

DCNN-Based Automatic Segmentation and Quantification of Aortic Thrombus Volume: Influence of the Training Approach

Karen López-Linares^{1,2,4(✉)}, Luis Kabongo^{1,4}, Nerea Lete^{1,4},
Gregory Maclair^{1,4}, Mario Ceresa², Ainhoa García-Familiar³, Iván Macía^{1,4},
and Miguel Ángel González Ballester^{2,5}

¹ Vicomtech-IK4 Foundation, San Sebastián, Spain
klopez@vicomtech.org

² Universitat Pompeu Fabra, Barcelona, Spain

³ Hospital Universitario Donostia, San Sebastián, Spain

⁴ Biodonostia Health Research Institute, San Sebastián, Spain

⁵ ICREA, Barcelona, Spain

Abstract. Computerized Tomography Angiography (CTA) based assessment of Abdominal Aortic Aneurysms (AAA) treated with Endovascular Aneurysm Repair (EVAR) is essential during follow-up to evaluate the progress of the patient along time, comparing it to the pre-operative situation, and to detect complications. In this context, accurate assessment of the aneurysm or thrombus volume pre- and post-operatively is required. However, a quantifiable and trustworthy evaluation is hindered by the lack of automatic, robust and reproducible thrombus segmentation algorithms. We propose an automatic pipeline for thrombus volume assessment, starting from its segmentation based on a Deep Convolutional Neural Network (DCNN) both pre-operatively and post-operatively. The aim is to investigate several training approaches to evaluate their influence in the thrombus volume characterization.

Keywords: AAA · EVAR · Thrombus · Segmentation · DCNN · Volume

1 Introduction

An abdominal aortic aneurysm (AAA) is a focal dilation of the aorta that exceeds its normal diameter in more than 50%. If not treated, it tends to grow and may rupture, with a high mortality rate [1]. Lately, AAA treatment has shifted from open surgery to a minimally invasive alternative, known as Endovascular Aneurysm Repair (EVAR) [2]. This technique consists in the transfemoral insertion and deployment of a stent using a catheter. Although better peri-operative outcomes are achieved, long-term studies show similar mortality rates between patients treated with EVAR and patients treated with open surgery [3]. This is due to the appearance of EVAR related complications, known as endoleaks,

which translate into a recurrent blood-flow into the thrombus area that causes its continuous growing, with the associated rupture risk and possible re-intervention. Thus, post-operative surveillance is required to evaluate changes and detect possible complications, for which Computed Tomography Angiography (CTA) is the preferred imaging modality. This follow-up is traditionally based on the observation of CTA scans at different times and the manual measurement of the maximum aneurysm diameter, although AAA volume has been reported as a better predictor of the disease progression [4]. In [5], a fully-automatic thrombus segmentation approach based on a Deep Convolutional Neural Network (DCNN) was proposed, specifically designed for post-operative thrombus segmentation. Our aim is to extend that work by providing segmentation for both pre-operative and post-operative scenarios and to provide a full pipeline for thrombus volume assessment, investigating the influence of network training strategies in the automatic segmentation quality and volume quantification.

2 State-of-the-art

Historically, aneurysm size, measured through its largest diameter, has been the most commonly employed rupture risk indicator. This evaluation is done both pre-operatively, to determine if an intervention is required, and post-operatively, to assess the patient’s progression. Thrombus volume seems to be a better rupture risk indicator [4], but it is hardly used in the clinical practice due to the lack of automatic thrombus segmentation methods. The thrombus appears as a non-contrasted structure in the CTA, its shape varies and its borders are fuzzy, which makes it difficult to develop robust automatic segmentation approaches. Thus, the subsequent precise and automatic thrombus characterization is unfeasible.

Currently there are only few dedicated software that provide assistance to EVAR-treated aneurysm follow-up: VesselIQ Xpress (GE) [6] and Vitrea Imaging (Toshiba) [7] allow the semi-automatic segmentation and volume quantification of the thrombus. Hence, recent research work aims at obtaining a robust, automatic thrombus segmentation algorithm easily applicable in the clinical practice. Traditionally proposed methods combine intensity information with shape constraints to minimize a certain energy function [8–10]. Machine learning approaches have also been proposed [11], as well as radial model methods [12]. In [13] a deformable model-based approximation was introduced and recently another deformable model approximation, validated in a large number of pre-operative and post-operative datasets has been presented in [14]. However, these algorithms require user interaction and/or prior lumen segmentation along with centerline extraction and their performance highly depends on multiple parameter tuning, affecting their robustness and clinical applicability.

Lately, DCNNs have gained attention in the scientific community for solving complex segmentation tasks, surpassing the previous state-of-the-art performance in many problems. In [15] a novel and automatic patch-based approach to pre-operative AAA region detection and segmentation is described, based on Deep Belief Networks (DBN). Comparison with ground truth segmentation was

not provided. In [5], a DCNN for automatic post-operative thrombus segmentation and evaluation was presented. Our goal is to extend that work by providing pre-operative and post-operative AAA segmentation and volume quantification, training the network with more datasets and evaluating the influence of the training approximation in the subsequent thrombus volume measurement.

3 Methods

We propose an automatic approach to thrombus segmentation and volume quantification. Segmentation is based on a DCNN specifically designed to segment the thrombus in post-operative datasets, initially presented in [5]. The network is based on Fully Convolutional Networks [16] and Holistically-Nested Edge Detection network [17] and combines low-level features with coarser representations that ensure the smooth contour of the thrombus is kept. To evaluate the influence of the training strategy in the segmentation and volume quantification results, we carry out three experiments: first we train and test the network with mixed pre-operative and post-operative datasets; then, a separate training approach using only pre-operative or only post-operative data is included to compare the results and draw conclusions. Since the number of annotated quality data is limited, we train in a 2D slice-by-slice manner. Training in 2D provides advantages regarding speed, lower memory consumption and the ability to use pretrained networks and fine-tuning. These advantages are leveraged and the 3D coherence of the output binary segmentation is provided in a subsequent post-processing step. Finally, segmentation quality is evaluated by comparison with manually delineated ground truth segmentations. Thrombus volume is computed from the ground truth segmentation and the post-processed automatic segmentation to check the ability of the proposed approach to characterize the thrombus. A visual representation of these steps is shown in Fig. 1. Each step is further explained in the following subsections.

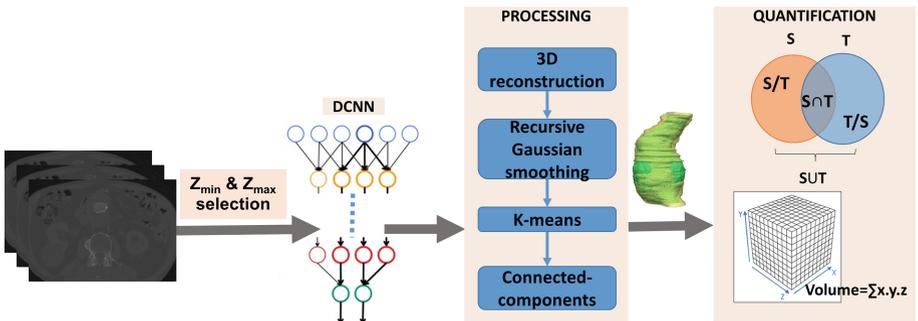


Fig. 1. Pipeline for automatic thrombus segmentation and volume quantification.

3.1 Abdominal Aortic Aneurysm Datasets

A total of 38 contrast-enhanced CTA datasets from different patients that present infrarenal aneurysms were employed for our experiments. 20 of them are post-operative datasets, while 18 of them correspond to pre-operative scans. These datasets have been obtained with scanners of different manufacturers and have a spatial resolution ranging from 0.725 to 0.977 in x and y, and 0.625-1 in z. They also have varying contrast agent doses. The patient is always located in supine position and the CTA starts around the diaphragm and expands to the iliac crest. The data have been divided into training and testing sets. Training data consists in 20 datasets, 11 post-operative and 9 pre-operative. Testing data is composed of 18 datasets, 9 post-operative and 9 pre-operative. None of the datasets correspond to the same patient. We did not discard datasets of patients with outlying characteristics, so the variability in the data is relatively large in terms of thrombus size and shape or noise. In the post-operative datasets of patients with a favorable evolution, endotension cases and datasets where a leak is visible have been included. For all the patients manually obtained segmentations are available and used as ground truths for the current study. Note that the number of pixels corresponding to the thrombus is much smaller than the number of pixels corresponding to background, with a mean ratio of approximately 1:8.

3.2 Experimental Setup: Thrombus Segmentation

As mentioned above, thrombus segmentation is based on a DCNN network, trained slice-by-slice. Figure 2 is a visual representation of the network architecture. Our goal is to investigate the influence of the training approach by performing three experiments. In the three of them, we train the same network architecture, with the same hyperparameters and try to minimize the Softmax loss, which reduces the influence of extreme values or outliers in the data and provides the probability of each pixel corresponding to a certain class. Learning rate is set $10e-3$, with a step down policy of 33% and gamma equal to 0.1. The Stochastic Gradient Descent solver is employed and training is done during 100 epochs, with a batch size of 4 images and no batch accumulation.

In the first experiment, we train our network with pre-operative and post-operative data, all together. The network is trained with 2D slices of 11 post-operative datasets and 2D slices of 9 pre-operative datasets. None of the datasets correspond to the same patient. Data augmentation is applied in the form of 90 rotations and mirroring to enlarge the datasets and to prevent the network from failing if a rotated dataset is introduced for testing. For testing, datasets of additional different 18 patients are employed, 9 pre-operative and 9 post-operative. The total number of slices for each stage is summarized in Table 1.

In the second and third experiments, 2 networks are trained separately, one only with pre-operative data, the other one just with post-operative. For the pre-operative, 9 pre-operative datasets are used for training and validation, and 9 different datasets are saved for testing purposes. These datasets are identical

to those of the first experiment, and the same data augmentation is applied. Testing is done on slices of patients not included in the training phase, as in the first experiment. In the third experiment, corresponding to the post-operative data, the same approximation is followed. The 11 post-operative datasets used for training on the first experiment are utilized to train this network, and the same 9 post-operative datasets are employed for testing. Data augmentation is also equally applied. Table 1 summarizes the data for these experiments.

Table 1. Training, validation and testing slices used in each experiment.

	Experiment 1	Experiment 2	Experiment 3
Train	4380	1835	2545
Validation	772	323	449
Testing	2878	1447	1431

3.3 Post-processing and Quantification

The output provided by the DCNN are 2D probability maps, where each intensity value is the probability of that pixel being thrombus or not. Thus, an automatic processing of these maps is included as the last step to obtain the 3D binary mask segmentation. First, we reconstruct the 3D prediction map volume and apply Gaussian filtering in the z-direction to ensure some continuity in this direction. We set the sigma value to $\sigma = 2 * Spacing_z$. Then, K-means clustering of the 3D probability map is employed, where the number of clusters is fixed to 6, experimentally. The output cluster image is filtered and binarized, by removing the class with the lowest probability of being thrombus. A subsequent connected component analysis is used to keep the largest object, which in our experiments always corresponds to the thrombus. The Volume is measured based on the Divergence Theorem Algorithm (DTA), by estimating the volume of the thrombus from its point-list, as explained in [18]. Finally, a comparison between the automatic thrombus binary segmentation (source, S) and the expert delineated ground truth (target, T) is included to evaluate segmentation quality in terms of total overlap, Dice coefficient, false negative rate (FN) and false positive rate (FP), as proposed in [19]. The volume difference between both segmentations is also included.

$$\begin{aligned}
 \text{Total overlap for thrombus region } (r) &: | S_r \cap T_r | / | T_r | \\
 \text{Dice coefficient for thrombus region } (r) &: 2 | S_r \cap T_r | / (| S_r | + | T_r |) \\
 \text{False negative error for thrombus region } (r) &: | T_r / S_r | / | T_r | \\
 \text{False positive error for thrombus region } (r) &: | S_r / T_r | / | S_r | \\
 \text{Volume difference} &: | V_T - V_S | / V_T
 \end{aligned}$$

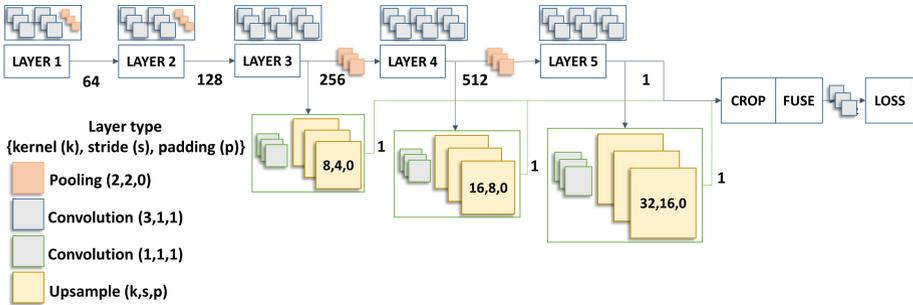


Fig. 2. Deep convolutional neural network architecture for thrombus segmentation.

4 Results

Table 2 summarizes the results for the first experiment, where the network is trained and tested both with pre-operative and post-operative data. The mean Dice similarity coefficient is 81.4%, being this coefficient higher in the post-operative than in the pre-operative. Since the number of pre-operative slices is smaller than the number of post-operative slices and the network was initially designed for the post-operative scenario, a reduction in the accuracy in the pre-operative could be expected. This also impacts the volume difference between the automatically segmented thrombus and the ground truth, being this difference larger in the pre-operative than in the post-operative. The mean volume difference is 12.8%, where the over-estimation of the volume is of 10.9% and the sub-estimation is of 13.9%. Sub-estimation occurs in the double of cases where over-estimation occurs. Qualitative results of this experiment are shown in Fig. 3.

In the second and third experiments, we trained the same network but only with pre-operative or post-operative data. Results are reported in Table 2. In the pre-operative, contrary to our initial hypothesis that an improvement should be observed when training two networks separately, a reduction in the Dice coefficient and an increase of the volume difference is obtained compared to the first experiment. We attribute these results to the reduction in the number of training samples, being only slices extracted from 9 different datasets. Hence, the ability of the network to generalize diminishes. When testing, the variability in the aneurysm shape affects more notably the segmentation quality, and the results for one testing dataset have a stronger impact on the global mean. The worst result for a pre-operative dataset corresponds to the case depicted in Fig. 4, where there is contrasted blood inside the aneurysm area. Probably, the network does not expect to find high-contrasted areas that are not lumen, stent or calcifications, which always set the limits for the segmentation, and thus, it understands that there is a border in this contrasted area and sub-segments the aneurysm; without this dataset the mean Dice coefficient would be equal to 72.8% and the volume difference would be 14%, which approximates to the results of the first experiment, but with half the number of training images.

Table 2. Testing results for the three experiments: 1) the network is trained and tested both with pre-operative and post-operative data, 2) the network is trained only with pre-operative data, 3) only post-operative data is employed.

Experiment		Total overlap	Dice	FN	FP	Volume difference
1	Pre	0.784 ± 0.127	0.790 ± 0.102	0.216 ± 0.127	0.193 ± 0.103	0.134 ± 0.093
	Post	0.817 ± 0.066	0.837 ± 0.062	0.183 ± 0.066	0.133 ± 0.095	0.121 ± 0.081
	Mean	0.801 ± 0.103	0.814 ± 0.087	0.199 ± 0.103	0.163 ± 0.103	0.128 ± 0.088
2	Pre	0.616 ± 0.171	0.697 ± 0.132	0.384 ± 0.171	0.154 ± 0.076	0.265 ± 0.220
3	Post	0.886 ± 0.058	0.855 ± 0.065	0.134 ± 0.058	0.152 ± 0.087	0.086 ± 0.080

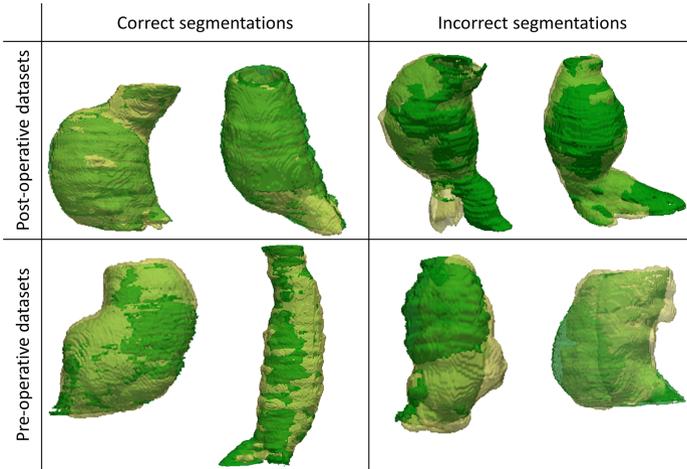


Fig. 3. Qualitative segmentation results of the first experiment. The manual ground truth is shown in green and the automatic segmentation in yellow. (Color figure online)

With respect to the third experiment, related to the post-operative, an increase of the Dice coefficient and a decrease on the volume difference is observed, which agrees with our hypothesis that by training both scenarios separately, better results can be expected. A 33.9% improvement in the volume difference is achieved, although only half the number of training images have been utilized.

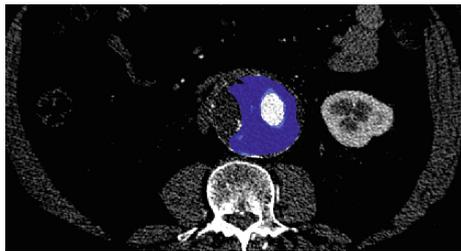


Fig. 4. Incorrect pre-operative AAA segmentation due to contrasted blood inside it.

5 Conclusions

In this paper, we have investigated the influence of the DCNN training strategy in the automatic segmentation and quantification of the AAA volume. Three experiments have been performed: first, the network has been trained and tested with both pre- and post-operative datasets; then, two networks have been trained separately, only with pre- or post-operative data. The same training and testing datasets are used for all the experiments, which correspond to different patients. The results showed that by training separately for the pre-operative and the post-operative scenarios, similar or even better results could be obtained compared to training everything together. However, the number of training samples is smaller when training separately and thus, outlier datasets impact more negatively the results than when training together. We conclude that by training separately with a larger number of cases more precise results could be obtained. Each network would adapt to the specificities of each scenario, such as the appearance of the stent in the post-operative, the bigger thrombus size in the post-operative when the evolution is unfavorable or the appearance of leaks. Fine-tuning from network weights learned from medical images could also improve the results.

Regarding volume quantification, the goal is to utilize thrombus volume during follow-up to assess disease progression. EVAR reporting standards [20] state that an increase in the aneurysm volume of 5% is considered clinically relevant and a clinical failure after EVAR. Intra-observer and inter-observer variability for volume measurements have ranged between 3% and 5% from semi-automatically segmented aneurysms [20–22]. From a clinical perspective, our pipeline produces significant volume differences between the ground truth and the automatic segmentation. These differences vary in the range of 8% to 12% in the post-operative and 13% to 26% in the pre-operative and can equally correspond to over-estimation or sub-estimation of the volume. The automatic segmentation results are reasonably good, but the measured volume values still need to be refined to be directly applicable in the clinical practice for quantitative progression assessment. Future work aims at reducing the volume difference between ground truth and automatically segmented thrombus, by adapting our method to that purpose and analyzing the volume quantification results with more data.

References

1. Pearce, W.H., Zarins, C.K., Bacharach, J.M.: Atherosclerotic peripheral vascular disease symposium II: controversies in abdominal aortic aneurysm repair. *Circulation* **118**(25), 2860–2863 (2008)
2. Moll, F.L., Powell, J.T., Fraedrich, G., Verzini, F., Haulon, S., Waltham, M., Ricco, J.B.: Management of abdominal aortic aneurysms clinical practice guidelines of the European society for vascular surgery. *Eur. J. Vasc. Endovasc. Surg.* **41**(1), 1–58 (2011)
3. Stather, P.W., Sidloff, D., Dattani, N.: Systematic review and meta-analysis of the early and late outcomes of open and endovascular repair of abdominal aortic aneurysm. *J. Vasc. Surg.* **58**(4), 1142 (2013)

4. Renapurkar, R.D., Setser, R.M., O'Donnell, T.P., Egger, J., Lieber, M.L., Desai, M.Y., Stillman, A.E., Schoenhagen, P., Flamm, S.D.: Aortic volume as an indicator of disease progression in patients with untreated infrarenal abdominal aneurysm. *Eur. J. Radiol.* **81**(2), 87–93 (2012)
5. López-Linares, K., Aranjuelo, N., Kabongo, L., Maclair, G., Lete, N., Leskovsky, P., Garca-Familiar, A., Macía, I., González Ballester, M.A.: Fully automatic segmentation of abdominal aortic thrombus in post-operative CTA images using deep convolutional neural networks. *Proc. Int. J. CARS* **12**(1), 29–30 (2017)
6. http://www.3gehealthcare.com/en/products/categories/advanced_visualization/applications/autobone_and_vesseliq_xpress
7. <https://www.vitalimages.com/clinicalapplications/tabs/ct-endovascular-stent-planning>
8. Duquette, A.A., Jodoin, P.M., Bouchot, O., Lalande, A.: 3D segmentation of abdominal aorta from CT-scan and MR images. *Comput. Med. Imaging Graph.* **36**(4), 294–303 (2012)
9. Egger, J., Freisleben, B., Setser, R., Renapuraar, R., Biermann, C., O'Donnell, T.: Aorta segmentation for stent simulation. In: *MICCAI Workshop on Cardiovascular Interventional Imaging and Biophysical Modelling*, p. 10 (2009)
10. Freiman, M., Esses, S.J., Joskowicz, L., Sosna, J.: An iterative model-constrained graph-cut algorithm for abdominal aortic aneurysm thrombus segmentation. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 672–675 (2010)
11. Maiora, J., Ayerdi, B., Graña, M.: Random forest active learning for AAA thrombus segmentation in computed tomography angiography images. *Neurocomputing* **126**, 71–77 (2014)
12. Macía, I., Legarreta, J.H., Paloc, C., Graña, M., Maiora, J., García, G., Blas, M.: Segmentation of abdominal aortic aneurysms in CT images using a radial model approach. In: Corchado, E., Yin, H. (eds.) *IDEAL 2009*. LNCS, vol. 5788, pp. 664–671. Springer, Heidelberg (2009). doi:10.1007/978-3-642-04394-9_81
13. Demirci, S., Lejeune, G., Navab, N.: Hybrid deformable model for aneurysm segmentation. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 33–36 (2009)
14. Lalys, F., Yan, V., Kaladji, A., Lucas, A., Esneault, S.: Generic thrombus segmentation from pre- and post-operative CTA. *Int. J. Comput. Assist. Radiol. Surg.* **12**(9), 1501–1510 (2017)
15. Hong, H., Sheikh, U.: Automatic detection, segmentation and classification of abdominal aortic aneurysm using deep learning. In: *2016 IEEE 12th International Colloquium on Signal Process Its Application*, pp. 242–246 (2016)
16. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
17. Xie, S., Tu, Z.: Holistically-nested edge detection. CoRR abs/1504.06375 (2015)
18. Alyassin, A.M., Lancaster, J.L., Downs, J.H., Fox, P.T.: Evaluation of new algorithms for the interactive measurement of surface area and volume. *Med Phys.* **21**(6), 741–752 (1994)
19. Tustison, N., Gee, J.: Introducing dice, jaccard, and other label overlap measures to ITK. *Insight J.* (2009)
20. Chaikof, E., Blankensteijn, J., Harris, P., White, G., Zarins, C., et al.: Reporting standards for endovascular aortic aneurysm repair. *J. Vasc. Surg.* **35**(5), 1048–1060 (2002)

21. Parr, A., Jayaratne, C., Buttner, P., Golledge, J.: Comparison of volume and diameter measurement in assessing small abdominal aortic aneurysm expansion examined using computed tomographic angiography. *Eur. J. Radiol.* **79**(1), 42–47 (2011)
22. van Prehn, J., van der Wal, M.B., Vincken, K., Bartels, L.W., Moll, F.L., van Herwaarden, J.A.: Intra- and interobserver variability of aortic aneurysm volume measurement with fast CTA postprocessing software. *J. Endovasc. Ther.* **15**(5), 504–510 (2008)