cases it is presented with point-to-point relative error, average point-to-point distance, or evaluating the mean error for each interface separately.

Figure 9 shows qualitative examples of the CIM segmentation results and plaque detection for four CCA and four Bulb images. The first and third columns show examples of CIM region segmentation, outlined in green, in non-plaque images; whereas the second and the fourth columns show examples of images with plaque, outlined in red.

Finally, it is important to note that the processing time to estimate the CIMT and detect a plaque is only 0.79 seconds, as can be also seen in Table 1

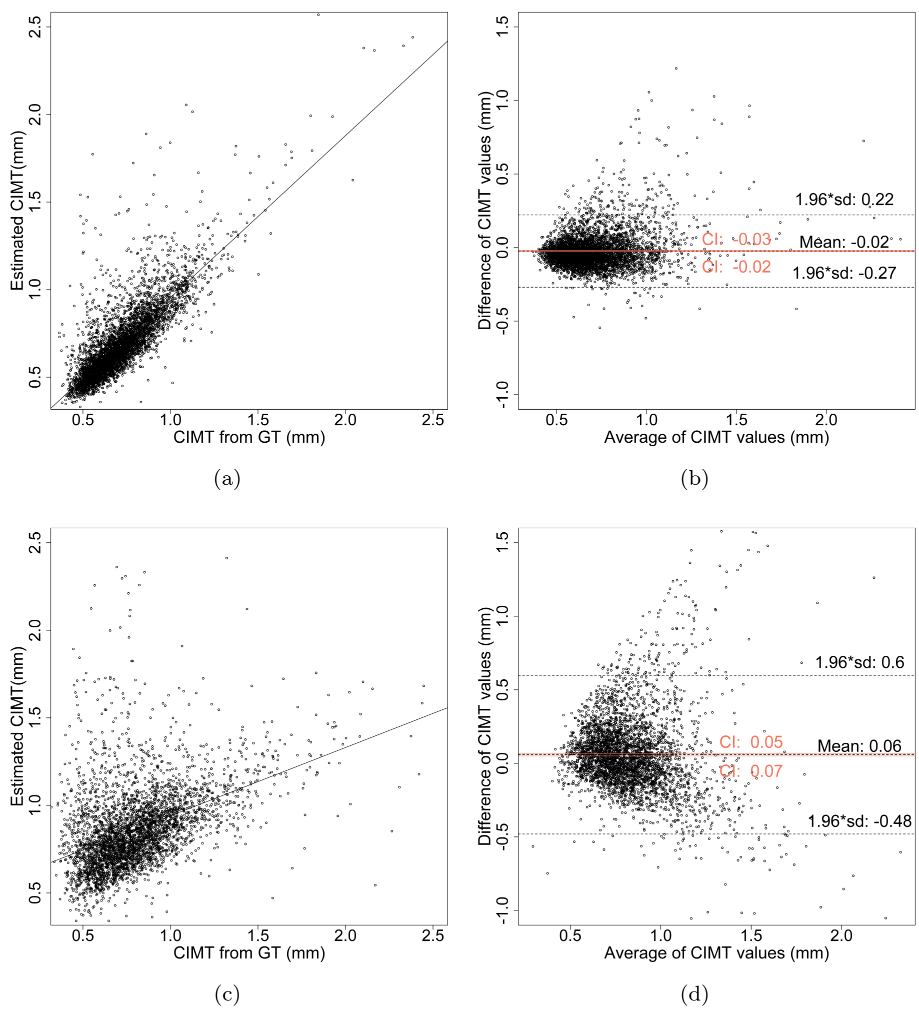
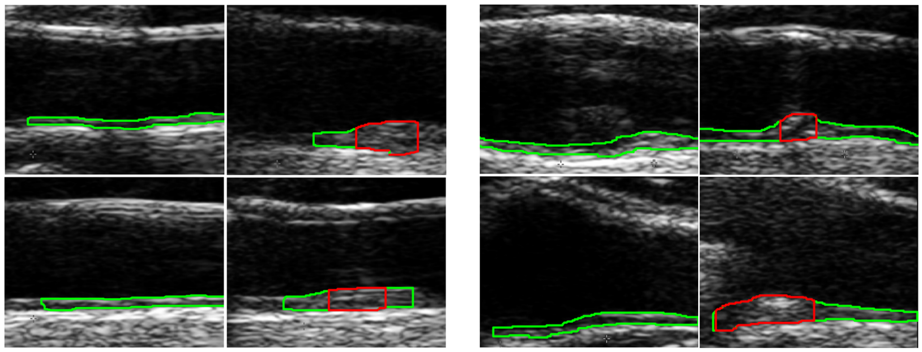


Figure 8: Correlation between CIMT values (left), and Bland-Altman analysis (right). Both plots show the relation between GT and the estimated values in CCA images, (a) and (b); and in Bulb images, (c) and (d). Red solid lines show the confidence intervals (CI) for the "mean of the differences" line.

(column “Proc. Time per Frame”).



(a) CCA images

(b) Bulb images

Figure 9: Qualitative results of the CIM segmentation for eight different images. Green lines are the CIM region boundaries and red lines the detected plaque boundaries. Images are cropped for visualization purpose.

Experiment 3: Plaque detection. Table 3 includes the plaque detection results in CCA and Bulb images, showing a promising performance, mostly in CCA. The smaller number of plaques in the data set gives lower sensitivity values than specificity values. Regarding Bulb images, there is still large room for improvement, probably due to the poorer quality of these images, as commented before. Note that the RF2 method needs a sophisticated post-processing to achieve similar results to our NN method. Figure 9 shows qualitative examples of the plaque detection results.

Territory Images	Method	# Plaques/ Total images	Acc	Sens	Spec
CCA	RF2	50/4, 722	50.05%	100.00%	49.00%
	RF2-PP	50/4, 722	94.08%	86.00%	94.16%
	Our proposal	50/4,751	96.45%	80.00%	96.63%
Bulb	RF2	240/3, 539	35.09%	98.33%	30.49%
	RF2-PP	240/3, 539	78.50%	69.58%	79.15%
	Our proposal	264/3,733	78.09%	78.32%	75.00%

Table 3: Results of plaque detection in REGICOR images for different methods, the number of plaques in each territory and the following validation measures: Accuracy (Acc), Sensitivity (Sens), and Specificity (Spec).

Experiment 4: Generalization power. Figure 10 illustrates qualitative results of the segmentation method in some NEFRONA images. It shows the CIM region segmentation result (in green) together with the CIM region from NEFRONA GT (in yellow). We can observe that, generally, the CIM region is slightly over-segmented. According to this, Figure 11 (right) shows an overestimation of the CIMT in the Bland-Altman plot (mean 0.29, note that the bias is statistically significant). Despite this error, Figure 11 (left) shows that the obtained values have a good correspondence with the CIMT values of the NEFRONA database, with a cc of 0.58.

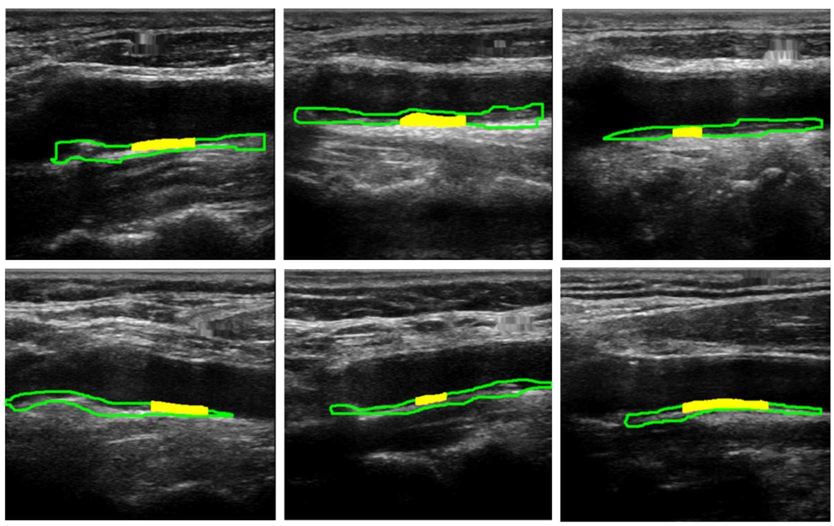


Figure 10: Qualitative segmentation results for NEFRONA CCA images. In green, delimitation of CIM region segmentation. In yellow, the CIM region from NEFRONA GT.

4. Conclusions and Future Work

In this paper, we have presented, for the first time in the literature, a single-step approach, based on DenseNets, for semantic CA segmentation. The proposed method accurately localizes the CIM region in CCA. Given the segmentation, we have validated the CIMT estimation and the detection of atherosclerotic plaque with a large data set of more than 8,000 images. We have compared the results obtained by the proposed method with those of other DL models and

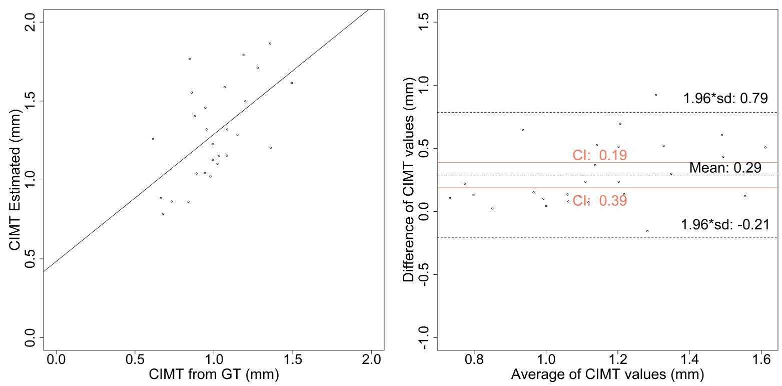


Figure 11: Correlation between CIMT values (left), and Bland-Altman analysis (right). Both plots show the relation between GT and the estimated values in CCA images from NEFRONA data set. Red solid line shows the confidence intervals for the "mean of the differences" line.

shallow approaches, demonstrating more accurate results of the segmentation, more general CIMT measurement and good plaque detection results. This superior performance is attributed to the effective use of SS together with the CIMT estimation approach. Moreover, we have proven the generalization capability of the method applying the model previously trained with one data set (REGICOR) in a new test data set (NEFRONA).

The proposed study has some limitations that are summarized below, and that will be considered in our future work. These limitations mostly arise from the number of images used in some of the experiments,thus the increase in the size of some datasets constitutes the first point of improvement in our study. On the one hand, the segmentation GT only includes a representative subset of the REGICOR images (159 CCA, 172 Bulb). On the other hand, the generalization power test was carried out using a small dataset composed of only 27 images (NEFRONA). Additionally, the proposed method is not applied on image sequences, which could improve reliability by measuring hundreds of images for each subject. Regarding CIMT estimation, we propose a pre-processing step that uses a smoothing algorithm based on mathematical morphology. Taking into account the unpredictable effect of this type of algorithms on segmentation results, a more detailed study is required to evaluate the impact of our proposed algorithm and to compare it with other pre-processing techniques. In this part of the methodology, it is also worth noting that the criteria of considering CIMT values higher than 0.4 mm (see Section 2.2) could exclude real cases with a low CIMT. Finally, the division of the CIM region in vertical columns could overestimate the CIMT values in case of oblique forms of the CA; thus, this methodological issue could be carefully addressed as suggested by Bianchini et al. [25].

Additionally, we want to further improve the segmentation results in terms of an adequate generalization to other data sets, by exploring new domain transfer techniques. We also plan to add information indicating the presence of plaque into the neural network in a way that it can learn the differences in shape between images of healthy subjects (thin CIM region shape) and images of subjects with atherosclerosis (irregular CIM region shape).

Acknowledgments

This work was partially supported by: the Spanish Ministry of Economy and Competitiveness through the Instituto de Salud Carlos III-FEDER (CIBERCV and FIS CPII17/00012), the Spanish Ministry of the Economy and Competitiveness (TIN2015-65069-C2-2-R and TIN2015-66951-C2), the Principado de Asturias Regional Government (IDI-2018-000176), all grants through the ERDF; and the Catalan Agència de Gestió d'Ajuts Universitaris de Recerca (2017-SGR-222 and 2017-SGR-1742).

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU. Finally, the authors would also like to acknowledge Virtudes Maria from UDETMA for her support in the CA image labeling.

References

[1] P.-J. Touboul, M. Hennerici, S. Meairs, H. Adams, P. Amarenco, N. Bornstein, L. Csiba, M. Desvarieux, S. Ebrahim, R. H. Hernandez, et al.,

Mannheim carotid intima-media thickness and plaque consensus (2004–2006–2011), Cerebrovascular Diseases 34 (4) (2012) 290–296.

[2] F. Molinari, C. S. Pattichis, G. Zeng, L. Saba, U. R. Acharya, R. Sanfilippo, A. Nicolaides, J. S. Suri, Completely automated multiresolution edge snapper—a new technique for an accurate carotid ultrasound imt measurement: clinical validation and benchmarking on a multi-institutional database, IEEE Transactions on Image Processing 21 (3) (2012) 1211–1222.

[3] F. Faita, V. Gemignani, E. Bianchini, C. Giannarelli, L. Ghiadoni, M. Demi, Real-time measurement system for evaluation of the carotid intima-media thickness with a robust edge operator, Journal of Ultrasound in Medicine 27 (9) (2008) 1353–1361.

[4] M. C. Bastida-Jumilla, R.-M. Menchón-Lara, J. Morales-Sánchez, R. Verdú-Monedero, J. Larrey-Ruiz, J.-L. Sancho-Gómez, Frequency-domain active contours solution to evaluate intima–media thickness of the common carotid artery, Biomedical Signal Processing and Control 16 (2015) 68–79.

[5] C. P. Loizou, T. Kasparis, C. Spyrou, M. Pantziaris, Integrated system for the complete segmentation of the common carotid artery bifurcation in ultrasound images, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, 2013, pp. 292–301.

[6] S. Zhao, Z. Gao, H. Zhang, Y. Xie, J. Luo, D. Ghista, Z. Wei, X. Bi, H. Xiong, C. Xu, S. Li, Robust segmentation of intima-media borders with different morphologies and dynamics during the cardiac cycle, IEEE Journal of Biomedical and Health Informatics 22 (5) (2017) 1571–1582.

[7] C. Qian, X. Yang, An integrated method for atherosclerotic carotid plaque segmentation in ultrasound image, Computer Methods and Programs in Biomedicine 153 (2018) 19–32.

[8] C. Zhang, M. M. Vila, P. Radeva, R. Elosua, M. Grau, A. Betriu, E. Fernandez-Giraldez, L. Igual, Carotid artery segmentation in ultrasound images, in: CVII-STENT: Computing and Visualization for Intravascular Imaging and Computer Assisted Stenting in conjunction with MICCAI, 2015.

[9] F. Molinari, G. Zeng, J. S. Suri, A state of the art review on intima–media thickness (IMT) measurement and wall segmentation techniques for carotid ultrasound, Computer Methods and Programs in Biomedicine 100 (3) (2010) 201–221.

[10] C. P. Loizou, A review of ultrasound common carotid artery image and video segmentation techniques, Medical & Biological Engineering & Computing 52 (12) (2014) 1073–1093.

[11] R.-M. Menchón-Lara, J.-L. Sancho-Gómez, Fully automatic segmentation of ultrasound common carotid artery images based on machine learning, Neurocomputing 151 (Part 1) (2015) 161–167.

[12] J. Y. Shin, N. Tajbakhsh, R. T. Hurst, C. B. Kendall, J. Liang, Automating carotid intima-media thickness video interpretation with convolutional neural networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2526–2535.

[13] G. Huang, Z. Liu, K. Q. Weinberger, L. van der Maaten, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[14] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (4) (2017) 640–651.

[15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[16] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, The One Hundred Layers Tiramisu: Fully convolutional DenseNets for Semantic Segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 1175–1183.

[17] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: International Conference on Machine Learning, 2015, pp. 448–456.

[18] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: 27th International Conference on Machine Learning, 2010, pp. 807–814.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research 15 (1) (2014) 1929–1958.

[20] M. Grau, I. Subirana, D. Agis, R. Ramos, et al., Grosor íntima-media carotídeo en población española: valores de referencia y asociación con los factores de riesgo cardiovascular, Revista Española de Cardiología 65 (12) (2012) 1086–1093.

[21] E. de Groot, G. K. Hovingh, A. Wiegman, P. Duriez, A. J. Smit, J.-C. Fruchart, J. J. Kastelein, Measurement of arterial wall thickness as a surrogate marker for atherosclerosis, Circulation 109 (23 suppl 1) (2004) III–33.

[22] O. Ronneberger, P. Fischer, 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.

[23] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.

[24] T. Tieleman, G. Hinton, RMSProp adaptive learning, COURSERA: Neural Networks for Machine Learning.

[25] Functional and Structural Alterations of Large arteries: Methodological Issues, Current Pharmaceutical Design 19 (13) (2013) 2390–2400.