European Journal of Radiology 121 (2019) 108713

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Fully automatic segmentation of type B aortic dissection from CTA images enabled by deep learning

Long Cao[a,1], Ruiqiong Shi[b,c,1], Yangyang Ge[a,1], Lei Xing[d], Panli Zuo[c], Yan Jia[c], Jie Liu[a], Yuan He[a], Xinhao Wang[a], Shaoliang Luan[a], Xiangfei Chai[c,**], Wei Guo[a,*]

[a] Department of Vascular and Endovascular Surgery, Chinese PLA General Hospital, Beijing, PR China
[b] Institute of Information Science, Beijing Jiaotong University, Beijing, PR China
[c] Huiying Medical Technology Co., Ltd., Dongsheng Science and Technology Park, Beijing, PR China
[d] Department of Radiation Oncology Stanford University School of Medicine, Stanford, CA, USA

ARTICLE INFO

Keywords:
Type B aortic dissection
CTA
Deep learning
Convolutional neural network
Automatic segmentation

ABSTRACT

Purpose: This study sought to establish a robust and fully automated Type B aortic dissection (TBAD) segmentation method by leveraging the emerging deep learning techniques.

Methods: Preoperative CTA images of 276 patients with TBAD were retrospectively collected from January 2011 to December 2018. Using a reproducible manual segmentation protocol of three labels (whole aorta, true lumen (TL), and false lumen (FL)), a ground truth database (n = 276) was established and randomly divided into training and testing sets in a rough 8:1 ratio. Three convolutional neural network (CNN) models were developed on the training set (n = 246): single one-task (CNN1), single multi-task (CNN2), and serial multi-task (CNN3) models. Performance was evaluated using the Dice coefficient score (DCS) and lumen volume accuracy on the testing set (n = 30). Pearson correlation, Intra-class correlation coefficients and Bland–Altman plots were used to evaluate the inter-observer measurement agreement.

Results: CNN3 performed the best, with mean DCSs of 0.93 ± 0.01, 0.93 ± 0.01 and 0.91 ± 0.02 for the whole aorta, TL, and FL, respectively ($p < 0.05$). Each label volume from CNN3 showed excellent agreement with the ground truth, with mean volume differences of −31.05 (−82.76 to 20.65) ml, 4.79 (−11.04 to 20.63) ml, and 8.67(−11.40 to 28.74) ml for the whole aorta, TL, and FL, respectively. The segmentation speed of CNN3 was 0.038 ± 0.006 s/image.

Conclusion: Deep learning-based model provides a promising approach for accurate and efficient segmentation of TBAD and makes it possible for automated measurements of TBAD anatomical features.

1. Introduction

Type B aortic dissection (TBAD) is a life-threatening event with bleeding within and along the wall of the aorta, resulting in a separation of a true lumen (TL) and false lumen (FL) by the intimal flap [1]. Currently, computed tomography angiography (CTA) is routinely performed for the diagnosis and follow up of TBAD, and also provide the anatomical feature assessment before thoracic endovascular aortic repair (TEVAR) [2,3].

A significant step in the prognostication of TBAD and TEVAR planning is fast and accurate quantification of anatomical features, especially for the lumen size [4]. Aortic lumen segmentation is the paramount first step to evaluate the anatomical features. A number of computerized segmentation methods have been proposed in the past to automate the process [5–8]. Up to this point, only limited success in distinguishing the boundaries of TL and FL is accomplished with purely

Abbreviations: TBAD, type B aortic dissection; TEVAR, thoracic endovascular aortic repair; TL, true lumen; FL, false lumen; CTA, computed tomography angiography; CNN, convolutional neural network; 3D, three-dimensional; DCS, dice coefficient score; ICC, intra-class correlation coefficient

* Corresponding author at: Department of Vascular and Endovascular Surgery, Chinese PLA General Hospital, 28 Fu Xing Road, 100853, Beijing PR China.
** Corresponding author.
E-mail addresses: caolong@301hospital.com.cn (L. Cao), Ruiqiong_Shi@bjtu.edu.cn (R. Shi), geyangyang@301hospital.com.cn (Y. Ge), lei@stanford.edu (L. Xing), zuopanli@huiyihuiying.com (P. Zuo), jiayan@huiyihuiying.com (Y. Jia), liujie3514@163.com (J. Liu), heyuan@301hospital.com.cn (Y. He), wangxinhao@301hospital.com.cn (X. Wang), luanshaoliang@126.com (S. Luan), chaixiangfei@huiyihuiying.com (X. Chai), guowei@301hospital.com.cn (W. Guo).

[1] Three first authors: Long Cao, Ruiqiong Shi and Yangyang Ge contributed equally to this work.

https://doi.org/10.1016/j.ejrad.2019.108713
Received 22 June 2019; Received in revised form 4 September 2019; Accepted 13 October 2019
0720-048X/ © 2019 Published by Elsevier B.V.

intensity-based techniques. Currently, the intensity-based technique is the most-commonly used method in CTA workstation, which allows rough aortic segmentation and 3D reconstruction by volume rendering method. However, this technique has limited ability to distinguish the boundary between the opacified aorta and periaortic organs and tissue of similar intensity. Additionally, there are numerous reentry tears along the dissection lamella, and therefore, it is difficult to separate the TL and FL by the intensity-based technique. As such, the measurement process in current practice is a labor intensive and time-consuming, involving a substantial amount of manual interventions, such as centerline point adjustment and lumen boundary editing to aid the semi-automated computer methods [4].

Recent advancement in deep learning [9] provides an unique opportunity for automated and accurate segmentation of the lumen (the whole aorta, TL, and FL). Noteworthy, a few studies have been carried out in using convolution neural networks (CNNs) for the segmentation of abdominal aortic aneurysms (AAAs) and demonstrated the potential of deep learning-based segmentation [10]. In reality, TBAD segmentation on CTA data shares much similarity with the problem of AAAs, which lead us to hypothesize that deep learning could also be a useful approach for automatically segmenting the lumens in TBAD. The goal of this study is to develop three CNN models to segment simultaneously the whole aorta, TL, and FL, and to demonstrate the high accuracy and efficiency of the deep learning-based approach.

2. Methods

2.1. Data

The pipeline of this study is illustrated in Fig. 1. All CTA scans in Digital Imaging and Communications in Medicine (DICOM) format were from the retrospective cohort (n = 340) of the Registry Of type B aortic dissection with the Utility of STent graft (ROBUST) study, which is a multicenter, observational study (registration number: ChiCTR-POC-17011726) [11]. The local institutional review boards approved the use of the CTA data. Written informed consent was waived due to the retrospective nature of the enrolled cohort. Patients with CTA scans with slice thickness > 3 mm (n = 39), severe artifacts (n = 24), or incomplete coverage of the aorta (n = 1) were excluded in the present study. Finally, 276 out of 340 patients imaged from January 1, 2011, to December 31, 2018, were included in the present study. The individual information in DICOM header files of all preoperative CTA images were replaced by a continuous number.

2.2. Ground truth database

Manual segmentations were performed by using 3D Slicer (Harvard Medical School, USA; [12]. To ensure a reliable protocol for manual segmentation, we defined three labels, namely, Label 1, Label 2, and Label 3, to represent the whole aorta, TL, and FL, respectively (Fig. 2). More details regarding ground truth establishment process involving manual segmentation protocol and protocol validation are described in the Supplemental Materials.

2.3. Model

An established CNN architecture called 3D U-net [13] was used for the CTA analysis. Prior to training our model, the ground truth dataset (n = 276) was randomly divided into two groups in 8:1 ratio, namely, a training set (n = 246) and testing set (n = 30). Fig. 3 illustrates the entire model, including the preprocessing, U-net, and post-processing.

2.3.1. Image preprocessing

(a) **Normalization of Images from Different Manufactures**: Fixed values out of the display's field of view were normalized to −1024 HU. (b) **Voxel Size Normalization:** For each input volume, the voxel size was resampled to 1 mm^3 by adjusting the pixel size to 0.707 mm and slice thickness to 2 mm using cubic spine interpolation. (c) **Voxel Value Normalization**: These values were normalized so that they had a zero mean and unit variance. (d) **Removal of table and raw radiation areas:** Only the anatomical regions were retained in the image (details seen in supplement). (e) **Resizing**: The images were down-sampled to 128 × 128 × 256 pixels for input to the CNN network. **(f) Image Augmentation:** data augmentation by rotation within the range -10° and 10° is applied in training to solve limitation of small training set.

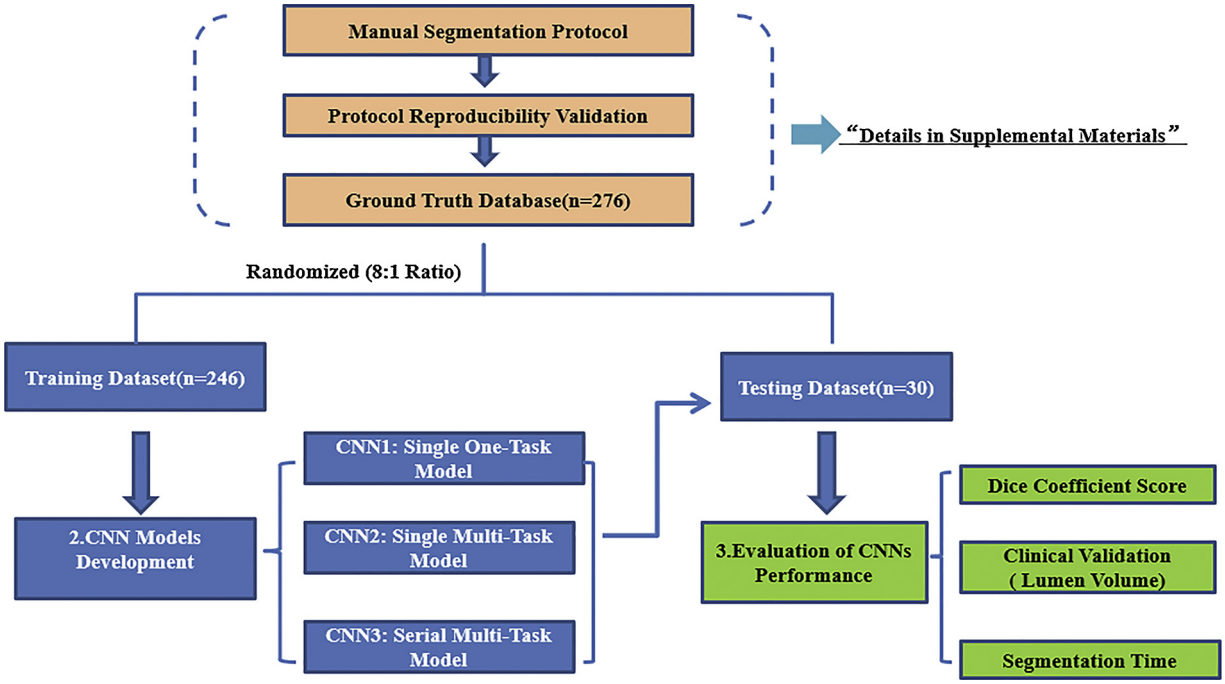

Fig. 1. The complete pipeline of this study.
CNN = convolutional neural network.

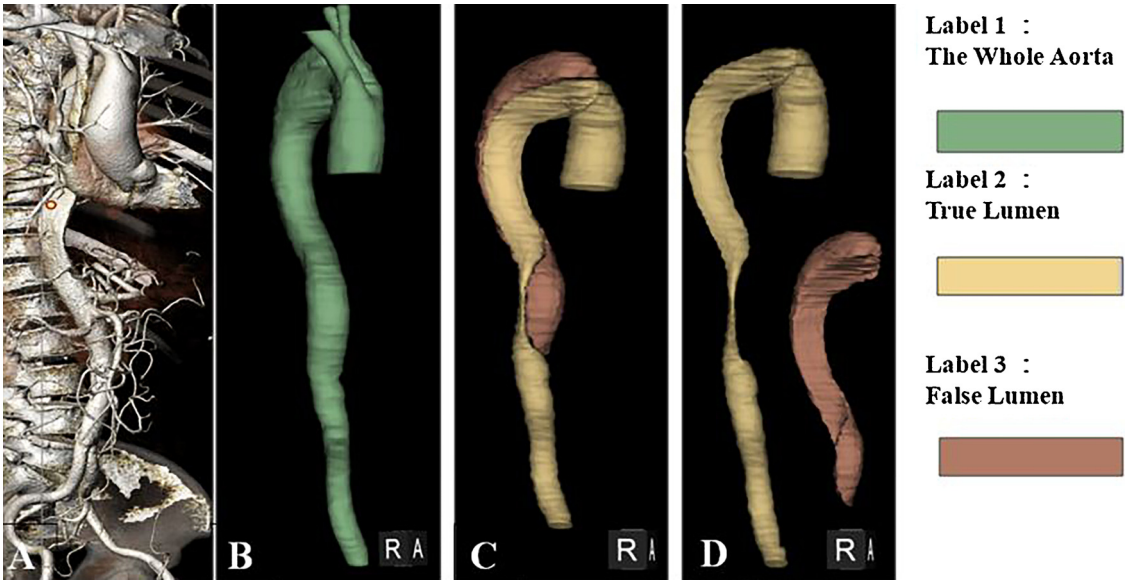

Fig. 2. Manual segmentation model.
(A) Native CTA in VR. **(B)** manual segmentation result of the whole aorta, **(C)** and **(D)** TL and FL as 3-dimensional visualizations. VR = volume rendering; TL = true lumen; FL = false lumen.

2.3.2. CNN training

Three CNN networks with different configurations were developed and trained. Fig. 4 illustrates the framework of the CNNs. CNN1 was trained using a single-task network framework to segment the whole aorta, TL, and FL separately. CNN2 was trained on a single multi-task network framework, which segments the whole aorta, TL, and FL simultaneously. CNN3 employed a serial multi-task network framework containing two 3D U-net models. The first network was used for the whole aorta segmentation and provides a bounding box for the true/false lumen segmentation in the next network. The two networks were trained separately. The first network will be trained twice, one is for rough segmentation of the whole aorta, the second is for fine segmentation of whole aorta with the first model's bounding box. More details regarding CNN3 are described in the Supplemental Fig. S2.

3-fold cross validation was used in this experiment. In the training set, we randomly sample 30 cases as validation set and the other 216 cases as training images. All CNN networks were trained with an initial learning rate of 0.001 that was divided by 10 every 2000 epochs. The batch size was set to 1 due to the calculation capabilities. The weighted Dice loss was used as the loss function (seen in the Supplement). For training, 10,000 epochs were performed. The label image for training consists of four labels from the manual annotation: the whole aorta, TL, FL, and background. All networks were trained and tested on the Google TensorFlow platform using a Nvidia DGX station with 265 GB memory, Intel Xeon E5-2698v4 CPU, and four NVIDIA Tesla V100 GPUs with 32 GB of memory.

2.3.3. Image post-processing

We propose a simple strategy to combine the voxel masks of the whole aorta, TL, and FL masks. The TL label has the highest priority in the combination, the FL label has the second-highest priority, and the remainder of the whole aorta mask was defined as the aorta wall. Hence, the final output mask of segmentation contains the aorta wall, TL, and FL, and this mask was up-sampled to its original image size.

2.4. CNN segmentation performance

After training, all CNN models were tested and evaluated on the 3 validation sets (for each set, n = 30) and testing data set (n = 30). Using the ground truth as a reference, the CNN segmentation

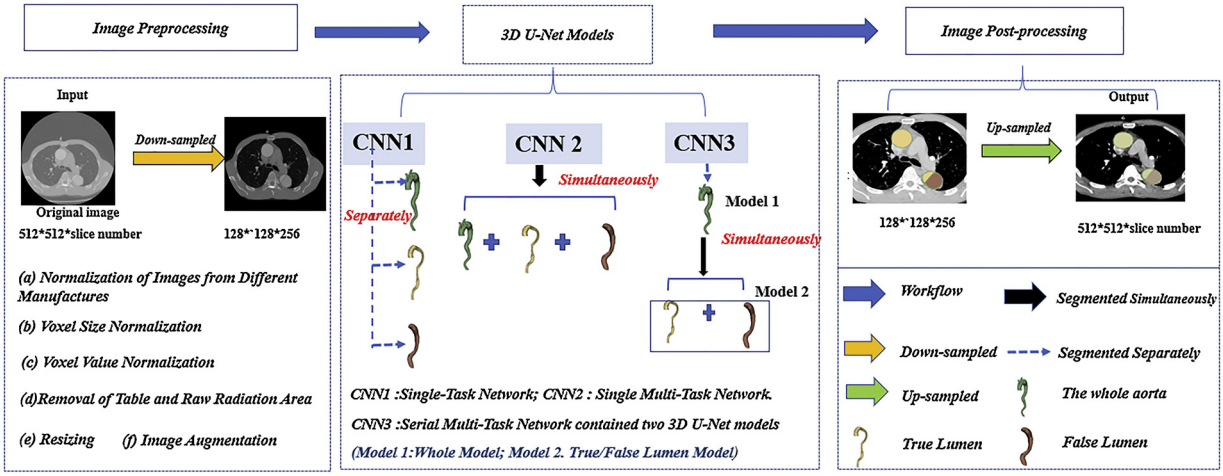

Fig. 3. Entire model of automatic TBAD segmentation, including preprocessing, U-net, and postprocessing.
TBAD = type B aortic dissection; CNN = convolutional neural network.

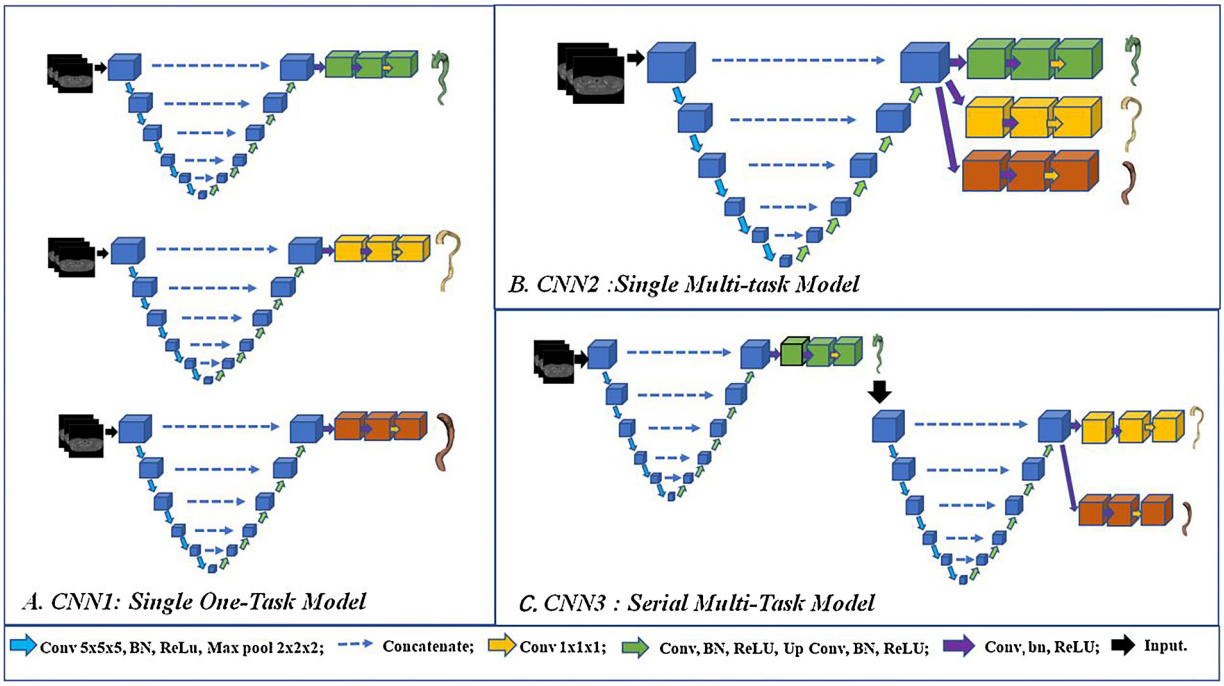

Fig. 4. CNN frameworks used in the study.
(A) CNN1, **(B)** CNN2, and (C) CNN3 frameworks. Colored arrows indicate different operations: Blue arrows = convolution with a 5 × 5 × 5 kernel, batch normalization, rectified linear unit, and 2 × 2 × 2 maximum pooling. Dashed blue arrows = concatenation from the down-sampling path to the up-sampling path. Yellow arrows = convolution with a 1 × 1 × 1 kernel. Green arrows = convolution, rectified linear unit, up-convolution, batch normalization, and rectified linear unit. Purple arrows = convolution, batch normalization, and rectified linear unit. Black arrows = input. CNN = convolutional neural network.

performances were assessed using three metrics successively : (a) the DCS of each label; (b) Precision, Recall, Sensitivity and Specificity of each label; (c) ASSD (Average symmetric surface distance) of each label (details seen in the Supplement); (d) the difference in measured volume with respect to the ground truth and volume measurement variability $\left(\left|\frac{\text{CNNa} - \text{Ground truth}}{\text{Ground truth}}\right|*100\%\right)$; and (f) CNN segmentation time.

2.5. Statistical analysis

Continuous variables are reported as means ± standard deviations. The Wilcoxon signed-rank test was used to compare two paired variables. Pearson correlation coefficient (r), intra-class correlation coefficients (ICCs), and Bland–Altman plots were used to evaluate the agreement of the inter-observer (or inter-method) measurements. ICCs > 0.8 were considered to indicate excellent agreement and P-values < 0.05 were considered statistically significant. All statistical analyses were performed using SPSS version 22.0 (IBM SPSS Inc, Chicago, Illinois).

3. Results

3.1. Patient demographics and CTA parameters

The mean age of the study patients (n = 276) was 52 years (range 31–83 years), with a significant male predominance (n = 216). Patient characteristics and acquisition parameters taken from the DICOM data are shown in Table 1.

3.2. CNN model performances

In the 3-fold cross validation sets, the mean DCSs of CNN3 showed significantly higher than other two CNN networks, either for segmenting the whole aorta, TL or for FL (Table 2). Using the 3 models generated by 3-fold cross validation for each CNN, we deployed model fusion strategy (voting strategy, that means at least 2 models determine the pixel belongs to foreground, then it will be labeled to foreground) to evaluate the testing set (n = 30).

In testing set, for the whole aorta, the mean DCS of CNN3 was 0.93 ± 0.01, which is higher than that of CNN1 (0.88 ± 0.03; P < 0.05) and CNN2 (0.83 ± 0.05; P < 0.05). Similar results were observed for the TL and FL segmentations derived from CNN3. The mean DCSs of CNN3 were 0.93 ± 0.01 and 0.91 ± 0.02 for TL and FL, respectively, indicating a significant improvement over the DSCs generated by the other two CNNs (P < 0.05). An example segmentation result from the testing set is shown in Fig. 5.

Because CNN3 obtained the best DCS performance, we evaluated its performance using the additional evaluation metric mentioned on the section "*CNN Segmentation Performance*" (seen in Supplement table S2). Besides, clinically, we further evaluated its ability to quantify the lumen volume of the whole aorta, TL, and FL by comparing its results with the ground truth of the testing set (n = 30). The ICCs and r for the volume of each label obtained by CNN3 and manual segmentation were all above 0.95 (p < 0.05; Table 3). The mean difference of label volume between the CNN3 and manual segmentation was −31.05 (−82.76 to 20.65) ml for the whole aorta volume, 4.79 (−11.04–20.63) ml for TL volume, and 8.67 (−11.40 to 28.74) ml for FL volume (Fig. 6). Compared with the ground truth on the testing set, the mean volume measurement variabilities of CNN3 were 6.66 ± 3.67%, 4.81 ± 2.48% and 5.37 ± 3.48% for the whole aorta, TL, and FL, respectively. The greatest variabilities were 13.75%, 10.44%, and 13.29% for the whole aorta, TL, and FL, respectively (see Supplemental Fig. S4).

3.3. CNN3 segmentation time

Segmentation time depended on the number of slices. Supplemental Table S3 lists all the results of the testing set. The average segmentation time per slice was 0.038 ± 0.006 s. Looking at one case as an example (case 160, 793 slices), the total computation time was 31.06 s, which includes the preprocessing with down-sampling time of 6.68 s, model

Table 1
Patient Characteristics and CTA Parameters.

		Overall (n = 276)	Training Set (n = 246)	Testing Set (n = 30)	Data Sets for segmentation protocol (n = 30)
Patient Characteristics	Male[b]	243 (88)	216 (89)	27 (90)	19 (64)
	Age range (y)	30–83	30–83	31–75	32–73
	Mean age(y)[a]	52.55 ± 11.35	53.06 ± 11.27	50.03 ± 11.54	49 ± 12.05
CTA Parameters					
	CT tube current (mA)[a]	397.19 ± 308.08	366.92 ± 274.62	525.33 ± 402.36	420.42 ± 142.24
	Peak tube voltage range (kVp)	80–140	80–130	100–140	80–120
	Pixel spacing range(mm)	0.42–0.98	0.42–0.98	0.59–0.90	0.64-0.95
	Pixel spacing(mm)[a]	0.74 ± 0.11	0.74 ± 0.12	0.73 ± 0.07	0.76 ± 0.09
	Slice thickness range(mm)	0.4–1.5	0.4–1.5	0.45–1.3	0.5-1.3
	Slice thickness(mm)[a]	0.89 ± 0.30	0.90 ± 0.31	0.86 ± 0.27	0.84 ± 0.29
	Voxel size range(mm^3)	0.09–1.15	0.09–1.15	0.24–0.78	0.21-0.96
	Voxel size(mm^3)[a]	0.50 ± 0.25	0.51 ± 0.26	0.45 ± 0.17	0.51 ± 0.24
	Voxel value range	−3067–21,660	−3067–21,660	−3024–19,435	−3067–21,660
	Manufactures[b]				
	GE Medical Systems	42(15)	37(15)	5(17)	9 (30)
	Siemens	152(55)	132(53)	18(60)	14 (47)
	Philips	67(24)	62(25)	6(20)	3 (10)
	Toshiba	15(5)	15 (6)	1 (3)	4 (13)

Note. —Voxel Size (mm^3) = Pixel spacing (mm) × Pixel spacing (mm) × Slice Thickness (mm).
[a] Data are mean ± standard deviation.
[b] Data are number of patients and data in parentheses are percentages.

computation time of 6.28 s, and postprocessing with up-sampling time of 18.1 s.

4. Discussion

Using the proposed deep learning-based models and workflow [14], we successfully demonstrated the feasibility of automatic TBAD segmentation. First, based on high-quality ground truth annotated by experts, we revealed that a CNN network with a multi-task output is capable of segmenting the TBAD into the whole aorta, TL, and FL simultaneously. CNN3 achieved the best DCSs of (0.93 ± 0.01, 0.93 ± 0.01, and 0.91 ± 0.02 for the whole aorta, TL, and FL, respectively), and calculated aortic lumen volumes that closely correspond to those of manual segmentation with only 0.038 ± 0.006 s per slice. These results indicate a significant step forward towards the automated measurement of TBAD.

The challenge of obtaining high-quality, labeled data for supervised machine learning is one of the key barriers to its adoption in radiology practice [14]. We established a rigorous manual segmentation protocol to create a ground truth database with annotations from the experts. Our validation study showed that the manual segmentation results of the experts are highly robust. Considering that the segmentation of TBAD lumens is the first step for measurement [4], we designed a manual protocol consisting of three labels to identify the lumens of interest (Label 1: the whole aorta, Label 2: TL, and Label 3: FL). The annotation region of each label completely meets the needs of TEVAR planning and standard measurement of TBAD [2,4]. In detail, Label 2 (TL) has obvious clinical significance, because its diameter directly determines the final size of the stent-graft. In addition, Label 3 (FL) is a powerful risk predictor factor for aortic remodeling. Moreover, the annotation region of Label 1 was clinically valuable because it involves the supra-arch branches, which provide the location of the proximal landing zone. In addition, it can be used as a base for intimal tear detection by subtracting Labels 2 and 3 according to our protocol (as shown in the Supplemental Fig. S1). Then, the length of the proximal landing zone, number of tears, and tear location may be calculated automatically in the future.

Using the ground truth database, we further developed and compared three CNN models. Our results show that all the CNN networks were able to segment the whole aorta, TL, and FL. However, compared with the other two methods, the serial multi-task model of CNN3 achieved the highest DCSs. This is easy to understand because the first model in the CNN3 only performs a rough segmentation that provides an accurate bounding box for the further fine segmentations of TL and FL. On the other hand, the parameters of the convolutional and deconvolutional layers in the multi-task network are shared by the two segmentation tasks for TL and FL in the second model of the CNN3, which improves the performance of each correlated task when compared with the use of several task-specific models to solve each task separately.

To further assess the segmentation accuracy and clinical performance, we compared each labeled volume generated by CNN3 to the corresponding ground truth in the testing set. The Bland–Altman plots show that, for each label, 28 out of 30 patients were within the limits of agreement. An in-depth analysis shows that the limit of agreement for TL (−11.05, 20.63) ml and FL (−11.40, 28.74) ml almost fell within

Table 2
CNN Performance—Dice Coefficient Scores Obtained by Different CNN Models on the Validation and Testing Sets.

Compartment	Label information	Different CNN Networks			*P Value*[a]	
		CNN1	CNN2	CNN3	CNN3 VS 1	CNN3 vs 2
Validation Set	Label 1: The Whole Aorta	0.86 ± 0.03	0.83 ± 0.04	0.93 ± 0.01	P < 0.01*	P < 0.01*
	Label 2: True Lumen	0.70 ± 0.07	0.68 ± 0.09	0.92 ± 0.02	P < 0.01*	P < 0.01*
	Label 3: False Lumen	0.63 ± 0.09	0.62 ± 0.06	0.91 ± 0.02	P < 0.01*	P < 0.01*
Testing Set	Label 1: The Whole Aorta	0.88 ± 0.03	0.83 ± 0.05	0.93 ± 0.01	P < 0.01*	P < 0.01*
	Label 2: True Lumen	0.71 ± 0.08	0.67 ± 0.09	0.93 ± 0.01	P < 0.01*	P < 0.01*
	Label 3: False Lumen	0.63 ± 0.09	0.63 ± 0.07	0.91 ± 0.02	P < 0.01*	P < 0.01*

Note. —Data are mean ± standard deviation. ***P Valuea***: Wilcoxon signed-rank test. CNN1 = convolutional neural network (CNN) trained with a single one-task; CNN2 = CNN trained with single multi-task; CNN3 = CNN trained by serial multi-task.
* Statistically significant difference (P <0.05).

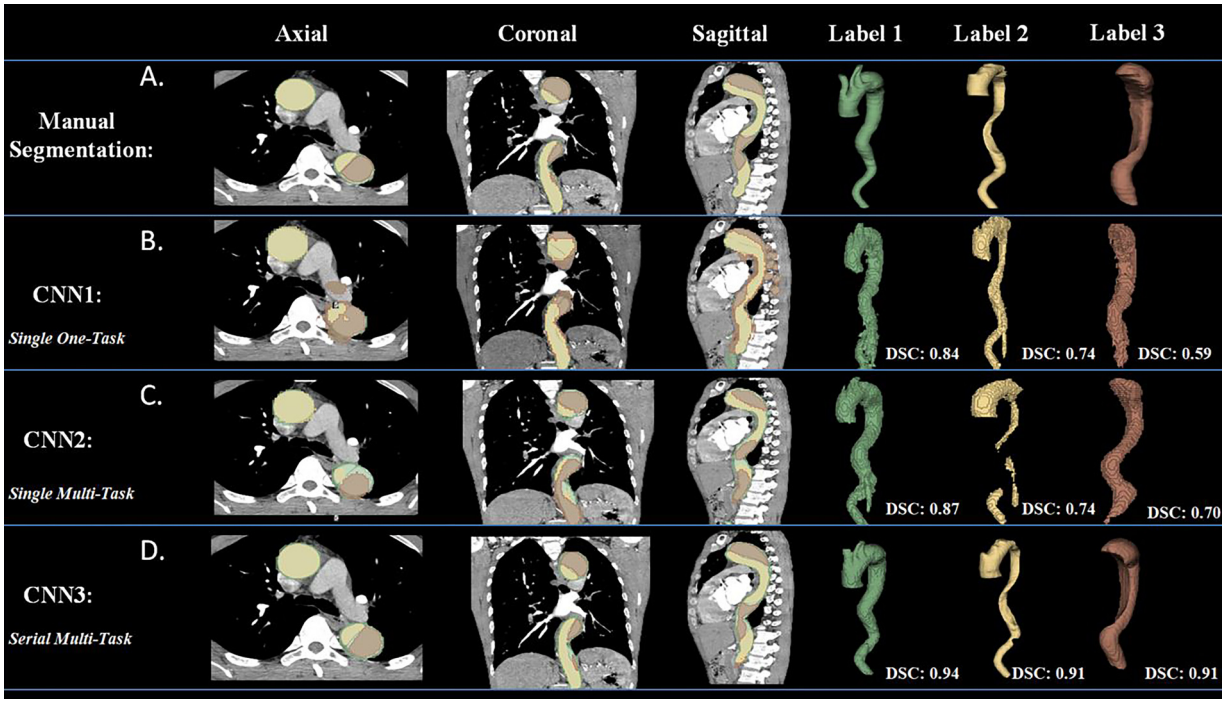

Fig. 5. Example of an automated TBAD segmentation result derived from different CNNs.
Original CTA image of a 48-year-old man obtained at GE Medical Systems (case 160 in the database). (A) Manual segmentation result. Automatic segmentation result by (B) CNN1, (C) CNN2, and (D) CNN3. CNN1 = convolutional neural network (CNN) trained using three independent one-task models; CNN2 = CNN trained using a single multi-task model; CNN3 = CNN trained using serial multi-task models. DCS = Dice coefficient score. Label 1 = whole aorta, Label 2 = true lumen, and Label 3 = false lumen.

the currently used a priori "acceptance" boundaries (−20.00, 20.00) ml for the TL volume and (−30.00, 30.00) ml for FL [4]. Clinically, the currently recognized intra-observer and inter-observer variability for volume measurements can range from 3% to 8% [2]. Our pipeline produces an acceptable variability from the manually measured volume: the means is 6.66%, 4.81%, and 5.37% for the whole aorta, TL, and FL, respectively. If 10% is used as the maximum acceptable limit in variability, as provided by a currently used workstation (3 Mensio, The Netherlands), the number of out-of-limit patients is reduced from 11 to 6, 3 to 1, and 7 to 4, for the whole aorta, TL and FL, respectively. Considering that the TL has the highest priority for clinical use, the result here is competitive. The results of the deeper analysis are also encouraging and promising. We observed that the large volume variability (above 10%) mainly occurs in dataset with uncommon anatomical features, such as huge aneurysmal dissection, extremely tortuous descending aorta and anatomic abnormalities. We believe that the result can be improved if more unusual images are collected to enhance the dataset for model training. Additionally, automatic segmentation using CNN3 is fast and makes it a valuable tool for clinical segmentation use. In our present study, the segmentation time of the maximum number of 2331 slices takes only about 62.3 s.

Compared with the current intensity-based segmentation methods, the biggest advantage of a CNN-based method is the automatic detection of different aortic lumen boundaries. Using the established lumen boundaries, the centerline of the lumen and the corresponding curved planar reformation can be easily generated without any manual adjustments. Subsequently, the lumen diameter, area, aortic length, and other anatomical features can be extracted automatically using image-processing algorithms, which accelerates the measurement process and clinical decision making.

4.1. Study limitations

First, given the low prevalence of TBAD, our training and testing datasets are small. More training data, especially for the instances of uncommon morphology that caused large volume deviations, should further improve the performance of the deep learning model. Second, although the CNN3-measured volumes lie within the experts' inter-

Table 3
CNN Performance—Volumes Derived from CNN3 and Ground Truth on the Testing Set.

Compartment	Ground Truth (ml)	CNN3 (ml)	ICC and (95%CI)	r	Mean difference(ml) and (95%CI)	Measurement variability (%)
Label 1: The Whole Aorta	500.88 ± 184.68	531.93 ± 201.93	0.978 (0.731,0.994)	0.99*	−31.05 (-82.76,20.65)	6.66 ± 3.67
Label 2: True Lumen	168.95 ± 41.12	164.15 ± 39.19	0.973 (0.919,0.989)	0.98*	4.79 (-11.04,20.63)	4.81 ± 2.48
Label 3: False Lumen	169.54 (134.69-228.43)†	166.76 (131.43-220.82)†	0.996 (0.974,0.999)	0.99*	8.67 (-11.40,28.74)	5.37 ± 3.48

Note. —Unless otherwise indicated, data are mean ± standard deviation.
Data in parentheses are 95% confidence intervals.
ICC = intra-class correlation coefficients, CI = confidence intervals, CNN3 = convolutional neural network trained by serial multi-task.
Measurement variability (%) = $\left(\left|\frac{\text{CNNa} - \text{Ground truth}}{\text{Ground truth}}\right| * 100\%\right)$.
† Data are median (interquartile range).
* Statistically significant difference (P <0 .05) was using Pearson correlation coefficients(r).

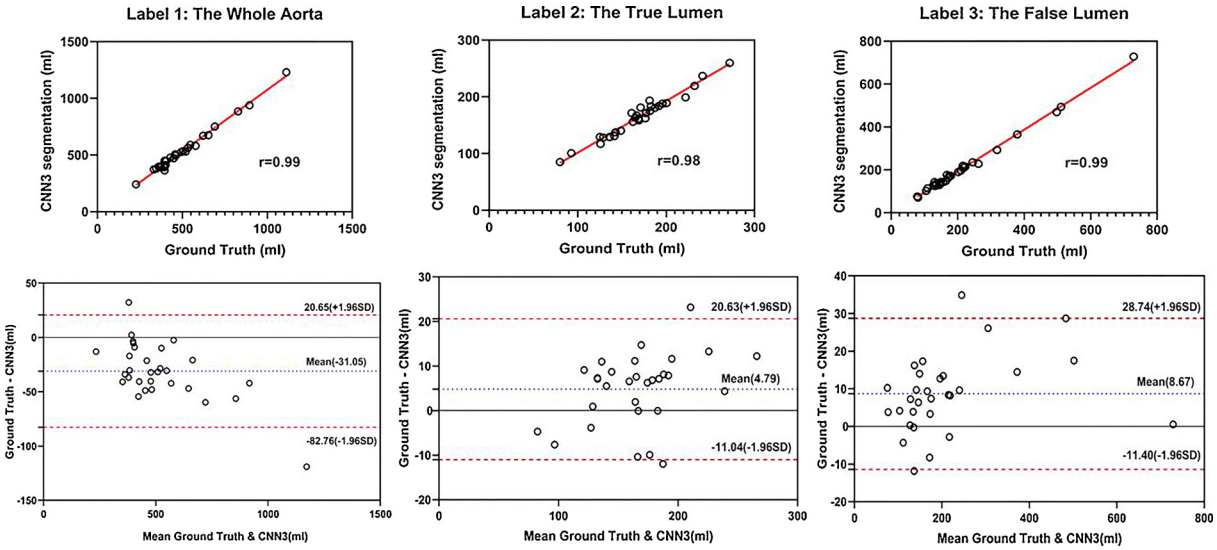

Fig. 6. Correlation (upper) and Bland–Altman (down) plots of volume parameters generated from ground truth and CNN3 segmentation.
From left to right shows Label 1 (the whole aorta), Label 2 (true lumen), and Label 3 (false lumen). CNN3 = convolutional neural network trained using serial multi-task models

observer rate, over- and under-segmentation appeared on some slices. This is partly due to the current limit of computational capability as down-sampling must be done before the data can be imported into the CNN model. For the TL, which is compressed seriously by the FL into a crescent shape, the down-sampling can lead to a disconnected TL and deteriorate the performance of deep learning models. Finally, more morphological parameters and external validation dataset outside Robust study should be used in the process of CNN development and evaluation.

In conclusion, our preliminary study demonstrated that deep learning provides a valuable tool for accurate segmentation of the whole aorta, TL, and FL in TBAD, and makes it possible for automated measurements of TBAD anatomical features. Future work will be aimed at extending this model to a larger training dataset and integrating more automatic measurement algorithms to simplify the measurement process.

Role of the funding source

This work was supported by the Beijing Municipal Natural Science Foundation, BJNSF (7181010).

Declaration of Competing Interest

The authors state that they have no conflicts interest or relationships with industry to declare.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2019.108713.

References

[1] R. Erbel, V. Aboyans, C. Boileau, E. Bossone, R.D. Bartolomeo, H. Eggebrecht, et al., ESC Guidelines on the diagnosis and treatment of aortic diseases: document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC), Eur. Heart J. 35 (41) (2014) 2873–2926.
[2] M.F. Fillinger, R.K. Greenberg, J.F. McKinsey, E.L. Chaikof, T. R. S. Society for Vascular Surgery Ad Hoc Committee on Reporting standards for thoracic endovascular aortic repair (TEVAR), J. Vasc. Surg. 52 (4) (2010) 1022–1033 1033 e15.
[3] F. Saremi, C. Hassani, L.M. Lin, C. Lee, A.G. Wilcox, F. Fleischman, et al., Image predictors of treatment outcome after thoracic aortic dissection repair, Radiographics 38 (7) (2018) 1949–1972.
[4] A.V. Kamman, J.A. van Herwaarden, M. Orrico, F.J. Nauta, R.H. Heijmen, F.L. Moll, et al., Standardized protocol to analyze computed tomography imaging of type B aortic dissections, J. Endovasc. Ther. 23 (3) (2016) 472–482.
[5] A.A. Duquette, P.M. Jodoin, O. Bouchot, A. Lalande, 3D segmentation of abdominal aorta from CT-scan and MR images, Comput. Med. Imaging Graph. 36 (4) (2012) 294–303.
[6] Y. Zheng, M. John, R. Liao, A. Nottling, J. Boese, J. Kempfert, et al., Automatic aorta segmentation and valve landmark detection in C-arm CT for transcatheter aortic valve implantation, IEEE Trans. Med. Imaging 31 (12) (2012) 2307–2321.
[7] O.C. Avila-Montes, U. Kurkure, R. Nakazato, D.S. Berman, D. Dey, I.A. Kakadiaris, Segmentation of the thoracic aorta in noncontrast cardiac CT images, IEEE J. Biomed. Health Inform. 17 (5) (2013) 936–949.
[8] J.A. Martinez-Mera, P.G. Tahoces, J.M. Carreira, J.J. Suarez-Cuenca, M. Souto, A hybrid method based on level set and 3D region growing for segmentation of the thoracic aorta, Comput. Aided Surg. 18 (5–6) (2013) 109–117.
[9] L. Saba, M. Biswas, V. Kuppili, E. Cuadrado Godia, H.S. Suri, D.R. Edla, et al., The present and future of deep learning in radiology, Eur. J. Radiol. 114 (2019) 14–24.
[10] K. Lopez-Linares, N. Aranjuelo, L. Kabongo, G. Maclair, N. Lete, M. Ceresa, et al., Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks, Med. Image Anal. 46 (2018) 202–214.
[11] D. Rong, Y. Ge, Y. Xue, F. Liu, K. Lu, P. Liu, et al., Protocol for the ROBUST (Registry Of type B aortic dissection with the Utility of STent graft) study: an ambispective, multicentre, open cohort study, BMJ Open 7 (12) (2017) e019317.
[12] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, et al., 3D slicer as an image computing platform for the Quantitative Imaging Network, Magn. Reson. Imaging 30 (9) (2012) 1323–1341.
[13] F.P. Ronneberger O, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, MICCAI, 2015.
[14] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A.E. Samir, O.S. Pianykh, et al., Current applications and future impact of machine learning in radiology, Radiology 288 (2) (2018) 318–328.