2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Deep Learning-Based 3D Segmentation of True Lumen, False Lumen, and False Lumen Thrombosis in Type-B Aortic Dissection

Liana D. Wobben[1,2], Marina Codari[1], Gabriel Mistelbauer[3], Antonio Pepe[4], Kai Higashigaito[1], Lewis D. Hahn[5], Domenico Mastrodicasa[1], Valery L. Turner[1], Virginia Hinostroza[1], Kathrin Bäumler[1], Michael P. Fischbein[6], Dominik Fleischmann[1], and Martin J. Willemink[1]

Abstract— **Patients with initially uncomplicated type-B aortic dissection (uTBAD) remain at high risk for developing late complications. Identification of morphologic features for improving risk stratification of these patients requires automated segmentation of computed tomography angiography (CTA) images. We developed three segmentation models utilizing a 3D residual U-Net for segmentation of the true lumen (TL), false lumen (FL), and false lumen thrombosis (FLT). Model 1 segments all labels at once, whereas model 2 segments them sequentially. Best results for TL and FL segmentation were achieved by model 2, with median (interquartiles) Dice similarity coefficients (DSC) of 0.85 (0.77-0.88) and 0.84 (0.82-0.87), respectively. For FLT segmentation, model 1 was superior to model 2, with median (interquartiles) DSCs of 0.63 (0.40-0.78). To purely test the performance of the network to segment FLT, a third model segmented FLT starting from the manually segmented FL, resulting in median (interquartiles) DSCs of 0.99 (0.98-0.99) and 0.85 (0.73-0.94) for patent FL and FLT, respectively. While the ambiguous appearance of FLT on imaging remains a significant limitation for accurate segmentation, our pipeline has the potential to help in segmentation of aortic lumina and thrombosis in uTBAD patients.**

Clinical relevance— **Most predictors of aortic dissection (AD) degeneration are identified through anatomical modeling, which is currently prohibitive in clinical settings due to the time-intense human interaction. False lumen thrombosis, which often develops in patients with type B AD, has proven to show significant prognostic value for predicting late adverse events. Our automated segmentation algorithm offers the potential of personalized treatment for AD patients, leading to an increase in long-term survival.**

*LDW was supported by the Hendrik Mullerfonds, a Marco Polo grant, the De Cock foundation and the Groningen University Fund. AP was supported by the TU Graz LEAD Project *Mechanics, Modeling and Simulation of Aortic Dissection*. DM was supported by the NIBIB (5T32EB009035). MJW was supported by the American Heart Association (18POST34030192).

[1]Liana D. Wobben, Marina Codari, Domenico Mastrodicasa, Valery L. Turner, Virginia Hinostroza, Kathrin Bäumler, Dominik Fleischmann and Martin J. Willemink are with the Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA

[2]Liana D. Wobben is with the Department of Biomedical Engineering, University of Groningen, Groningen, The Netherlands wobbenliana@gmail.com

[3]Gabriel Mistelbauer is with the Department of Simulation and Graphics, University of Magdeburg, Magdeburg, Germany

[4]Antonio Pepe is with the Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria

[5]Lewis D. Hahn is with the Division of Cardiothoracic Radiology, UC San Diego School of Medicine, San Diego, CA, USA

[6]Michael P. Fischbein is with the department of Cardiothoracic Surgery, Stanford University School of Medicine, Stanford, CA, USA

I. INTRODUCTION

Medical management of initially uncomplicated type-B aortic dissection (uTBAD) is associated with a poor long-term survival of only 60% at five years, due to a high rate of late adverse events (LAEs) [1]. Early identification of patients who may potentially benefit from preventative thoracic endovascular aortic repair (TEVAR) is thus highly desirable. Several studies suggest that morphological features extracted from computed tomography angiography (CTA) might predict LAEs in patients with uTBAD [2], [3]. False lumen thrombosis (FLT) also plays an important role, although its developing mechanism is incompletely understood [3]. Automated and robust segmentation of the true lumen (TL) and false lumen (FL) – the defining features of aortic dissection – is a prerequisite for extracting and exploring the predictive power of any other morphologic feature, including the presence and extent of false lumen thrombosis (FLT). Manual segmentation of TL, FL, and FLT is difficult, poorly reproducible, and prohibitively time consuming to be considered for clinical application.

Machine learning techniques may overcome these limitations. Hahn et al. proposed a 2D network for segmenting the TL and FL [4]. Although this work reported good results, it is reasonable to expect that the addition of a third dimension is an important requisite for FLT segmentation, since it is hard to differentiate between thrombus and slow blood flow without evaluating adjacent slices. Cao et al. applied a 3D multi-task framework to segment the TL and FL, by taking an entire scan volume as input, with considerable downsampling in the z-dimension [5]. Downsampling overcomes limitations in memory and computational capability, but quality loss and image deterioration occur in both the CT and ground truth data [5], [6]. The recent publication of Chen et al. proves that better results are achievable without downsampling, which opens up the possibility of accurate FLT segmentation [7]. However, no attempts at automatic segmentation of FLT have been made [6].

The first aim of this work was to develop an automated pipeline to segment the TL and FL, without forcing the input to be downsampled in the z-direction. The second aim was to also segment the FLT.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

II. METHODS

A. Patients and imaging data

We retrospectively identified 164 CTA scans from 43 patients with uTBAD who underwent baseline and surveillance imaging at our institution between 2003 and 2017. After excluding incomplete or technically inadequate CT scans from 6 patients, our final study dataset consisted of 147 scans from 40 patients (13 women), with a median (interquartiles) age of 48 (36-59) years. FLT was present in 93/147 scans from 27/40 patients.

B. Ground-truth segmentation

All scans were manually segmented by a trained radiologist (KH, six years of experience) using Intuition software (TeraRecon, Foster City, CA). Image voxels were labeled as TL, patent FL (PFL; non-thrombosed part of the FL), FLT, and background. In order to establish the suitability of the manual segmentations as ground truth and to set a benchmark for network performance, eight scans of the study population, belonging to six patients, were segmented by a second observer (MJW, six years of experience). Dice similarity coefficients (DSCs) were calculated to assess similarity between the segmentations of both observers.

C. Preprocessing

Our segmentation approach uses multiplanar reformation images (MPRs) centered along and perpendicular to the aortic centerline, as described by Hahn et al. [4]. MPR stacks were resampled to obtain an isotropic resolution of 1 mm, and subsequently cropped in-plane to a size of 80×80 mm. Unity-based normalization was applied to the voxel intensities of all scans.

We split the dataset into training, validation, and test sets, with a ratio of 70:15:15, resulting in a number of scans ratio of 103:22:22, number of patients ratio of 28:6:6, and number of scans with FLT ratio of 65:14:14, respectively.

Augmentation was performed on 25% of training data. This value was empirically determined to optimize model performance. The following transformations were randomly employed: (1) flipping along the transverse axis, (2) translation along the transverse and sagittal axis, (3) uniform scaling, (4) non-uniform scaling, or (5) gamma correction of density values.

D. Models

We investigate the efficacy of (1) a single-step multi-task model, (2) a sequential multi-task model and (3) a single-step single-task model (Fig. 1).

1) Model 1: In the first model, a single network classifies image voxels as TL, FL, and background (for all cases) or TL, PFL, FLT, and background (for cases containing FLT).

2) Model 2: In the second model, a cascade of networks performs stepwise segmentation. After the initial segmentation of the whole aorta, the TL and FL are segmented in a second step. Then, PFL and FLT were segmented in cases containing FLT. In the second (and third) step of this cascade, the prediction of the previous step is used as an additional input channel.

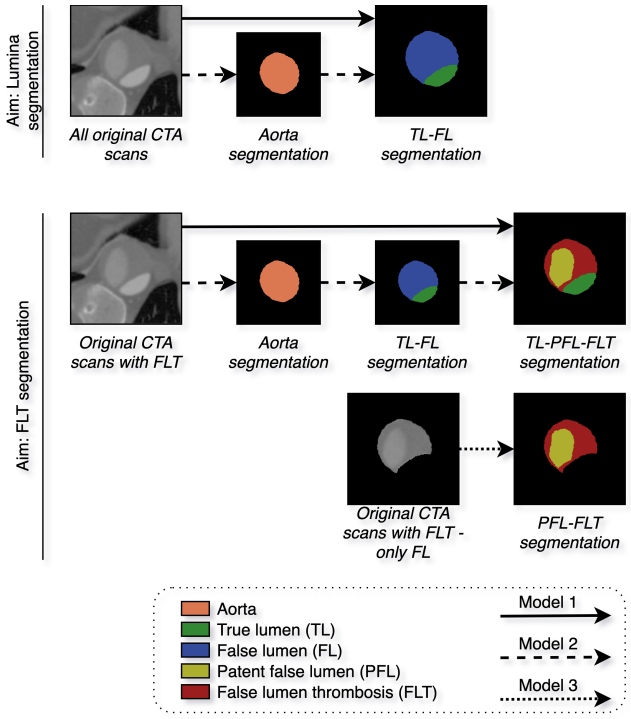


Fig. 1: Visualization of the segmentation pipeline. Each arrow corresponds to a 3D residual U-Net. Model 1 segments all labels at once, whereas model 2 utilizes a cascade of networks. Model 3 is a single-step model that starts from the false lumen ground truth segmentation. CTA = computed tomography angiography.

3) Model 3: In the third model, the FLT is segmented directly from the manually segmented FL. This enables us to independently evaluate the performance of the network to segment FLT given a high quality FL segmentation.

E. Architecture

All networks used in our three models share the same architecture, data input details and optimization mechanism.

The networks were built on top of Lee et al.'s residual symmetric U-Net, with max-pooling layers in the encoder, strided transposed convolution layers in the decoder and large-scale skip-connection between the encoder and decoder[8]. The basic blocks, used in each level, consist of a single convolution, followed by a residual block consisting of two convolutions and a residual skip-connection.

The proposed networks take a slab composed of 16 consecutive MPR slices as input, with a batch size of 8. During training, slabs were randomly selected within the image volume. During each epoch a different subset of slabs was selected.

A cosine decay scheme without restarts, was chosen for the learning rate (LR). The maximum LR value of 10^{-4} was estimated using the LR range test [9], and used as starting LR value. We used Adam as optimization algorithm and the Generalized Dice loss (GDL) as loss function [10].

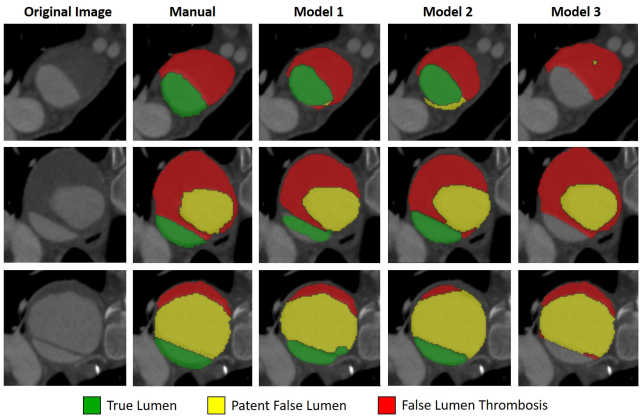


Fig. 2: Automated segmentation outcome in one patient with false lumen thrombosis (FLT). The 3D rendering shows the ground truth manually segmented aorta, while 2D images are sampled along the center-line at the proximal, middle, and distal segments of the FLT.

III. RESULTS

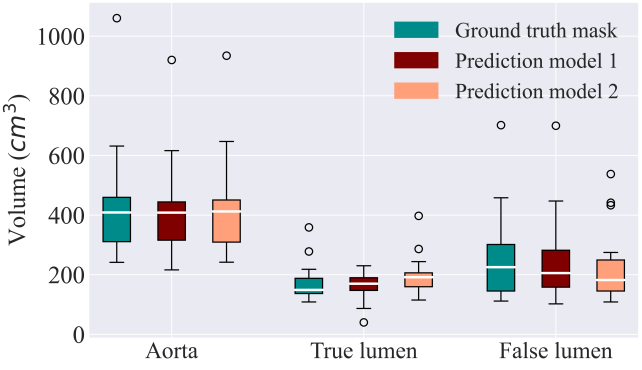
A. Interobserver variability

The median (interquartiles) DSCs between segmentations performed by the two observers were 0.92 (0.91-0.93) for the TL, 0.91 (0.88-0.94) for the PFL, and 0.78 (0.74-0.80) for the FLT label, respectively.

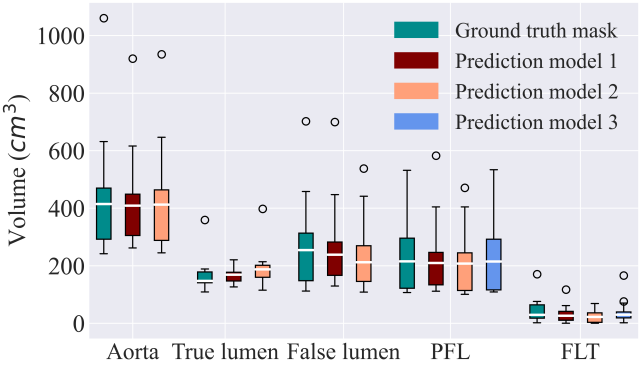
B. Network performance

Table I summarizes the performance of model 1 and 2 for TL and FL segmentation using the entire database. Both models were also applied to the subset of scans that contained FLT to segment TL, PFL, and FLT (Table II). This table also shows the results of the application of model 3 to the manually segmented FL mask of the scans with FLT.

Fig. 2 shows the segmentation outcome for a patient with FLT. The segmented volume distributions in the test set are shown in Fig. 3.



(a) Segmented volumes of all testing scans.



(b) Segmented volumes of testing scans with FLT.

Fig. 3: Model performance: segmented volumes per model. PFL = patent false lumen; FLT = false lumen thrombosis.

IV. DISCUSSION

We developed three models for segmentation in CTA-scans of patients with uTBAD.

The TL and FL were successfully segmented using model 2, which utilizes two sequential residual U-nets, with results superior to those of model 1, a single-step network. The superiority of the second model was expected, since the presence of the aorta segmentation from the first step benefits the second step. We hereby essentially divide the segmentation problem into sub problems. It is likely that the additional input information obtained from the previous segmentation layer improved network convergence during training.

DSC scores seem comparable to those of Hahn et al, who used a pre-trained 2-dimensional U-net with a VGG11 encoder and reported mean $\pm$ standard deviation values of 0.87 ± 0.06 and 0.89 ± 0.04 for the TL and FL respectively. Although this study was based on the same dataset, a fair comparison of developed models is not possible, due to the difference in training, validation and testing data partitions. While Hahn et al. reported good results, the use of a 3D model is required in the current work, due to small FLT volumes that make information from adjacent slices necessary for accurate segmentation. Cao et al., who utilizes a sequence of 3D U-nets without residual skip-connections, reports Dice values of 0.93 ± 0.01 and 0.91 ± 0.02 for the TL and FL, respectively. Their model, however, forces the scans to be downsampled in the z-direction [5], which can result in partial volume artifacts, a method that would not be feasible for FLT segmentation, due to the small volumes of thrombi. The recent study of Chen et al. achieves excellent results without downsampling, but the authors state that their straightening algorithm might not be suitable for scans with FLT [7]. For the FLT results, comparison to other methods is not possible, since no other attempts of automatic segmentation of FLT in AD patients have been made.

Although the successful segmentation of the TL and FL is promising, we reported suboptimal results for FLT segmentation with model 1 and 2. After qualitative examination, we observed that both models performed better on scans with a larger amount of thrombosis. Notably, the single-step model performed better than the serial model in the testing set at segmenting PFL and FLT (Table II). This result may be caused by the high sensitivity of the last segmentation step to a suboptimal segmentation of the FL. Our third

TABLE I: Model performance: Dice similarity coefficients of developed models in the complete dataset. Data are presented as median and interquartile range.

Model	Phase	Aorta	True lumen	False lumen
1	Training	0.93 (0.92 - 0.94)	0.81 (0.73 - 0.85)	0.82 (0.74 - 0.88)
	Validation	0.92 (0.91 - 0.94)	0.74 (0.69 - 0.83)	0.78 (0.67 - 0.84)
	Testing	0.93 (0.92 - 0.94)	0.74 (0.68 - 0.76)	0.79 (0.73 - 0.84)
2	Training	0.97 (0.96 - 0.97)	0.94 (0.92 - 0.95)	0.94 (0.92 - 0.96)
	Validation	0.95 (0.94 - 0.96)	0.90 (0.85 - 0.92)	0.90 (0.85 - 0.92)
	Testing	0.95 (0.95 - 0.96)	0.85 (0.77 - 0.88)	0.84 (0.82 - 0.87)

TABLE II: Model performance: Dice similarity coefficients of developed models in scans containing false lumen thrombosis (FLT). Data are presented as median and interquartile range.

Model	Phase	Aorta	True lumen	False lumen	Patent false lumen	False lumen thrombosis
1	Training	0.94 (0.92 - 0.94)	0.84 (0.77 - 0.87)	0.85 (0.78 - 0.90)	0.84 (0.73 - 0.90)	0.75 (0.66 - 0.79)
	Validation	0.93 (0.90 - 0.94)	0.81 (0.69 - 0.85)	0.82 (0.69 - 0.87)	0.80 (0.68 - 0.87)	0.72 (0.66 - 0.77)
	Testing	0.93 (0.91 - 0.94)	0.74 (0.71 - 0.77)	0.83 (0.79 - 0.84)	0.82 (0.79 - 0.86)	0.63 (0.40 - 0.78)
2	Training	0.96 (0.96 - 0.97)	0.93 (0.93 - 0.94)	0.94 (0.91 - 0.95)	0.93 (0.88 - 0.94)	0.77 (0.66 - 0.83)
	Validation	0.95 (0.94 - 0.96)	0.92 (0.86 - 0.93)	0.91 (0.85 - 0.92)	0.90 (0.85 - 0.93)	0.76 (0.68 - 0.79)
	Testing	0.96 (0.95 - 0.96)	0.86 (0.77 - 0.88)	0.86 (0.84 - 0.88)	0.85 (0.83 - 0.86)	0.50 (0.19 - 0.65)
3	Training				0.98 (0.96 - 0.99)	0.92 (0.83 - 0.95)
	Validation				0.95 (0.93 - 0.97)	0.91 (0.87 - 0.92)
	Testing				0.99 (0.98 - 0.99)	0.85 (0.73 - 0.94)

developed model, that segmented PFL and FLT starting from the FL ground truth, promisingly proved that, once the FL is segmented correctly, the network is able to discriminate between PFL and FLT with excellent performance. Taking into consideration recent advances in FL segmentation [7], and the possibility of combining this technique with the current study, the realization of accurate FLT segmentation can reasonably be expected in the near future.

A limitation of the study is the small sample size. A larger training dataset may improve the performance of the network. Secondly, the interobserver variability for the FLT segmentations is suboptimal. While the TL and FL interobserver analyses show DSC values higher than 0.90, this is not the case for FLT. A possible explanation is the difficulty to distinguish between slow blood flow and FLT or a thick flap and thrombus around the flap. The suboptimal reproducibility of FLT, even by expert observers, poses a limitation to the development of supervised models. Nevertheless, identification of large amounts of FLT might be more relevant than that of small portions of clot.

Future studies are needed to improve segmentation performance and investigate the voxel-level accuracy required to provide meaningful prognostic information. Since multiple CNN models have been proposed for segmenting TL and FL in TBAD patients, future studies should rigorously compare their performances and generalizability using a common, independent and multicentric dataset. Moreover, future effort should be made to create ensemble models in order to overcome the limitations of individual segmenters.

V. CONCLUSION

We developed models for segmentation of TL and FL on CTA-scans of patients with uTBAD, which are the key features of aortic dissection and prerequisites for extraction of other features. In addition, this study is the first to present a step towards FLT quantification, which may improve individual prognostication of patients with uTBAD.

REFERENCES

[1] Y.-L. Qin, G. Deng, T.-X. Li, W. Wang, and G.-J. Teng, "Treatment of acute type-B aortic dissection: thoracic endovascular aortic repair or medical management alone?," JACC: Cardiovascular Interventions, vol. 6, no. 2, pp. 185–191, Feb 2013.

[2] A.M. Sailer et al., "Computed tomography imaging features in acute uncomplicated stanford type-B aortic dissection predict late adverse events," Circ: Cardiovascular Imaging, vol. 10, no. 4, pp. e005709, Apr 2017.

[3] D. Spinelli et al., "Current evidence in predictors of aortic growth and events in acute type B aortic dissection," Journal of Vascular Surgery, vol. 68, no. 6, pp. 1925–1935.e8, Dec 2018.

[4] L.D. Hahn et al., "CT-based true- and false-lumen segmentation in type B aortic dissection using machine learning," Radiology: Cardiothoracic Imaging, vol. 2, no. 3, pp. e190179, Jun 2020.

[5] L. Cao et al., "Fully automatic segmentation of type b aortic dissection from CTA images enabled by deep learning," European Journal of Radiology, vol. 121, pp. 108713, Dec 2019.

[6] A. Pepe et al., "Detection, segmentation, simulation and visualization of aortic dissections: A review," Medical Image Analysis, vol. 65, pp. 101773, Oct 2020.

[7] D. Chen, et al., "Multi-stage learning for segmentation of aortic dissections using a prior aortic anatomy simplification.," Medical Image Analysis, vol. 69, pp. 101931, Apr 2021.

[8] K. Lee, J. Zung, P. Li, V. Jain, and H. Seung, "Superhuman accuracy on the SNEMI3D connectomics challenge," May 2017, arXiv:1706.00120.

[9] L.N. Smith, "Cyclical learning rates for training neural networks," in IEEE Winter Conference on Applications of Computer Vision, 2017, pp. 464–472.

[10] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2017, pp. 240–248.