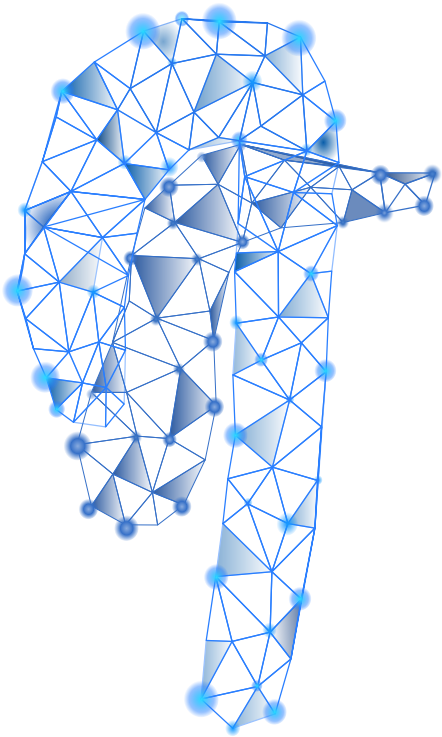
Automatic Quantification of the Aorta and Pulmonary Artery in Chest CT

methods and validation in lung screening

Zahra Sedghi Gamechi

Automatic Quantification of the Aorta and Pulmonary Artery in Chest CT

methods and validation in lung screening



Zahra Sedghi Gamechi

The work presented in this thesis was conducted at the Department of Radiology & Nuclear Medicine of the Erasmus MC, Rotterdam, the Netherlands. In addition, the work was carried out in the ASCI and COEUR graduate schools.

For financial support for the publication of this thesis, the following organizations are gratefully acknowledged: the Department of Radiology & Nuclear Medicine (Erasmus MC) and Almende B.V.

ISBN	978-94-6423-355-1
Cover Design	Zahra Sedghi Gamechi
Thesis Layout	Zahra Sedghi Gamechi
Printing	ProefschriftMaken

Automatic Quantification of the Aorta and Pulmonary Artery in Chest CT

methods and validation in lung screening

Automatische kwantificatie van de aorta en longslagader in CT-thorax

methoden en validatie bij longscreening

Thesis

to obtain the doctoral degree at the
Erasmus University Rotterdam
upon the authority of the
Rector Magnificus

Prof. dr. F.A. van der Duijn Schouten

and in accordance with the decision of the Doctorate Board.
The public defense shall be held on

Tuesday 21 September 2021 at 15.30 hours

by

Zahra Sedghi Gamechi
born in Urmia, Iran

Erasmus University Rotterdam

Doctoral Committee

Promotors	Prof. dr. M. de Bruijne	
	Prof. dr. W.J. Niessen	
Other members	Prof. dr. A. van der Lugt	(Erasmus Medical Center)
	Prof. dr. I. Išgum	(Amsterdam University Medical Center)
	Dr. R. San José Estépar	(Harvard Medical School)

To my precious Father,
my lovely Mother,
and my dear Husband.

Contents

Contents **i**

1 General Introduction **4**
1.1 Anatomy 4
1.2 Associated Diseases 7
1.3 Thoracic imaging 8
1.4 Challenges in Quantitative Analysis of the Aorta & Pulmonary artery 11
1.5 Segmentation Methods 11
1.6 Outline and Contributions of this Thesis 13

2 Automated 3D segmentation and diameter measurement of the thoracic aorta on non-contrast enhanced CT **18**
2.1 Introduction 19
2.2 Materials and Methods 19
2.2.1 Study Population & Image Acquisition 19
2.2.2 Manual Annotation 20
2.2.3 Automatic Aorta Segmentation Approach 22
2.2.4 Aortic Diameter Measurement 23
2.2.5 Validation and Statistical analysis 24
2.3 Results 24
2.4 Discussion 28
2.A Supplemental Material 34

3 Growth of the thoracic aorta in the smoking population: The danish lung cancer screening trial **40**
3.1 Introduction 41
3.2 Methods 41
3.2.1 Study Population 41
3.2.2 Computed Tomography Imaging 42
3.2.3 Measurements of Aortic Diameter 42
3.2.4 Statistical Analysis 43
3.3 Results 44
3.3.1 Study population 44
3.3.2 Aortic Diameters and Aortic Growth 44
3.3.3 Determinants of Aortic Growth 45

3.4 Discussion . . . 46
3.4.1 Determinants of Aortic Growth . . . 48
3.4.2 Limitations . . . 49
3.5 Conclusion . . . 50
3.A Supplemental Material . . . 51

4 Optimal surface graph cuts to segment the pulmonary artery and aorta on non-contrast CT 56
4.1 Introduction . . . 57
4.2 Methods . . . 58
4.2.1 Pre-processing . . . 58
4.2.2 Seed point Extraction . . . 58
4.2.3 Centerline Extraction . . . 60
4.2.4 Vessel Segmentation . . . 61
4.2.5 Diameter Measurement and PA:AA Ratio . . . 62
4.3 Experiments . . . 63
4.3.1 Dataset and Manual Annotation . . . 63
4.3.2 Parameter Optimization . . . 64
4.3.3 Seed point and Centerline Extraction . . . 64
4.3.4 Segmentation . . . 65
4.3.5 Repeatability, Diameters, and PA:AA Ratio . . . 65
4.4 Results . . . 65
4.5 Discussion . . . 67
4.6 Conclusion . . . 72
4.A Supplemental Material . . . 73

5 An end-to-end approach to segmentation in medical images with CNN and Posterior-CRF 76
5.1 Introduction . . . 77
5.2 Related Work . . . 80
5.3 Methods . . . 80
5.3.1 CNN Model . . . 82
5.3.2 Conditional Random Fields . . . 82
5.3.3 CRF with Predefined Features . . . 83
5.3.3.1 CRF as Post-processing . . . 84
5.3.3.2 End-to-end Training CRF . . . 84
5.3.4 Proposed CRF with Learning-based Features . . . 84
5.3.4.1 Back-propagation of the Learning-based CRF . . . 84
5.4 Experiments . . . 85
5.4.1 Implementation . . . 85
5.4.1.1 CNN Implementation . . . 85
5.4.1.2 CRF Implementation . . . 86
5.4.2 CRF Settings . . . 86
5.4.2.1 Post-processing CRF . . . 86
5.4.2.2 End-to-end CRF with Predefined Features . . . 86
5.4.2.3 End-to-end CRF with Learned Features . . . 86

5.4.2.4 CRF Parameters . . . 87
5.4.3 Datasets and Preprocessing . . . 88
5.4.3.1 CT Arteries Dataset . . . 88
5.4.3.2 MRI White Matter Hyperintensities (WMH) Dataset 88
5.4.3.3 MRI Ischemic Stroke Lesions (ISLES) Dataset . . . 88
5.4.4 Data Augmentation and Training Details . . . 90
5.4.5 Evaluation Metrics . . . 90
5.5 Results . . . 90
5.5.1 Segmentation Results . . . 90
5.5.2 Optimization of the End-to-end CRF . . . 94
5.5.3 Influence of CRF Hyperparameters . . . 94
5.5.4 Influence of Hierarchical CNN Features as CRF Reference Maps 95
5.5.5 Evolution of CNN and CRF Outputs . . . 96
5.6 Discussion . . . 96
5.6.1 Interaction between CRF and CNN . . . 97
5.6.2 Posterior-CRF vs. Mean-field Network . . . 99
5.6.3 Posterior-CRF vs. Graph Neural Networks . . . 100
5.6.4 Limitations . . . 100
5.7 Conclusions . . . 101
5.A Mean-field Inference . . . 102
5.B SHAP Analysis of Post-processing CRF . . . 102

6 General Discussion 108
6.1 Summary . . . 108
6.2 Optimal Surface Graph Cut based Segmentation Method . . . 108
6.3 Posterior-CRF Segmentation Method . . . 109
6.4 Diameter Measurement . . . 110
6.5 General Discussion & Future Directions . . . 112
6.6 Conclusion . . . 115

Bibliography 117

Acronyms 131

Lay Summary (English Summary) 133

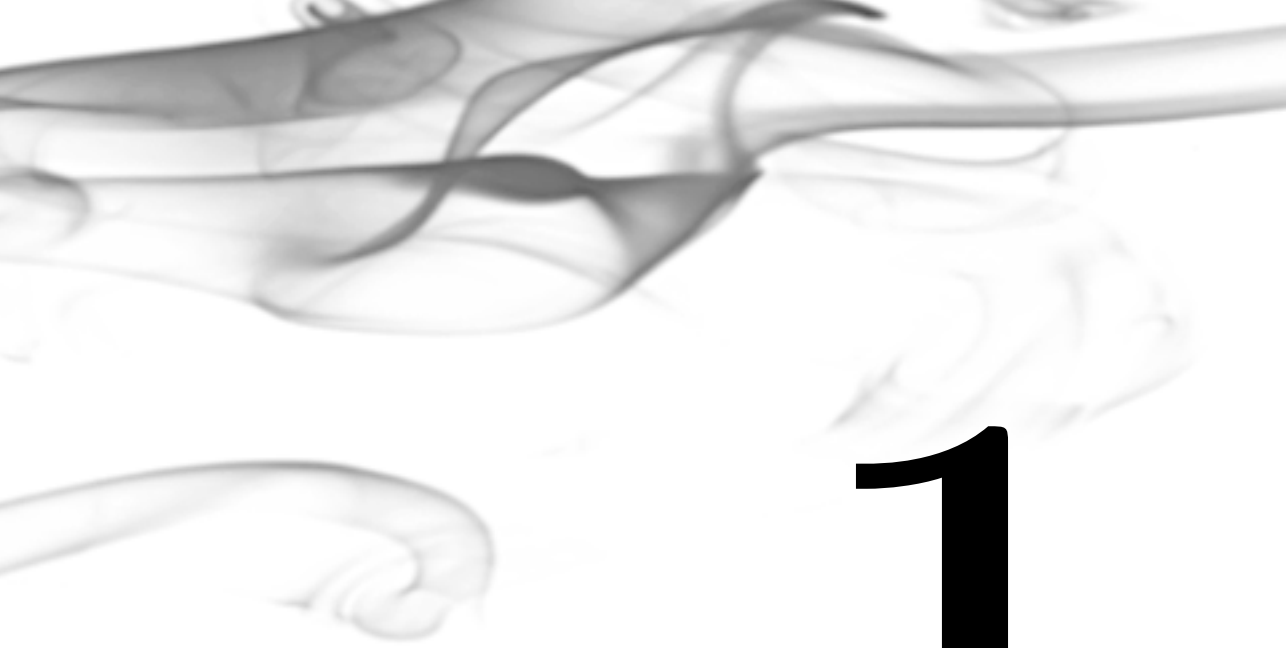
Samenvatting (Dutch Summary) 137

Acknowledgement 141

Publications 145

PhD Portfolio 147

About the author 151



1

General Introduction

Cardiovascular diseases and Chronic Obstructive Pulmonary Disease (COPD) are among the major leading causes of death globally [1]. Patients with COPD are at increased risk of cardiovascular disease [2, 3], while the prognosis in COPD is greatly affected by the presence of cardiovascular disease [4, 5]. Noninvasive imaging techniques such as Computed Tomography (CT), together with quantitative image analysis, play an increasingly important role in investigating the clinical and pre-clinical stage of cardiovascular disease. In CT scans used for lung cancer screening [3, 6–18], both the heart and the lungs are visualized. Therefore in such scans, it might be possible to identify both cardiovascular disease and COPD.

The aorta and the pulmonary artery are the two largest arteries in the chest. Changes in the shape and the size of these arteries are associated with several cardiovascular diseases [19–24] and with the increased risk of severe exacerbations and mortality in patients with COPD [9, 25, 26]. Assessing abnormalities in the size and shape of these arteries requires diameter measurements in which measurements derived from 3D segmentations are most reliable. However, performing such measurements manually is labor-intensive and time-consuming. Therefore fully automated 3D segmentation and subsequent diameter analysis are desirable.

This thesis focuses on automated image analysis to characterize the shape and diameters of the aorta and pulmonary artery to facilitate clinical and epidemiological research, assist in the early-stage diagnosis of aortic aneurysm, and extract risk factors for the exacerbation of COPD. This chapter provides a background on the anatomy of the aorta, pulmonary artery, and surrounding structures (Section 1.1). Subsequently, the disease associated with these vessels (Section 1.2), chest imaging modalities (Section 1.3), image processing challenges (Section 1.4), and vascular segmentation techniques (Section 1.5) are discussed, followed by summarizing the contributions and outlining the contest of this thesis (Section 1.6).

1.1 Anatomy

The aorta and pulmonary artery are the two major arteries in the human body that carry blood away from the heart. The aorta is the biggest artery and is responsible for transporting oxygenated blood from the heart's left ventricle to the rest of the body. The pulmonary artery is responsible for carrying the deoxygenated blood from the heart's right ventricle to the lungs for oxygenation.

The aorta begins at the bulb-shaped root originating from the left ventricle at the aortic valve level and then courses through the chest and abdomen in a candy cane–shaped configuration (Figure 1.1). The thoracic aorta is part of the aorta located in the chest (thorax) and includes the aortic root, ascending aorta, aortic arch, and descending aorta. The part of the aorta that passes the diaphragm and goes through the abdomen is called the abdominal aorta. A depiction of the aorta anatomy is illustrated in Figure 1.1.

The aorta normally has a diameter of approximately 2 cm. It has a slightly larger diameter at the aortic root. As the aorta descends into the abdomen, it narrows progressively. The aortic root consists of three sinuses of Valsalva, also known as aortic sinuses, which give rise to coronary arteries. The junction of the aortic root to the tubular part of the ascending aorta is called the sinotubular junction. The ascending

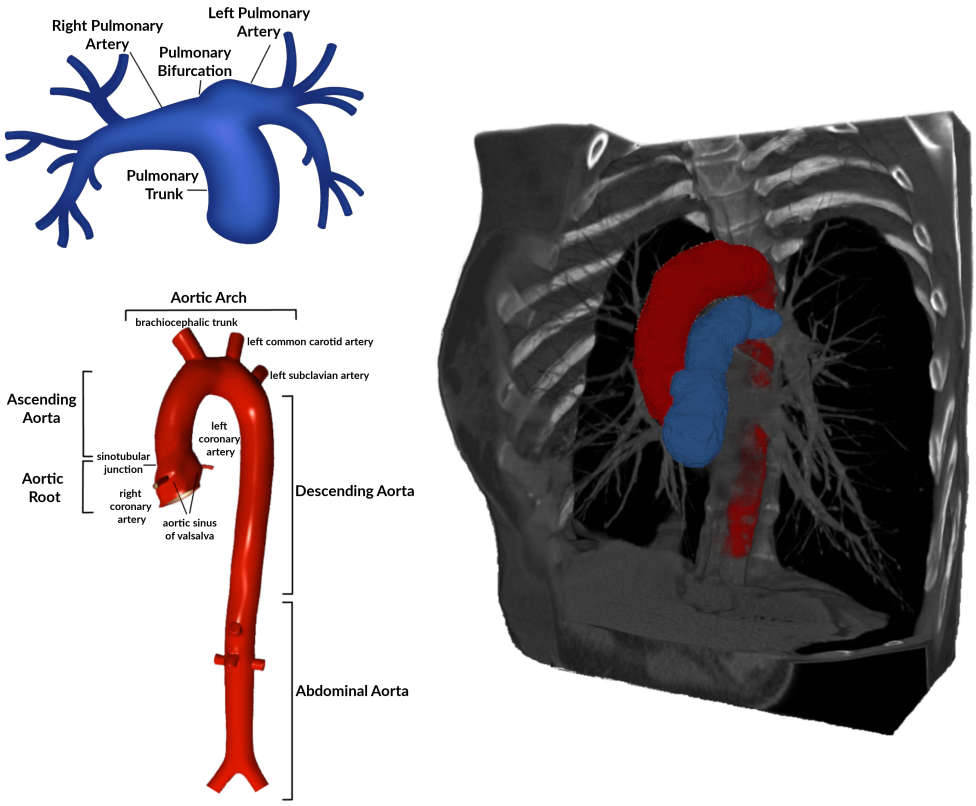


Figure 1.1: Illustration of the aorta and pulmonary artery anatomy and their position in the body in a 3D view.
source of the images in left: Radiopaedia.org;
source of the image in right: this author.

aorta arises from the sinotubular junction, wherefrom the left side is adjacent to the pulmonary artery trunk, and arches back over the right pulmonary artery to the posterior part of the chest and becomes the aortic arch. Normally three major branches arise from the aortic arch, the brachiocephalic artery, the left common carotid artery, and the left subclavian artery. These vessels supply blood to the upper body. The descending aorta extends from the aortic arch and descends downwards towards the diaphragm. Behind the descending thoracic aorta is the vertebral column.

The pulmonary artery consists of the trunk, left, and right pulmonary arteries, which are relatively wide and short. The pulmonary trunk or the main pulmonary artery normally has a diameter of approximately 2-3 cm and is approximately 5 cm long. The pulmonary trunk originates from the bottom of the heart's right ventricle and ascends towards the aortic arch, where it is adjacent to the ascending aorta. Below the aortic arch, the pulmonary trunk bifurcates in a "Y" shape into the left

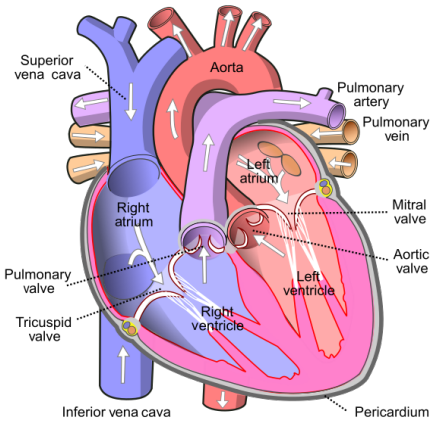


***Figure 1.2: Cardiac anatomy**. Source: Wikimedia: Diagram of the human heart. https://commons.wikimedia.org.*

and right pulmonary arteries, each of which directs the blood to the corresponding lung. This main branching (pulmonary bifurcation) is located above the heart to the left of the ascending aorta. Both left and right pulmonary arteries divide into smaller branches after they enter the lungs. In this thesis, we will only consider the left and right pulmonary arteries before their secondary bifurcation and mention them as the main branching arteries of the pulmonary artery. A depiction of the pulmonary artery anatomy is illustrated in Figure 1.1.

The aorta and pulmonary artery are surrounded by other structures. Under the aortic arch, the carina exists, which is the end of the trachea and where the trachea branches. The pulmonary trunk lies to the left of the carina, and to the right lies the descending aorta. Behind the descending aorta is the vertebral column. The descending aorta is located between the vertebral column and the heart with close proximity to the left atrium of the heart see Figure 1.3. The heart consists of four chambers; two lower chambers named ventricles, where the aorta and pulmonary artery arise from, and two upper chambers named atria, where the superior vena cava and pulmonary veins are connected. From the right atrium superior vena cava arises to the right of the ascending aorta. On the surface of each atrium, there is a small muscular flap named auricle, which resembles an ear or earlobe. The auricle on the upper wall of the left atrium is adjacent to the pulmonary artery trunk and where it bifurcates to the right pulmonary artery. Figure 1.2 illustrates an overview of the cardiac anatomy where the valve of the aorta and pulmonary artery and the surrounding structures in the heart can be seen.

Below the pulmonary artery bifurcation, the pulmonary veins enter the posterior part of the heart's left atrium. Besides these anatomical structures, epicardial fat exists around the heart and its arteries. In an axial view as shown in Figure 1.3 aorta has almost a circular shape and the pulmonary artery generally has a round-elliptic shape.

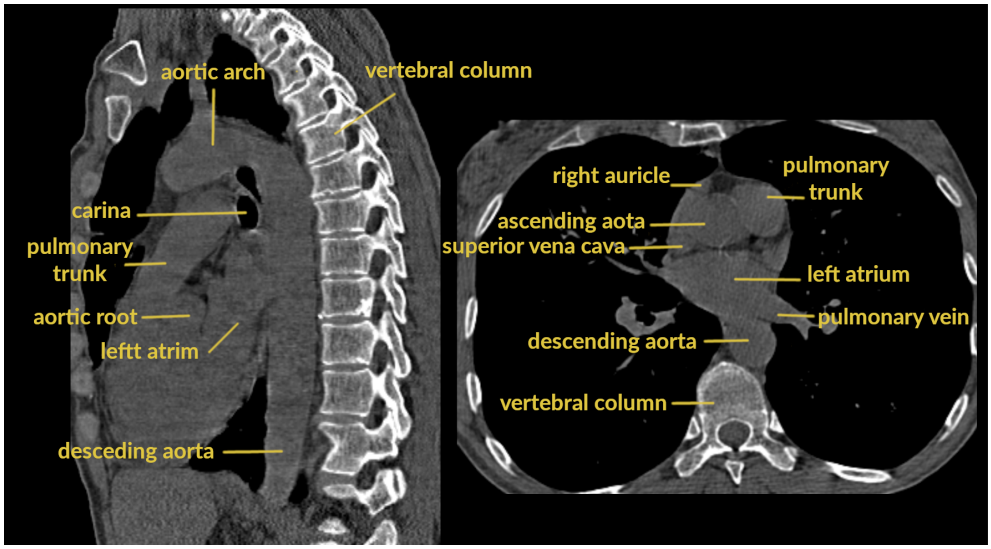


Figure 1.3: Sagittal and axial view of a non-ECG-gated non-contrast CT.

1.2 Associated Diseases

Chronic Obstructive Pulmonary Disease (COPD) is a group of lung diseases characterized by chronic airflow obstruction associated with enhanced inflammation in the airways and lungs. Cardiovascular disease is a general term for conditions affecting the heart or blood vessels. COPD and cardiovascular disease frequently occur together, and their coexistence is associated with worse outcomes than either condition alone [27].

Changes in the size of the aorta may indicate aortic dilatation, aortic aneurysm [21], and coarctation of the aorta [24]. The aorta is considered dilated if its diameters exceed the norms for a given age and body size. An aortic aneurysm is a permanent, localized abnormal dilatation or bulging of the aorta, having at least a 50% increase in diameter compared with the expected normal diameter [21]. In patients with aortic aneurysms, the aortic size has a profound impact on the risk of dissection [28, 29]. Most patients with a dilated aorta or aortic aneurysm are asymptomatic, and these conditions are not often detected by physical examination. The diagnosis usually is made during screening for aortic aneurysm in the context of positive family history or by coincidence on imaging examinations performed for other purposes like lung cancer screening or when a complication occurs, such as aortic dissection or rupture [30, 31]. In this last group of patients, the aortic dissection is often the first presentation and often results in death. Aortic dilatation and aneurysm are cardiovascular diseases observed in patients with COPD, with smoking as a common risk factor.

Due to this silent process with high risks associated with aortic aneurysm, detecting the aortic dilatation at an early stage is desired. Accurate assessment of the aortic diameter is a key component in detecting aortic aneurysms and guiding therapeutic

decisions in which the risk of dissection, rupture, and death is estimated [22]. In addition, aortic diameter is a major criterion for recommending elective operation where detecting aortic dilatation at an early stage enables preventive surgery, which might save lives.

Narrowing or constriction of the aorta, typically in the aortic arch or descending aorta, is called coarctation of the aorta. Coarctation of the aorta is a common congenital anomaly where long-term complications such as aortic aneurysms can develop from untreated or treated coarctation [23, 24]. Aortic aneurysm associated with coarctation in adults could remain asymptomatic for a prolonged time and can threaten their lives [32]; therefore, early detection of aortic coarctation is important [33].

Pulmonary hypertension is a common complication of COPD; such a complication is associated with increased risks of exacerbation and decreased survival [34]. Pulmonary hypertension represents a chronic condition characterized by increased blood pressure in the pulmonary circulation, which can cause structural problems like aneurysm or dissection of pulmonary arteries. Therefore, the location and severity of enlargement in the diameter of the pulmonary artery on CT play an important role in diagnosis and guide the clinician in the management of pulmonary hypertension [35].

Dilatation of the main pulmonary artery or its main branching arteries is associated with a pulmonary aneurysm [36, 37] and is an important metric for the presence of pulmonary hypertension [19, 20, 35]. Dilatation of the pulmonary artery is often the first imaging finding to suggest the diagnosis. The dilatation and aneurysm in the pulmonary artery are rare abnormalities with life-threatening complications such as pulmonary artery dissection. Like aortic aneurysms, pulmonary artery aneurysms can frequently be asymptomatic and are incidentally diagnosed on imaging performed for other reasons [36].

Moreover, the ratio of the diameter of the pulmonary artery to the diameter of the ascending aorta at the level of pulmonary artery bifurcation (PA:AA) is associated with the presence of pulmonary hypertension [20] and is shown to be a strong predictor for severe exacerbation [9, 25], and increased mortality [26] in patients with COPD. It is shown that PA:AA is a better predictor for pulmonary artery pressure in patients with primary pulmonary hypertension than only the diameter of the main pulmonary artery and is used to measure the dilatation of the pulmonary artery [38].

1.3 Thoracic imaging

Medical imaging plays a significant role in disease prevention, early detection, diagnosis, and treatment. Computed tomography (CT) and magnetic response imaging (MRI) are the common noninvasive imaging modalities used to visualize cardiac structures.

CT is a projection-based imaging modality that uses tomographic reconstruction algorithms on X-ray projections to generate 3D images of the body. In CT, cross-sectional images generated from the combination of multiple X-ray projections acquired at many different orientations around the body are reconstructed into a 3D image volume. Based on the attenuation of X-rays in various tissue types, CT provides detailed gray-scale images and an accurate density of any part of the body, including the bones, muscles, fat, organs, and blood vessels. These densities are expressed using

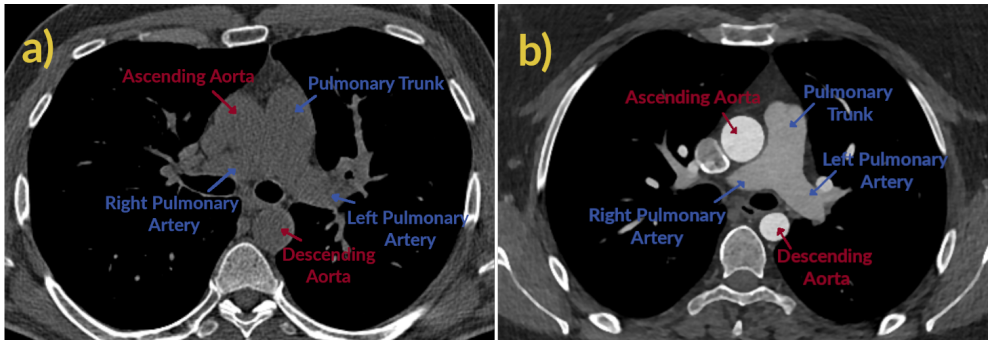


Figure 1.4: CT vs. CTA. *(a) Axial slice of a low-dose non-ECG-gated non-contrast CT scan where the boundaries between the arteries are blurred with a high noise level. (b) Axial slice of a CTA scan where the boundary of the arteries is clearly distinguishable.*

the Hounsfield Unities (HU), with water with zero HU value. Tissues with a higher density than water, such as bone (700 to 3000 HU) and blood (45 to 65 HU), have positive HU and are visualized in brighter colors, while structures with less density than water such as air (-1000 HU) and fat (-30 to -70 HU) appear darker with negative HU. Currently, with the use of multi-detector and multi-source CT (MDCT), the temporal and special resolution of CT imaging has increased [39].

Chest CT often is the imaging modality of choice for diagnosis and follow-up of patients with aortic pathology and the study of lung disease and pulmonary vasculature. CT with faster acquisition time, better isotropic spatial resolution, convenience, and easier access than MRI, is widely used in clinical practice. Despite the vast advantages of CT imaging, the risk of exposure to ionizing radiation is a potential limitation. Ionizing radiation can damage cells and slightly increase the risk of developing cancer. Therefore, current CT imaging and image analysis improvements are often aimed at significant dose-reduction for providing a safer imaging procedure [40]. Lowering the dose leads to the reduction of the image quality by increasing the noise level. CT angiography (CTA) improves the depiction of the vasculature by the administration of iodinated contrast media. However, though rare, the contrast agent can produce undesired side effects such as allergic reactions and kidney damage. Therefore, it is desired to avoid contrast when the vasculature is likely to be visible on non-contrast-enhanced CT. Low-dose non-contrast CT provides an acceptable image quality with a lower risk of exposure to ionizing radiation and fewer side effects of contrast agents.

An axial slice of a CTA and a low dose non-contrast CT are shown in Figure 1.4. In CTA, the anatomical structures are clearly defined and delineated, whereas in the non-contrast CT, there is more noise and the border between the vessels and the surrounding structures is unclear.

MRI is an imaging modality that utilizes strong magnetic fields and radiofrequency waves to construct detailed 3D image volumes of the body. Even though MRI has no risk of ionizing radiation and provides better soft-tissue contrast, it has a lower in-plane spatial resolution and is a more expensive modality. Therefore, CT is generally

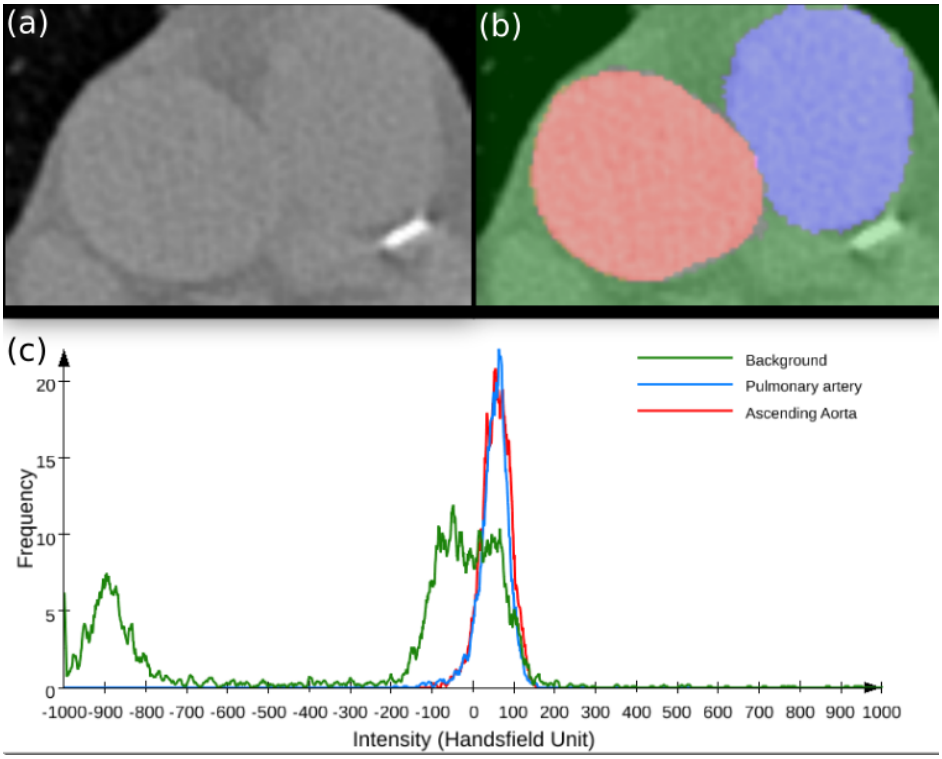


Figure 1.5: Intensity distribution in non-contrast CT. *(a) Axial slice at a region close to the heart chamber. (b) Scan overlaid with manual annotations of the aorta in red, the pulmonary artery in blue, and the surrounding structures as background in green. (c) The intensity distribution of voxels in the aorta (red), pulmonary artery (blue), and background (green).*

preferred and widely used in clinical practice.

Many patients with COPD or at risk of developing cardiovascular disease undergo a low-dose, non-ECG-gated, non-contrast thoracic CT for lung cancer screening. Electrocardiogram (ECG) gating is the act of monitoring and collecting heart rate during the scan to ensure scanning only between the heartbeats to achieve a motion-free scan. However,this procedure requires the acquisition of highly overlapping slices, which exposes patients to a higher radiation dose. Therefore it is common to perform low-dose, non-ECG-gated, non-contrast CT for lung cancer screening to reduce the risk of radiation exposure [6, 18]. With the growing use and widespread availability of such scans, there is an opportunity to measure the aorta and pulmonary arteries in such scans to investigate the presence of early-stage cardiovascular disease and/or predict complications in patients with COPD.

1.4 Challenges in Quantitative Analysis of the Aorta & Pulmonary artery

Manual analysis of the vessels in cardiac CT requires clinical expertise to identify and locate the correct structures. Since the aorta and pulmonary artery are large vessels that can be affected by pathology at multiple locations along all their length, the entire vessel in the chest should be imaged and measured. A complete 3D analysis obtains the most accurate quantification of the aorta and pulmonary artery dimensions. Manual 3D analysis requires drawing the vessel contours in a large number of reformatted slices perpendicular to the vessel axis. This is a very labor-intensive and time-consuming process and subject to intra- and inter-observer variability. Also, the human workload is a limiting factor when performing quantitative measurements in large-scale imaging studies and clinical practice. Therefore, automated analysis is desirable and could add significant value in the standardization and repeatability of vessel segmentation and subsequent diameter analysis.

Automatic segmentation of the aorta and pulmonary artery in non-ECG-gated, non-contrast CT scans is a challenging task. The main challenge is the unclear vessel boundaries due to the lack of contrast between blood pool regions, muscle walls, and fat (see Figure 1.5). In addition, the proximity of the aorta and pulmonary artery to each other and other structures with similar intensity values, such as the heart, intensifies the difficulty in vessel boundary detection. Figure 1.5 illustrates the intensity similarity between the vascular area and the surrounding voxels in the region close to the heart chamber. The obvious overlap of the intensity distributions of the background and voxels within the aorta and pulmonary artery shows the difficulty in segmenting the vascular area from the background. Furthermore, the similarity in the aorta and pulmonary artery intensity distribution and the unclear boundary between them (Figure 1.5.a) demonstrates the difficulty in detecting the correct border of each vessel. Additionally, in non-ECG-gated CT scans, the existence of motion artifacts caused by the heart's motion during the cardiac cycle at the regions close to the heart makes the automatic segmentation even more challenging where manual measurements are difficult even for experienced radiologists.

1.5 Segmentation Methods

Automatic and semi-automatic segmentation of the aorta and pulmonary artery has been the topic of studies in MRI [41] and CTA [42–52]. Most of these studies obtain reasonable segmentations in the high contrast scans. However, these methods do not translate well to non-contrast CT scans where the vessel boundaries are not well defined in many places. Therefore, relatively few studies can be found on non-contrast CT scans, mostly on the segmentation of the aorta [53–63] and a few on pulmonary arteries containing the pulmonary trunk, left, and right pulmonary arteries [64–66].

Different segmentation methods for the aorta and pulmonary artery have been presented in the literature, such as multi-atlas based models [53], level-set based methods [60], active shape models [55], and deep learning based methods [61–63, 66]. Among these presented segmentation methods, methods that use shape priors in some form generally have gained a better performance.

Also, graph cut methods have been applied to different imaging modalities for artery segmentation [67, 68] and have obtained promising results in many tasks. Deng et al. [67] proposed a graph-cut method using random forest based discriminative features on non-contrast CT for aorta segmentation. They achieved a high segmentation performance in the abdominal aorta; however, they have not applied their method to the thoracic aorta and pulmonary artery, which are more challenging to segment. Another approach is to use graph cut methods with shape priors. Graph cuts can achieve a global optimum with low processing times, and it is possible to incorporate shape or smoothness constraints when designing the graph structure. This thesis will present an optimal surface graph cut approach similar to Petersen et al. [69], initially proposed for airway segmentation, for pulmonary artery and aorta segmentation. This method incorporates a shape prior via constructing the graph based on flow lines traced from an initial, smoothed segmentation. The non-intersecting flow lines guarantee non-self-intersecting surfaces and make it possible to segment high curvature areas such as the bifurcation of the pulmonary artery and the aortic arch while guaranteeing a shape that is similar to the initialization shape.

Besides the high-performing graph-cut techniques for medical image segmentation, deep learning techniques are being used widely in recent years [51, 61, 62, 66, 70–72]. Deep learning based algorithms have the advantage of directly learning from data in an end-to-end fashion using a general-purpose learning procedure. With the increasing amount of data to learn from, deep learning based algorithms have gradually outperformed previous methods. Chen et al. [70] have compared current techniques for segmenting cardiovascular structures in different imaging modalities where deep Convolutional Neural Networks (CNNs) have achieved remarkable performance and shown great success in many segmentation tasks. However, few studies are performed on aorta segmentation on non-contrast CT scans [61, 62, 66], and to the best of our knowledge, there are no deep learning based algorithms presented for the segmentation of the pulmonary artery trunk and its major branches on non-contrast CT scans.

CNNs are supervised neural networks that are powerful in extracting local features and perform good predictions. However, the lack of context information for modeling interactions and relations between nearby objects can result in poorly segmented boundaries. To address this challenge, in this thesis, we will present an end-to-end training method based on the combination of a CNN (such as U-net) with a Conditional Random Field (CRF). CRFs are probabilistic graphical models that model the correlations and dependencies among the voxels being predicted. Although it is common to combine CNN with a CRF to refine the voxel-level predictions made by CNN, CRF based on predefined features such as intensity similarity was often used as a post-processing technique. However, intensity-based information alone provides a low-quality feature space for the CRF due to the intensity similarity between vessels and surrounding structures. Therefore, in this thesis, the presented end-to-end method named Posterior-CRF allows the CRF to use features learned by a CNN, optimizing the CRF and CNN parameters concurrently. In this thesis, the proposed Posterior-CRF is applied for jointly segmenting the aorta and pulmonary artery on non-contrast CT scans.

1.6 Outline and Contributions of this Thesis

This thesis focuses on developing and validating techniques to automatically segment the aorta and pulmonary artery and, subsequently, measure diameters on non-ECG-gated, non-contrast CT scans. The main contributions of this thesis can be divided into methodological and translational contributions:

– ***Methodological***

– *Chapter 2, 4* Adapt an optimal surface graph cut approach for segmenting the aorta and pulmonary artery in non-contrast non-ECG-gated CT scans.

– *Chapter 2, 4* Present a method to fully automatically extract a landmark for the pulmonary artery bifurcation level and the seed points of the aorta, including the ascending and descending aorta, and pulmonary artery including the pulmonary trunk, left and right pulmonary arteries.

– *Chapter 4* Develop a robust, reproducible, and fully automatic 3D volumetric diameter measurement technique for PA:AA biomarker extraction.

– *Chapter 5* Develop a new end-to-end trainable algorithm for image segmentation that uses CNN-learned features in a CRF and simultaneously optimizes the CRF and CNN parameters.

– ***Translational***

– *Chapter 2* develop an automatic tool to measure the aortic diameters perpendicular to vessel centerline with no human interaction in several cross-sectional levels.

– *Chapter 2* Validate the accuracy of the optimal surface graph cut method and the repeatability of the aortic diameter measurement technique on a large cohort from the Danish Lung Cancer Screening Trial.

– *Chapter 3* On longitudinal data, study the sex-specific distribution of the aortic diameters and the aortic growth in a large population of current or former smokers.

– *Chapter 3* study the association of aortic growth with clinical characteristics regarding smoking status, history of stroke, hypertension or hypercholesterolemia, and Agatston calcium scores.

The outline and the structure of this thesis are as follows:

Chapter 2 presents a fully automatic method based on optimal surface graph-cuts for the segmentation of the aorta on non-ECG-gated, non-contrast CT scans.

The method is developed and evaluated using data from the Danish lung cancer screening trial. From the extracted 3D aorta segmentation, the diameter of the ascending and descending aorta are calculated at cross-sectional slices perpendicular to the extracted centerline, at multiple, fixed levels relative to the pulmonary artery bifurcation level. The method's accuracy is then evaluated by comparing the automatic 3D aorta segmentation and diameter measurements with the manual measurements. Finally, the repeatability of the diameter measurements is evaluated on scan-rescan pairs. The results show that this method is a promising technique to accurately and reproducibly assess subtle signs of aorta dilatation in non-ECG-gated, non-contrast CT scans without any human interaction.

Chapter 3 uses the method presented in Chapter 2 to extract aortic diameters to investigate the growth of the thoracic aorta in a large population. This study presents longitudinal data on sex-specific growth of the ascending and descending aorta in a large population of current or former smokers, a subgroup of the general population. The measured growth rate is 0.1 mm/year for this population, which is consistent with numbers reported for growth in the general population. In addition, aortic growth is comparable between current and ex-smokers, and aortic growth is not associated with pack-years.

Chapter 4 presents a 3D fully automatic method for segmenting the pulmonary artery and the aorta, extending the method presented in Chapter 2. The method extracts a landmark for the level of the pulmonary artery bifurcation. With the 3D volumetric average diameter measurement technique presented in this chapter, the ratio of the diameter of the pulmonary artery to the diameter of the ascending aorta at the level of the pulmonary artery bifurcation is automatically extracted. The diameters extracted by the presented 3D volumetric diameter measurement technique show high scan-rescan repeatability. This chapter presents the qualitative and quantitative analysis showing that our method provides robust, accurate, and repeatable measurements of the pulmonary artery and aorta diameters and the PA:AA ratio.

Chapter 5 presents a deep learning based algorithm called Posterior-CRF that uses CNN-learned features in a CRF. This method is validated on multiple modalities and medical image segmentation tasks. The aorta and pulmonary artery are segmented in non-contrast CT, and white matter hyperintensities and ischemic stroke lesions are segmented in multi-modal MRI. With high accuracy in all three segmentation tasks, the segmentation results showed that spatial coherence or intensity features alone are not sufficient. The CNN features in the last layer, as presented in Posterior-CRF, are more informative for CRF than the intensity features in the original images.

Chapter 6 provides a general discussion of the contribution and achievements of this thesis and presents possible future research directions.

2

Automated 3D Segmentation and Diameter Measurement of the Thoracic Aorta on Non-contrast Enhanced CT

Zahra Sedghi Gamechi*, Lidia R. Bons*, Marco Giordano, Daniel Bos, Ricardo P.J. Budde, Klaus F Kofoed, Jesper Holst Pedersen, Jolien W. Roos-Hesselink, Marleen de Bruijne

European Radiology. September, 2019

** joint first authorship*

Abstract

Objectives: To develop and evaluate a fully automatic method to measure diameters of the ascending and descending aorta on non-ECG-gated, non-contrast computed tomography (CT) scans.

Material and methods: The method combines multi-atlas registration to obtain seed points, aorta centerline extraction, and an optimal surface segmentation approach to extract the aorta surface around the centerline. From the extracted 3D aorta segmentation, the diameter of the ascending and descending aorta was calculated at cross-sectional slices perpendicular to the extracted centerline, at the level of the pulmonary artery bifurcation, and at $1-cm$ intervals up to $3cm$ above and below this level. Agreement with manual annotations was evaluated by dice similarity coefficient (DSC) for segmentation overlap, mean surface distance (MSD), and intra-class correlation (ICC) of diameters on 100 CT scans from a lung cancer screening trial. Repeatability of the diameter measurements was evaluated on 617 baseline one-year follow-up CT scan pairs.

Results: The agreement between manual and automatic segmentations was good with $0.95/pm0.01$ DSC and $0.56/pm0.08mm$ MSD. ICC between the diameters derived from manual and automatic segmentations was 0.97, with the per-level ICC ranging from 0.87 to 0.94. An ICC of 0.98 for all measurements and per-level ICC ranging from 0.91 to 0.96 were obtained for repeatability.

Conclusions: This fully automatic method can assess diameters in the thoracic aorta reliably even in non-ECG-gated, non-contrast CT scans. This could be a promising tool to assess aorta dilatation in screening and clinical practice.

Key Points:

- Fully automatic method to assess thoracic aorta diameters.
- High agreement between fully automatic method and manual segmentations.
- Method is suitable for non-ECG-gated CT and can therefore be used in screening.

2.1 Introduction

Aortic aneurysm with the risk of acute dissection is an important cause of mortality in the western world [73]. The prevalence of thoracic aortic aneurysms is estimated at around 0.3 percent in the normal population [74, 75]. Most patients with a dilated aorta or aortic aneurysm are asymptomatic. The diagnosis can be made as during screening in the context of positive family history or by coincidence on imaging examinations performed for other purposes like lung cancer screening [12]. However, acute dissection is often the first presentation, in which case over 50% of all patients die within 30 days [30].

Because of this silent process with high risks, screening programs using non-contrast computed tomography (CT) could be considered. In patients with aortic aneurysms, the aortic size has a profound impact on the risk of dissection [28, 29]. Detecting aortic dilatation at an early stage enables preventive surgery, which might save lives. CT imaging of the thoracic aorta could become available as part of a comprehensive assessment of CT imaging performed for screening purposes including also other organs (lungs, coronary calcium, vertebral bone density, etc.) [12].

By measuring aortic dimensions in such screening cohorts, we will also gain more information on normal values of aortic diameters, the normal increase in diameters over time, risk factors for dilatation, and a better insight into prognosis.

Besides its potential in screening, non-contrast CT is frequently used to diagnose and follow-up patients in clinical practice. It plays a central role in imaging the thoracic aorta because of the short time required for image acquisition, the ability to obtain a complete 3D view of the entire aorta, and its widespread availability. In addition, CT scans can be used for follow-up of patients with dilatation, especially in cases where echocardiography does not adequately visualize the dilatation. The ESC Guidelines and ACCF/AHA guidelines [21, 76] describe standard anatomical landmarks for reporting aortic diameters in CT in clinical practice.

Performing measurements of the aorta manually are labor-intensive and subject to inter-observer variability. Therefore, to assess aortic dilatation in screening settings and clinical practice, automated aorta segmentation and subsequent diameter analysis are desirable. While automatic solutions for aortic measurements in CT angiography (CTA) exist [42–45, 47], automatic aorta segmentation in non-contrast CT scans is more challenging due to the lack of contrast between blood pool regions and surrounding soft tissue [53, 54, 57, 59, 60].

This chapter aims to develop and validate an automatic method to robustly assess diameters of the ascending and descending aorta in non-ECG-gated, non-contrast CT without human interaction.

2.2 Materials and Methods

2.2.1 Study Population & Image Acquisition

The CT scans used in this chapter are from the Danish Lung Cancer Screening Trial (DLCST) [6]. A Multi-Detector CT scanner ($M \times 8000$ IDT 16 row scanner, Philips Medical Systems) was used to acquire CT scans at 120 kV / 40 mAs at

Table 2.1: Clinical characteristics of 100 subjects used in validation. *Values are expressed as mean ± standard deviation and (range).*

Validation set (n = 100)	**Male**	**Female**
Number of CT scans (n)	50	50
Age ($years$)	58.5 ± 5.4 (50 - 70)	58.3 ± 4.8 (50 - 70)
Weight (kg)	84.0 ± 12.0 (60 - 120)	67.6 ± 12.2 (48 - 103)
Height (cm)	179.8 ± 6.3 (163 - 195)	167.0 ± 6.1 (155 - 179)
BMI	26.0 ± 3.6 (18.7 - 37.0)	24.3 ± 4.7 (16.2- 41.3)
Agatston score at ascending aorta & arch	231.3 ± 416.7 (0 - 2190)	193.3 ± 274.4 (0 - 1128)
Agatston score at descending aorta	53.5 ± 116.2 (0 - 483)	81.4 ± 316.4 (0 - 2139)

maximum inspiration breath-hold and without cardiac gating. This protocol leads to an effective dose of around 1 mSv [77]. The scans were reconstructed with a sharp kernel (Philips D), in-plane isotropic resolution of $0.78 \times 0.78mm$, and $1mm$ slice thickness. Participants were current or former smokers between 50 and 70 years of age. For this chapter, 742 participants were randomly selected, which were divided into three non-overlapping sets: (see supplementary Table 2.A.1 for clinical characteristics of the entire data)

- baseline scans of 25 subjects for parameter optimization of the proposed method;
- baseline scans of 100 subjects for evaluation of the method's accuracy (see Table 2.1);
- baseline and first year follow up scans of 617 subjects to evaluate the repeatability of the method;

Therefore, aortic diameter measurements were performed in 1334 CT scans in total.

2.2.2 Manual Annotation

Manual annotations were made using an in-house annotation tool developed in MeVisLab[1]. 100 CT scans were annotated by a physician (LB) for validation and an additional 25 scans by an experienced observer (ZSG) for method development. The annotation tool was similar to that described previously for carotid artery segmentation in [78]. First, the window level/width was adjusted to $200HU/600HU$ for all cases. Then, the aortic centerlines were drawn manually using the axial, coronal, and sagittal views, starting from the sinotubular junction of the ascending aorta and ending at the diaphragm level of the descending aorta. Subsequently, the centerlines were checked and modified in reformatted cross-sectional views perpendicular to the drawn centerline. The obtained centerlines were used to generate curved multiplanar reformatted

[1] https://www.mevislab.de/

Figure 2.1: Screenshot of the manual annotation tool (left). *The middle image shows two manually drawn longitudinal contours (yellow) and a few cross-sectional contours (red), which are perpendicular to the manual centerline (blue). A cross-sectional slice at the ascending aorta and the corresponding contour is shown as well. The corresponding 3D surface of the aorta are shown on the right image.*

images of the entire aorta, with longitudinal views at six different angles equally spaced every 30ř and cross-sectional views every $1mm$ along the centerline. Longitudinal contours were drawn manually, whereupon cross-sectional contours were computed using spline interpolation through the intersection points of the longitudinal contours with the cross-sectional planes. Finally, after checking the cross-sectional contours in all cross-sections and adjusting them if required, the contours were converted to a 3D binary image using variational interpolation [79]. An example of manual annotation is shown in Figure 2.1.

To manually locate the pulmonary artery bifurcation level, an experienced physician (DB) checked the scans in axial view and annotated the pulmonary artery bifurcation level where the left and right pulmonary arteries and the bifurcation from the pulmonary trunk were all visible.

2.2.3 Automatic Aorta Segmentation Approach

To extract a full 3D segmentation of the aorta and a landmark point for the pulmonary artery bifurcation level, we applied a combination of image processing techniques. First, to avoid the segmentation to attract to the heart-lung or bone borders, we applied preprocessing as proposed in our previous work [80].

Subsequently, a multi-atlas registration method was applied [81] to localize the aorta, the pulmonary artery trunk, and the left and right pulmonary arteries. In this method, 25 preprocessed CT scans were non-rigidly registered to the scan in which the segmentation was required (target image). From these 25 registered images, ten CT scans with the highest similarity to the target image were selected. The corresponding manual annotations of these ten scans were then deformed and combined using a per voxel majority voting procedure to obtain a coarse initial segmentation of the aorta and pulmonary arteries. The initial segmentation of the pulmonary arteries was then skeletonized, and the slice where the main pulmonary artery bifurcates into the left and right pulmonary arteries was extracted as the pulmonary artery bifurcation level. This level is used as the landmark level.

To start tracing the centerline of the aorta, aortic seed points were extracted as the center of mass of the coarse initial aorta segmentation at the axial slice 3 cm beneath the landmark level for the ascending aorta and 6 cm beneath the landmark level for the descending aorta. The aortic centerline was then extracted between these seed points by a minimum cost path tracking algorithm similar to [80]. In this algorithm, the cost function was based on the maximum output of a multi-radius medialness filter in coronal and axial views multiplied with a lumen intensity similarity metric. Next, the centerlines were refined by re-computing the minimum cost path after curved multiplanar reformatting perpendicular to the previous centerline [82]. Failure in the centerline extraction was automatically detected by using the landmark level and the initial pulmonary artery segmentation. Centerlines that did not reach the landmark level or were inside the pulmonary artery segmentation were considered failed extractions and were excluded.

To obtain a first estimate of the aorta, the extracted centerline was dilated using a spherical structuring element with its radius defined by the estimated radius of the aorta obtained from the medialness filter. Subsequently, an optimal surface graph cut

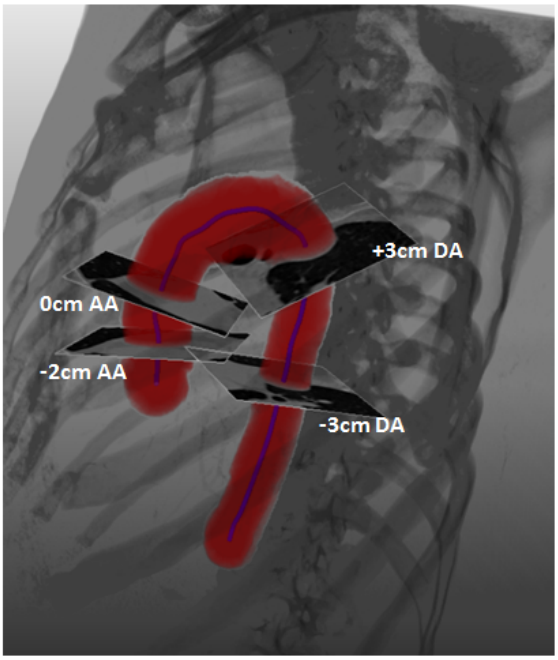


Figure 2.2: 3D automatic segmentation of the aorta *and the corresponding automatic centerline showing cross-sections at the ascending aorta at the pulmonary artery bifurcation level (0 cm AA) and 2 cm below this level (-2 cm AA) and the descending aorta at 3 cm above (+3 cm DA) and below (-3 cm DA) the pulmonary artery bifurcation level.*

segmentation method[2] [69], initialized by the dilated centerline, was used to accurately extract the surface of the aorta. The parameters for atlas registration, centerline extraction, and graph cut segmentation were tuned to maximize the similarity with manual annotations on 25 CT scans.

2.2.4 Aortic Diameter Measurement

Aortic diameters were assessed at multiple, fixed levels relative to the pulmonary artery bifurcation level. Based on the extracted pulmonary bifurcation level, thirteen cross-sectional slices were defined perpendicular to the extracted aortic centerline, located at 1-cm intervals around the bifurcation level from 2 cm below this level to 3 cm above for the ascending aorta and from 3 cm above to 3 cm below this level for the descending aorta. For the ascending aorta, the cross-sectional slice at 3 cm below the pulmonary artery bifurcation level was sometimes in the aortic root below the sinotubular junction which the aorta boundaries at the sinus of Valsalva are very

[2]Available at https://bitbucket.org/opfront/opfront

unclear due to the lack of gating and contrast. Therefore, no measurements were performed at this level. Figure 2.2 shows an example of 3D segmentation with the corresponding centerline and four of the measured cross-sections.

The cross-sectional average aortic diameter at each of the 13 cross-sectional slices was computed from manual and automatic segmentations. For the manual segmentations, diameter measurements were performed perpendicular to the manual centerlines and at levels relative to the manually indicated pulmonary artery bifurcation level. For the automatic segmentations, the automatically extracted centerlines and pulmonary artery bifurcation level were used instead.

2.2.5 Validation and Statistical analysis

The method was validated on 100 CT scans with manual annotations. The segmentation accuracy was assessed by the dice similarity coefficient (DSC) and mean surface distance (MSD). DSC [83] measures the degree of spatial overlap of the automatic segmentation with the manual segmentation, and it ranges between 0 and 1, where higher values indicate higher similarity. MSD shows the symmetric mean surface distance in millimeters between the manual and automatic segmentation surfaces, where a lower value is better. The agreement between the manual and automatic segmentations was assessed from $3cm$ beneath the landmark level at the ascending aorta to 6 cm beneath this level at the descending aorta. DSC, MSD, aortic diameters, and the error in the diameter were expressed as mean $\pm$ standard deviation (range).

The error in the extracted landmark level was assessed by the distance between the manually extracted pulmonary artery bifurcation level and the automatically extracted level in millimeters. In addition, the aortic centerlines were automatically checked for failed extractions.

The agreement between the manual and automatic diameter measurements was assessed by

- Intra-class correlation (ICC) based on a single-rating, absolute-agreement, two-way mixed-effects model [84];
- R^2 Pearson's correlation;
- Bland-Altman analysis;

The repeatability of the method was assessed by comparing the automatically extracted diameters of two scans of 617 subjects with a time period of 1 year in between. Within 1 year, changes in aortic diameters are expected to be small, with 0.1-0.2 mm growth per year in a healthy population [28, 75]. All statistical analyses were done in MATLAB.

2.3 Results

Figure 2.3 shows examples of segmentation results. Out of all 1334 CT scans only in two cases, the seed points at the descending aorta were extracted incorrectly. Centerline extraction further failed in seven cases, all of which were easily detected automatically. Average DSC for the entire aorta was 0.95 ± 0.01 (0.92–0.96) and MSD

Table 2.2: Average aortic diameters *from the automatic and manual segmentations for each measuring level from the 100 CT scans. Values are expressed as mean ± standard deviation.*

	Measuring Level	**Female (*n=50*)**		**Male (*n=50*)**	
		Automatic	**Manual**	**Automatic**	**Manual**
Ascending Aorta	**-2 cm**	33.3 ± 3.5	34.2 ± 3.5	36.2 ± 3.9	37.0 ± 3.7
	-1 cm	34.2 ± 3.2	35.1 ± 3.4	36.9 ± 3.7	37.7 ± 3.6
	0 cm	33.9 ± 3.3	35.0 ± 3.3	36.7 ± 3.5	37.7 ± 3.5
	+1 cm	33.5 ± 3.0	34.7 ± 3.2	36.3 ± 3.6	37.4 ± 3.5
	+2 cm	33.3 ± 2.7	34.3 ± 2.9	35.4 ± 3.3	36.7 ± 3.2
	+3 cm	33.0 ± 2.9	33.4 ± 3.1	35.2 ± 3.0	36.0 ± 3.2
Descending Aorta	**+3 cm**	27.6 ± 2.6	28.7 ± 2.5	29.5 ± 2.4	30.8 ± 2.5
	+2 cm	27.3 ± 2.5	28.2 ± 2.4	29.0 ± 2.3	30.2 ± 2.1
	+1 cm	26.7 ± 2.3	27.5 ± 2.2	28.5 ± 2.1	29.4 ± 2.0
	0 cm	26.3 ± 2.3	27.1 ± 2.2	28.0 ± 2.0	28.9 ± 1.9
	-1 cm	26.1 ± 2.1	26.9 ± 2.2	27.7 ± 2.1	28.9 ± 1.9
	-2 cm	25.8 ± 2.2	26.9 ± 2.2	27.6 ± 2.0	28.6 ± 1.8
	-3 cm	25.5 ± 2.3	26.7 ± 2.2	27.4 ± 2.1	28.3 ± 2.0

0 cm is the pulmonary artery bifurcation level, where minus is a level below this level and plus is a level above the pulmonary bifurcation level.

was 0.56 ± 0.08 (0.43–0.93) *mm*. The mean absolute distance between the manual and automatic landmark level of the pulmonary artery bifurcation was 2.55 ± 1.94 *mm*, with almost no bias (mean signed distance 0.45 ± 3.18 *mm*).

Box plots for the average manual and automatic diameters for each measuring level are shown in Figure 2.4. Diameters measured at the different levels, for men and women separately, are shown in Table 2.2. High agreement between manually and automatically measured diameters was obtained, with an overall ICC and R^2 Pearson's correlation of 0.97. The level-wise correlations together with the correlations separated per gender are shown in Table 2.3 (see supplementary Figure 2.A.1 for scatter plots of each measuring level).

An average absolute diameter error of 1.09 ± 0.6 *mm* between manual and automatic diameters was obtained over all measuring levels, which showed a slight underestimation of the automated measurements compared to manual measurements (mean signed error -0.97 ± 0.8 *mm*). As shown in box plots of the level-wise diameter errors in Figure 2.5, larger errors (more than 3 *mm*) were extracted in 8 out of 100 scans. In four cases, a large error occurred due to motion artifacts at the ascending aorta (beneath the landmark level), and in three cases, it occurred at the aortic arch due to branching arteries. In one case, the error was along the entire aorta due to a 6-*mm* difference between the automatic and manual landmark levels. Bland-Altman plots of manual and automated diameter measurements are given in Figure 2.6.

From the 617 subjects used to assess repeatability, 7 subjects had failed centerline or seed point extraction. From the remaining 610 subjects, ICC between the automatic

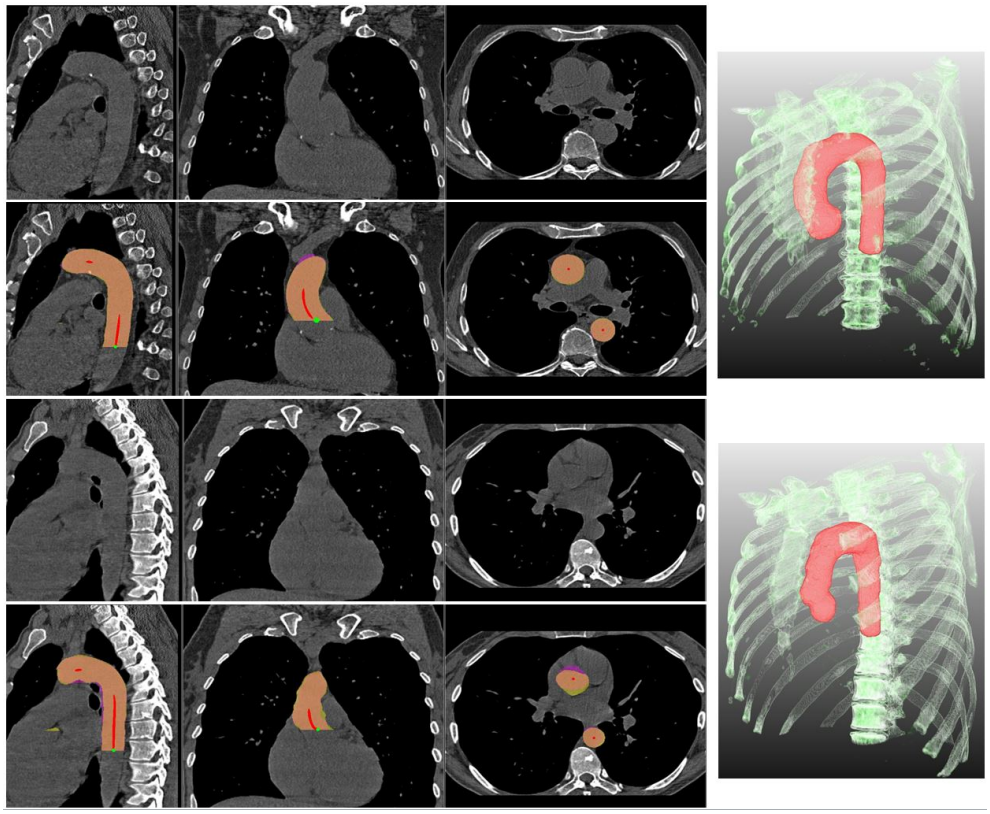


Figure 2.3: Segmentation Examples; *Two samples with the best (top two rows) and the worst (bottom two rows) automatic segmentation results. The columns from left show the sagittal, coronal, and axial views, respectively. The right column shows the 3D visualization of the automatic segmentation in red. The first and third rows are the original CT scans, while the second and fourth rows show the CT scan with the overlap of the corresponding manual and automatic segmentations with DSC = 0.96 and MSD = 0.60 mm for the first sample and DSC = 0.92 and MSD = 1.44 mm for the second sample. Orange shows the regions where the manual and automatic segmentations overlap. Magenta is the region included in the automatic segmentation, but not in the manual segmentation, and yellow is the region that is inside the manual segmentation, but not in the automatic segmentation. Centerline points are indicated in red and seed points in green.*

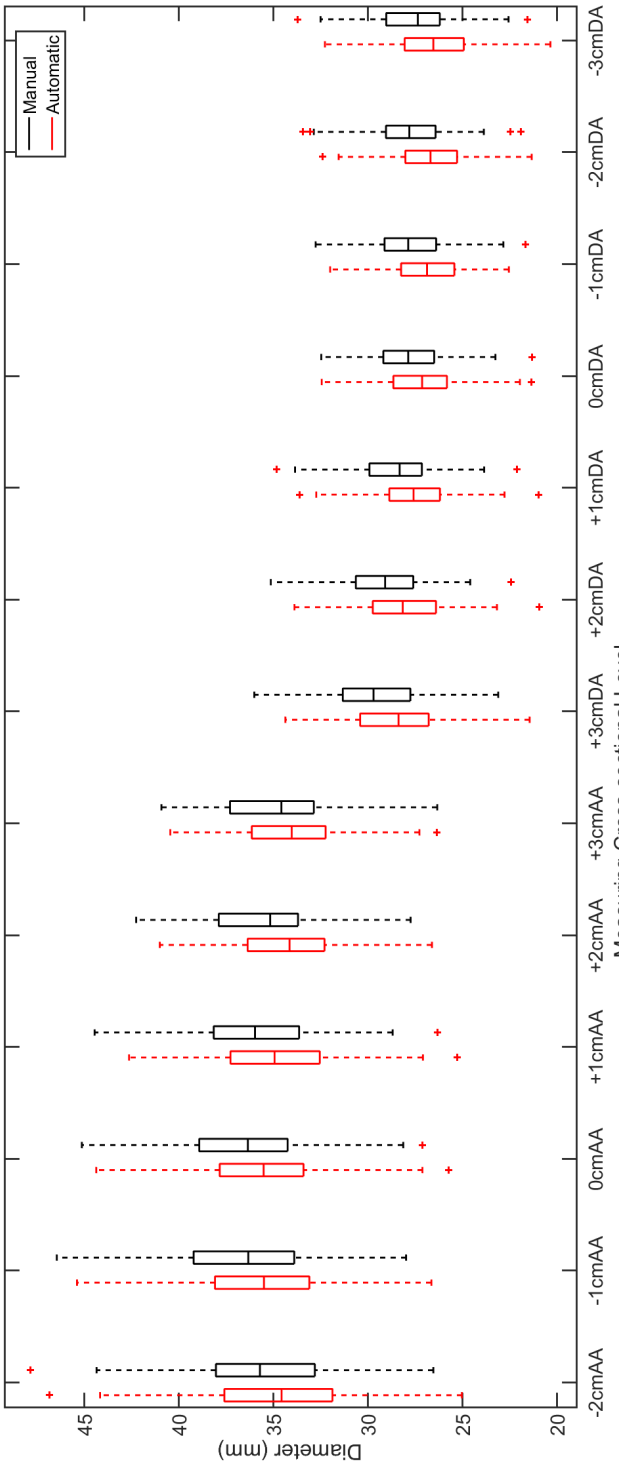


Figure 2.4: ***Average manual (black) and automatic (red) diameter per measuring level.*** *From left to right, diameters measured at the different levels along the aorta from 2 cm below the pulmonary artery bifurcation level (0 cm) at the ascending aorta (AA) to 3 cm below the pulmonary artery bifurcation level at the descending aorta (DA).*

Table 2.3: ICC and R^2 Pearson's correlation between the automatic and manual diameters for the 100 CT scans.

	Measuring Level	ICC *(n=100)*	R^2 Pearson *(n=100)*	ICC female *(n=50)*	ICC male *(n=50)*
Ascending Aorta	**-2 cm**	0.93	0.90	0.89	0.94
	-1 cm	0.94	0.94	0.91	0.95
	0 cm	0.94	0.95	0.92	0.94
	+1 cm	0.92	0.94	0.89	0.92
	+2 cm	0.92	0.95	0.91	0.90
	+3 cm	0.94	0.93	0.94	0.93
Descending Aorta	**+3 cm**	0.88	0.92	0.88	0.85
	+2 cm	0.88	0.91	0.88	0.84
	+1 cm	0.89	0.92	0.88	0.87
	0 cm	0.90	0.93	0.90	0.87
	-1 cm	0.89	0.94	0.90	0.83
	-2 cm	0.87	0.93	0.86	0.83
	-3 cm	0.89	0.95	0.86	0.88

ICC : Intra class correlation;
Measuring levels as in Table 2.2;

diameters of the scan and rescan of each subject is shown in Table 2.4. From these 610 subjects, 72 subjects (12%) had an absolute diameter difference larger than 3 *mm* between the two time points at any of the measuring levels. In 35 out of 72 cases (48.6%), the segmentations appeared visually correct in both time points. In 17 cases of these 35 cases, a 2- or 3-*mm* difference between the extracted landmark level in one of the time points resulted in big diameter differences at 2 cm below the landmark level at the ascending aorta (in average 3.7 ± 0.5 *mm*). This is due to the aortic anatomy at the sinotubular junction where the aorta below this level is on average 3 *mm* larger than above [85]. In 5 out of 35 cases, there was more than 6-*mm* difference between the extracted landmark levels from the two time points, leading to a diameter measurement at very different levels along the entire aorta being compared (in average 3.4 ± 0.5 *mm*). The remaining 13 out of 35 cases appeared to have a slightly larger diameter at one of the time points (in average 3.7 ± 0.7 *mm*), possibly due to the aortic size changes during the cardiac cycle. In 37 out of 72 cases (51.4%), the average diameter difference (3.6 ± 0.6 *mm*) was due to segmentation error which mainly occurred at the aortic arch which was due to branching arteries, or was at the ascending aorta below the pulmonary artery bifurcation level which was due to heart motion artifacts caused by the non-ECG-gated data.

2.4 Discussion

We presented a fully automatic method to segment the thoracic aorta and measure aortic diameters. In our evaluation on 100 non-ECG-gated, non-contrast CT scans, the

3D segmentation algorithm performed well with an average segmentation overlap of 0.95 ± 0.01 and a mean surface distance between manual and automatic segmentations of less than 1 voxel (0.56 *mm*).

The agreement with diameters obtained from manual segmentations was high, with an overall ICC of 0.97 and an average per-level ICC of 0.91 ± 0.03, which is similar to the agreement reported between observers in [26] (ICC = 0.94). The manual diameters were on average approximately 1 *mm* larger than automatic diameters. This bias is similar to inter-observer bias reported in [86] for mid-ascending aorta diameter measurement on CTA. Scan-rescan repeatability was high, with an overall ICC of 0.98 and an average per-level ICC of 0.94 ± 0.01.

The mean ascending aorta diameters measured at the pulmonary artery bifurcation level were 36.7 ± 3.5 *mm* for males and 33.9 ± 3.3 *mm* for females. These values are similar to those reported by Kalsch et al [75] (37.1 ± 4 *mm* for males and 34.5 ± 4 *mm* for females), while they were slightly greater than those reported by Wolak et al [22] (33.5 ± 4 *mm* for males and 31.4 ± 3 *mm* for females). These differences may be due to differences in the study populations, CT scan protocol, and measurement approach.

A significant diameter increase of on average 0.11 ± 1.0 *mm* was measured in repeated scans after 1 year. This agrees well with reported natural yearly aortic diameter growth of 0.1 – 0.2 *mm* per year in the healthy population [75, 87]. In 12% of repeat scan pairs (72 subjects), diameter changes larger than 3 *mm* were observed. In the majority of these cases (44 subjects), large diameter differences occur at the ascending aorta beneath the landmark level which is due to the anatomy and the difficulty of measuring these regions. Due to motion artifacts in the non-ECG-gated scans, segmentation of the proximal part of the aorta including the aortic root is difficult even for experienced radiologists. However, although isolated aortic root aneurysms are seen in patients with Marfan syndrome [76], it is less common than aneurysms of the ascending aorta more distal to the aortic root. Therefore, the aortic root segmentation is less important in our application than the ascending aorta. In the remaining 28 cases, the large diameter difference was either in the aortic arch (15) or in the descending aorta (8), or at multiple locations due to error in the extraction of the pulmonary artery bifurcation level (5). Diameters measured at the aortic arch were visually correct; however, slightly larger diameters were measured at the location of branching arteries. In descending aorta, the large diameter differences were mainly due to segmentation error. An example of the diameter profile and absolute diameter difference profile between the manual and automatic diameters along the aortic centerline is shown in supplementary Figure 2.A.2.

In contrast with our study in this chapter, in literature, most methods for automatic aorta segmentation were evaluated on CTA in which the aortic lumen is much more clearly visible [42–45, 47]. Few methods were proposed to segment the aorta in non-contrast CT [53, 54, 57, 59, 60]. Compared to these previous works, shown in Table 2.5, our proposed method is evaluated on a larger dataset and shows better performance.

We proposed to measure aortic dimensions at fixed intervals with respect to a single anatomical landmark level, the pulmonary artery bifurcation. In clinical practice, multiple anatomical landmarks including locations in the aortic arch are used instead for reporting aortic diameters in CTA [21, 76, 88]. However, consistently extracting

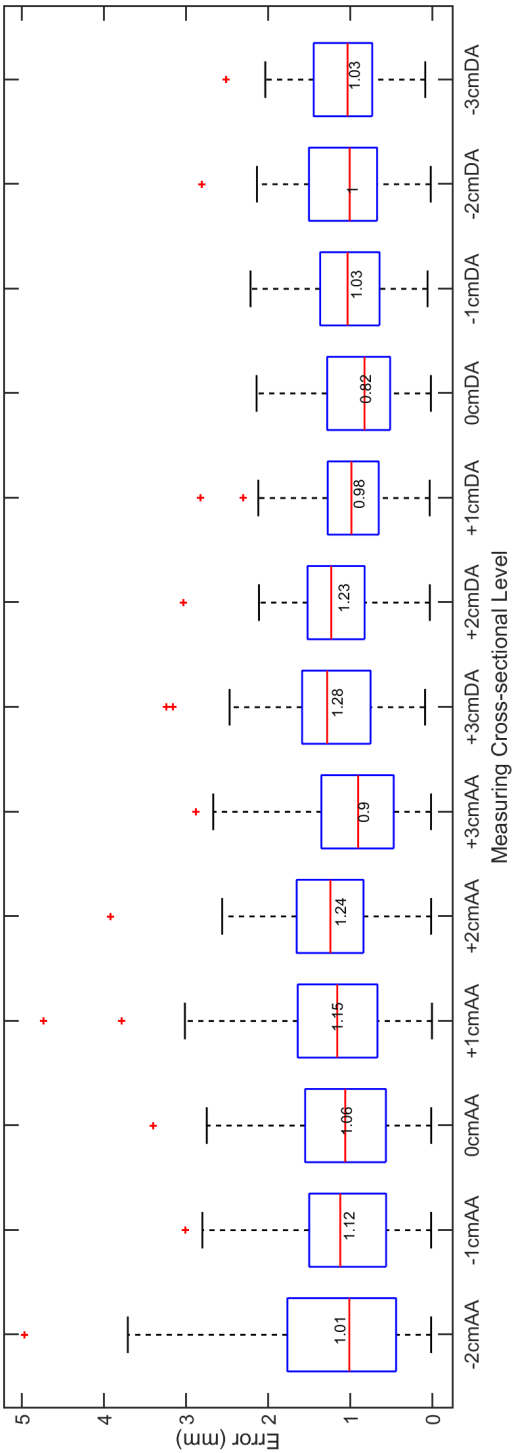


Figure 2.5: The absolute difference between the aortic diameters obtained from automated and manual 3D segmentations. *From left to right, diameters measured at the different levels along the aorta from 2 below the pulmonary artery bifurcation level (0 cm) at the ascending aorta (AA) to 3 cm below the pulmonary artery bifurcation level at the descending aorta (DA). The box plot shows the median (red line), interquartile range (boxes), the 99.3% coverage of the data (whiskers), and the outliers (+ symbol).*

***Table 2.4: Repeatability**: ICC between the automatic diameters of the scan and rescan of 610 subjects.*

	Ascending aorta						**Descending aorta**						
	-2cm	-1cm	0cm	+1cm	+2cm	+3cm	+3cm	+2cm	+1cm	0cm	-1cm	-2cm	-3cm
ICC	0.91	0.95	0.96	0.96	0.95	0.95	0.94	0.94	0.93	0.94	0.93	0.94	0.94

ICC : Intra class correlation;
Measuring levels as in Table 2.2;

***Table 2.5: Performance comparison of methods for the aorta segmentation on non-contrast CT.** Values are expressed as mean ± standard deviation.*

Author [ref. no.]	Evaluation Data size	DSC	Jaccard coefficient	MSD (mm)
Kitasaka et al. [57]	7 CT	0.93 ± 0.03	–	0.90 ± 0.33
Avila-Montes et al. [54]	45 CT	0.84 ± 0.10	0.74 ± 0.13	–
Kurugol et al. [60]	45 CT	0.92 ± 0.01	0.85 ± 0.02	0.62 ± 0.09
Isgum et al. [53]	29 CT	–	0.78 ± 0.04	–
Xie et al. [59]	60 CT	0.93 ± 0.01	–	1.39 ± 0.19
Proposed Method	**100 CT**	0.95 ± 0.01	0.90 ± 0.01	0.56 ± 0.08

DSC : Dice Similarity Coefficient;
MSD : Mean Surface Distance.

these landmarks especially in non-ECG-gated CT is difficult. Moreover, the aorta diameter is poorly defined at the locations of the brachiocephalic artery, left-common carotid artery, and left-subclavian artery. Consistent measurements in the arch require landmark points in between branches that are not affected by this issue; however, detecting such points automatically and robustly in non-contrast CT scans is difficult. Furthermore, aortic dilatation is less common in the arch than in the ascending and descending aorta. Therefore, in this chapter, we focus on the ascending and descending aortas which clinically are of more interest. In non-contrast CT, diameters have been mainly measured at the pulmonary artery bifurcation level [22, 74, 75, 89, 90]. The measuring levels used in this chapter approximately cover the same area used in CTA [21, 76, 88] but are easier to extract reliably in non-contrast and non-ECG-gated CT.

A limitation of our study in this chapter is that the method was validated only on a relatively healthy screening population. Further investigation would be required to evaluate the performance of abnormal aortic shapes or large aneurysms. However, in all cases with aortic dilatation as indicated in the original radiology reports, the obtained segmentation was correct. In our data, calcification in the aorta was assessed by the Agatston score [91]. Visual inspection of the scans with Agatston score higher

than 1500 for the entire aorta (58 out of 742 subjects) showed that the proposed method segmented the calcifications correctly inside the vessel wall in all cases.

The proposed automatic method is a promising technique to accurately and reproducibly assess subtle signs of aorta dilatation in non-ECG-gated, non-contrast CT scans without any human interaction and could be used for efficient screening for aortic dilatation as well as for monitoring of aortic change in clinical practice as part of a comprehensive CT analysis, including lung screening.

Funding

This chapter has received funding by Iranian Ministry of Science Research and Technology (MSRT), the Netherlands Organisation for Scientific Research (NWO), Dutch Heart Foundation (The Hague, The Netherlands, grant number: 2013T093), and Danish Ministry of Interior and Health.

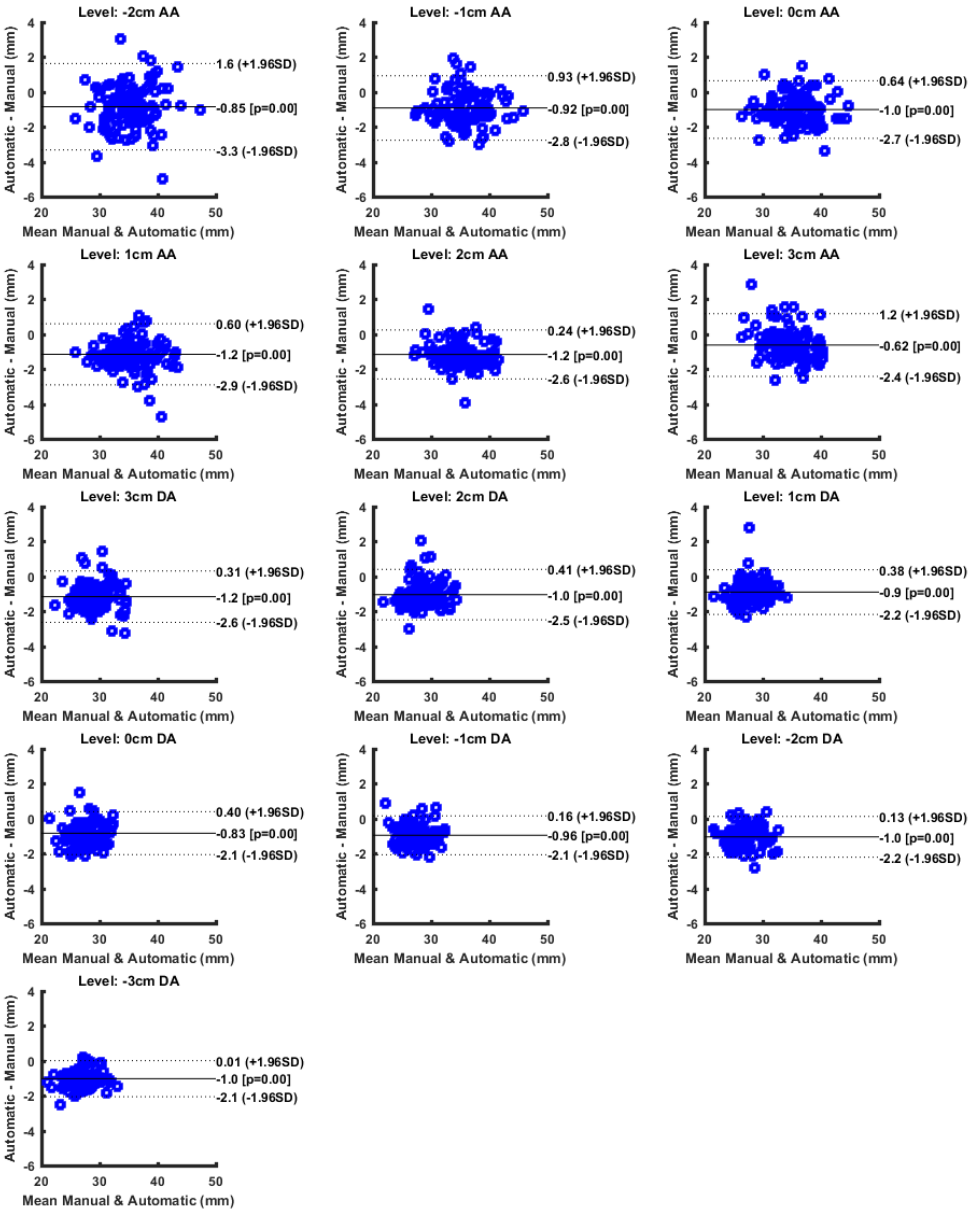


Figure 2.6: ***Bland-Altman plots*** *for each measuring level from 2 cm below the pulmonary artery bifurcation in the ascending aorta (AA) until 3 cm below this level in the descending aorta (DA). The measuring level, limits of agreement, and the mean difference are displayed on the plots.*

Appendix

2.A Supplemental Material

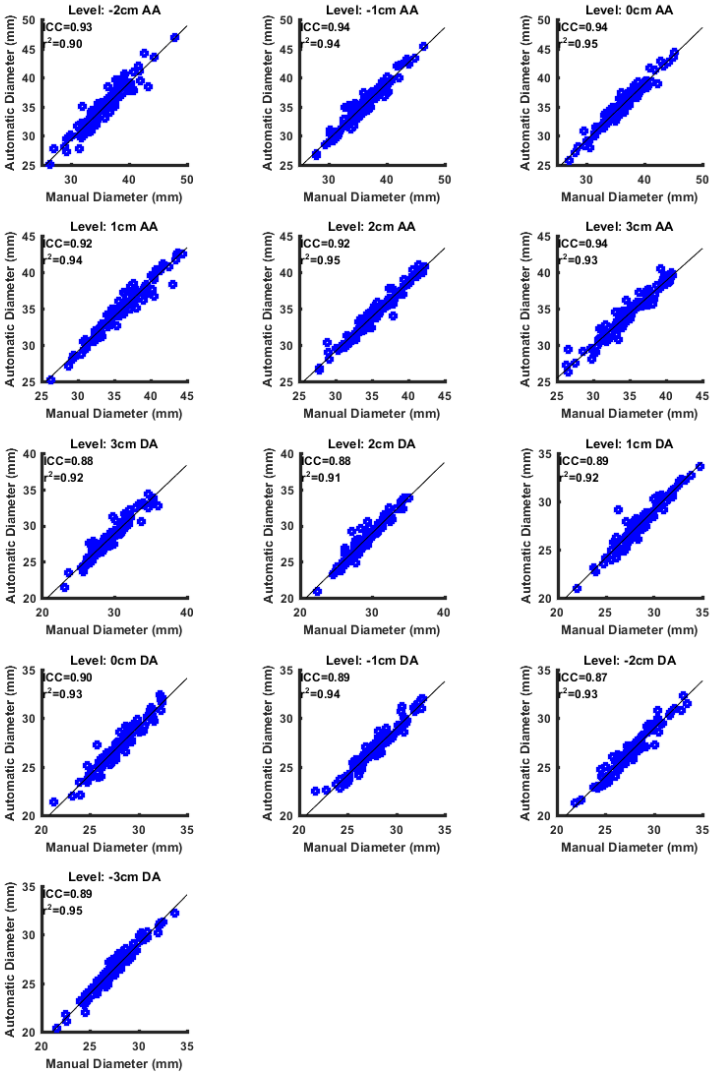


Figure 2.A.1: Scatter plots *for each measuring level from 2 cm below the pulmonary artery bifurcation in the ascending aorta (AA) until 3 cm below this level in the descending aorta (DA). The measuring level, ICC, and the R2 Pearson correlation.*

Table 2.A.1: Clinical characteristics of the data in all three sets.
Values are expressed as mean ± standard deviation and (range).

	Method Development (n=25)		Validation (n=100)		Repeatability (n=617)	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
n	12	13	50	50	378	239
Age, years	57.4 ± 4.0 (51.3-67.8)	58.3 ± 5.2 (51.5-67.8)	58.5 ± 5.4 (50.0-70.0)	58.3 ± 4.8 (50.0-70.0)	58.9 ± 5.2 (49.3-71.0)	58 ± 5.3 (49.8-70.9)
Weight, kg	82.3 ± 10.4 (66-102)	75.8 ± 9.2 (60-90)	84.0 ± 12.0 (60-120)	67.6 ± 12.2 (48-103)	83.9 ± 12.8 (57-130)	68.6 ± 12.6 (42-126)
Height, cm	177.1 ± 5.0 (167-185)	167.6 ± 6.3 (158-180)	179.8 ± 6.3 (163-195)	167.0 ± 6.1 (155-179)	179.4 ± 6.5 (163-200)	166.7 ± 6.1 (150-182)
BMI	26.3 ± 3.5 (23.3-33.4)	27.1 ± 3.9 (20.1-35.2)	26.0 ± 3.6 (18.7–37.0)	24.3 ± 4.7 (16.2-41.3)	26.1 ± 3.7 (17.3-40.9)	24.7 ± 4.1 (16.4-41.2)
AS at AA & Arch	50.4 ± 126.4 (0-435)	259.1 ± 425.4 (0-1298)	231.3 ± 416.7 (0-2190)	193.3 ± 274.4 (0-1128)	304.3 ± 749.6 (0-6492)	342.4 ± 742.7 (0-5563)
AS at DA	53.3 ± 99.1 (0-256)	37.8 ± 98.7 (0-336)	53.5 ± 116.2 (0-483)	81.4 ± 316.4 (0-2139)	166.7 ± 696.2 (0-9705)	224.1 ± 692.2 (0-5261)

n : number of participants; ***AS*** : Agatston Score; ***AA*** : Ascending Aorta; ***DA*** : Descending Aorta

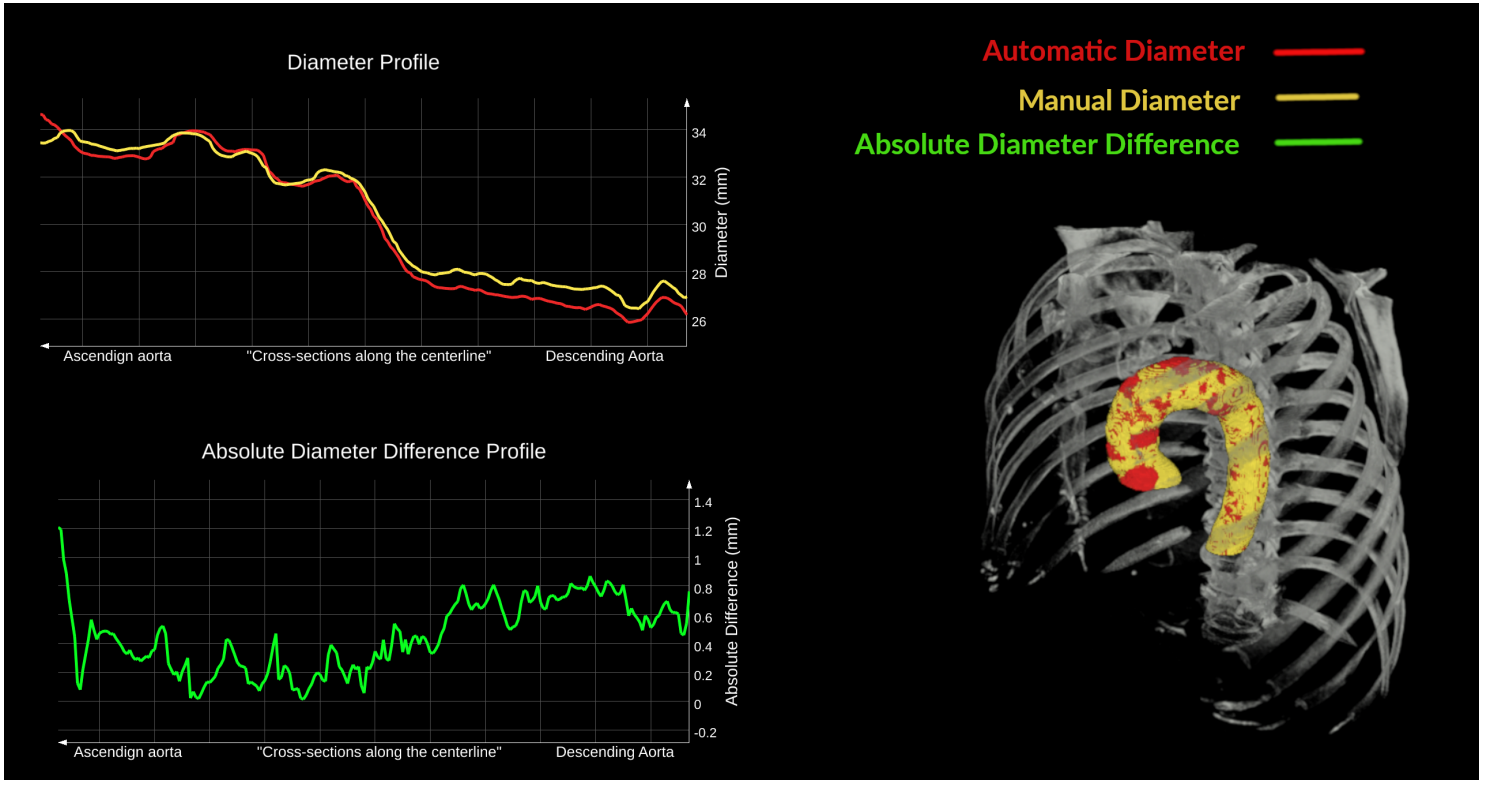


Figure 2.A.2: An example of the diameter profile & absolute diameter difference profile along the aorta centerline. *The diameter profile of the automatic (red) and manual (yellow) diameters extracted in cross-sectionals perpendicular to the vessel centerline are shown in the top-left diagram. The absolute diameter difference along the aorta centerline (green) is shown in the bottom left diagram. A 3D automatic segmentation of the aorta (red), overlaid with the manual segmentation (yellow), is shown on the right.*

3

Growth of the Thoracic Aorta in the Smoking Population: The Danish Lung Cancer Screening Trial

Lidia R. Bons*, **Zahra Sedghi Gamechi***, Carlijn G.E. Thijssen, Klaus F Kofoed, Jesper Holst Pedersen, Zaigham Saghir, Johanna J.M. Takkenberg, Isabella Kardys, Ricardo P.J. Budde, Marleen de Bruijne, Jolien W. Roos-Hesselink

International Journal of Cardiology. January, 2020

** joint first authorship*

Abstract

Background: Although the descending aortic diameter is larger in smokers, data about thoracic aortic growth is missing. This chapter aims to present the distribution of thoracic aortic growth in smokers and compare it with the literature of the general population.

Methods: Current and ex-smokers aged 50-70 years from the longitudinal Danish Lung Cancer Screening Trial were included. Mean and 95^{th} percentile of annual aortic growth of the ascending aortic (AA) and descending aortic (DA) diameters were calculated with the first and last non-contrast computed tomography scans during follow-up. Determinants of change in aortic diameter over time were investigated with linear mixed models.

Results: A total of 1987 participants (56% male, mean age 57.4±4.8 years) were included. During a median follow-up of 48 months, mean AA and DA growth rates were comparable between males (AA 0.12 ± 0.31 mm/year and DA 0.10 ± 0.30 mm/year) and females (AA 0.11 ± 0.29 mm/year and DA 0.13 ± 0.27 mm/year). The 95^{th} percentile ranged from 0.42 to 0.47 mm/year, depending on sex and location. Aortic growth was comparable between current and ex-smokers and aortic growth was not associated with pack-years. Our findings are consistent with aortic growth rates of 0.08 to 0.17 mm/years in the general population. Larger aortic growth was associated with lower age, increased height, absence of medication for hypertension or hypercholesterolemia, and lower Agatston scores.

Conclusions: This longitudinal study of smokers in the age range of 50-70 years shows that ascending and descending aortic growth is approximately 0.1 mm/year and is consistent with growth in the general population.

Highlights:

- In current and ex-smokers, the ascending and descending aorta grows on average 0.1 mm/year
- Aortic growth was comparable between current and ex-smokers and was not associated with pack-years
- Thoracic aortic growth in smokers is consistent with cross-sectional data from the general population
- Annual growth did not statistically significantly differ between males and females
- Based on 95^{th} percentiles, aortic growth of 0.5 mm/year can be considered the upper limit of normal

3.1 Introduction

Dilatation of the thoracic aorta is associated with an increased risk of aortic dissection [87], with high mortality rates of up to 50% in the first 30 days [30]. In addition to the absolute diameter, the fast growth of 3-5 mm/year is mentioned in the guidelines on the diagnosis and treatment of aortic diseases as an important risk factor for dissection and is, therefore, an additional indication to perform preventive surgery [76, 92]. However, data about risk factors for fast aortic growth is scarce. It has been shown that patients with a bicuspid aortic valve or Marfan syndrome show larger aortic growth rates than the general population [93]. Smoking is associated with a larger diameter of the aortic arch and descending aorta [22, 94, 95] and with larger aortic growth of the abdominal aorta [96]. Whether smoking is associated with faster thoracic aortic growth is still unknown. With the use of a large prospective longitudinal cohort study, the Danish Lung Cancer Screening Trial (DLCST), we aimed to investigate whether aortic growth is larger in current or former smokers when compared to the available cross-sectional studies of the general population. With our longitudinal data of the thoracic aortic growth, we will also be able to identify risk factors for fast growth in this subgroup of the population.

3.2 Methods

3.2.1 Study Population

Participants were recruited from DLCST (www.ClinicalTrials.gov, registration number: NCT 00496977), a randomized controlled trial conducted between 2004 and 2010. Participants in the DLCST volunteered in response to local media advertisements. Current and former smokers aged 50-70 years with at least 20 pack-years and forced expiratory volume in the first second (FEV 1) of $> 30\%$ of predicted value were included. Participants with body weight above 130 kg, previous treatment for any kind of cancer within 5 years, tuberculosis within 2 years, and any serious illness with life expectancy < 10 years were excluded. The primary aim of this Randomized Control Trial (RCT) was to investigate the effect of computed tomography screening on lung cancer mortality. No statistically significant effects of CT screening on lung cancer mortality were found. The study was approved by the National Ethics Committee of Denmark (identification no. H-KA-02045, supplementary protocol 20148) and all participants gave written informed consent. The study design is explained in more detail before [6].

In the DLCST study, 2052 participants were randomized to the screening group, which received annual multidetector computed tomography (MDCT) during a 5 year period. These MDCT scans provided the opportunity to perform a post-hoc analysis in which the aortic growth was measured over a long period. For this study, we excluded participants with $<$ 1-year follow-up between the first and last CT scans (n=65) because this follow-up period was too short to measure growth accurately. Therefore, overall, 1987 participants were included in the current study.

Clinical characteristics regarding smoking status, history of stroke and ischemic heart disease, medical treatment for diabetes, hypertension or hypercholesterolemia,

and Agatston calcium scores (of the ascending aorta + arch and of the descending aorta) were collected at baseline as previously defined and described [91]. The Agatston calcium score is a measure of arterial calcium on computed tomography. The calculation is based on the weighted density score given to the highest attenuation value (HU) multiplied by the volume of the calcification. The Agatston score of the ascending aorta, aortic arch, and descending aorta were assessed by one observer using Vitrea v.6.0 (Vital Images, Inc., MN, USA). A standardized procedure for calcium scoring with a threshold of 130 Hounsfield units (HU) was used to identify aortic calcifications.

3.2.2 Computed Tomography Imaging

All non-ECG-gated, non-contrast CT scans were performed in a single institution with a 16-row Philips $M \times 8000$ MDCT scanner, Philips Medical Systems, Eindhoven, the Netherlands. Scans were performed in supine position after full inspiration in the caudocranial scan direction including the entire rib cage and upper abdomen with 120 kV and 40 mAs. Scans were performed with spiral data acquisition with the following parameters: section collimation, 16×0.75 mm; pitch, 1.5; and rotation time of 0.5 s. The obtained data were reconstructed with a slice thickness of 1 mm and a hard reconstruction algorithm (Philips D kernel).

3.2.3 Measurements of Aortic Diameter

Aortic diameters were measured with the use of an automatic method, which is validated in 100 participants showing a good agreement with manual aortic diameter measurements [97]. The method combines multi-atlas registration to obtain seed points, aorta centerline extraction, and an optimal surface segmentation approach [69] to extract the aorta surface around the centerline. From the extracted 3D aorta segmentation, the average diameters of the ascending aorta and descending aorta at the level of the pulmonary artery bifurcation were computed from the cross-sectional area measured at cross-sectional slices perpendicular to the extracted centerline (Section 3.2.3). The aortic wall with possible calcification was included in the measurements. In 29 participants, an error occurred in the automatic method for centerline extraction, and therefore no aortic diameters were automatically computed. The ascending and descending aortic diameters for these cases were measured manually by drawing the centerline and cross-sectional vessel contour perpendicular to the centerline at the pulmonary bifurcation level as described in detail in our previous work [97]. In the remaining 1958 subjects with accurate centerline extraction, we visually checked the following cases to identify inadequate measurements as a result of the automatic method: (1) all outliers of the aortic diameter at baseline and follow-up defined as 2.7 standard deviation above or beneath the median; (2) all subjects who showed aortic growth or decline of > 3.5 mm; and (3) a random sample of 200 images (100 baseline and 100 follow-up scans in the same subjects). From the randomly selected 200 scans, only 3 (1.5%) at the ascending aorta and 4 (2%) at the descending aorta showed a slight over or under segmentation. Overall, in 68 subjects, adequate measurements of the automatic method were not available due to inadequate segmentation. Also, the aortic diameters for these 68 cases were measured manually

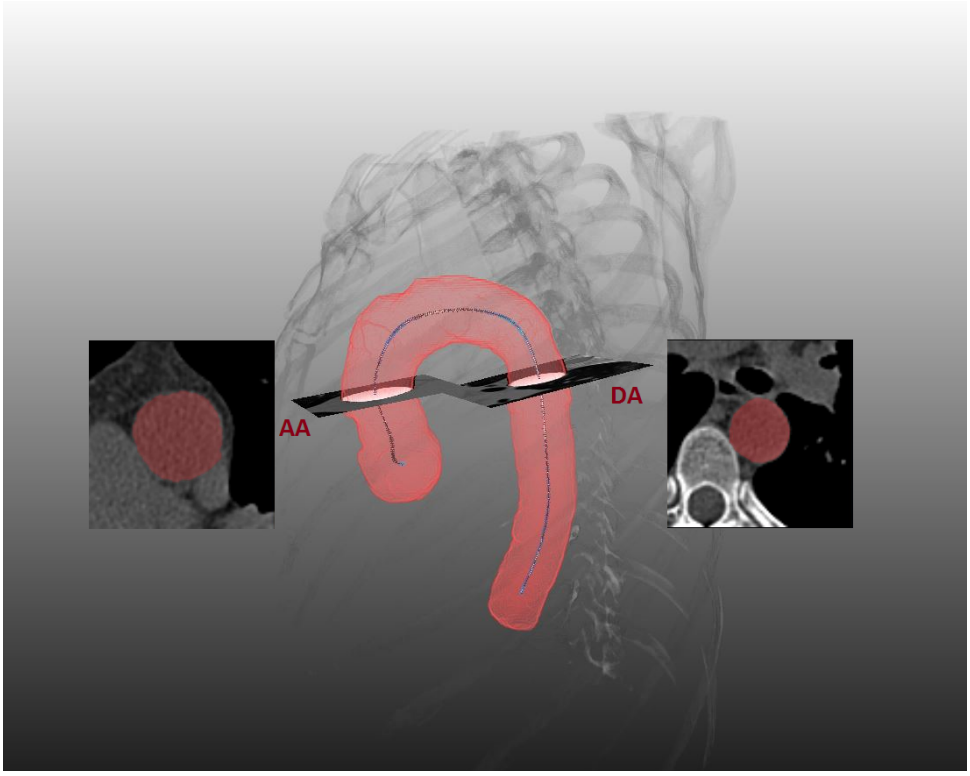


Figure 1: Measurements of the average ascending and descending aortic diameter. *A 3D image of the automatic tool which extracts the centerline (blue) and the surface of the aorta (red) to compute the ascending (AA) and descending aortic (DA) diameters at the level of the pulmonary artery bifurcation. Cross-sectional views of the ascending and descending aorta are shown left (AA) and right (DA). Both cross-sections are overlaid with the automatically extracted aortic area (in red). The average diameter is computed as* $Diameter = 2\sqrt{Area/\pi}$.

for both the ascending and descending aorta diameter. As a result, ascending and descending aortic diameters and aortic growth were available in all 1987 participants.

3.2.4 Statistical Analysis

Data are expressed as mean ± SD or as median ± interquartile range in case the distribution was not normal. Data distribution was checked using histograms. Categorical variables are presented as frequencies with percentages. To present the distribution of annual aortic growth, the annual growth rate was calculated by subtracting the aortic diameter measured on the baseline CT scan from the aortic diameter measured on the last scan during follow-up and subsequently dividing this value by the number of years

between the baseline and last follow-up scan. The Student's t-test or Mann-Whitney test was used to compare means between two groups at baseline. Comparison of categorical variables was made using the Chi-square test or the Fisher's exact test. For the analyses of Agatston scores, we used natural log-transformed values and added 1.0 mm^3 to the nontransformed Agatston values (Ln(calcification volume + 1)) to deal with values of zero. For pack-years, we used the log-transformed values.

To investigate whether change in aortic diameter was associated with baseline characteristics, linear mixed-effects (LME) models were used. The ascending and descending aortic diameters were consecutively used as the dependent variable. Time was entered as a random effect. First, all baseline variables were entered concomitantly as independent variables to identify whether they were independently associated with the aortic diameter (while considering that the aortic diameter was measured twice in each participant by using the LME). All baseline characteristics (i.e. age, height, weight, sex, medical treatment, medical history, pack-years, Agatston scores) were deemed clinically relevant based on previous research [22, 95, 98]. Second, interaction terms of each of the baseline variables with time were entered consecutively into the multivariable model to assess the independent effect of each of these variables on the change of aortic diameter over time. We also examined the interaction term between time and large aortic diameter (ascending aorta > 40 *mm* and descending aorta > 30 *mm*) to assess whether participants with larger aortic diameters show larger changes in aortic diameter over time. All interaction terms that were found to be significant were presented in the figures. We checked whether the assumptions underlying linear mixed-effects modeling (linearity and homoscedasticity) were satisfied.

The IBM SPSS® statistics 21.0 software was used to analyze the data, and a p-value of < 0.05 was considered significant.

3.3 Results

3.3.1 Study population

The baseline characteristics of the 1987 included participants are presented in Table 1 Table 1 for the total group and separately for males and females. The mean age of our cohort was 57.4 ± 4.8 years. Antihypertensive medication was used by 14.8% of the participants.

3.3.2 Aortic Diameters and Aortic Growth

The distribution of the aortic diameters for both males and females can be found in Supplemental Figure 3.A.1. The ascending and descending aortic diameters at baseline were significantly larger in males (ascending aorta 36.0 ± 3.5 *mm* and descending aorta 28.2 ± 2.2 *mm*) than in females (ascending aorta 33.6 ± 3.2 *mm* and descending aorta 26.1 ± 2.2 *mm*). A baseline aortic diameter of $\geqslant 40$ *mm* at the ascending aorta was found in 167 (8%) participants. For the descending aorta, a baseline aortic diameter of $\geqslant 40$ *mm* was found in 1 (0%) participant and $\geqslant 30$ *mm* in 257 (13%) participants. The distribution of annual aortic growth for both males and females is shown in Figure 2 and was calculated during a median follow-up of 48 months (IQR $47 - 50$ months).

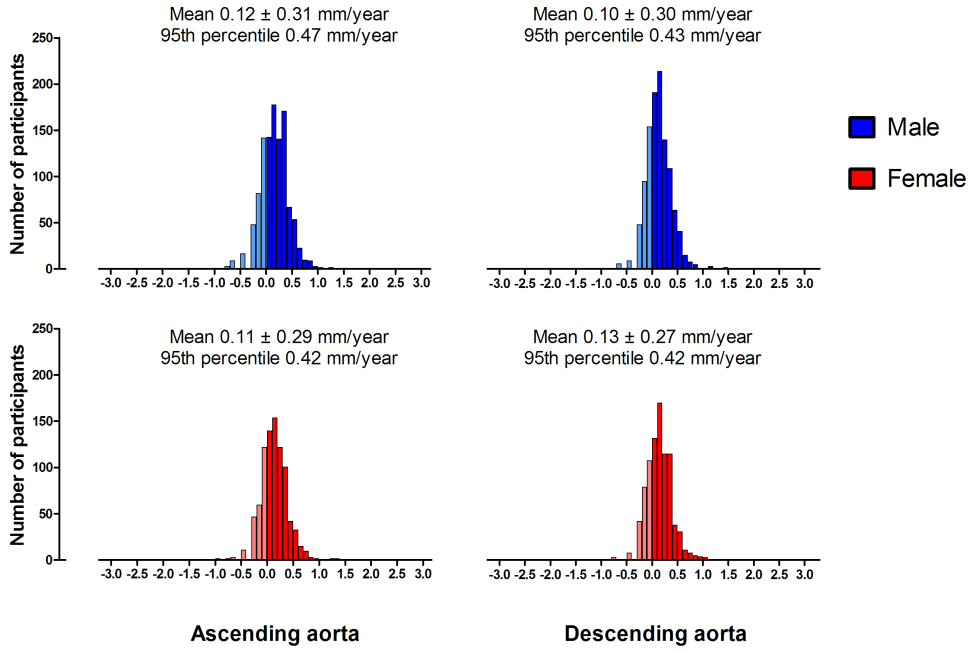


Figure 2: Annual growth of the ascending and descending aorta. *Lighter bars represent a decrease in diameter, and darker bars represent an increase in diameter. No differences were found between males and females in ascending aortic growth ($p = 0.394$) and descending aortic growth ($p = 0.087$).*

Annual growth did not statistically significantly differ between males and females for the ascending aorta (males 0.12 ± 0.31 mm/year and females 0.11 ± 0.29 mm/year) and descending aorta (males 0.10 ± 0.30 mm/year and females 0.13 ± 0.27 mm/year). In addition, the aortic growth did not differ significantly between current or former smokers for the ascending aorta (current 0.12 ± 0.30 mm/year and former 0.13 ± 0.29 mm/year) and descending aorta (current 0.11 ± 0.30 mm/year and former 0.11 ± 0.25 mm/year). In total, 621 (31%) participants showed decrease of the ascending aortic diameter in time and 604 (30%) of the descending aortic diameter. Eighteen people (1%) had an aortic growth of > 1 mm/year, which in 9 persons only occurred in the ascending aorta (2 former and 7 current smokers), in 6 persons only in the descending aorta (all current smokers) and in 3 persons in both the ascending and descending aorta (one former and two current smokers). In two people (0.1%) > 2 mm/year (both descending aorta) was found and only one (0.05%) showed > 3 mm/year.

3.3.3 Determinants of Aortic Growth

The association between the baseline characteristics and the aortic diameter is shown in Supplemental Table 3.A.1. Higher age, larger height and weight, hypertension, and

Table 1: Baseline characteristics. *Values are presented as mean (SD) or median (IQR) for continuous variables and N (%) for dichotomous variables. Missing values were present for age (n=1,* 0.0%*), weight (n=2,* 0.0%*) and Agatston scores of the aorta (n=8,* 0.4%*).*

	Total (n=1987)	**Males (n=1111)**	**Females (n=876)**	**p-value**
Age *(years)*	57.4 ± 4.8	57.8 ± 4.8	56.9 ± 4.8	< 0.001
Height *(cm)*	173.8 ± 8.8	179.4 ± 6.3	166.7 ± 6.0	< 0.001
Weight *(kg)*	76.5 ± 14.2	83.2 ± 12.1	68.0 ± 12.0	< 0.001
Medical treatment				
Hypertension, ***N*** *(%)*	294(14.8%)	158(14.2%)	136(15.5%)	0.416
Hypercholesterolemia, ***N*** *(%)*	168(8.5%)	109(9.8%)	59(6.7%)	0.014
Diabetes, ***N*** *(%)*	39(2.0%)	30(2.7%)	9(1.0%)	0.008
History of stroke, ***N*** *(%)*	34(1.7%)	24(2.2%)	10(1.1%)	0.082
History of ischemic heart disease, ***N*** *(%)*	40(2.0%)	36(3.2%)	4(0.5%)	<0.001
Current smoking, ***N*** *(%)*	491(24.7%)	274(24.7%)	217(24.8%)	0.955[c]
Pack-years[a]	34 (27-42.5)	36 (29-46)	31 (25.5-39)	< 0.001[b]
Agatston score ascending aorta + arch[a]	36 (0-273)	33 (0-247)	39.5 (0-303.8)	0.795[b]
Agatston score descending aorta[a]	0 (0-38)	0 (0-45)	0 (0-25)	0.005[b]
Baseline ascending aortic diameter	35 ± 4	36 ± 3	34 ± 3	< 0.001
Baseline descending aortic diameter	27 ± 2	28 ± 2	26 ± 2	< 0.001

[a]Nontransformed median score with interquartile range
[b]Mann-Whitney test
[c]Fisher's exact test

higher Agatston scores were associated with larger ascending aortic diameters, while female and diabetes were associated with smaller ascending aortic diameters. For the descending aorta, higher age, height, weight and Agatston scores were associated with larger aortic diameters, while female and hypercholesterolemia were associated with smaller aortic diameters.

Figure 3 shows the significance of the interaction terms between baseline variables and time from the linear mixed-effects models. Larger height was associated with a larger increase in aortic diameter over time. Higher age, hypertension, hypercholesterolemia, and Agatston scores were associated with a smaller increase in ascending aortic diameter over time. For the descending aorta, higher age, hypertension, and higher Agatston score of the descending aorta were associated with a smaller change of the descending aorta over time.

3.4 Discussion

This is the first study presenting longitudinal data on sex-specific growth of the ascending and descending aorta in a large population of current or former smokers

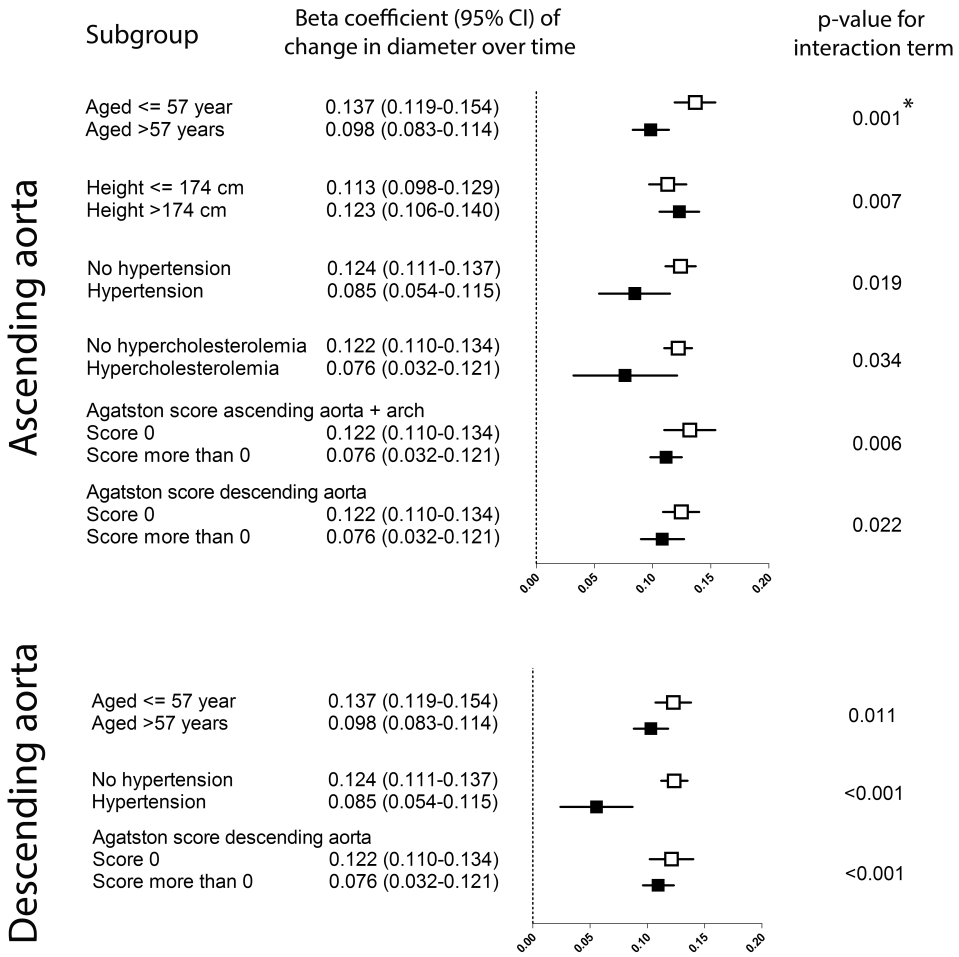


Figure 3: Mixed models including interaction terms between baseline variable and time in years. *All models were adjusted for sex, age, height, weight, hypertension, hypercholesterolemia, diabetes, history of stroke, history of ischemic heart disease, pack-years, Agatston score of the ascending aorta and aortic arch, and the Agatston score of the descending aorta. The continuous variables age and height were dichotomized at their median value. Age, height, and Agatston scores were added as a continuous variable in the interaction terms.*
**Interpretation: A higher age is associated with less increase in the ascending aortic diameter over time (in years).*

with at least 20 pack-years. Males showed a growth of 0.12 ± 0.31 mm/year for the ascending aorta and 0.10 ± 0.30 mm/year for the descending aorta. In females, we found a growth of 0.11 ± 0.29 mm/year for the ascending aorta and 0.13 ± 0.27 mm/year for the descending aorta. Previous studies showed that smoking is associated with larger diameter of the aortic arch or descending aorta [22, 94, 95], suggesting faster growth. As such, it would be expected that the descending aortic growth will also be faster in our study compared to the general population. Nevertheless, our study showed comparable or even smaller growth rates than the two largest cross-sectional cohort studies that reported on the association between age and descending aortic diameter. Kalsch et al. [75] calculated an increase of 0.17 mm for males and 0.16 for females per 1-year increase in age. Wolak et al. [22] showed that the descending aortic diameter was 0.13 mm larger each 1-year increase in age, which is comparable to our results. Only one study, the Framingham Heart Study [98], measured the thoracic aortic growth longitudinally in a healthy population, but they solely measured the growth at the level of the aortic root. In addition, we have also found no association between pack-years and descending aortic growth. Therefore, we can conclude that our data do not support the hypothesis that descending aortic growth would be larger in current or former smokers compared to the general population. Since there is no association found previously between the ascending aortic diameter and smoking, we did not expect any effect of smoking on the ascending aortic growth, which was also confirmed by our results.

The conclusions must be interpreted with caution, consiering the measurement variability of non-ECG-gated, non-enhanced CT. In previous literature, the mean intra-observer variation between two measurements of the ascending aorta found in contrast CT scans is found to be $0.1 - 0.3$ mm for manual measurements [86, 99, 100]. Possibly for non-contrast CT scans it is larger. The decrease in AA and DA diameter in 31% and 30% of the participants, respectively, is in part caused by this measurement variability. However, the absolute mean difference between the first and last CT scan, not divided by the amount of years in between the two scans, was 0.46 ± 1.05 mm for the ascending aorta and 0.44 ± 0.97 mm for the descending aorta. This is higher than we would expect based on the intra-observer variability of $0.1 - 0.3$ mm and therefore, our change in aortic diameter could not only be explained by measurement variability. Moreover, the use of identical CT scanners and automated segmentation for both baseline and follow-up measurements is an important strength of this study because it prevented us from additional inter-observer and inter-modality variability.

3.4.1 Determinants of Aortic Growth

From previous literature, we know that body measurements are important in the assessment of aortic diameters [28, 101]. For instance, in Turner patients with typically a short stature, the use of the aortic size index (ASI) is advised, which corrects for body surface area [102]. For aortic growth, little data is available on the effect of body measurements. The Framingham Heart Study [98] included a slightly younger population (mean age 50 ± 14 years) with comparable BMI (25.5 ± 4.4 kg/m^2) and showed that BMI was correlated with change in aortic root diameter over time. We examined height and weight separately and showed that the effect of body measures

on the ascending aortic growth is mainly based on height.

Higher age was associated with both less ascending and descending aortic growth. Aortic remodeling over the adult life is accompanied by reduced aortic elasticity [103] and reduced tortuosity with increased curvature [104]. Because of these changes, one may expect that aortic growth will decrease at older age, and aortic diameters will stabilize. Treatment for hypertension was associated with slower ascending and descending aortic growth. Since higher blood pressure is associated with larger descending aortic diameters [95], we would assume that participants with hypertension would show larger descending aortic growth. However, patients being treated for hypertension may represent the group with controlled blood pressure, and the group of patients who are not receiving treatment may contain patients with uncontrolled blood pressure. Because we had no information about the exact blood pressure, which is a limitation of this study, we could not verify this assumption. This could also be the case with hypercholesterolemia because patients with treatment for hypercholesterolemia showed a smaller increase in aortic diameter. An ascending aortic diameter of > 40 mm or descending aortic diameter of > 30 mm was not associated with change in diameter over time in our cohort. Although patients with aortic aneurysms show larger growth rates [105], we may have had too few patients with aortic dilatation in our cohort to prove this.

A recent systematic review, which included all causes of thoracic aortic aneurysms, showed a mean growth rate in patients from 0.2 to 2.8 mm/year for ascending aorta and aortic arch, while those for descending and aorta ranged from 1.9 to 3.4 mm/year [105]. Detaint et al. [93] observed at the level of the ascending aorta an aortic growth of 0.12 ± 1.0 mm/year in Marfan syndrome and 0.42 ± 0.6 mm/year in bicuspid aortic valve (BAV). These growth rates of patients with a bicuspid aortic valve or degenerative aortapathy are larger than found in our cohort with current and ex-smokers. In current guidelines for thoracic aortic diseases, different definitions are used for extensive growth ($\geqslant 3$ mm or $\geqslant 5$ mm), which warrants preventive surgery. Our study showed only two cases with growth > 2 mm/year and only one with > 3 mm/year, which suggests that extensive growth, defined by the guidelines, is relatively rare in the general smoking population. Based on our 95^{th} percentiles, annual aortic growth of 0.5 mm is the upper limit of normal in current or former smokers.

3.4.2 Limitations

One large limitation of our study is that we did not include our own reference group of healthy subjects. The literature only contained cross-sectional data with the mean thoracic aortic growth rate of the general population, and therefore we could not compare the distribution (95^{th} percentile) of aortic growth rates in our group with a reference group. Another limitation is the lack of information about diseases related to aortic pathology, such as connective tissue disease and bicuspid aortic valve. This information was not available because the primary aim of this RCT was to investigate the effect of computed tomography screening on lung cancer mortality. Because this study was a post-hoc analysis, thoracic aortic growth was neither a primary nor a secondary outcome measure of the original trial. The limited age range of $50-70$ years also prevents the generalization of our results to the total population. In addition,

the limited aortic growth may have limited the power of our analysis. However, our cohort was large enough to find a significant aortic growth for both the ascending and descending aorta, and also several determinants were found to be significant associated with the change in thoracic aortic diameter over time. Another limitation of this study is the use of non-ECG-gated, non-contrast CT scans. Non-ECG-gated CT scans show significantly more motion artifacts than ECG-gated CT scans [106], which likely affect aortic measurements. The use of contrast-enhanced CT is preferred for thoracic aortic measurements but could cause unnecessary complications. However, both baseline and follow-up measurements were made in the same manner.

An issue that warrants consideration in our study is the fact that we examined a total of 13 variables. If we were to account for multiple testing using a Bonferroni correction, only age would remain statistically significant for the ascending aorta, while hypertension and Agatston score would remain statistically significant for the descending aorta. However, our study was not data-driven but hypothesis-driven; the choice of variables we investigated was based on previous findings from the literature. These variables were thus already implicated in the disease process by earlier studies. Correcting for multiple testing in spite of this hypothesis-driven approach could result in failure to recognize potentially interesting factors. In any case, our findings may be considered as indicative of a potential association, and these hypothesis generating findings merit validation in other large studies.

3.5 Conclusion

This longitudinal study of current and ex-smokers shows that the ascending and descending aorta grows on average 0.1 mm/year in both males and females in the age range of $50 - 70$ years. The aortic growth rates are consistent (or even smaller) with the numbers available in cross-sectional studies of the general population. According to the 95^{th} percentile, an aortic growth of > 0.5 mm/year can be considered the upper limit of normal. Larger change of aortic diameters in time was associated with lower age, increased height, absence of medication for hypertension or hypercholesterolemia, lower Agatston score, and a large thoracic aortic diameter.

Funding

This chapter was supported by grants from the Dutch Heart Foundation (2003B179) and ZonMw (84920001). In addition, this study was funded by the Iranian Ministry of Science Research and Technology (MSRT), the Netherlands Organisation for Scientific Research (NWO). The Danish Lung Cancer Screening Trial (DLCST) was funded by the Danish Ministry of Interior and Health. The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Appendix

3.A Supplemental Material

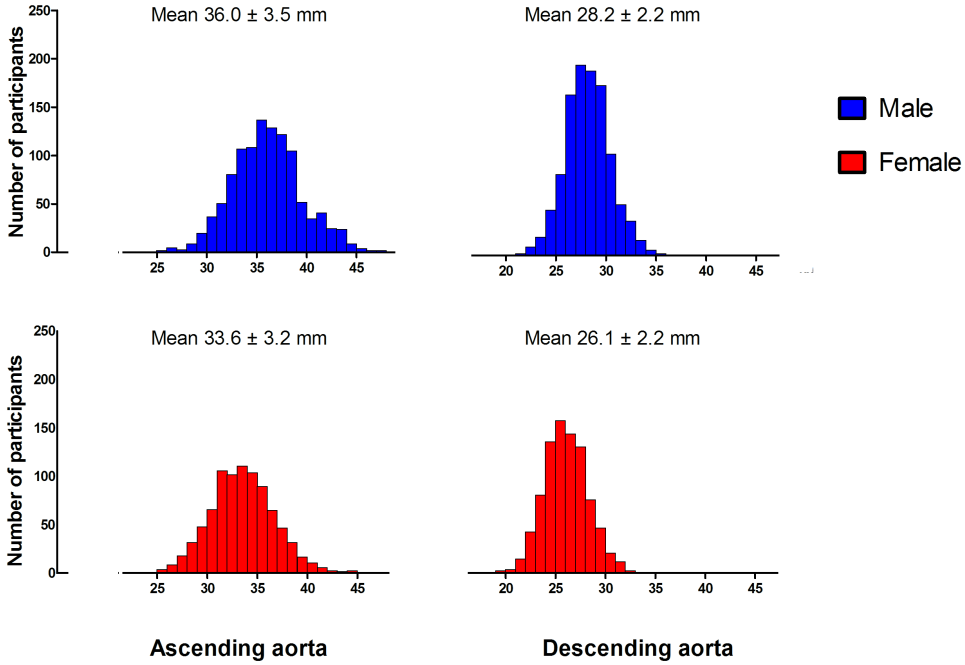


Figure 3.A.1: Baseline ascending and descending aortic diameters. *Males show a larger baseline diameter than females for both the ascending and descending aortic diameters ($p < 0.001$).*

Table 3.A.1: Mixed models for ascending and descending aortic diameters.

		Regression coefficient (95% CI)	*p-value*
Ascending Aorta	**Time of measurement**	0.12 (0.11;0.13)	< 0.001
	Female†	-0.94 (-1.35;-0.52)	< 0.001
	Age, 10 years	1.37 (1.04;1.70)	< 0.001
	Height, cm	0.06 (0.04;0.09)	< 0.001
	Weight, kg	0.05 (0.03;0.09)	< 0.001
	Hypertension	0.61 (0.20;1.03)	0.004
	Hypercholesterolemia	-0.40 (-0.95;0.15)	0.151
	Diabetes	-1.30 (-2.32;-0.28)	0.012
	History of stroke	0.72 (-0.35;1.79)	0.187
	History of ischemic heart disease	-0.37 (-1.43;0.69)	0.490
	Pack-years‡	0.03 (-0.41;0.47)	0.894
	Agatston score ascending aorta + arch‡	0.08 (0.02;0.14)	0.012
	Agatston score descending aorta‡	0.09 (0.03;0.16)	0.006
Descending Aorta	**Time of measurement**	0.11 (0.10;0.12)	< 0.001
	Female	-0.85 (-1.10;-0.59)	< 0.001
	Age, 10 years	1.52 (1.32;1.73)	< 0.001
	Height, cm	0.04 (0.02;0.05)	< 0.001
	Weight, kg	0.04 (0.03;0.05)	< 0.001
	Hypertension	0.15 (-0.11;0,41)	0.251
	Hypercholesterolemia	-0.52 (-0.86;-0.18)	0.003
	Diabetes	-0.28 (-0.91;0.35)	0.387
	History of stroke	-0.49 (-0.15;0.17)	0.149
	History of ischemic heart disease	0.29 (-0.36;0.95)	0.383
	Pack-years‡	0.10 (-0.18;0.37)	0.487
	Agatston score ascending aorta + arch‡	0.04 (0.00;0.08)	0.048
	Agatston score descending aorta‡	0.12 (0.08;0.16)	< 0.001

∗ All baseline variables were entered concomitantly as independent variables to identify whether they were independently associated with the aortic diameter (while taking into account that the aortic diameter was measured twice in each participant by using the LME)
† Interpretation: the ascending aortic diameter is on average 0.94 *mm* smaller in females than in males.
‡ Log transformed

4

Optimal Surface Graph Cuts to Segment the Pulmonary Artery and Aorta on Non-contrast CT

Zahra Sedghi Gamechi, Andres M. Arias-Lorza, Zaigham Saghir, Daniel Bos, Marleen de Bruijne

Under review

Abstract

Purpose: Accurate segmentation of the pulmonary arteries and aorta is important due to the association of the diameter and the shape of these vessels with several cardiovascular diseases and with the risk of exacerbations and death in patients with COPD. We propose a fully automatic method based on an optimal surface graph cut algorithm to quantify the full 3D shape and the diameters of the pulmonary arteries and aorta in non-contrast computed tomography (CT) scans.

Methods: The proposed algorithm first extracts seed points in the right and left pulmonary arteries, the pulmonary trunk, and the ascending and descending aorta using multi-atlas registration. Subsequently, the centerlines of the pulmonary arteries and aorta are extracted by a minimum cost path tracking between the extracted seed points, with a cost based on a combination of lumen intensity similarity and multi-scale medialness in 3 planes. The centerlines are refined by applying the path tracking algorithm to curved multi-planar reformatted scans and are then smoothed and dilated non-uniformly according to the extracted local vessel radius from the medialness filter. The resulting coarse estimates of the vessels are used as initialization for a graph-cut segmentation. Once the vessels are segmented, the diameters of the pulmonary artery (PA) and the ascending aorta (AA) and the PA:AA ratio are automatically calculated both in a single axial slice and in a 10 *mm* volume around the automatically extracted pulmonary artery bifurcation level. The method is evaluated on non-contrast CT scans from the Danish Lung Cancer Screening Trial. Segmentation accuracy is determined by comparing with manual annotations on 25 CT scans. Intra-class correlation (ICC) between manual and automatic diameters, both measured in axial slices at the pulmonary artery bifurcation level, is computed on an additional 200 CT scans. Repeatability of the automated 3D volumetric diameter and PA:AA ratio calculations (perpendicular to the vessel axis) are evaluated on 118 scan-rescan pairs with an average in-between time of 3 months.

Results: We obtained a Dice segmentation overlap of 0.94 ± 0.02 for pulmonary arteries and 0.96 ± 0.01 for the aorta, with a mean surface distance of 0.62 ± 0.33 *mm* and 0.43 ± 0.07 *mm* , respectively. ICC between manual and automatic in-slice diameter measures was 0.92 for PA, 0.97 for AA, and 0.90 for the PA:AA ratio, and for automatic diameters in 3D volumes around the pulmonary artery bifurcation level between scan and rescan was 0.89, 0.95, and 0.86, respectively.

Conclusions: The proposed automatic segmentation method can reliably extract diameters of the large arteries in non-ECG-gated non-contrast CT scans such as are acquired in lung cancer screening.

4.1 Introduction

Cardiovascular diseases and Chronic Obstructive Pulmonary Disease (COPD) are among the major leading causes of death globally [1]. In the search for early identification of individuals at risk of cardiovascular disease in COPD [2, 4], imaging-based assessments of the shape and size of the pulmonary artery and aorta have rapidly gained interest. Changes in these two large arteries may indicate cardiovascular diseases including pulmonary hypertension [19, 20], aortic dilatation and aortic aneurysm [22], and coarctation of the aorta [23]. The pulmonary artery to ascending aorta diameter ratio (PA:AA) at the level of pulmonary artery bifurcation is shown to be associated with the presence of pulmonary arterial hypertension [20] and is associated with poorer health status [107], increased risk of severe exacerbations [9, 25], and increased mortality [26] in patients with COPD. Performing diameter measurements manually is labor-intensive and time-consuming and has high intra- and inter-observer variability. Diameter measurements derived from 3D segmentations are more reliable but are even more time-consuming to obtain manually. To accurately assess the pulmonary arteries and aorta, automatic 3D segmentation is therefore desirable.

With the growing use of low-dose non-contrast thoracic CT scans for lung cancer screening [6, 16, 18], there is an opportunity to measure the pulmonary arteries and aorta in these scans in order to investigate the presence of early-stage cardiovascular disease and/or predict complications in patients with COPD. However, in non-contrast CT, segmentation of the aorta and especially the pulmonary artery is challenging due to proximity to other structures with similar intensity values. Furthermore, in non-ECG-gated CT as is commonly used in lung screening, additional challenges are motion artifacts and unclear vessel boundaries at the regions close to the heart.

In the literature, automated segmentation methods of the pulmonary arteries and aorta have been presented mainly for Magnetic Resonance Imaging [41] and contrast-enhanced CT Angiography (CTA) [48, 49] which have high contrast between vessels, fat, and surrounding muscles. However, these methods do not translate well to non-contrast CT, where the vessel boundaries are not well defined in many places. Therefore, relatively fewer studies can be found on non-contrast CT scans, on the segmentation of the aorta [53–56, 58–63] and especially the pulmonary artery containing the pulmonary trunk, left, and right pulmonary arteries [64, 65].

Among existing segmentation methods, those using a shape prior [53, 59, 60, 65] generally obtain good segmentation results on non-contrast CT scans. Xie et al. [59, 65] employed a cylinder matching method to extract the centerline of the pulmonary artery trunk and aorta. To segment the vessels, they used geometric constraints from adjacent organs obtained from a pre-computed anatomy label map. Although the obtained results are good, an anatomy map is not always available. Another approach is to use graph cut methods with shape priors. Graph cuts can achieve a global optimum with low processing times and it is possible to incorporate shape constraints in the graph structure. Graph cut methods have been applied to different imaging modalities for artery segmentation [67, 68] and have obtained promising results in many tasks. Deng et al. [67] proposed a graph-cut method using random forest based discriminative features on non-contrast CT for aorta segmentation. They achieved a high segmentation performance in the abdominal aorta, however, they have not

applied their method to the pulmonary artery, which is more challenging to segment.

This chapter presents an optimal surface graph cut-based method to segment the pulmonary arteries and aorta. We adopt the optimal surface graph cut approach by Petersen et al. [69], originally proposed for airway segmentation, which incorporates a shape prior via constructing the graph based on flow lines traced from an initial, smoothed segmentation. The non-intersecting flow lines guarantee non-self-intersecting surfaces and make it possible to segment high curvature areas such as the bifurcation of the pulmonary artery and the aortic arch while guaranteeing a shape that is similar to the initialization shape.

A preliminary version of the work presented in this chapter is presented in [108]. In the current work, the proposed method is fully automated and includes a multi-atlas registration technique to automatically extract seed points and a landmark for the level of the pulmonary artery bifurcation. At the level of the extracted landmark, the PA and AA diameters and PA:AA ratio is automatically extracted. The current work also provides an extensive validation using a larger data set, comparison with manual measurements, and scan-rescan repeatability assessment. Unlike our previous work in Chapter 2 focusing on aorta segmentation [97], in this chapter, we segment both the pulmonary artery and aorta with a more robust landmark detection technique.

4.2 Methods

The main steps of our proposed method are (1) Pre-processing; (2) Automatic seed point and landmark extraction with multi-atlas registration; (3) Centerline extraction for vessel localization; (4) Vessel segmentation using an optimal surface graph cut algorithm; (5) 3D diameter measurement and biomarker extraction. An overview of our method is shown in Figure 1.

4.2.1 Pre-processing

To reduce the unnecessary computational cost a bounding box is calculated around the lungs. The lungs are segmented using thresholding and morphological smoothing similar to Lo et al.[109]. Thereafter, the scans are cropped using a bounding box around the lungs Figure 2 (a). Then, low and high intensity values are clipped to prevent the centerline extraction and vessel segmentation attracting to the heart-lung or bone-lung borders, which often have a higher gradient than the boundary of the vessels of interest. Truncating intensities higher than 150HU, such as those presented in bones, or lower than -150HU, generally presented in the lungs, makes the vessel borders relatively stronger and easier to detect. The gradient magnitude of an axial slice shown in Figure 2 (b-e) illustrates the effect of pre-processing on enhancing the edges of the pulmonary arteries and aorta.

4.2.2 Seed point Extraction

Seed points are obtained with a multi-atlas registration method similar to the one presented in Tang et al. [82]. A set of N atlas images A_i, $i = 1, 2, ..., N$ and their corresponding label images La_i, Ls_i are used. La_i includes the manual segmentation of

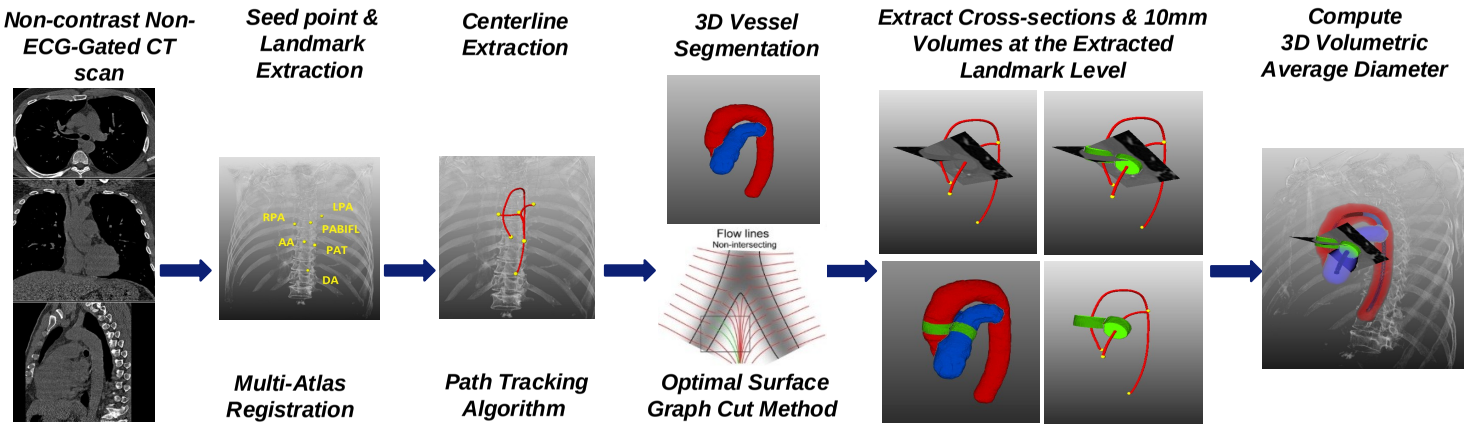


Figure 1: An illustration of the method overview. *The 3D volumetric segment of the segmented aorta (red) and pulmonary artery (blue) is shown in green. Abbreviations are: pulmonary artery trunk (PAT), left (LPA) and right pulmonary artery (RPA), pulmonary artery bifurcation level (PABIFL), ascending aorta (AA), descending aorta (DA). .*

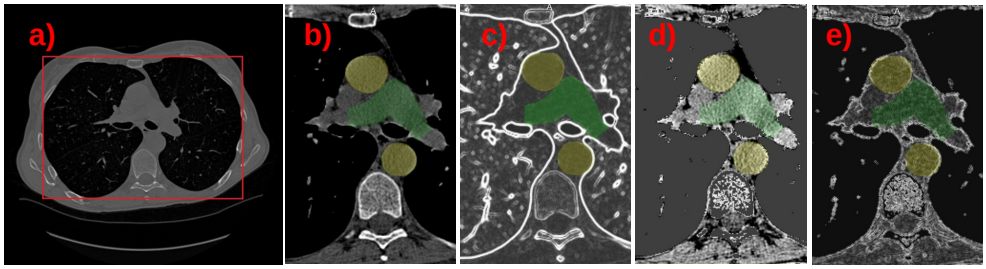


Figure 2: ***CT pre-processing.*** *An axial view of a CT scan with the bounding box in red is in(a), zoomed-in view in (b), and the pre-processed scan in (d). The corresponding gradient magnitude of (b) is in (c), and the gradient magnitude of the pre-processed scan is in (e). Strong edges at lung and bone borders (b, c) are removed during pre-processing (d, e). Images are overlaid with the manual PA (green) and aorta (yellow) annotations.*

the pulmonary arteries and aorta and Ls_i includes manual seed points at the left and right pulmonary arteries before the secondary bifurcations, pulmonary artery trunk, pulmonary artery bifurcation point, ascending aorta at the sinotubular junction, and descending aorta at the diaphragm level. To address the large variation in the shape and the size of these arteries N is set to 25 atlases.

The multi-atlas registration approach consists of three stages. First, each of the atlas images A_i is registered to the target image T using an affine transformation followed by a nonrigid registration using a B-spline transformation model. Normalized mutual information is used as the similarity metric. The ten registered atlas images with the highest similarity (also defined by normalized mutual information) to T are selected, and the corresponding transformations are applied to Ls_i images to propagate the seed points to T. Finally, deformed Ls_i images are combined, and the final seed points are obtained by averaging the seed point locations per label.

The seed points of the aorta, left, right, and the trunk of the pulmonary artery are used in the next step to initialize the centerline extraction. The seed point of the pulmonary artery bifurcation level is the landmark level for measuring the pulmonary artery and aorta diameters as well as the PA:AA ratio.

4.2.3 Centerline Extraction

We applied a minimum path tracking algorithm to extract the vessel centerlines between the automatically extracted seed points. For the cost function, we used a weighted combination of a medialness filter [110], $m(x)$, and a lumen intensity similarity filter $s(x)$, similar to Tang et al. [82]. Medialness, $m : \Omega \rightarrow [0, 1]$, is a multi-scale filter which uses the circularity assumption and accumulates edge responses along different circle sizes, within a defined radius range of $[R_{min}, R_{max}]$. This gives strong responses in the voxels in the center of the vessel and drops rapidly towards the vessel boundary. Simultaneously, we extracted an estimate of the vessel radius at

each voxel, $r(x) : \Omega \rightarrow [R_{min}, R_{max}]$ by extracting the radius of the circle with the strongest edge response at each voxel location. We applied the medialness filter in the axial, coronal, and sagittal planes, with a different radius range for the pulmonary arteries and for the aorta. The medialness for the pulmonary arteries is defined to be the maximum medialness of axial, coronal, and sagittal medialness filters and for the aorta, it is defined to be the maximum medialness of axial and coronal medialness filters.

In regions with unclear vessel boundaries, medialness alone is not sufficient to ensure correct centerlines. Therefore, to prevent the vessel centerline from moving outside the vessel lumen, we added a lumen intensity similarity term $s : \Omega \rightarrow [0, 1]$ defined as:

$$s(x) = \begin{cases} 1 & (\mu - \delta) \leq I(x) \leq (\mu + \delta) \\ e^{-\left(\frac{I(x)-\mu}{\sqrt{2}\delta}\right)^2} & \text{elsewhere} \end{cases} \tag{4.1}$$

where I is the intensity of the voxel at position x. Tang et al.[82] defined μ and δ as the mean and standard deviation of intensity in small regions around the seed points. In our case, they are selected as the mean and standard deviation of intensity in the manual annotations of the pulmonary arteries and aorta in 25 CT scans, which is $\mu = 52HU$ and $\delta = 46HU$. Structures with higher intensity, such as bones, or lower intensity, such as lungs, get a low response.

From both $m(x)$ and $s(x)$ high responses were obtained in the vessel center and low responses in the background. The cost function $C(x)$ was defined by an inverted combination of weighted $m(x)$ and $s(x)$ where the factors α and β control the importance of each term, respectively ($C(x) = \frac{1}{\epsilon + m(x)^{\alpha} s(x)^{\beta}}$). From the constructed cost function, the minimum cost path $C(x)$ was obtained by applying Dijkstra's algorithm between the automatically extracted seed points for each vessel, i.e. one path between the endpoints of the aorta and two paths between the two endpoints of the left and right pulmonary arteries and its trunk.

Finally, to improve centerline accuracy in areas with high curvature, the centerlines were smoothed with a Gaussian filter with standard deviation σ_c and then were refined by re-computing the minimum cost path after curved multi-planar reformatting (CMPR) perpendicular to the previous centerline [82]. An improved estimate of vessel radius $r(x)$ was extracted from medialness at the CMPR step. The centerline for the entire pulmonary was obtained as the union of centerlines for the right and left pulmonary arteries: $C^{PA} = C^{RP} \cup C^{LP}$.

4.2.4 Vessel Segmentation

To segment the vessels, we applied an optimal surface graph cut algorithm [69], that finds a globally optimal solution of a given cost function. By using non-intersecting columns based on flow lines from a predefined initial shape, (self)intersecting surfaces can be avoided and the topology of the prior shape is preserved. This makes segmentation of high curvature surfaces possible. To construct a graph with nodes and edges, an initial coarse segmentation is used to generate the graph columns. We used a non-uniform morphological dilation of the vessel centerlines as the initial segmentation.

Each centerline point was dilated with a spherical structuring element, with a radius extracted from the radius map $r(x)$. This non-uniform centerline dilation provides information about the shape of the vessel and results in a more accurate vessel surface than the uniform centerline dilation used in Arias et al. [68].

Once we computed the initial segmentation, we converted it to a mesh and constructed graph columns based on flow lines. The resolution of the initial mesh was set to $0.5mm \times 0.5mm \times 0.5mm$. The flow lines were obtained by tracing the gradient of a smoothed version of the initial segmentation. A Gaussian kernel with a standard deviation σ was used to smooth the initial segmentation. The flow lines were traced from each mesh node, inward in the gradient direction and outward in the negative gradient direction. Graph nodes were sampled in graph columns at regular arc length intervals along the flow lines with a sampling interval set to 0.3 *mm*. The graph nodes represent the possible image positions the vessel surface can take. A set of edges which represent the association between nodes connects the nodes in the graph. The two consecutive nodes in the same column are connected by directed edges named intra-column edges. The cost of these edges represents local image information associated with the border location and is chosen as the first-order derivative of the image intensity along the graph column. In CT scans, the vessels have a higher image intensity than the background therefore the intensity transitions from high to low and the cost gets to its minimum in the border.

To encourage a smooth segmentation similar in shape to the initialization, "smoothness penalty" edges (penalizing non-smoothness) were added as in [69]. These inter-column edges connect the nodes in adjacent columns with a constant cost P, penalizing solutions that deviate from the original shape. A minimum graph cut, minimizing the total cost of edges being cut and separating the graph vertices into vessel and background, provided the final segmentation. This minimization was solved with a min-cut/max-flow optimization algorithm. Details of the optimal surface graph segmentation approach can be found in Petersen et al. [69]

4.2.5 Diameter Measurement and PA:AA Ratio

In clinical practice, the diameter of the pulmonary artery (PA) and ascending aorta (AA) and the PA:AA ratio are measured manually, usually in an axial slice at the level of pulmonary artery bifurcation. This level is defined as the axial slice where the pulmonary trunk bifurcates, ideally where the right and left pulmonary arteries to appear to be of similar size. The selected slice may be one of a few axial slices that fit this criterion, which is likely to lead to inconsistent measurements between annotators and also between baseline and follow-up scans. Therefore, to ensure a consistent measurement, we calculated the diameters from a 3D vessel segment of 10 *mm* length from the segmented pulmonary artery and aorta. The segment is selected perpendicular to the vessel centerline from 5 *mm* before to 5 *mm* after the automatically extracted landmark for the level of the pulmonary artery bifurcation. The average diameter assuming a circular cross-section is computed from this segment as $Diameter = 2\sqrt{\frac{0.1 \times volume}{\pi}}$. Subsequently, from the average diameters, the PA:AA ratio is calculated. A 3D view of the segmented pulmonary arteries and aorta with

the 10 *mm* segment and the cross-sections in the extracted landmark level is shown in Figure 1.

For a meaningful comparison with diameters measured manually in axial slices, the automatic diameters were also calculated in the axial slice at the same level and in the same direction as the manual diameter measurement. Subsequently, the automatically extracted PA:AA ratio in the axial view was compared to the manually measured PA:AA ratio in the same view.

4.3 Experiments

4.3.1 Dataset and Manual Annotation

The 471 CT scans used in this study are from the Danish Lung Cancer Screening Trial (DLCST) [6]. The study was approved by the Ethical Committee of Copenhagen County and funded by the Danish Ministry of Interior and Health. A Multi-Detector CT scanner (M×8000 IDT 16 row scanner, Philips Medical Systems) was used to acquire scans at 120 kV/ 40 mAs at maximum inspiration breath-hold and without cardiac gating. The scans have an in-plane isotropic resolution of 0.781×0.781 and 1 *mm* slice thickness. Participants were current or former smokers between 50 and 70 years of age. Baseline CT scans of 235 participants, randomly selected, were used for parameter optimization, method development, and method evaluation. In addition, to assess the repeatability of the proposed method, 118 additional participants were selected who had a baseline scan and a repeat scan after on average 3 months [minimum 2, maximum 5].

Manual annotations of the centerline and contours of the pulmonary arteries and aorta were made using an in-house annotation tool developed in MeVisLab[1] with a similar framework as was described previously for carotid artery segmentation [78]. With this tool, first, the centerline of the vessels was drawn manually using the axial, coronal, and sagittal views to compute a CMPR. Then the centerlines were checked and modified if needed in the axial view of the CMPR generated every 1 *mm* along the centerline. In longitudinal views at six different angles, equally spaced every 30°, longitudinal contours of the vessel were drawn manually. Subsequently, cross-sectional contours were computed using spline interpolation through the intersection points of the longitudinal contours with the cross-sectional planes. Finally, after checking the cross-sectional contours in all cross-sections and adjusting them if required, the contours were converted to a 3D binary image using variational interpolation [79]. The window level/width for annotation was set to 200 HU/600 HU for all cases. For the pulmonary arteries, a centerline and binary segmentation image were first created for the pulmonary trunk + left pulmonary artery and trunk + right pulmonary artery individually. Then, the segmentation for the entire pulmonary was obtained as the union of these two segmentations.

With this tool centerlines of 35 CT scans (10 for optimizing the parameters of the centerline extraction and 25 for full manual segmentation) and volumes of 25 CT scans were annotated by an experienced observer (ZSG). Besides, an experienced physician

[1] https://www.mevislab.de/

(DB) indicated the diameter of the pulmonary artery and ascending aorta at the level of the pulmonary artery bifurcation in an additional 200 CT scans. This was done in the axial view, in the slice where both the right and left pulmonary arteries appear to be of similar size at the axial view.

4.3.2 Parameter Optimization

The centerline extraction parameters were optimized in a grid search on 10 CT scans with manually drawn centerlines. The objective was to minimize the Mean Centerline Distance (MCD) between the manual and automatic centerlines. The radius range of the medialness filter, the cost function weights α and β, and the sigma σ_c for centerline smoothing were optimized. From radius range of $[5mm, 30mm]$ we obtained a medialness radius range of $[8mm, 16mm]$ for pulmonary arteries, and $[12mm, 24mm]$ for the aorta which agreed well with the radius range reported for these vessels in the literature [22, 111]. From range $[1, 15]$, the weights $\alpha = 4$ for pulmonary arteries and $\alpha = 10$ for the aorta and $\beta = 2$ for both vessels were obtained, with smoothing of $\sigma_c = 9$ mm for pulmonary arteries, and $\sigma_c = 11$ mm for the aorta.

A five-fold cross-validation on 25 CT scans (independent of those used for centerline parameter tuning) was performed in which the best parameters determined from 20 CT scans were used to segment the 5 left out scans. In this, the best parameter set was selected as the parameter set giving the maximum average Dice similarity coefficient (DSC). The parameters to optimize were σ of the Gaussian Kernel used to smooth the initial segmentation, and the smoothness penalty P of the optimal surface graph cuts. From the 5 parameter sets extracted in the cross-validation the most frequent parameter set was selected for the rest of the experiments. The obtained optimal parameter set was σ of 4.4 mm and 2.4 mm and P of 32 and 40 for the pulmonary arteries and aorta, respectively.

4.3.3 Seed point and Centerline Extraction

The centerline and the seed point extraction were validated on 25 CT scans using Mean Centerline Distance (MCD). MCD is the average symmetric Euclidean distance of all points of the automatically extracted centerline to the manual centerline.

Seed point extraction was assessed by computing the MCD between the automatic centerlines traced from manually placed seed points with those traced from automatically extracted seed points. Seed points that were placed on the border or outside the vessel resulted in a centerline with large MCD compared to the centerlines traced from manual seed points. We also used the non-parametric Mann-Whitney U test to assess whether there is a significant difference between the MCD computed between the manual centerlines and the automatic centerlines traced from manually placed seed points and MCD between the manual centerlines and the automatic centerlines traced from automatically extracted seed points.

Failure in the centerline extraction could in all cases (471 CT scans) be automatically detected as follows. For the aorta, centerlines that never reached above the automatically extracted pulmonary artery bifurcation point, or that went through the coarse initial pulmonary artery segmentation, were considered as failed extractions.

For the pulmonary arteries, an additional centerline for only the pulmonary trunk was traced between the seed points at the pulmonary artery trunk and the pulmonary artery bifurcation point. Subsequently, the MCD between this centerline and the main centerlines was extracted and the centerlines that had a large MCD, or the ones which went through the coarse initial aorta segmentation, were considered as failed extractions.

4.3.4 Segmentation

The segmentation accuracy was assessed on 25 CT scans by comparing it with manual segmentations. The Dice similarity coefficient (DSC) was computed to assess the degree of spatial overlap of the automatic segmentation with the manual segmentation. The mean symmetric surface distance (MSD) was computed in millimeters between the manual and automatic segmentation surfaces. For a larger scale, qualitative assessment, the pulmonary arteries and aorta were segmented on 436 additional CT scans which had no manual annotations of the full volume. Based on visual inspection, we separated the segmentations into three groups: high-quality segmentation; segmentation with minor errors (max 3 *mm* SD); segmentation requiring correction (more than 3 *mm* SD);

4.3.5 Repeatability, Diameters, and PA:AA Ratio

The automatic diameter of the pulmonary artery and the ascending aorta at the level of the pulmonary artery bifurcation as well as the PA:AA ratio, were computed on 436 CT scans. Out of 436 CT scans, the accuracy of the in-slice diameters and the PA:AA ratio were assessed on 200 CT scans by comparing them with the in-slice manual diameters and manual PA:AA ratio.

In the remaining 236 CT scans, the average volumetric diameters and the PA:AA ratio extracted from the 10 *mm* segment around the landmark level were computed. These 118 short-term repeat scan pairs (236 scans) were used for assessing the repeatability of the method. Changes of the main pulmonary artery diameters and the aortic diameters within the three month period are expected to be negligible since the changes in the main pulmonary artery diameter within 8 months is 0.5 ± 0.18 *mm* [112] and the annual change in aortic diameters is $0.1 - 0.2$ *mm* [75].

The diameters were expressed as mean $\pm$ standard deviation (range). The repeatability and the agreement between the manual and automatic diameter and the PA:AA ratio was assessed by the intra-class correlation (ICC) [95% confidence interval], based on a single-rating, absolute-agreement, two-way mixed-effects model [84]. The quality of the vessel segmentation on all 436 CT scans was assessed quantitatively with the visual inspection as explained in Section 4.3.4.

4.4 Results

MCD between the manual and automatic centerlines, for both manual and automatically extracted seed points, are shown in Table 1. The centerlines were always extracted inside the vessel and close to the vessel center. The average distance was

***Table 1: MCD (mm) between automatic and manual centerlines**, in 25 CT scans for the left (LPA), and right pulmonary arteries (RPA) starting from pulmonary trunk (PAT), and for the aorta from both manual and automatically extracted seed points. P-values are computed using non-parametric Mann-Whitney U test.*

	PAT to RPA	**PAT to LPA**	**Aorta**
Manual SP[a]	2.14 ± 0.63	2.38 ± 0.66	1.59 ± 0.51
Automatic SP[a]	2.46 ± 0.81	2.77 ± 1.41	1.54 ± 0.33
P-Value	0.06	0.27	0.35

[a] *SP* : Seed Points

***Table 2: Qualitative assessment of the PA and aorta segmentation** of 419 CT scans with a visual inspection. Segmentations are with no obvious error (high quality), with max 3 mm SD (minor error), or with more than 3 mm SD (correction required).*

Visual Inspection	**Aorta**	**PA**
High Quality	394 (94%)	387 (92%)
Minor Error	18 (4%)	24 (6%)
Correction Required	7 (2%)	8 (2%)

less than 0.5 *mm* between the automatic centerlines traced from the automatically extracted seed points and the ones traced from the manually placed seed points. The non-significant difference in MCD to the manual centerlines between these two sets infers the reliability of the seed point extraction method.

Out of all 471 CT scans used in this study, centerline extraction failed in only 17 cases (17 out of 942 vessels including the pulmonary artery and aorta (1.8%)) and all failure cases were detected automatically as described in Section 4.3.3. In 5 cases, the centerline failure was due to wrong seed point extraction in which the seed points were extracted in the border or outside of the vessel. In 4 cases, the aortic centerline made a shortcut through the pulmonary artery, and in the remaining 8 cases, the pulmonary artery centerline was either at the border of the vessel or made a shortcut through the background. These failures were mainly due to unclear vessel borders where the pulmonary artery is adjacent to the aorta or other structures. Scans with failed centerlines were excluded from subsequent analysis.

The vessel segmentation of the 436 CT scans with no manual annotation, was quantitatively assessed with visual inspection. Table 2 shows the assessed quality of the aorta and pulmonary arteries segmentation of the 419 CT scans with no failure in the centerline extraction.

Box plots of segmentation DSC and MSD, for the pulmonary arteries and aorta

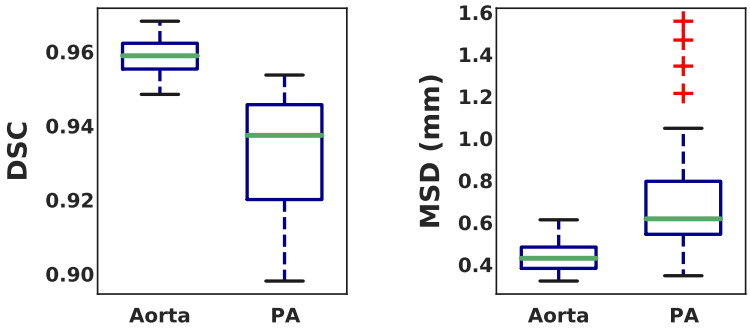


Figure 3: ***Box plots of the DSC and MSD*** *between the manual and automatic PA and aorta segmentation. The plot shows the median(green), interquartile range(boxes),* 99.3% *coverage of the data(whiskers), and the outliers (+ symbol in red).*

obtained in 5-fold cross-validation on 25 CT scans with manual segmentation, are shown in Figure 3, with an average $DSC = 0.94 \pm 0.02$ and average $MSD = 0.62 \pm 0.33$ mm for pulmonary arteries and $DSC = 0.96 \pm 0.01$ and $MSD = 0.43 \pm 0.07$ mm for the aorta. Figure 5 illustrates 3 examples of segmentation results overlaid with manual annotations.

From 200 CT scans used to compare manual and automatic in-slice diameters, 11 scans failed the centerline or seed point extraction. High agreement between the diameters was obtained on the remaining 189 CT scans. The scatter plots of these diameters are illustrated in the first row of Figure 4. The 3 first rows in Table 3 present the diameters, diameter difference, and ICC between the automatic and manual measurements.

From 118 subjects used to assess the repeatability, 6 subjects had failed centerline or seed point extraction. For the remaining 112 subjects (224 CT scans), the automatic average diameter extraction from 3D volume showed a high correlation between scan and rescan of each subject with an ICC of 0.89, 0.95, and 0.86 for the pulmonary artery, aorta, and the PA:AA ratio, respectively. The diameters, diameter difference, and ICC between the scan-rescan pairs of 112 subjects (224 CT scans) is shown in the last 3 rows of Table 3 and the scatter plots are illustrated in the second row of Figure 4.

4.5 Discussion

In this chapter, we presented a fully automatic segmentation and diameter measurement method to segment the pulmonary arteries and aorta and to measure their diameters in non-ECG-gated, non-contrast CT scans. Automatic extraction of the level of the pulmonary artery bifurcation allowed automatic measurement of the PA:AA ratio. We verified the quality of the segmentations on 25 CT scans, by comparing them with full 3D manual annotations. The segmentation algorithm performed well with an average

Table 3: Diameters, diameter difference, and ICC *between manual and automated measurements (first 3 rows) and between the short term scan-rescan pairs (last 3 rows), of the pulmonary artery, ascending aorta, and PA:AA ratio.*

Method		**Diameter (mm)**		**Difference**	**ICC [95% CI[b]]**
		Manual	**Automatic**		
2D Axial n=189	**PA**	26.39 ± 3.28 (18.32, 36.19)	25.79 ± 3.53 (18.06, 36.72)	0.60 ± 1.36	0.92 [0.89-0.94]
	Aorta	34.83 ± 3.63 (25.55, 44.53)	34.39 ± 3.74 (25.21 , 43.81)	0.44 ± 0.96	0.97 [0.95-0.97]
	PA:AA	0.76 ± 0.09 (0.51,1.01)	0.75 ± 0.10 (0.50 , 1.09)	0.01 ± 0.04	0.90 [0.87-0.93]
		Scan (baseline)	**Rescan**		
3D Volume n=112	**PA**	28.24 ± 3.37 (21.58, 38.14)	28.40 ± 3.41 (21.66, 38.48)	0.16 ± 1.45	0.89 [0.87-0.93]
	Aorta	34.72 ± 3.34 (27.65, 44.22)	34.76 ± 3.35 (27.28 , 45.37)	0.04 ± 0.97	0.95 [0.93-0.97]
	PA:AA	0.82 ± 0.10 (0.60, 1.12)	0.82 ± 0.09 (0.61 , 1.14)	0.01 ± 0.05	0.86 [0.82-0.90]

[b] *CI* : Confidence interval

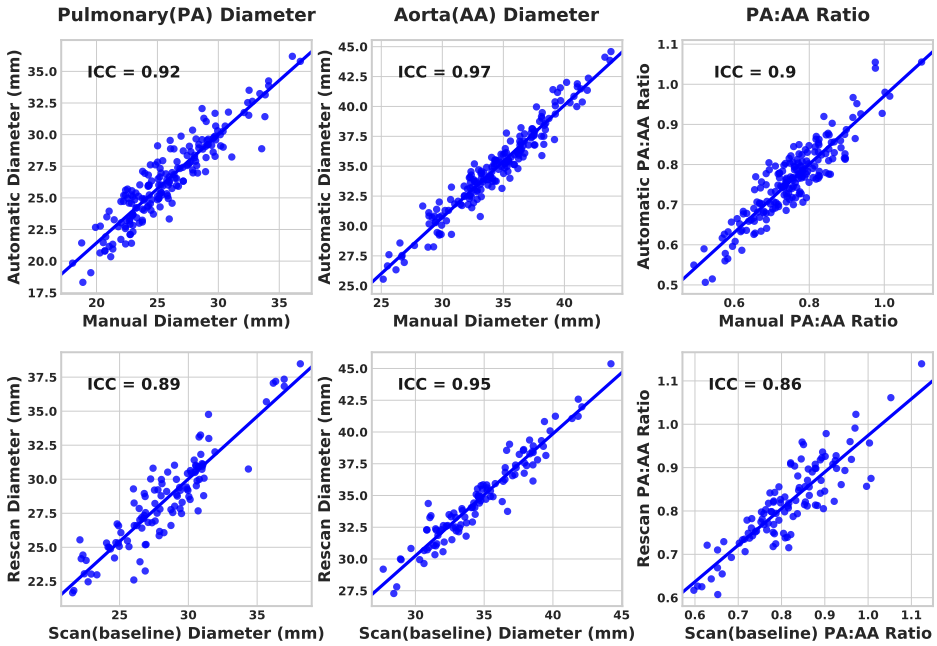


Figure 4: ***Scatter plots and ICC*** *of the manual and automatic diameters measured at the axial slice at the level of PA bifurcation (first row), and of the automated average volumetric diameter estimates in scan and short term rescans (second row) for, from left to right, the PA, aorta, and PA:AA ratio.*

Dice overlap of 0.93 and 0.96 and mean surface distance of 0.62 *mm* and 0.43 *mm*, less than the in-plane voxel size, for the pulmonary arteries and aorta, respectively.

Visual inspection indicated that 92% of pulmonary arteries and 94% of the aorta were segmented with high quality and with no obvious error. In the remaining almost 6-8%, segmentation errors mainly occurred in regions close to the heart. In this region, the pulmonary artery trunk and the aortic root are adjacent and have similar intensity. This makes the vessel boundaries unclear which results in segmentation errors. Furthermore, motion artifacts caused by the motion of the heart during the cardiac cycle increase the ambiguity of the vessel boundaries and make the segmentation difficult, even for experienced radiologists. Figure 5 shows the 3 cases, where the cases at row 2 and 3 have the largest segmentation errors. Segmentation errors with respect to the manual segmentation are visible in an axial slice close to the heart (column b), whereas the segmentation has high accuracy at the level of the pulmonary bifurcation (column d).

Table 4 presents comparative segmentation results on the methods proposed in the literature to segment the pulmonary artery and aorta on non-contrast CT scans. We report higher Dice overlap and lower mean surface distance than previous methods for

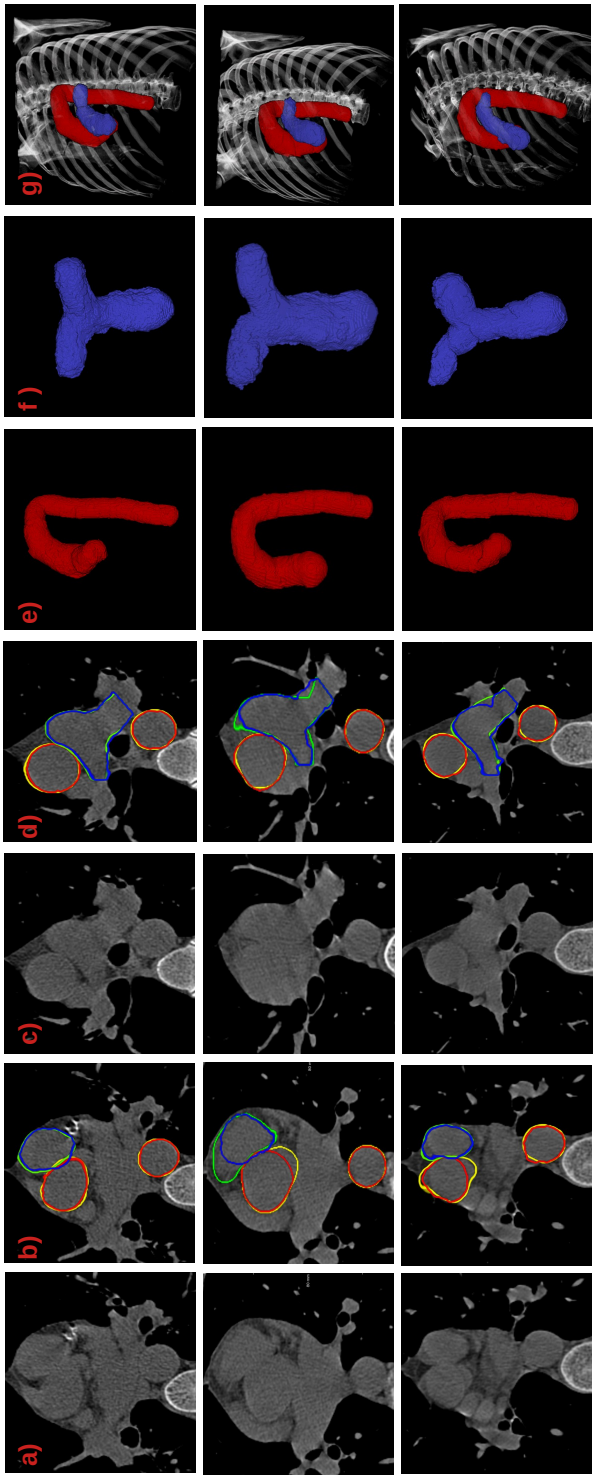


Figure 5: ***Examples of segmentation obtained.*** *First row is a good segmentation with $DSC = 0.95$ for PA and $DSC = 0.97$ for the aorta. The next 2 rows are the two cases with the largest segmentation errors (lowest DSC) for the PA and aorta, respectively, with $DSC = 0.90$, $DSC = 0.96$ and for row two and $DSC = 0.92$, $DSC = 0.95$ for row three. Column (a) shows an axial slice close to the heart and column (c) shows a slice at the pulmonary artery bifurcation level. In columns (b) and (d) the manual PA (green) and aorta (yellow) are overlaid with the automatic segmentation of the PA (blue) and aorta (red). The 3D segmentation of the aorta (e) and PA (f) are shown together with the 3D view of the CT scan in (g).*

Table 4: Segmentation results compared with literature.

	Method	DSC	MSD
PA	Xie et al.[65]	0.88	- -
	Presented Method	**0.94 ± 0.02**	**0.62 ± 0.33**
Aorta	Isgum et al.[53]	0.87 ± 0.03	- -
	Avila-Montes et al.[54]	0.88 ± 0.05	- -
	Dasgupta et al.[55]	0.88 ± 0.06	- -
	Tahoces et al.[56]	0.95	- -
	Xie et al.[59]	0.93 ± 0.01	1.39 ± 0.19
	Kurugol et al.[60]	0.92 ± 0.01	0.62 ± 0.09
	Trullo et al.[62]	0.89 ± 0.04	- -
	Noothout et al.[61]	0.91 ± 0.04	1.32 ± 0.85
	He et al.[63]	0.95	- -
	Sedghi Gamechi et al.[97]	0.95 ± 0.01	0.56 ± 0.08
	Presented Method	**0.96 ± 0.01**	**0.43 ± 0.07**

both arteries. Compared to our work in Chapter 2, in this chapter besides segmenting the pulmonary arteries, we evaluate the method on a larger region for the aorta, from the diaphragm level at the descending aorta to the aortic root.

We achieved a high agreement between the automatic and manual diameters for the pulmonary artery, aorta, and PA:AA ratio in axial slices, with an ICC of 0.92, 0.97, and 0.90 respectively. This is higher than the inter-observer agreement reported by Terzikhan et al. [26] of 0.91 and 0.94 for the pulmonary artery and aorta diameter. Table 3 indicates that the manual diameters are slightly (on average 0.5 *mm*) larger than automatic diameters for both pulmonary artery and aorta. This may be explained by the fact that observers do not necessarily choose the point of maximum intensity gradient as the boundary. The difference between the automated method and observer annotations is similar to the inter-observer bias of 0.4 *mm* reported by Tonelli et al. [112] for pulmonary artery diameter on non-contrast CT scans, and inter-observer bias of 0.5 *mm* reported by Quint et al. [86] for mid-ascending aorta diameter on CTA scans.

The manual PA:AA diameter measurements such as presented in [20, 25, 26, 107] are subjective based on slice location. Measurements that are further away from the bifurcation show smaller diameters than the ones close to the pulmonary bifurcation. Linguraru et al. [49] show that a small shift along the pulmonary artery centerline can lead to diameter changes up to 20%. Moreover, the orientation of the vessels with respect to the patient and with respect to the axial plane may vary, leading to variability in axial diameter measurements. Also, determining the bifurcation level is difficult and is prone to variability. Therefore, the 3D volumetric average diameter measurement in segments perpendicular to the vessel centerlines as proposed in Section 4.2.5 is less subjective and is a more robust and reproducible technique than diameter measurement in 2D axial slices, which potentially decreases discordant

measurements. The diameters extracted by the presented 3D method showed high scan-rescan repeatability (Figure 4).

Considering that lung cancer screening with non-contrast CT is becoming more common, our method can be used to screen for (mild) pulmonary artery and aortic dilatation as biomarkers of cardiovascular disease in the same populations.

A limitation of this study is that the method is evaluated on data from a single scan protocol and a relatively healthy screening population. For application in data from very different scan protocols, parameters may need to be adjusted. However, we have successfully applied the same method, with identical parameter settings, on other data including CTA scans of patients with abnormal aortic shape due to Turner syndrome (see supplementary Figure 4.A.1).

4.6 Conclusion

A fully automatic method is presented to segment the pulmonary arteries and aorta on non-ECG-gated, non-contrast CT scans. Qualitative and quantitative analysis demonstrates that our method provides robust, accurate, and reproducible measurements of the pulmonary artery and aorta diameters and the PA:AA ratio. Automatically extracting a full 3D shape and size of the vessel, the vessel diameters, and biomarkers with high accuracy, in non-ECG-gated non-contrast CT scans such as are acquired in lung cancer screening, can provide important prognostic information and enable the early-stage diagnosis of cardiovascular disease and provide factors for risk assessment in patients with COPD.

Appendix

4.A Supplemental Material

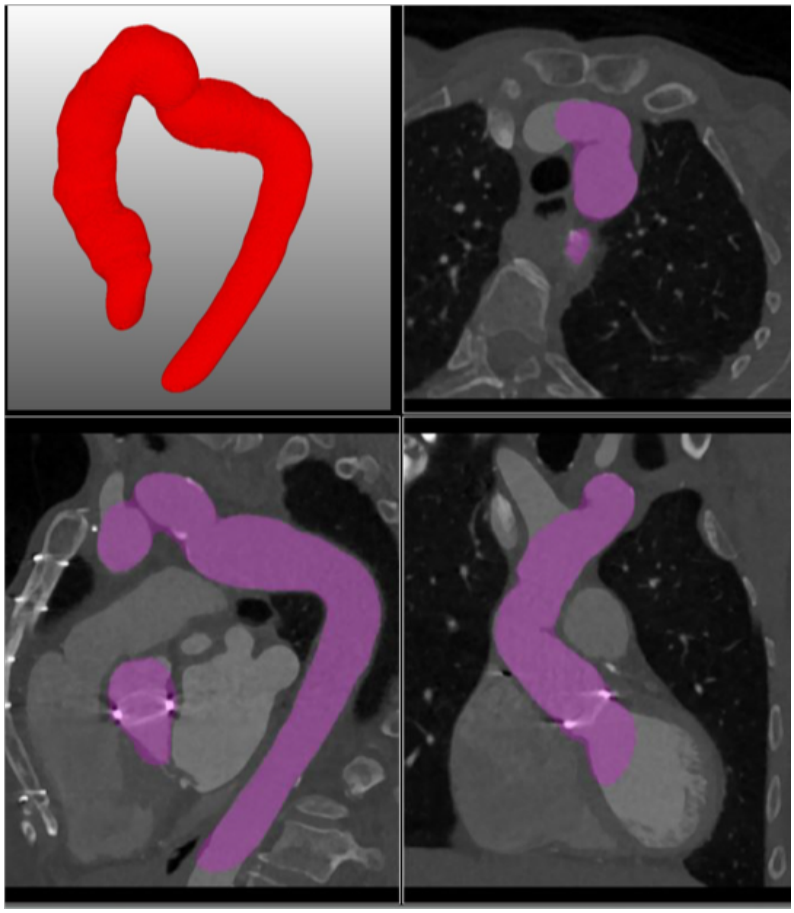


Figure 4.A.1: *Example of the presented method applied on CTA scan with an abnormal aorta shape due to Turner syndrome. The top row illustrates the 3D aorta segmentation and the axial view of the arch. the lower row shows the coronal and sagittal view.*

5

An End-to-end Approach to Segmentation in Medical Images with CNN and Posterior-CRF

Shuai Chen, **Zahra Sedghi Gamechi**, Florian Dubost, Gijs van Tulder, Marleen de Bruijne

Under review

Abstract

Conditional Random Fields (CRFs) are often used to improve the output of an initial segmentation model, such as a convolutional neural network (CNN). Conventional CRF approaches use manually defined features, such as intensity to improve appearance similarity or location to improve spatial coherence. These features work well for some tasks, but can fail for others. For example, in medical image segmentation applications where different anatomical structures can have similar intensity values, an intensity-based CRF may produce incorrect results. As an alternative, we propose *Posterior-CRF*, an end-to-end segmentation method that uses CNN-learned features in a CRF and optimizes the CRF and CNN parameters concurrently. We validate our method on three medical image segmentation tasks: aorta and pulmonary artery segmentation in non-contrast CT, white matter hyperintensities segmentation in multi-modal MRI, and ischemic stroke lesion segmentation in multi-modal MRI. We compare this with the state-of-the-art CNN-CRF methods. In all applications, our proposed method outperforms the existing methods in terms of Dice coefficient, average volume difference, and lesion-wise F1 score.

5.1 Introduction

After the breakthrough of deep learning in computer vision [113–115], deep convolutional neural networks (CNNs) and their variants [116–118] quickly started to dominate medical image segmentation, outperforming traditional machine learning methods in many applications [119–122]. To refine the prediction from the CNN, it is common to combine CNN with a conditional random field (CRF) [123]. By modeling pairwise relationships and interactions between voxel-wise variables over the whole image, the CRF can improve the coherence of the segmentation. In previous work, CRFs based on predefined features such as intensity similarity and spatial coherence have been used as an efficient post-processing technique or trained in an end-to-end manner in a recurrent neural network to refine the CNN outputs [71, 118, 124, 125].

Most often, a CRF uses a combination of voxel intensity and voxel location as pairwise potentials. Although this works well in several computer vision applications [125, 126], there can be challenges in other applications. The approach assumes that voxels that have similar intensity and are close to each other in the image are likely to belong to the same class. There are many applications among others in medical image analysis in which this assumption does not hold. For example, the intensity-based features of the CRF are not sufficient for problems where the intensity is not informative enough to identify object boundaries, such as the artery segmentation problem in Figure 2.a. The spatial component of the CRF, on the other hand, requires extra careful tuning when the CRF is applied to data with isolated small objects, such as the white matter hyperintensities in Figure 2.b, which may be erroneously removed by excessive smoothing. In stroke lesion segmentation, a large appearance difference between lesion objects of the same class also goes against the CRF assumption that the same class objects should have similar intensity (see Figure 2.c).

In this chapter, we propose *Posterior-CRF*, a new learning-based CRF approach for image segmentation that allows the CRF to use features learned by a CNN, optimizing the CRF and CNN parameters concurrently. The learning-based CRF makes the CNN features update to work best with CRF in an end-to-end manner. During training, the CRF inference works in the CNN feature space, which is likely to be more informative for segmentation than the original intensity values of the image.

We demonstrate our method in three medical image analysis applications. Our first application is the segmentation of the aorta and pulmonary artery in non-ECG-gated, non-contrast chest CT scans. In these images, the aorta and the pulmonary artery share similar intensity values, which goes against the CRF assumption that similar classes should share similar intensity [59, 108]. The boundaries between the objects are not recognizable by intensity alone, making a standard CRF less effective (Figure 2.a). Our second application is the segmentation of white-matter hyperintensities in brain MRI. These small objects are sparsely distributed in the brain (see Figure 2.b) and may be removed by the CRF, which optimizes for the spatial coherence of segmentation. Our third application is the segmentation of ischemic stroke lesions in brain MRI, which have very heterogeneous intensities and shapes within the same lesion class (Figure 2.c).

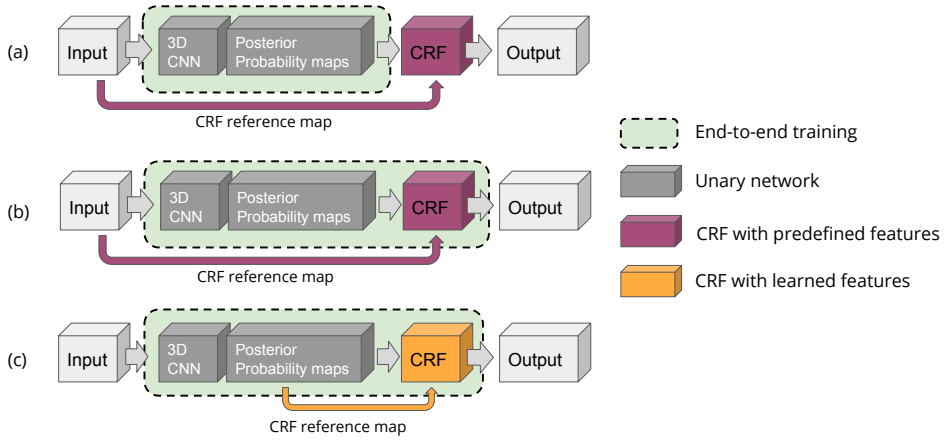


Figure 1: Different CRF-based approaches *For each graph: (a) Post-processing CRF [118, 124]; (b) End-to-end training CRF with predefined features [125]; (c) Proposed Posterior-CRF, which uses CNN feature maps as CRF reference maps.*

Contributions

1. We present a new end-to-end trainable algorithm for image segmentation called *Posterior-CRF* using learnable features in CRF pairwise potentials. We explore how the proposed method affects CNN learning during training.

2. We compare the performance of a fully-connected CRF in several settings: post-processing, end-to-end training with predefined features, and end-to-end training with learned features. Ablation experiments are conducted to investigate the influence of CRF parameters and which level of the CNN feature maps are more informative for the CRF inference. We found that the features in the last CNN feature maps provide a more consistent improvement than features in early CNN layers and predefined intensity features.

3. We evaluate our methods in three applications: aorta and pulmonary artery segmentation in non-contrast CT, which can be used to compute important biomarkers such as the pulmonary artery to aorta diameter ratio [108]; white matter hyperintensities segmentation in multi-sequence MRI, which is of key importance in many neurological research studies [121]; and ischemic stroke lesion segmentation in multi-sequence MRI, which can provide biomarkers for stroke diagnosis [122]. In the experiments, the proposed Posterior-CRF outperforms CNN without CRF, post-processing CRF, end-to-end intensity-based CRF, and end-to-end spatial-based CRF.

A preliminary version of this work, focused on a single application and with less validation, appeared as an extended abstract in [127].

Figure 2: Difficult cases for conventional CRF inference in medical image segmentation. *(a) Segmentation of arteries in CT: first row shows two axial slices of the non-contrast CT scan with red arrows indicating indistinguishable boundaries; second row shows the corresponding ground truth of the aorta (yellow) and pulmonary artery (green); (b) White matter hyperintensities segmentation in MRI: four examples are shown with the ground truth of the lesions (green), red arrows indicate small isolated lesions that can be easily removed by CRF; (c) Ischemic stroke lesions segmentation in MRI: first row shows the ground truth of the lesions (green) where large appearance difference between lesions can be observed (red arrows); second row shows a close-up view of the lesions. Best viewed in color with zoom.*

5.2 Related Work

End-to-end Training of CRF and CNN. CRF is widely used as an efficient post-processing method to refine the output of CNN segmentation models (for example, [71, 118, 124]). However, applying a CRF as post-processing means that the CNN is not able to adapt its output to the CRF. Zheng et al. [125] proposed to optimize CNN and CRF jointly by reformulating the CRF inference as a recurrent neural network (RNN) operation, such that the CRF weights can be learned together with the CNN. This approach makes the unary potentials and the kernel weights in pairwise potentials trainable, which saves the computational cost of grid search for other approaches to tune these weights, although the CRF still works in the predefined fixed feature space. In this chapter, we focus on a new CRF approach where the CRF inference works in a learning-based CNN feature space.

Locally-connected CRFs with Learned Potentials. While conventional CRFs use predefined Gaussian edge potentials, the potentials can also be learned through a neural network. Vemulapalli et al. [128] learn the pairwise potentials of a Gaussian CRF in a bipartite graph structure. This approach uses a simpler continuous CRF model which provides better convergence of mean-field inference than the conventional discrete CRF models. In this chapter, we focus on the most widely used discrete CRF model which is a natural fit for the dense segmentation problem. Lin et al. [129] and Li et al. [130] learn pairwise CRF potentials to model patch-wise relationships. The patch-wise potentials provide a better ability to model the semantic compatibility between image regions and have different effects compared to our approach, where we do not consider patch-wise relationships. Our method uses traditional Gaussian edge potentials [123] similar to Zheng et al. [125] which are easier to compute in a fully-connected manner. Unlike Zheng et al., we derive the potentials from the feature space learned by a CNN. This allows us to model global interactions between voxel-wise variables using learning-based features.

Other Methods Related to CRF. Next to CRF, there are several other approaches that aim to model interactive relationships or add global information to neural networks. Graph neural networks (GNN) [131, 132] model interactions between variables by applying graph convolution filters, which allow them to learn global relationships between voxels. We further address GNN in the Discussion. The recently proposed non-local CNN [133] uses layer-wise self-attention [134–136] to make each layer in the network focus on the areas that encoded the most non-local information in the preceding layer. While this allows non-local CNNs to model long-range dependencies, they are unable to model the interactions that can be learned by a CRF or GNN. In this chapter, we focus on the fully-connected CRF model which is an efficient approach of modeling both interactive relationships and global information.

5.3 Methods

Our method consists of two parts that are optimized jointly: 3D CNN and 3D CRF. In Section 5.3.1, we describe the CNN model, which provides unary potentials for

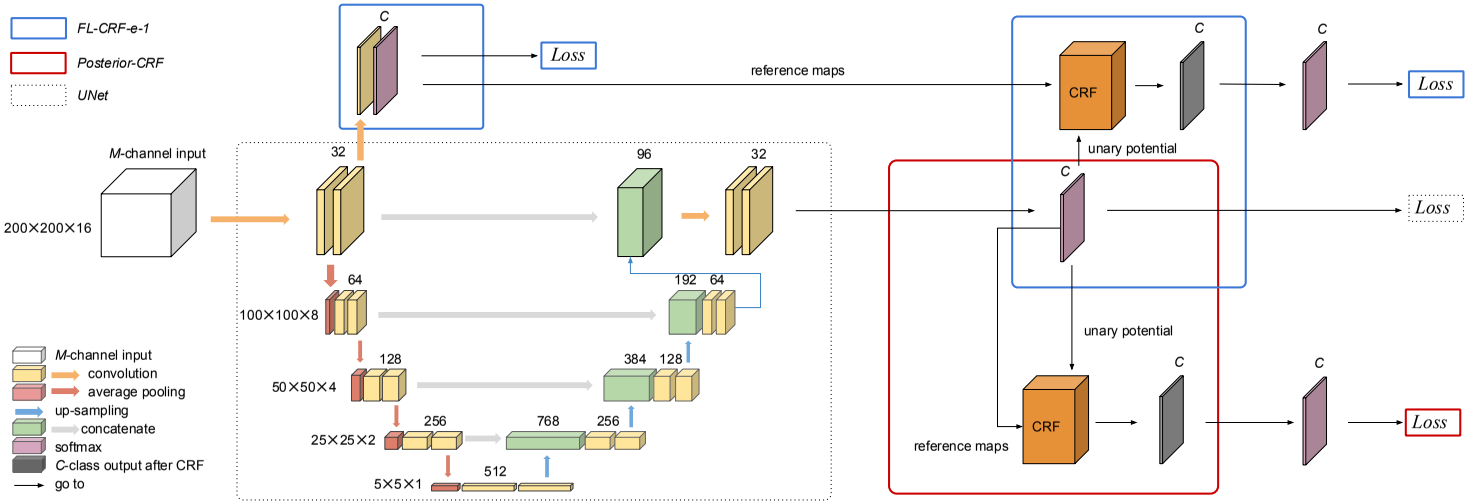


Figure 3: Proposed feature-learning-based CRF using early/later CNN feature maps. *The backbone architecture is based on 3D UNet. The skip-connections concatenate the feature maps from the encoder path with the upsampled ones from the decoder path. The CRF module is placed on top of the CNN and infers the most likely posterior class probability conditioned on the CRF features. M is the number of input imaging modalities. C is the number of output classes. Two proposed CRF variants are shown in this figure: 1. Posterior-CRF (red rectangle), which uses the last CNN layer as CRF reference maps; 2. FL-CRF-e-1 (blue rectangles), which uses the first level CNN layer as CRF reference maps. Best viewed in color with zoom.*

the CRF inference as well as features for the pairwise potentials for the proposed Posterior-CRF. Then we introduce the CRF in Section 5.3.2. We show two previously proposed ways to perform CRF inference using predefined features: post-processing (Section 5.3.3.1) and end-to-end training with predefined features (Section 5.3.3.2). Our proposed end-to-end training with learned features is presented in Section 5.3.4, followed by Section 5.3.4.1 about the back-propagation of the proposed learning-based CRF. The mean-field inference algorithm used in the proposed method is explained in Appendix in Section 5.A.

5.3.1 CNN Model

Our CNN model is based on UNet [116], the most widely used network architecture for medical image segmentation. It has a multi-scale design with skip-connections that connect the encoding and decoding parts of the network, which allow the decoding path to use the early, high resolution feature maps without losing information through pooling. We use 3D UNet as the basic CNN architecture to provide the unary potentials for CRF inference as well as features for the pairwise potentials for the proposed Posterior-CRF. Details of the network layout used in our experiments are given in Figure 3.

5.3.2 Conditional Random Fields

In this section, we describe the CRF as proposed in [123]. In image segmentation, a CRF models voxel-wise variable x_i taking values in $\{1, ..., C\}$ as a set of random variables $\mathcal{X} = \{x_1, ..., x_N\}$, where C is the number of classes and N is the number of voxels in the image. During training, x_i is converted into a soft classification vector of length C, indicating for each class the probability that the ith voxel belongs to that class, with the L_1 norm $|x| = 1$. x_i obey a Markov property conditioned on a global observation, the image $\mathbf{I}$ consisting of variables $\mathcal{I} = \{I_1, ..., I_N\}$. In this chapter, $\mathbf{I}$ is the observed 3D CT/MRI scans, with its length given by the number of imaging modality channels M times the number of voxels per channel N.

Consider a fully-connected pairwise CRF model ($\mathbf{X}$,$\mathbf{I}$) characterized by a prior Gibbs distribution:

$$P(\mathbf{X}|\mathbf{I}) = \frac{1}{Z(\mathbf{I})}\exp(-\sum_{c\in\mathcal{C}_\zeta}\phi_c(\mathbf{X}_c|\mathbf{I})) \tag{5.1}$$

where $\zeta = (\mathcal{V}, \mathcal{E})$ is an undirected graph describing the random field $\mathbf{X}$. Each clique c in a complete set of unary and pairwise cliques $\mathcal{C}_\zeta$ in ζ, and ϕ is the potential for each clique. We seek a maximum a posteriori probability (MAP) estimation $\mathbf{x}$ that minimizes the corresponding Gibbs energy $E(\mathbf{X} = \mathbf{x}|\mathbf{I})$:

$$E(\mathbf{X} = \mathbf{x}|\mathbf{I}) = \sum_i \varphi_u(x_i|\mathbf{I}) + \sum_{i<j}\varphi_p(x_i, x_j|\mathbf{I}) \tag{5.2}$$

$$MAP(P(\mathbf{X}|\mathbf{I})) : \mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} E(\mathbf{X} = \mathbf{x}|\mathbf{I}) \tag{5.3}$$

where i and j range from 1 to N. The first term $\varphi_u(x_i)$ in Equation 5.2 is the unary potential, which in our case is the current C length vector of voxel i representing the

class probabilities in the CNN posterior probability maps. The second term $\varphi_p(x_i, x_j)$ is the pairwise potential:

$$\varphi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{K} \omega_m k_m \tag{5.4}$$

where $\mu(x_i, x_j)$ is the label compatibility function that describes the interactive influences between different pairs of classes, ω_m is the linear combination weight of different pre-defined kernels k_m and K is the total number of kernels. Each k_m is a modified Gaussian kernel with specific feature vector $\mathbf{f}$:

$$k(\mathbf{f}_i, \mathbf{f}_j) = \prod_{s=1}^{S} \mathbf{exp}(-\frac{1}{2}(f_i^s - f_j^s)^{\mathrm{T}} \mathbf{\Lambda}^s (f_i^s - f_j^s)) \tag{5.5}$$

The feature vector $\mathbf{f}$ is defined from S arbitrary feature spaces. $\mathbf{\Lambda}$ is a symmetric positive-definite precision matrix that defines the shape of each kernel. In semantic segmentation, typically a combination of intensity (I) and position features (p) has been used [118, 123, 125]:

$$\begin{aligned} \varphi_p(x_i, x_j) =& \mu(x_i, x_j)[\omega_1 \mathbf{exp}(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}) \\ &+ \omega_2 \mathbf{exp}(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2})] \end{aligned} \tag{5.6}$$

where the first kernel controlled by ω_1 is called *appearance kernel* and the second kernel controlled by ω_2 is called *smoothness kernel*. The parameters θ_α, θ_β and θ_γ control the influence of the corresponding feature spaces. The appearance kernel is inspired by the observation that nearby voxels with similar intensity are likely to be in the same class, while voxels that are either further away or have larger intensity difference are less likely to be in the same class. The smoothness kernel can remove isolated regions and produce smooth segmentation results [118, 123]. Note that the position feature appears in both appearance kernel and smoothness kernel, where spatial information has different contributions to each of the two kernels, depending on the spatial standard deviations θ_α and θ_γ.

5.3.3 CRF with Predefined Features

Conventional CRFs use predefined features, such as the image intensity and spatial position shown in Equation 5.6. These features are commonly used in CRFs to encourage intensity and spatial coherence, based on the assumption that voxels that have a similar intensity or are close together are likely to belong to the same class.

We evaluate two state-of-the-art approaches to combine CRFs with predefined features with a CNN:

1. Apply the CRF as post-processing to refine the CNN outputs (Section 5.3.3.1);
2. Implement the CRF as a neural network layer that can be trained together with the CNN in an end-to-end manner (Section 5.3.3.2).

5.3.3.1 CRF as Post-processing

After we train a CNN model and get its predictions, we can apply CRF as a post-processing method to refine the results [124]. We refer to this method as *Postproc-CRF* (Figure 1.a).

5.3.3.2 End-to-end Training CRF

The CNN and CRF can be combined more elegantly by optimizing them together in an end-to-end manner [125] (Figure 1.b), which allows the CRF to influence the CNN optimization. The end-to-end CRF uses the same pairwise potentials as that in the post-processing CRF (Equation 5.6). We refer to this variant as *Intensity-CRF*.

To investigate the spatial term in the end-to-end CRF, we can also use only the position features as the CRF feature space, which means that the CRF layer will only encourage nearby voxels to have the same class. We implement this CRF by setting the weight of the appearance kernel ω_1 to zero and make it not trainable. We refer to this method as *Spatial-CRF*.

5.3.4 Proposed CRF with Learning-based Features

Our proposed CRF uses a learning-based feature space. We replace the intensity feature vector I in the CRF kernel (Equation 5.6) with the new feature vector $F(\mathbf{I})$ from the CNN feature maps. The information in these CNN feature maps differs per level: in the first level of UNet the feature maps contain information close to the intensity, while in the last level of the UNet they contain more context for each voxel and potentially more class-discriminative information.

We refer to the CRF that uses features learned by CNN as *feature-learning-based CRF* (see Figure 1.c) and refer to the specific form of CRF using the features in the last CNN softmax layer as *Posterior-CRF* (see Figure 3).

Unlike the CRFs with predefined features, our CRF takes CNN feature maps as the reference maps and updates the random field $\mathbf{X}$ based on $F(\mathbf{I})$ instead of on $\mathbf{I}$ directly. Compared to the original CRF pairwise potential in Equation 5.6, the feature I is replaced with $F(\mathbf{I})$ and the new pairwise potential becomes:

$$\begin{aligned}\varphi_p(x_i, x_j) =& \mu(x_i, x_j)[\omega_1 \mathbf{exp}(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \\ & \frac{|F_i(\mathbf{I}) - F_j(\mathbf{I})|^2}{2\theta_\beta^2}) + \omega_2 \mathbf{exp}(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2})]\end{aligned} \tag{5.7}$$

5.3.4.1 Back-propagation of the Learning-based CRF

The back-propagation of the proposed end-to-end feature-learning-based CRF is shown in Figure 4. There are five steps within one optimization iteration. Steps 1∼3 are the forward process that generates the output of the CNN. In the 4^{th} step, CRF weights will adapt to the outputs calculated by the reference maps and unary maps, both given by CNN feature maps before back-propagation. In the 5^{th} step, CNN weights are

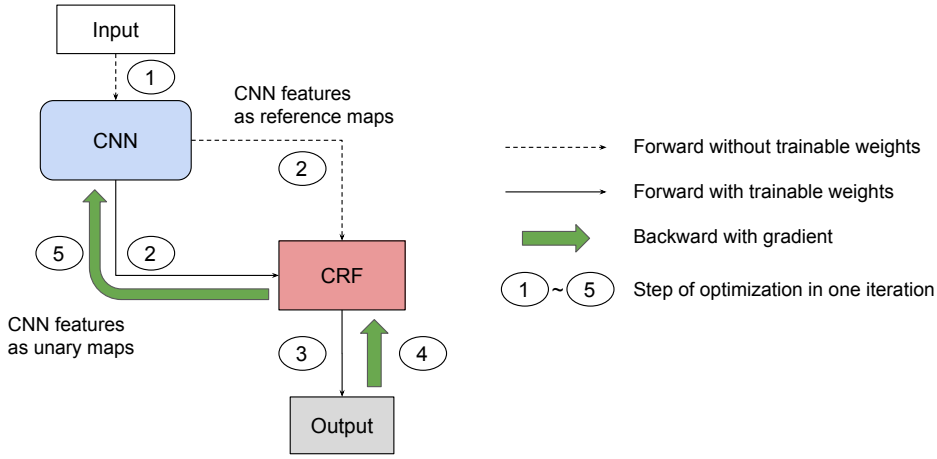


Figure 4: One end-to-end optimization iteration of the proposed CRF method. *Best viewed in color with zoom.*

updated to provide new unary maps and reference maps for CRF for the next iteration. When the optimization converges, both CNN and CRF weights become stable close to their optimal values. Note that the mean-field inference in CRF happens in the forward process (after step 2 and before step 3) and thus contributes to the gradient updates of both CNN and CRF weights. The derivation of the mean-field inference gradient is omitted due to the length of this chapter and can be found in Section 4.2 of the paper by Zheng et al. [125].

5.4 Experiments

In this section, we present experiments to evaluate the proposed method and compare it to the baseline methods: 3D UNet, Post-processing CRF, Intensity-CRF, and Spatial-CRF. Implementation details are discussed in Section 5.4.1, followed by the experimental settings (Section 5.4.2), the description of the datasets and pre-processing (Section 5.4.3), data augmentation and training details (Section 5.4.4) and evaluation metrics (Section 5.4.5).

5.4.1 Implementation

5.4.1.1 CNN Implementation

We implement all the algorithms in the TensorFlow framework. The detailed CNN architecture for the experiments is shown in Figure 3. All convolution layers use ReLU as the activation function except for the last output layer, which uses softmax to produce the final probability maps. For a fair comparison, all the methods in Table 3 use the same CNN architecture and hyperparameters.

We use the Dice coefficient as our segmentation loss function during training:

$$\mathcal{L}(q) = 1 - \frac{2|\sum_{i=1}^{N}\sum_{j=1}^{C} p_{i,j}q_{i,j}|}{\sum_{i=1}^{N}\sum_{j=1}^{C} p_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{C} q_{i,j}} \tag{5.8}$$

where $q_{i,j}$ is the predicted probability that voxel i belongs to the jth class. $p_{i,j}$ is the true label. The loss is minimized using the Adam optimizer [137].

5.4.1.2 CRF Implementation

In CRF, mean-field approximation can be used to calculate the maximum a posteriori probability (MAP) of the inference. We use an efficient approximation algorithm for mean-field inference [123, 138] built on a fast high-dimensional filtering using the permutohedral lattice [139] that allows voxel-wise fully-connected CRF to be iteratively computed in linear time. For a fair comparison, all the CRF methods in this chapter are implemented in 3D fully-connected manner.

5.4.2 CRF Settings

5.4.2.1 Post-processing CRF

For *Postproc-CRF*, we fix the label compatibility μ in Equation 5.6 to the identity matrix, which means that the CRF does not model label-specific interaction. In the case of multi-modal input, each imaging modality has a specific θ_β to control the strength of the intensity term.

5.4.2.2 End-to-end CRF with Predefined Features

We consider two forms of end-to-end CRFs with predefined features: *Intensity-CRF* uses intensity of the input image $\mathbf{I}$ and position information as its feature space. *Spatial-CRF* uses only the position information (the smoothness term in Equation 5.6). The label compatibility is a $C \times C$ parameter matrix which is optimized during training to allow the CRF to learn the label compatibility automatically. The weights ω_1 of the appearance kernel for *Intensity-CRF* and ω_2 of the spatial kernel for *Spatial-CRF* are $C \times C$ matrices, which we restrict to diagonal matrices because the relationship between classes is already covered by the label compatibility matrix. Inner product is calculated by multiplying the matrices. For simplicity, only one θ_β is applied for all modalities.

5.4.2.3 End-to-end CRF with Learned Features

The proposed *Posterior-CRF* uses the last softmax layer of the CNN as its reference map. The hyperparameters are the same as end-to-end CRF with predefined features. Note that Posterior-CRF is a special case of the feature-learning-based CRF. We can also use early CNN feature maps as CRF reference maps. An ablation study investigating other CRF variants can be seen in Section 5.5.4.

Table 1: Post-processing CRF parameters for each dataset. *Search range indicates the range of parameter values explored during grid search.*

Datasets	***CT Arteries***	***WMH***	***ISLES***	**Search range**
ω_1	6.39	3.85	9.75	(0.1, 10)
θ_α	4.09	4.46	8.74	(0.1, 10)
θ_β **for CT**	1.10	-	-	(0.1, 10)
θ_β **for T1**	-	7.01	9.26	(0.1, 10)
θ_β **for T2**	-	-	9.73	(0.1, 10)
θ_β **for FLAIR**	-	2.64	2.36	(0.1, 10)
θ_β **for DWI**	-	-	6.85	(0.1, 10)
ω_2	3.40	1.41	2.34	(0.1, 10)
θ_γ	4.83	0.11	1.35	(0.1, 10)
Iterations	3	1	2	(1, 5)

Table 2: Initial end-to-end CRF parameters for each dataset.

Methods	ω_1	θ_α	θ_β	ω_2	θ_γ	**Iterations**
			CT Arteries			
Spatial-CRF	-	-	-	3.40	4.83	3
Others	6.39	4.09	1.10	3.40	4.83	3
			WMH			
Spatial-CRF	-	-	-	1.41	0.11	1
Others	3.85	4.46	4.83	1.41	0.11	1
			ISLES			
Spatial-CRF	-	-	-	2.34	1.35	2
Others	9.75	8.74	7.05	2.34	1.35	2

5.4.2.4 CRF Parameters

Parameters in the post-processing CRF for each dataset were obtained by grid search on the validation set and are shown in Table 1. We computed results with 500 different configurations of Postproc-CRF on each dataset for grid-search. Parameters in the end-to-end CRFs (*Intensity-CRF*, *Spatial-CRF*, *Posterior-CRF*) are initialized with the same values as were used in post-processing CRF. Although the end-to-end CRF approaches have the ability to learn CRF weights automatically during training, we initialize all CRF approaches in the same way to facilitate visualization of the evolution of CRF parameters during training (see Figure 5). We study the sensitivity to different

CRF parameter initializations in Section 5.5.3.

The initial label compatibility matrix is set to an identity matrix and can be optimized during training. In the multi-modality case, the initial value of θ_β is averaged over all modalities. The initial values for each dataset are shown in Table 2.

5.4.3 Datasets and Preprocessing

We evaluate the proposed method on three segmentation problems: CT arteries, MRI white matter hyperintensities, and MRI ischemic stroke lesions. We chose these problems to study the generalizability of the method as these applications differ a lot in object shapes and appearances, imaging modalities, and suffer from different problems (see Figure 2).

5.4.3.1 CT Arteries Dataset

We use 25 non-contrast lung CT scans from 25 different subjects enrolled in the Danish Lung Cancer Screening Trial (DLCST) [6]. The aorta and pulmonary artery were manually segmented by a trained observer (ZSG). Images have an anisotropic voxel resolution of 0.78mm $\times$ 0.78mm $\times$ 1.00mm and are of size 512×512 with on average 336 slices (range 271-394). The 25 scans are split into three parts of 10, 5, and 10 scans for training, validation, and testing respectively. Due to the limitation of GPU memory, we first crop the original CT images and only keep the axial central part of 256×256 voxels for all slices. Then, 3D patches of the size $256 \times 256 \times 16$ are extracted from the cropped images. All training patches have 80% overlap in z-axis between neighboring patches to mitigate border effects. In total, there are 840 3D patches for training. We use the original CT intensities without normalization.

5.4.3.2 MRI White Matter Hyperintensities (WMH) Dataset

The White Matter Hyperintensities (WMH) Segmentation Challenge [121] provided images from 60 subjects (T1 and FLAIR) acquired from three hospitals and manually segmented for background and white matter hyperintensities. We randomly split these in 36 subjects for training, 12 for validation, and 12 for testing. For each subject, we cropped/padded MRI images into a constant size $200 \times 200 \times Z$, where Z is the number of slices in the image. We use Gaussian normalization to normalize the intensities inside the brain mask in each image to zero mean and unit standard deviation. We extract training patches of size $200 \times 200 \times 16$ with 80% overlap in z-axis between patches. In total, there are 528 3D patches for training.

5.4.3.3 MRI Ischemic Stroke Lesions (ISLES) Dataset

The ISLES 2015 Challenge [140] is a public dataset of diverse ischemic stroke cases. There are 4 MRI sequences available for each patient (T1, T2, FLAIR, and DWI). We use the sub-acute ischemic stroke lesion segmentation (SISS) dataset (28 subjects) with the lesion labels for experiments and randomly split them as 14 for training, 7 for validation and 7 for testing. The images are cropped/padded to the size $200 \times 200 \times Z$. Gaussian normalization is applied for normalizing the intensities in

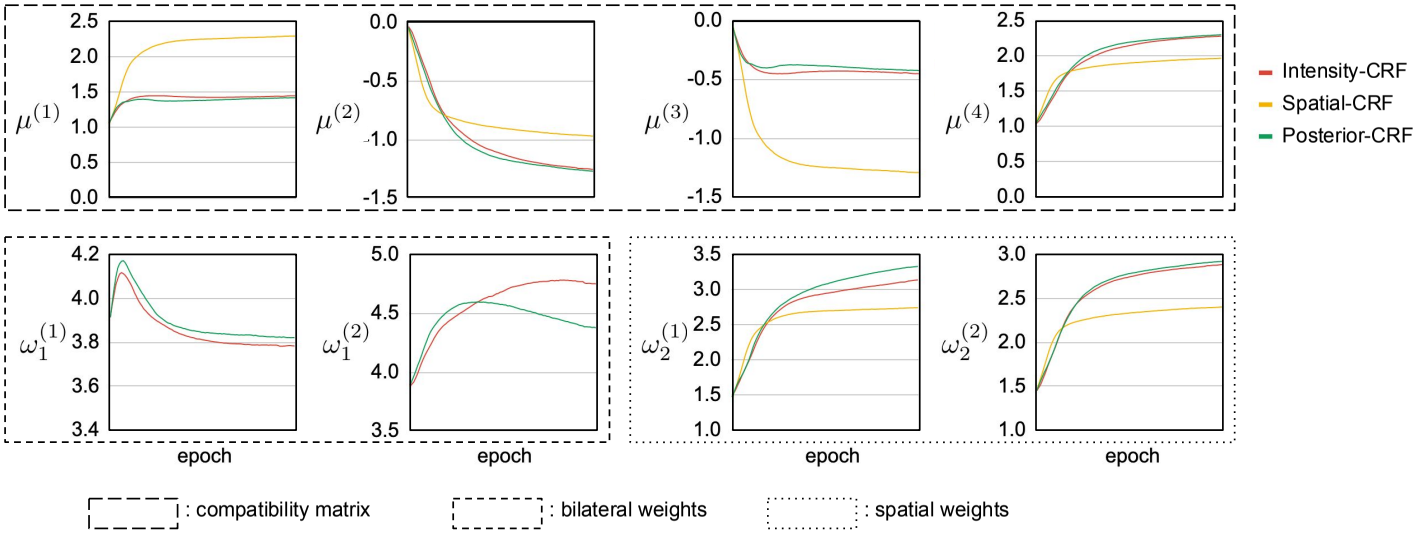


Figure 5: CRF parameters during training in WMH dataset. *The initial values of the CRF parameters can be found in Table 2. Best viewed in color with zoom.*

each image. Training patches of the size $200 \times 200 \times 16$ with 80% overlap in z-axis are extracted. In total, there are 560 3D patches for training.

5.4.4 Data Augmentation and Training Details

The network is trained on all mini-batches (each mini-batch contains one 3D patch). For each 3D patch in the current mini-batch we apply 3D random rotation sampled from ([-5,5],[-5,5],[-10,10]) degrees, shifting ([-24,24],[-24,24],[-7,7]) voxels, as well as random horizontal (left and right) flipping. We stopped training when the validation loss is not decreasing anymore and chose the model that achieved the best validation performance. The experiments are run on an Nvidia GeForce GTX1080 GPU. The average training time is 5∼10 hours for one CNN baseline model and 1∼2 hours more when the CRF layer is added.

5.4.5 Evaluation Metrics

We use four voxel-wise metrics of segmentation quality: Dice similarity coefficient (DSC), indicating the relative overlap with the ground truth (larger is better); 95^{th} percentile Hausdorff distance (H95), showing the extremes in contour distance from ground truth to the prediction (smaller is better); Average volume difference (AVD) as a percentage of the difference between ground truth volume and segmentation volume over ground truth volume (smaller is better), and Recall score (larger is better). For the lesion segmentations (WMH and ISLES), we additionally assess accuracy of lesion detection by computing the lesion-wise Recall and lesion-wise F1 score (larger is better). The lesion-wise metrics use the 3D connected components, while the voxel-wise metrics do not use 3D connected components. The correct detection of a lesion is determined by the overlap (at least one voxel) of the 3D components. F1 score is equivalent to lesion-wise Dice score and is calculated by 2*(precision*recall)/(precision+recall), where precision is calculated by true positives/(true positives+false positives).

5.5 Results

5.5.1 Segmentation Results

Table 3 shows the segmentation results for all three datasets. In most metrics, Posterior-CRF had the best performance in all datasets. For all datasets, CNN without CRF provides good baseline results, which indicates that 3D UNet is an efficient architecture to extract useful features for segmentation in these applications. Intensity-CRF performed worse on DSC than Posterior-CRF (statistically significant in aorta segmentation and WMH segmentation), which reveals the limitation of intensity features. Among all end-to-end CRF methods, Spatial-CRF performs worst for all datasets except ISLES. From these results, we conclude that spatial coherence alone is not sufficient and often detrimental to segmentation accuracy, and that the CNN features in the last layer are more informative for CRF than the intensity features in the original images.

Table 3: Results. *Mean (standard deviation). The best results are marked in bold. Each experiment is repeated 5 times with different random data split. The last two colomns are lesion-wise metrics. P-values are calculated by two-sided paired t-test. All CRF methods are implemented in 3D fully-connected manner and share the same CNN architecture and hyperparameters.*

Methods	DSC	H95 (mm)	AVD (%)	Recall	Recall (lesion)	F1 (lesion)
CT Arteries: Aorta						
CNN baseline	0.9291(0.02)$^\diamond$	5.5560(1.96)$^\diamond$	6.8780(4.17)$^\diamond$	0.8993(0.03)$^\diamond$	N/A	N/A
Postproc-CRF[118]	0.9264(0.02)$^\diamond$	5.1591(1.59)$^\diamond$	8.5326(4.81)$^\diamond$	0.8878(0.04)$^\diamond$	N/A	N/A
Intensity-CRF[125]	0.9457(0.01)*$^\diamond$	3.2802(0.77)*$^\diamond$	3.1967(2.58)	0.9548(0.02)*	N/A	N/A
Spatial-CRF	0.9188(0.02)$^\diamond$	7.6562(3.98)$^\diamond$	6.1013(5.13)$^\diamond$	0.8939(0.05)$^\diamond$	N/A	N/A
Posterior-CRF	**0.9538**(0.01)*	**2.8699**(0.86)*	**2.3688**(2.29)*	**0.9555**(0.02)*	N/A	N/A
CT Arteries: Pulmonary Artery						
CNN baseline	0.8510(0.05)$^\diamond$	10.3000(4.87)$^\diamond$	16.7687(12.60)$^\diamond$	0.8867(0.09)	N/A	N/A
Postproc-CRF[125]	0.8561(0.05)	10.0052(5.22)$^\diamond$	13.7071(10.26)$^\diamond$	0.8698(0.09)$^\diamond$	N/A	N/A
Intensity-CRF[125]	0.8773(0.04)*	8.9208(3.09)*	11.8671(8.66)*	**0.9079**(0.06)	N/A	N/A
Spatial-CRF	0.8558(0.06)$^\diamond$	10.5672(5.19)$^\diamond$	13.7399(13.47)	0.8603(0.09)$^\diamond$	N/A	N/A
Posterior-CRF	**0.8935**(0.04)*	**7.6635**(3.92)*	**8.9245**(7.07)*	0.8979(0.07)	N/A	N/A
WMH						
CNN baseline	0.7557(0.13)$^\diamond$	6.5015(9.87)$^\diamond$	28.3351(45.64)$^\diamond$	**0.7977**(0.14)	0.6476(0.14)	0.6648(0.11)$^\diamond$
Postproc-CRF[125]	0.6970(0.17)$^\diamond$	8.8659(7.79)$^\diamond$	35.0786(22.69)$^\diamond$	0.5947(0.20)$^\diamond$	0.3476(0.16)$^\diamond$	0.4831(0.16)$^\diamond$
Intensity-CRF[125]	0.7706(0.10)$^\diamond$	4.9403(4.58)	15.6263(16.44)*	0.7751(0.12)	0.6803(0.15)*	0.6705(0.10)$^\diamond$
Spatial-CRF	0.7602(0.11)$^\diamond$	5.8469(5.82)$^\diamond$	23.5154(25.76)$^\diamond$	0.7831(0.13)	**0.6876**(0.14)*	0.6569(0.11)$^\diamond$
Posterior-CRF	**0.7887**(0.09)*	**4.2972**(3.87)*	**14.8427**(12.66)*	0.7707(0.12)	0.6670(0.14)	**0.6952**(0.10)*
ISLES						
CNN baseline	0.5795(0.28)	27.6725(25.58)	72.3048(121.12)	0.6590(0.31)	0.7586(0.33)	0.4941(0.35)
Postproc-CRF[125]	0.5621(0.31)	**19.5302**(20.72)	59.1030(85.99)	0.6132(0.34)	0.6518(0.39)	0.5545(0.36)
Intensity-CRF[125]	0.5758(0.26)	46.6002(32.17)$^\diamond$	65.9278(68.98)	0.6397(0.30)	0.7350(0.33)	0.4094(0.31)$^\diamond$
Spatial-CRF	0.5898(0.26)	31.1519(29.50)	93.1006(171.83)	**0.6794**(0.28)	**0.7848**(0.31)	0.4945(0.34)
Posterior-CRF	**0.6075**(0.24)	25.1834(23.27)	**47.5171**(38.34)	0.6501(0.29)	0.7443(0.31)	**0.5625**(0.32)

*: significantly better than CNN baseline ($p<0.05$).
$^\diamond$: significantly worse than Posterior-CRF ($p<0.05$).

CRFs that depend strongly on intensity-based features have difficulties detecting objects that are similar in intensity. Examples of this problem can be observed in the segmentations for the CT arteries and ISLES datasets (Figure 6). In CT arteries segmentation, the aorta and pulmonary artery have very similar intensities, which causes most of the methods in our experiments to sometimes misclassify part of the aorta as pulmonary artery. This is especially true for Post-processing CRF but also for Intensity-CRF.

Posterior-CRF achieves a DSC segmentation overlap of 95.4% and an H95 lower than 2.87 mm in aorta segmentation, which is significantly better than all other methods on this dataset. We argue that this is because the features from the last CNN feature maps are more informative than the intensity-based features, which allows the CRF inference to focus on refining the object boundary without expanding into neighboring class voxels with similar intensities. The Posterior-CRF also gives a performance improvement in the segmentation of the pulmonary artery, but this is not always statistically significant. One reason is that the blurred boundary between the aorta and pulmonary artery often results in the oversegmentation of pulmonary artery, the errors in pulmonary artery are emphasized because the overall pulmonary artery volume is lower. Another reason could be the curved shape of the pulmonary artery, which makes the results vary a lot between patients.

We see similar behavior on the ISLES dataset. The intensity boundaries of the large ischemic stroke lesions are ambiguous and their appearance varies a lot between lesions. Most of the methods fail to segment the boundaries accurately (see Figure 6 ISLES). Post-processing CRF hardly solves the problem and performs slightly worse than CNN. Posterior-CRF achieves better (while less significant due to the large prediction variance between samples) segmentation performance on DSC, AVD, lesion-wise F1.

A properly tuned spatial component of the post-processing CRF can benefits CT arteries and ischemic stroke lesion segmentation (Appendix in Section 5.B, Figure 5.B.1 (a) and (c)). However, it can cause problems to white matter hyperintensities no matter how we try to tune it (Appendix in Section 5.B, Figure 5.B.1 (b)), where we can see a positive ω_2 always leads to a decreased performance since the spatial smoothing contributes to remove both isolated true positives and false positives if they are small enough. The complete SHAP analysis will be discussed in Appendix in Section 5.B.

The negative effect of the spatial smoothing results in the low average lesion-wise recall score in WMH segmentation for Postproc-CRF (34.8%) and can be observed in the WMH segmentation results (see Figure 6). In this case, Postproc-CRF is always worse than vanilla CNN (within our grid-search range). This is because the scenario where post-processing CRF has no influence (with both ω_1 and ω_2 set to zero) was not included in the grid search range (0.1,10). Intensity-CRF has a higher lesion-wise average recall than CNN baseline (68% to 64.8%) but a lower (not significantly) voxel-wise recall (77.5% to 79.8%): although it detects more correct lesions than CNN due to the intensity features, its use of spatial features causes it to undersegment individual lesions (see Figure 6). Spatial-CRF also suffers from this problem, with a high lesion-wise recall of 68.8% but low lesion-wise F1 of 65.7%.

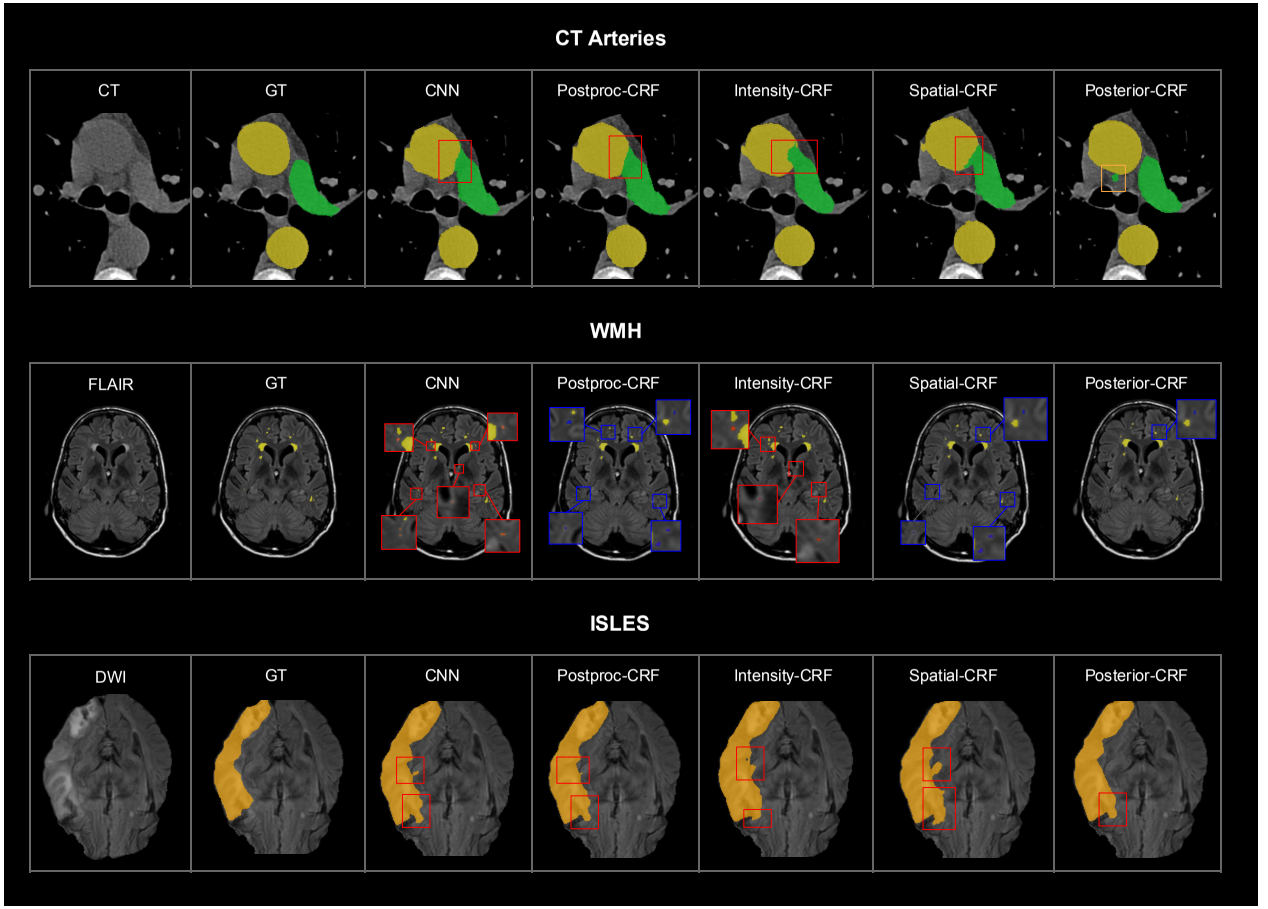


Figure 6: ***Example segmentation results.*** *From left for each row: (1) Original image (2) Manual annotation (3) CNN baseline (4) Postproc-CRF (5) Intensity-CRF (6) Spatial-CRF (7) Posterior-CRF. Aorta is colored with yellow and the pulmonary artery is green, white matter hyperintensities and ischemic stroke lesions in yellow. Red/blue rectangles indicate areas with over/under segmented voxels and the orange rectangle indicates another branch of pulmonary artery whose annotation starts in the next few slices and merged with the main branch gradually. In the WMH example (second row), only detections that do not overlap with any ground truth voxel (false positive lesions) or ground truth lesions for which no voxel is detected (false negative lesions) are highlighted, and in the zoomed patches red and blue voxels indicate false positive and false negative lesions respectively. Better viewed in color with zoom.*

5.5.2 Optimization of the End-to-end CRF

We show the evolution of the trainable CRF parameters in one data split of WMH dataset in Figure 5. For the four parameters in the 2×2 compatibility matrix μ and the two diagonal spatial kernel weights ω_2, Spatial-CRF falls into different local optimal values compared to other CRF methods, probably because different parameter scaling due to the lack of the appearance kernel. In contrast, Intensity-CRF and Posterior-CRF converged to similar optimal values for μ and ω_2. For the two diagonal bilateral kernel weights in ω_1 that control the appearance kernel, Intensity-CRF and Posterior-CRF converged to two different optimal values. This suggests that different CRF feature spaces contribute mostly through the appearance kernel and less through the compatibility matrix or the spatial kernel. Interestingly, for the second diagonal bilateral weight $\omega_1^{(2)}$, there is a different trend of Posterior-CRF compared to Intensity-CRF, which may indicate that at the early training stage Posterior-CRF uses similar feature space like that in Intensity-CRF, but at the later stage it finds and learns another set of features that may help categorize the lesion class better, which are more reliable than the original intensity features.

5.5.3 Influence of CRF Hyperparameters

We conduct experiments to investigate the influence of CRF hyperparameters on both end-to-end CRF with predefined features and the proposed CRF with learned features.

Trainable CRF parameters. The CRF weights μ, ω_1, and ω_2 in the end-to-end CRF learning can be automatically updated together with CNN weights. We run Intensity-CRF and Posterior-CRF using WMH datasets with five different initializations of CRF weights randomly sampled from the search scale with all other parameters the same as in Table 2. The CNN initializations are the same for all experiments. The results in Table 4 show that Intensity-CRF and Posterior-CRF converge to similar optimal points across different initializations. Spatial-CRF shows higher variances across experiments and is less stable to the change of initializations. Posterior-CRF is more robust to changes in initialization, achieving higher average performance and smaller standard deviations compared to Intensity-CRF and Spatial-CRF.

Empirically tuned parameters. The CRF standard deviation parameters θ_α and θ_γ, controlling the spatial terms, and θ_β controlling the appearance term, were tuned empirically to give the best results for post-processing CRF. We here test, for WMH segmentation, five different values of θ_α, θ_β, and θ_γ for Intensity-CRF and Posterior-CRF and five different values of θ_γ for Spatial-CRF within the search scale. All other parameters are the same as in Table 2. The results are shown in Figure 7. We can see that Posterior-CRF is more robust to θ_α and θ_β and has consistently better performance than Intensity-CRF within the search scale, suggesting that Posterior-CRF parameters are more easy to tune. All CRF methods degenerate performance when θ_γ becomes larger and show the best performance when using a similar value as that in the grid search for post-processing CRF. Spatial-CRF is more robust to θ_γ compared to other CRF methods and has similar performance as CNN baseline with

Table 4: Performance (Dice score) across 5 different initializations of CRF weights on WMH dataset.

Methods	Intensity-CRF	Spatial-CRF	Posterior-CRF
Mean (std)	0.7570 (0.008)	0.7507 (0.02)	**0.7833** (**0.003**)

larger θ_γ. This indicates that large θ_γ reduces the CRF effect and the spatial term may introduce more incorrect segmentation when there is also an appearance term in the end-to-end CRF like Intenity-CRF and Posterior-CRF.

5.5.4 Influence of Hierarchical CNN Features as CRF Reference Maps

We conduct experiments to investigate which level of features – earlier or deeper in the network – are more useful for the feature-learning-based CRF. We implement nine variants of feature-learning-based CRF with different levels of CNN feature maps as reference maps in the same 3D UNet architecture. For example, the method *FL-CRF-e-1* indicates the feature-learning-based CRF using the level 1 feature maps in the UNet encoder path as CRF reference maps. The implementation detail of FL-CRF-e-1 is shown in Figure 3. To reduce the computational cost and keep the same layer capacity as Posterior-CRF, the 32-channel (or more in deeper layers) feature maps are encoded into C-channel feature maps and go through a softmax layer as the CRF reference maps. Since there is no gradient flowing back through the reference map path, we optimize the softmax layer with the segmentation loss directly in order to preserve as much semantic information as possible. Note that for CRF methods that use deeper CNN layers as reference maps, such as FL-CRF-e-2 to FL-CRF-d-2, we upsample the reference maps to the original image scale and optimize them with the segmentation loss, similar to FL-CRF-e-1.

The results are shown in Figure 8. Note that if we use the CNN input as CRF reference maps, it turns into Intensity-CRF; if we use the last CNN layer as CRF reference maps, it turns into Posterior-CRF. In the figure, we can see that all feature-learning-based CRF approaches (including Posterior-CRF) outperform Intensity-CRF and the overall Dice performance in the decoder path is better than that in the encoder path, indicating that CNN learned features are more useful to the CRF inference than intensity is and later CNN features are more useful than early features. The performance degenerates towards the middle part of the UNet (from FL-CRF-e-1 to FL-CRF-e-5 and FL-CRF-d-1 to FL-CRF-d-4) but fluctuates at the $2^{\text{nd}}/3^{\text{rd}}$ level. We argue that this may be due to the pooling effect which enables CNN to extract higher-level features but loses the spatial information at the same time. Posterior-CRF achieves the best performance among all variants and we argue that this is because the last CNN layer are more likely to contain more useful information for CRF inference and it still keeps the same spatial scale as the original image.

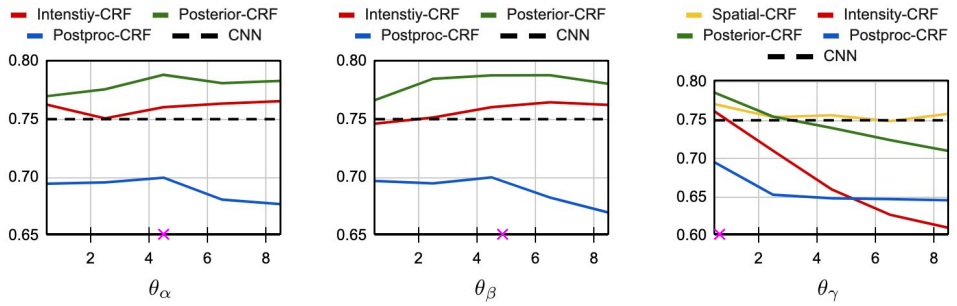


Figure 7: Dice performance of varying θ for CRF methods on WMH dataset. *CNN result is shown as the black dash line. Purple crosses indicate the values used in Table 4. Best viewed in color with zoom.*

5.5.5 Evolution of CNN and CRF Outputs

The concurrent optimization of CNN and CRF in our end-to-end models allows the CNN and CRF to interact during training. We observed that this has a strong effect on what the CNN learns in the early training epochs. Figure 9 shows the evolution of CNN and CRF outputs for three typical examples. The baseline CNN without CRF converges quickly and focuses on the large lesions, already producing a fairly sparse output after the first epoch. The end-to-end models converge more slowly, and in this case the output of the CNN is influenced by the choice of CRF mostly in the early stage of training. For example, the CNN in the Intensity-CRF model initially tends to highlight voxels with similar intensity as the foreground (1 to 20 epoch), while the CNN in the Spatial-CRF model preserves the spatial coherence between voxels and outputs many small groups of voxels (5 epoch). The CNN in the Posterior-CRF model first focuses on the coarse area that might contain the target lesions (1 to 5 epoch) and then refine the prediction gradually to the ground truth (5 to 20 epoch). Eventually, all models converge to a result close to the ground truth.

5.6 Discussion

In this chapter, we explored efficient methods to combine the global inference capabilities of a CRF with the feature extraction from a CNN. Our end-to-end approach optimizes the CRF and CNN at the same time, and allows the two components of the approach to cooperate in learning effective feature representations. This gives our method an advantage over traditional CRFs that only use the original image intensities and position information. Intensity-based features can be suboptimal for problems where the intensity does not provide sufficient information to find the object boundaries, for example because the contrast between objects is too small.

Unlike other CRF methods, our Posterior-CRF uses adaptive learning-based features that are learned by the CNN and can combine spatial and appearance information in a way that suits the CRF. The results show our method can achieve stable, good

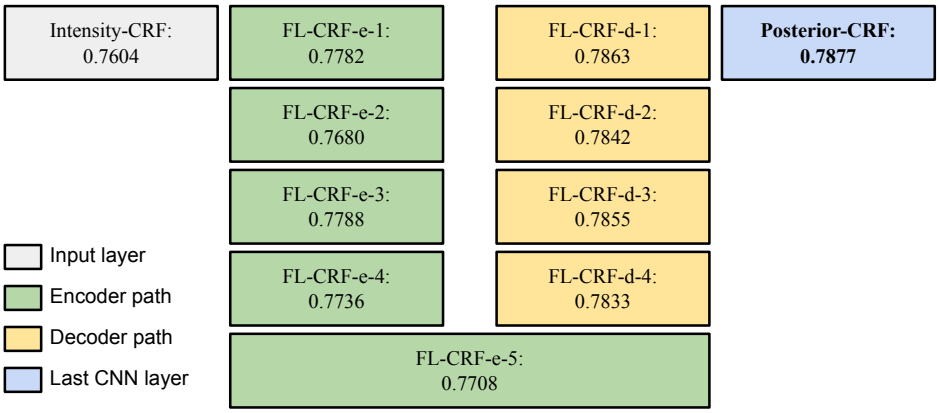


Figure 8: Dice performance of end-to-end CRFs using different CNN feature maps in an independent run on WMH dataset. *Different blocks indicate different level of CNN feature maps used as CRF reference maps. Best viewed in color with zoom.*

performance across a range of segmentation applications and imaging modalities. FL-CRF variants that use early CNN features in Section 5.5.4 achieve in-between performance between Intensity-CRF and Posterior-CRF, using learning-based features that range from more similar to intensity to more similar to posterior probability maps. Finally, we found that integrating learned features into the CRF model reduces the need to fine-tune CRF parameters, making the method easier to apply than CRF methods with predefined features.

5.6.1 Interaction between CRF and CNN

Figure 9 leads to the counter-intuitive observation that, at least initially, the CNNs in end-to-end models seem to imitate the CRF instead of complementing it. For example, the CNN output in Intensity-CRF highlights the ground truth, but also finds areas with similar intensities, producing something that looks very similar to the original image (20 epoch). The CNN output in Spatial-CRF selects the ground truth but also includes clusters of voxels in other areas (5 epoch).

This effect can be explained by the way the CNN and CRF interact during training. In Intensity-CRF and Spatial-CRF, the only interaction between CRF and CNN takes place through the unary map (Figure 4, step 5, green arrow). For example, consider how this works in the Intensity-CRF. In WMH segmentation, the ground truth is usually high-intensity area. However, for the voxels with high intensities but not the target lesions, it is difficult to get both low pairwise CRF potentials and low segmentation loss, since labeling them as non-lesion goes against the CRF assumption that voxels with similar high-intensities are more likely to be the lesion class. For

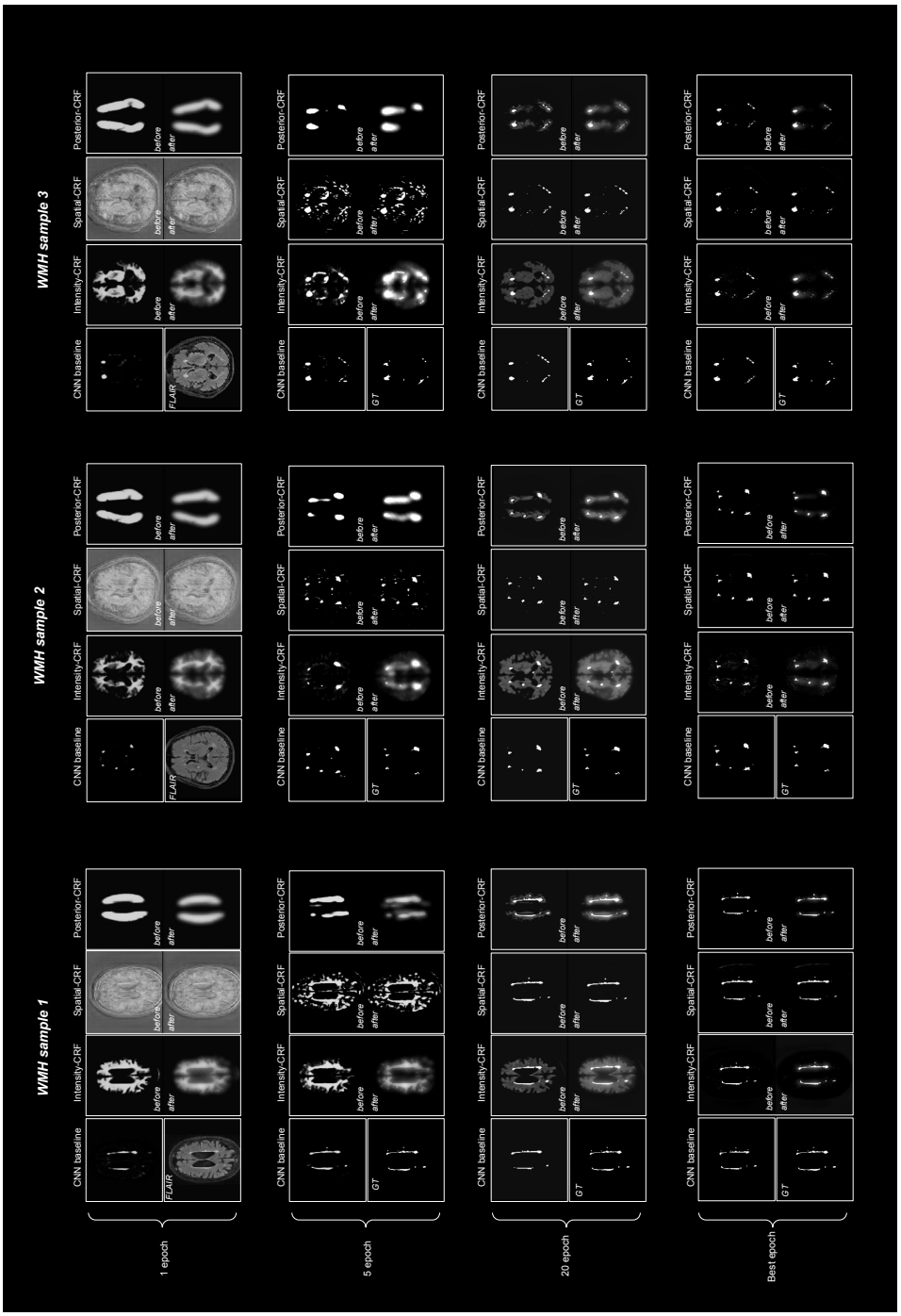


Figure 9: ***Evolution of CNN and CRF outputs during training.*** *The CNN output maps and CRF results for WMH segmentation in 3 different MRI images (columns) are shown at, from top row to bottom row, epoch 1, 5, 20, and the best epoch. The best epoch is chosen when the model shows the best validation performance till the end of training (usually at 50∼80 epoch).* ***FLAIR:*** *the input FLAIR image of the current training sample.* ***GT:*** *ground truth.* ***CNN baseline:*** *the last layer (softmax output) of CNN.* ***Intensity-CRF, Spatial-CRF, Posterior-CRF:*** *the probability maps before/after the CRF layer at different epochs during training.*

convenience, we call these voxels as *hard voxels*, indicating the voxels that do not fit the CRF assumption. In order to keep the correctly segmented lesions and reduce the CRF effect on the hard voxels at the same time, the CNN tends to provide unary maps that 1) highlight the ground truth area for lower segmentation loss, and 2) look similar to the CRF reference maps on the hard voxels for lower pairwise CRF potentials. In the later stage of training, CNN is encouraged to push the confidence of its outputs even further to minimize unary potentials and thus prevent CRF from undoing segmentation improvement on the hard voxels. From Figure 9, we can see that there are many hard voxels in Intensity-CRF (1 to 20 epoch, areas that look like the original image) and Spatial-CRF (5 epoch, clusters of voxels that do not belong to the ground truth) which may harm the segmentation. This indicates that the predefined features may not be the optimal feature space for the end-to-end CRF.

In the Posterior-CRF model, the CRF inference happens within the CNN feature space, which can improve the interaction between CNN and CRF. First, the features learned by CNN during training may contain information that is more useful for segmentation than that in the predefined features, which makes CRF benefit most from the CNN features. Second, using the learning-based features as CRF reference maps avoids the CRF assumption of the predefined features which may introduce many hard voxels, e.g., Intensity-CRF and Spatial-CRF, as discussed in the previous paragraph. With fewer hard voxels, the CNN in Posterior-CRF may provide better unary maps for the CRF inference.

5.6.2 Posterior-CRF vs. Mean-field Network

The mean-field approximation (MFA) in Posterior-CRF is somewhat similar to that in Mean-field networks (MFN) [141], since both methods use it to get the posterior probabilities of the variables. Therefore, MFN could be a promising alternative to the MFA process in our method. MFN has the advantage that it utilizes each layer of the network as an iteration of MFA, which has the advantage of allowing more relaxation on parameters and provides some efficiency improvements. This makes the idea of formulating Posterior-CRF as a feed-forward network like MFN very attractive. There are, however, a few limitations that would need to be solved.

The first limitation is in training. MFN is designed to provide a faster and more flexible way to obtain the prediction of MFA, by fitting a powerful function that predicts the real MFA result. To train an MFN, we first need to acquire the ground truth calculated by conventional mean-field iterations, which takes time during training but saves time during inference. On the other hand, Posterior-CRF provides a flexible and adaptive feature space for the conventional MFA, speeding up the procedure by applying Gaussian convolution in the message passing updates. As a result, the thing Posterior-CRF does is difficult to replicate with a MFN because the feature space of a Posterior-CRF changes during training, while MFN requires a predefined feature space to get the ground truth.

The second limitation is the tradeoff between dense inference and computation cost in the MFN. In its feed-forward network implementation, the computation cost increases exponentially when more neighbor nodes and number of layers are included, which limits its ability to model dense prediction problems such as segmentation tasks.

5.6.3 Posterior-CRF vs. Graph Neural Networks

The proposed Posterior-CRF shares some similarities with graph neural networks (GNN) [131, 132]: both approaches aim to model interactions between variables within a graph model. The difference is that Posterior-CRF pre-defines the global relationship between variables through the mean-field assumptions and solves the maximum a posteriori problem, whereas GNN learns the global variable relationship by applying graph convolution filters and mapping the input graph to the output graph [132].

It could be interesting to combine the global view of the Posterior-CRF and the more local view of the GNN. The Posterior-CRF might benefit from using a GNN to replace its CNN component for feature extraction. The graph-based network may extract better features for Posterior-CRF than a CNN, which is not designed to extract unary and pairwise features for a graphical model. Similarly, the GNN may benefit from the efficient message passing of the Posterior-CRF, which would allow it to use the local graph-based features as CRF features for global interactive modeling in a computationally efficient way.

5.6.4 Limitations

In this chapter, we show that the proposed Posterior-CRF method has benefits in the three medical imaging applications. Considering the medical imaging datasets are usually small largely because the manual annotations are very expensive to make, difference between Posterior-CRF and UNet may be smaller in larger training sets. But we know from literature that Intensity-CRF helps in some computer vision applications with large training sets (e.g., 10k 2D images or even more), it would be promising to test our method on these datasets. This is considered as our future work.

In Section 5.5.3, we show that Posterior-CRF is robust to different CRF initializations and hyperparameters. However, the standard deviation parameters still require careful tuning, especially for θ_γ in the spatial term. θ_γ is sensitive to the image scale of different datasets and the size of the target object in different applications. The optimal value from the grid search on post-processing CRF can be a good reference for the value used in the end-to-end CRF setting. Posterior-CRF is more robust to θ_α and θ_β compared to Intensity-CRF, which facilitates exhaustive tuning of these parameters.

The computational expense of the CRF also restricts the choice of applications. Compared to UNet, there is around 20% training time increased in average when applied CRF layer on top of the network. Given that Posterior-CRF uses posterior probability maps as its reference maps, it can become computationally expensive in multi-class segmentation problems. For a similar reason, Intensity-CRF and Postproc-CRF can become expensive when there are too many imaging modalities in the input channels M.

In the experiments, we use a plain 3D UNet as the backbone network for all methods. The training pipeline and hyperparameters are determined empirically and kept the same for all datasets, which could be suboptimal compared to elaborate automatic configuration strategies like nnU-Net [142]. On the WMH dataset we therefore checked the performance of nnU-Net (3D version without ensembling). Average Dice score of nnU-net (0.77) was slightly higher than our CNN baseline (0.76, difference not

statistically significant) but lower than the proposed posterior CRF using the CNN baseline as a backbone (0.79), which performed significantly better than the CNN baseline (see Table 3). Though our experiments have been limited to a standard 3D U-net architecture, We expect that posterior CRF can improve results of other segmentation architectures and other hyperparameter settings (such as nnU-net) as well.

5.7 Conclusions

In conclusion, we present a novel end-to-end segmentation method called Posterior-CRF that uses learning-based, class-informative CNN features for CRF inference. The proposed method is evaluated in three medical image segmentation tasks, including different MRI/CT imaging modalities and covering a range of object sizes, appearances and anatomical classes. In the quantitative evaluation, our method outperforms end-to-end CRF with early CNN features, end-to-end CRF approaches with predefined features, post-processing CRF, as well as a baseline CNN with similar architecture. In two of the three applications, our method significantly improves the segmentation performance. The qualitative comparison demonstrates that our method has good performance on segmenting blurred boundaries and very small objects.

Acknowledgment

The author would like to thank Raghavendra Selvan, Gerda Bortsova for their constructive suggestions for this chapter, Dr. Zaigham Saghir from DLCST for providing us with the chest CT scans, and organizers of WMH 2017 and ISLES 2015 Challenges for providing the public datasets. This work was partially funded by Chinese Scholarship Council, Iranian Ministry of Science, Research and Technology (MSRT), and The Netherlands Organisation for Scientific Research (NWO).

Appendix

5.A Mean-field Inference

Mean-field inference is an efficient approximation to computing distribution $Q(\mathbf{X})$ instead of the real CRF distribution $P(\mathbf{X})$, which could be done in an iterative algorithm 1 (see also Figure 5.A.1). X is the random field w.r.t the current 3D image patch $\mathbf{I}$.

There are three main steps inside the inference iteration. First is message passing, which is the most calculation-intense step that could be expressed as a convolution operation on all the pairwise kernels k and the initialized $Q(\mathbf{X})$. An efficient way to perform high-dimensional convolution is using permutohedral lattice algorithm [139]. In compatibility transform as the second step, all the convolution results $\hat{Q}_i^{(m)}(x_i)$ are weighted by $\omega^{(m)}$ in different sort of kernels and shared between labels to a varied extent, depending on the compatibility μ between these labels. At last, $Q(\mathbf{X})$ will be updated by the calculated pairwise potential and used as the input for the next iteration.

5.B SHAP Analysis of Post-processing CRF

We conduct SHAP (SHapley Additive exPlanations) [143] analysis on the post-processing CRF grid search results to investigate the contribution of each individual CRF parameter to the segmentation performance. With this analysis, we show that it is difficult to tune traditional CRF parameters to achieve a consistent performance

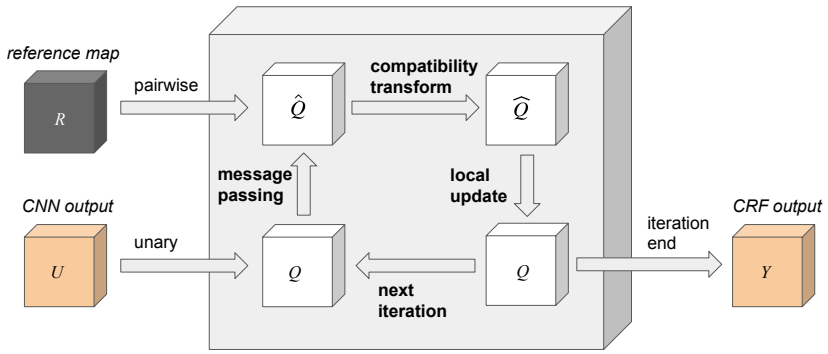


Figure 5.A.1: Mean-field approximation in the end-to-end CRF layer. *There are two inputs of the CRF layer, where U is the CNN probability maps as the unary maps and the pairwise distribution are calculated by the initialized distribution Q and the reference map* $\mathbf{I}$*. The updated distribution Y is the output of the layer at the end of the iteration.*

Algorithm 1 Mean-field inference in fully-connected CRF

$Q_i(x_i) \leftarrow U_i(x_i), i = 1, 2, ..., N$ ▷ **Initialize** $Q(\mathbf{X})$
while not reach max iteration number **do**
  $\hat{Q}_i^{(m)}(x_i) \leftarrow \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j) Q_j(x_i)$ for all m ▷ **Message Passing**
  $\widehat{Q}_i^{(m)}(x_i) \leftarrow \sum_{l \in \mathcal{L}} \mu^{(m)}(x_i, l) \sum_m \omega^{(m)} \hat{Q}_i^{(m)}(l)$ ▷ **Compatibility Transform**
  $Q_i(x_i) \leftarrow \exp\{ -\varphi_u(x_i) - \widehat{Q}_i(x_i) \}$ ▷ **Local Update**
  normalize $Q_i(x_i)$
end while

improvement on different applications, and our proposed method does not require tuning parameters. Moreover, the analysis shows the importance of each modality to each dataset, which can be automatically adapted in the proposed method but not in traditional methods. The model is trained using *XGBoost* [144] for 100 iterations using a learning rate of 0.5, 0.01, and 0.01 for CT Arteries, WMH, and ISLES respectively. Note that the SHAP analysis results can only be explained under the assumption of the current parameter search scales and XGBoost models.

The results are shown in Figure 5.B.1. The summary plot in the left sub-graph shows an overview of all parameter sets with the most important parameters on top of the list. For each dataset, the best and worst parameter settings are shown in the right sub-graph. For all datasets, the post-processing quality is affected most by the spatial parameters ω_2 and θ_γ, and less by the intensity parameters per modality θ_β.

The results on the CT arteries data (Figure 5.B.1a *left*) are more stable (with smaller SHAP values) than the results for WMH and ISLES, indicating that the post-processing CRF can hardly change the CNN output of the artery segmentation (see Figure 6 as an example).

In the WMH dataset, looking at independent parameter contributions, low values for spatial parameters ω_2, θ_γ (less smoothing), and a smaller number of iterations lead to an improved performance. This is not unexpected, because white matter lesions are sparsely distributed and spatial smoothing tends to remove small lesions. Too strong spatial correlations (either large weight ω_2 or small θ_γ) will remove true positives as well (see Figure 6). The summary plot (Figure 5.B.1.b *left*) shows, as expected, that the FLAIR image has a larger impact on the model than the T1 image. Table 1 also shows a smaller θ_β selected (corresponding to higher influence) for FLAIR.

Similar trends can be found for the ISLES dataset (Figure 5.B.1.c). Spatial parameters ω_2 and θ_γ are important to tune and high values can strongly harm the performance. The summary plot shows that the DWI image has a larger impact on the model than T1, T2, and FLAIR. In Table 1, θ_β for FLAIR and DWI are smaller than θ_β for T1 and T2, which means that FLAIR and DWI images are more informative for the segmentation of ischemic stroke lesions.

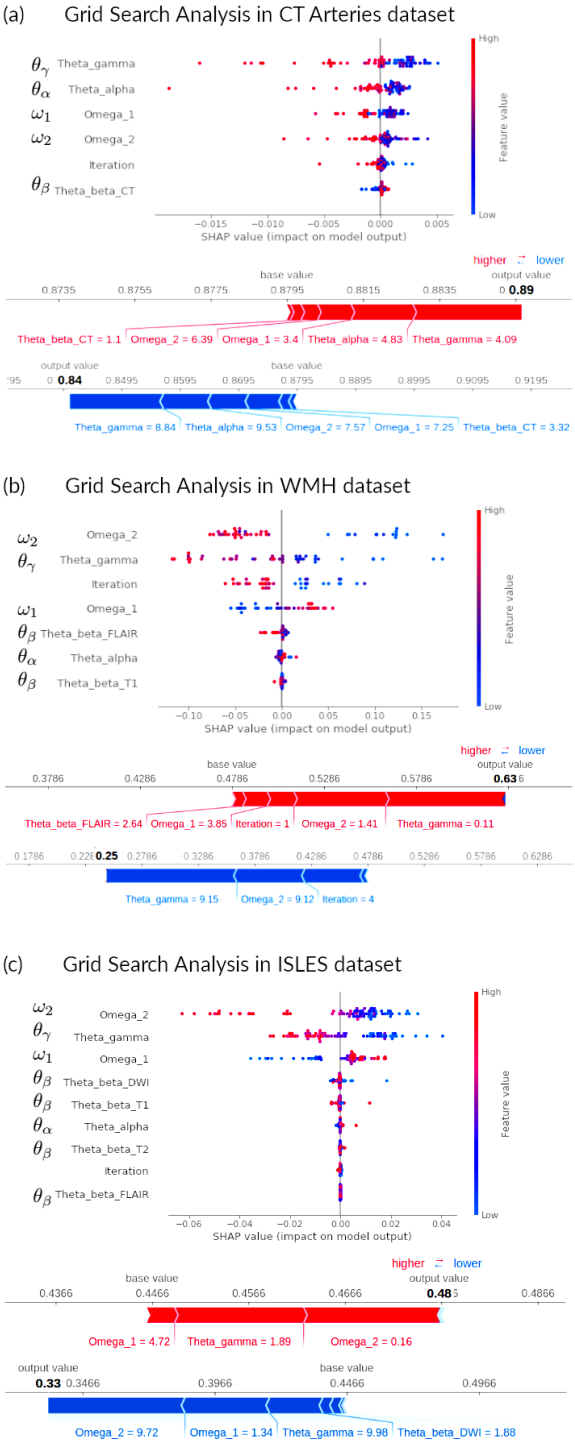


Figure 5.B.1: ***SHAP analysis of the grid search results.*** *See Section 5.B for an explanation.* ***Upper sub-graphs:*** *summary plots of all parameter sets evaluated during grid search. Positive SHAP values indicates a positive contribution to the performance and vice versa. The legend (feature value bar) shows the search range for each parameter. This reveals for example that lower values of ω_2 lead to better segmentation performance for all datasets.* ***Lower sub-graphs:*** *the best (1st row) and worst (2nd row) parameter sets for each dataset. Red bar represents positive contribution to the performance and blue bar is negative contribution. Base value is the average DSC of all grid search results and output value is the DSC in the parameter set depicted. Best viewed in color with zoom.*

6

General Discussion

6.1 Summary

Measurement of the dilatation of large blood vessels such as the aorta and pulmonary artery requires accurate delineation of the vessel anatomy. The most accurate quantification of their dimensions requires a full 3D analysis. Compared to manual measurement, automated 3D vessel segmentation and diameter measurement methods have the potential to objectively quantify vessel enlargement and provide a means of early-stage diagnosis. In this thesis, I develop and evaluate fully automatic methods for segmenting the aorta and pulmonary artery and measure their diameters in non-ECG-gated, non-contrast CT scans.

The presented segmentation methods in this thesis can be divided into two categories: optimal surface graph-cut based segmentation method presented in Chapter 2 and Chapter 4, and a deep learning based method presented in Chapter 5. This thesis also provides techniques for automatic diameter measurement in Chapter 2 and Chapter 4. Chapter 3 investigates the aortic growth rates in a population of current and ex-smokers using the automatic segmentation and diameter measurement technique presented in Chapter 2. This chapter summarizes the main contributions of our work and provides future research directions.

6.2 Optimal Surface Graph Cut based Segmentation Method

Chapter 2 and Chapter 4 present an optimal surface graph cut segmentation method to provide a 3D segmentation of the aorta and pulmonary artery. In these chapters, the method's accuracy and robustness are investigated on non-ECG-gated, non-contrast CT scans. Since graph cuts have the ability to achieve a global optimum in relatively short processing times, and they allow the incorporation of shape or smoothness constraints in the graph structure, a graph is constructed with non-intersecting graph columns based on flow lines generated from a predefined initial coarse segmentation. These graph columns resulted in non-self-intersecting surfaces and preserved the topology of the initial shape. In addition, smooth segmentation similar in shape to the initialization are encouraged by adding the "smoothness penalty" edges to the graph. This results in smooth segmentation even in the presence of motion artifacts caused by the motion of the heart during the cardiac cycle. To obtain an initial coarse segmentation providing information about the vessel's shape, the initial segmentation is defined as the non-uniform morphological dilation of the vessel centerlines based on a radius estimation map extracted by a multi-scale medialness filter.

Chapter 2 performed a 3D quantitative evaluation on 100 non-contrast CT scans for segmenting the aorta and obtained an average Dice Similarity Coefficient (DSC) of 0.95 ± 0.01 and an average mean surface distance (MSD) of 0.56 ± 0.08 *mm*. 3D quantitative evaluation on 25 non-contrast CT scans for the pulmonary artery and aorta segmentation presented in Chapter 4 resulted in a DSC of 0.94 ± 0.02 and 0.96 ± 0.01 with an MSD of 0.62 ± 0.33 *mm* and 0.43 ± 0.07 *mm*, respectively. Although unclear vessel boundaries and the similarity in the vessel intensities on non-contrast CT scans made the aorta and pulmonary artery segmentation a challenging task, the presented segmentation method performed well with a high segmentation accuracy. Additional qualitative assessment of the aorta and pulmonary artery segmentation on

419 non-contrast CT scans demonstrated the robustness of the presented method: 92% of the aorta and 94% of the pulmonary artery were segmented with very high quality and with no obvious errors, 4% of the aorta and 6% of the pulmonary artery had minor errors with a maximum 3 *mm* surface distance, and only 2% of both the aorta and pulmonary artery required corrections with errors larger than 3 *mm*. Chapter 4 showed that even though there are sometimes segmentation errors, these occur mainly in the pulmonary trunk or at the aortic root due to the presence of motion artifacts and adjacency of the vessels to each other and the surrounding structures. The method also has good segmentation accuracy in high curvature areas such as the pulmonary artery bifurcation, resulting in accurate diameter measurements of the pulmonary artery and aorta in the pulmonary artery bifurcation level and subsequently accurate measurements of the PA:AA ratio.

In Chapter 4, the accuracy of the aorta segmentation is improved by improving the landmark and seed point extraction accuracy compared to Chapter 2. In this chapter, the multi-atlas registration method is directly applied for seed point and landmark extraction, whereas in Chapter 2, the multi-atlas registration technique is used to extract a coarse segmentation of the pulmonary artery and the accuracy of landmark and seed points are dependent on the quality of this segmentation. In Chapter 4, the seed point for the descending aorta is extracted at the diaphragm level, whereas in Chapter 2, it is extracted 6 cm below the level of the pulmonary artery bifurcation. Having the seed point extracted at the diaphragm level resulted in larger aorta segmentation, which covers a larger part of the aorta anatomy and, therefore, is useful for diagnosing the descending aorta dilatation or aneurysm. Furthermore, the extracted seed points did no longer depend on the location of the extracted landmark for the level of the pulmonary artery bifurcation, which improved the reliability. The evaluation of seed point and centerline extraction on 942 vessels, including the pulmonary artery and aorta (471 CT scans), with 98.9% accurate vessel centerlines, demonstrated the accuracy and robustness of the method.

The advantage of this method, besides the smooth accurate segmentation result, is that it requires no human interaction for seed point or landmark placement, and thus the entire method is fully automated.

6.3 Posterior-CRF Segmentation Method

In the methods presented in Chapter 2 and Chapter 4, the aorta and pulmonary artery are segmented separately. With supervised neural networks such as Convolutional Neural Networks (CNN)s it is possible to convert the segmentation problem to a classification problem and segment the vessels jointly by a voxelwise classification. CNNs are powerful in extracting local features and perform good predictions. However, the lack of using context information for modeling interactions and relations between nearby objects can result in poorly segmented boundaries. To overcome this challenge and refine the prediction from the CNN, in Chapter 5, CNN is combined with a Conditional Random Field (CRF), where CRF models the correlations and dependencies among the voxels being predicted. It is common to use CRFs based on predefined features such as intensity as an efficient post-processing technique, based on the assumption that voxels with a similar intensity or close adjacency are likely to

belong to the same class. However, for adjacent vessels wiht similar intensities and unclear vessel boundaries in non-contrast CT scans, only intensity-based features of the CRF are not sufficient for identifying the vessel boundaries.

To address these challenges, in Chapter 5, a new end-to-end deep learning based segmentation technique, named Posterior-CRF, is presented that allows the CRF to use features learned by a CNN, optimizing the CRF and CNN parameters concurrently. During training, the CRF inference works in the CNN feature space, which is likely to be more informative for segmentation than the original intensity values of the image. The proposed Posterior-CRF method consists of two parts that are optimized jointly: 3D CNN and 3D CRF. A 3D UNet is used as the basic CNN architecture to provide unary potentials for the CRF inference as well as features for the pairwise potentials for the proposed Posterior-CRF. Posterior-CRF takes the last softmax layer of the UNet feature maps as the reference maps and updates the random field based on this new feature vector from the UNet feature maps instead of intensity features in the CRF kernel directly.

The proposed method is demonstrated and evaluated on three medical image analysis applications: segmentation of the aorta and pulmonary artery on non-ECG-gated, non-contrast chest CT scans; segmentation of white-matter hyperintensities on brain MRI; segmentation of ischemic stroke lesions on brain MRI. These three segmentation problems vary in object shape appearance, imaging modalities and pose different challenges such as isolated small objects in white matter hyperintensities or similar intensity values in adjacent arteries in non-contrast CT. Therefore validating the method on these three different applications presents the generalizability of the method.

The accuracy of the proposed network is compared with that of a baseline UNet, a postprocessing CRF, and two forms of end-to-end CRFs with predefined features: Intensity-CRF, which uses the original image intensities and position information, and Spatial-CRF, which uses only the position information. Among these networks, the proposed Posterior-CRF obtained a DSC of 0.95 for the aorta and 0.89 for the pulmonary artery, outperforming the baseline and other three networks. The DSC obtained with the proposed Posterior-CRF for the segmentation of white-matter hyperintensities and segmentation of ischemic stroke lesions in MRI also outperformed the other networks.

The presented method in Chapter 5 showed that spatial coherence or intensity features alone are not sufficient and often are detrimental to segmentation accuracy and that the CNN features in the last layer as presented in Posterior-CRF are more informative for CRF than the intensity features in the original images. In addition, this chapter showed that integrating learned features into the CRF model reduces the need to fine-tune CRF parameters, making the method easier to apply than CRF methods with predefined features.

6.4 Diameter Measurement

It is essential to provide precise, reliable, and reproducible aortic diameter measurements in almost all stages of the management of aortic pathologies as they are used in predicting the risk of rupture or decision-making for intervention. However, it is

very time-consuming to manually measure diameters at multiple levels perpendicular to the vessel centerline. Therefore automatic diameter measurement techniques are desired. In Chapter 2, an automatic method for aorta segmentation is presented. With this method, a 3D aorta segment is driven, and from that, aortic diameters are automatically measured at multiple, fixed levels relative to the landmark level extracted at the level of the pulmonary artery bifurcation. To improve the robustness of the diameter measurements, cross-sections are extracted at every 1 *mm* along the extracted aortic centerline, and subsequently, the average diameter from the segmented cross-sectional area is computed. Average aortic diameters are assessed at 13 cross-sections located at 1-cm intervals around the bifurcation level from 2 cm below this level to 3 cm above for the ascending aorta and from 3 cm above to 3 cm below this level for the descending aorta. The accuracy of the method is validated on 100 CT scans where high agreement between the manual and automatic aortic diameters is obtained with an overall intra-class correlation (ICC) of 0.97 and an average per-level ICC of 0.91 ± 0.03. The accuracy of the extracted landmark level of the pulmonary artery bifurcation is also compared to manual landmarks, and the method achieved a low mean absolute distance of 2.55 ± 1.94 *mm*, with almost no bias. Repeatability of the method is assessed by comparing the automatically extracted diameters of scan-rescan pairs of 617 subjects with a period of 1 year in between and achieved a per-level ICC of 0.94 ± 0.01. The results from the proposed automatic method in this chapter showed that it is an accurate and reliable technique to assess subtle signs of aorta dilatation or aortic diameter changes without any human interaction and, therefore, can be used in clinical practice.

The accuracy of the presented method in Chapter 2 led us to apply the presented method on a larger cohort to measure thoracic aorta diameters and assess the aortic growth rate in a subgroup of the general population, namely smokers. Therefore in Chapter 3, the ascending and descending aorta diameters are automatically measured at the level of the pulmonary artery bifurcation on the first and last non-contrast CT scan during the follow-up of almost 2000 current or former smokers from the Danish Lung Cancer Screening Trial (DLCST). A growth rate of approximately 0.1 *mm*/year in both males and females is found from the measured diameters, which is consistent with aortic growth rates of 0.08 to 0.17 *mm*/years measured in cross-sectional studies of the general population. Furthermore, with the presented 95^{th} percentiles of aortic growth ranged from 0.42 to 0.47 *mm*/year, we show that an aortic growth of > 0.5 *mm*/year can be considered the upper limit of normal.

With the growing use and widespread availability of non-contrast CT scans, such as those presented for lung cancer screening, there is an opportunity to measure the diameters of the aorta and pulmonary arteries to identify the presence of early-stage cardiovascular disease and/or predict complications in patients with COPD. The ratio of the diameter of the pulmonary artery to the diameter of the aorta at the level of the pulmonary artery bifurcation (PA:AA) has shown to be a strong predictor for severe exacerbations in patients with COPD and is associated with increased mortality. Clinically standardizing the measurement of the diameter of the main pulmonary artery is difficult since clinicians mainly measure the greatest diameter of the pulmonary artery in an axial slice, which the orientation of the vessels with respect to the patient and with respect to the axial plane may vary, leading to variability

in axial diameter measurements. Furthermore, measurements that are further away from the bifurcation show smaller/larger diameters than those close to the pulmonary bifurcation, leading to subjective measurements based on slice location. Determining the exact level of bifurcation is also difficult and is prone to variability. Therefore, in Chapter 4, a 3D volumetric average diameter measurement technique is presented, where the diameters are measured in 3D segments along the vessel centerlines around the extracted pulmonary artery bifurcation level. This proposed technique is less subjective and is more robust and reproducible than diameter measurement in 2D axial slices. We validated the repeatability of the presented measurement technique on scan-rescan pairs of 112 subjects (224 CT scans) with an average period of three months in between. Changes to the pulmonary artery and aorta diameters are expected to be negligible within three months. We obtained ICC of 0.89, 0.95, and 0.86 for the pulmonary artery, aorta, and the PA:AA ratio, respectively.

Besides the presented 3D volumetric diameter technique, for validating the accuracy of the method in 2D axial diameter measurement, the manual and automatic diameters of the pulmonary artery and aorta in axial slices at the level of pulmonary artery bifurcation are compared. A high agreement between the manual and automatic diameters is obtained with ICC of 0.92, 0.97, 0.90 for the pulmonary artery, aorta, and PA:AA ratio in axial slices. The extracted high agreements for both in-slice diameters and 3D volumetric diameters show the accuracy and repeatability of the segmentation and measurement method.

6.5 General Discussion & Future Directions

This thesis contributed to accurately and fully automatically segmenting the aorta and pulmonary artery and measuring their diameters. In this discussion, I compare the presented methods with each other, summarize the limitations, and present possible future directions for methodological improvements and translation to clinical practice. **Chapter 2** and **Chapter 4** proposed a fully automatic segmentation method based on optimal surface graph cuts. Due to the need for accurate aortic diameters, a fully automatic diameter measurement technique is presented in **Chapter 2**, and this method is applied on a large cohort in **Chapter 3**. **Chapter 3** studied the aortic growth rate in current and ex-smokers. In **Chapter 5**, a deep learning-based method is proposed, named Posterior-CRF, to jointly segment the aorta and pulmonary artery.

The optimal surface graph cut segmentation method proposed in **Chapter 2** and **Chapter 4** requires an initial segmentation prior to the vessel segmentation method. To this end, a multi-atlas registration technique is utilized for seed point and landmark detection, and a path-tracking algorithm is applied for centerline extraction followed by a morphological dilatation to construct an initial segmentation. Furthermore, the aorta and pulmonary artery are segmented separately, and to achieve the segmentation of both, two optimal surface graph problems are evaluated. This results in a relatively slow segmentation technique. Therefore, although this method's high accuracy and reliability make it suitable for clinical study, the relatively long processing time may limit the uptake of the method in clinical practice.

The learning-based Posterior-CRF segmentation method proposed in **Chapter 5**, once trained, can perform a joint segmentation of the aorta and pulmonary artery

in less than 10 seconds. However, note that only the vessels are segmented in this method, and no landmarks and centerlines are extracted for automatic diameter measurement and subsequent biomarker extraction. Therefore, although it has a minimal processing time for vessel segmentation, another study addressing landmark and biomarker extraction, validation, and adoption of the method to a large cohort, is still required for drawing a strong conclusion on the desired tool for clinical practice.

Besides the advantage of fast segmentation in the Posterior-CRF method, unlike the segmentation results from the optimal surface graph cut, the segmentation results do not suffer from overlapping areas. In the Posterior-CRF method, the segmentation problem is converted to a classification problem where the voxels are classified into one of the three classes of the aorta, the pulmonary artery, or the background and it is not possible to classify one voxel in more than one class. This results in having non-overlapping segmentations for the aorta and pulmonary artery. However, since the network has no prior information of the geometry of the vessels and no explicit shape constraints have been incorporated, in regions with very unclear boundaries, the method can classify a part of the aorta as pulmonary artery or vice-versa, resulting in unrealistic/incorrect surfaces.

The advantage of the optimal surface graph cut method is incorporating the topology constraints into the graph structure and defining "smoothness-penalty" edges to encourage a smooth segmentation similar in shape to the initialization. These constraints restrict the final segmentation to plausible vessel surfaces. However, the vessels are segmented separately. Therefore in regions where the vessels are adjacent and boundaries are unclear, such as the aortic root, the segmented aorta and pulmonary artery might slightly overlap.

In general, the segmentation accuracy of deep learning based methods is largely dependent on the size and quality of the ground truth dataset (manual annotations) used in its training. The number of samples and quality of the ground truth also affect parameter tuning of the optimal surface graph cuts method in **Chapter 4**. The ground truth used to train the Posterior-CRF in **Chapter 5** consists of only 10 CT scans. These scans are from relatively healthy subjects with no abnormalities in the vessels's shape and size, such as dilatation. This may reduce the generalizability of the methods. Comparing the Posterior-CRF method and the optimal surface graph cuts method on the same dataset shows that both methods have similar accuracy for the aorta segmentation with DSC of 0.953 and 0.959, respectively. However, for the pulmonary artery, the Posterior-CRF method achieves a DSC of 0.910, which is lower than that for the optimal surface graph cuts method with a DSC of 0.934 on the same dataset. This can be due to the small ground truth used for training the Posterior-CRF method. The errors mainly occur in the pulmonary artery trunk, where the vessel can have an ellipsoid or a circular shape. It is also important to note that segmenting the pulmonary artery in regions close to the heart chamber is more complicated than the aorta due to the complex anatomy of the pulmonary artery in this region. The ground truth is created by only one observer, where in regions close to the heart with very unclear boundaries, human error might exist. Although the process of manual annotations has a high cost, having more than one observer could help the methods to learn the correct boundaries even in regions difficult for experienced radiologists.

In this thesis, vessel centerline extraction is an essential step for analyzing the vessel geometry, and as an input for segmentation and diameter measurement. **Chapter 4** extracted the centerline of the left and right pulmonary artery separately without considering any interaction between these two arteries and any prior information about the pulmonary artery bifurcation level. Our proposed method in **Chapter 4** showed only less than 1% (8 out of 942) failures in the centerline extraction of the left and right pulmonary arteries, indicating the robustness of the method. The failed centerlines either were very close to the border, or the centerline of the right pulmonary artery made a shortcut through the ascending aorta and the aortic arch. Therefore, extracting both centerlines jointly and incorporating an extra seed point at the level of the pulmonary artery bifurcation as presented in [145] might result in no failures in centerline extraction.

Chapter 2 measured the aortic dimensions at fixed intervals with respect to a single anatomical landmark level, the level of the pulmonary artery bifurcation. In clinical practice, multiple standard anatomical landmarks as described in the ESC and ACCF/AHA guidelines [21, 76] are used instead for reporting aortic diameters. As a future perspective, it would therefore be interesting to design a method to automatically and robustly extract anatomical landmarks for aortic diameter measurements. However, consistently extracting these landmarks, especially on non-ECG-gated CT, is difficult. Although **Chapter 2** and **Chapter 4** proposed a landmark detection based on a multi-atlas registration technique, developing a deep learning-based technique for landmark detection may be a better approach. Since it is expected that a deep learning-based method is much faster than a multi-atlas registration technique and therefore, is more suitable for applying in clinical practice.

Chapter 2 also showed that the anatomical landmark at the level of pulmonary artery bifurcation, a standard landmark according to the ESC and ACCF/AHA guidelines, could be extracted reliably. Therefore **Chapter 3** proposed to measure the ascending and descending aorta diameters in this landmark on a larger cohort with both baseline and last year follow-up scans. Although we had enough information to find aortic growth rates with these two measured diameters, quantifying local changes to the surface of the entire thoracic aorta in a longitudinal study could provide more information such as the changes in the arch and the aortic root. Furthermore, longitudinal analysis of the vessel centerlines could provide information, such as the changes in the angle of the arch or the pulmonary artery bifurcation. This thesis only segmented the first, and last-year follow-up scans out of the five-year screening scans in DLCST. Longitudinally studying the surface of the aorta and pulmonary artery and their centerlines at all-time points simultaneously, not just at one landmark level, but locally at any point on the surface can provide more detailed in-time information. This can be an interesting topic to be considered in future studies.

Chapter 4 proposed a 3D volumetric diameter measurement technique and subsequently measured the PA:AA ratio. Although the robustness and repeatability of this technique are validated, the automatically measured PA:AA ratio is not validated as an imaging biomarker, for instance as a predictor metric for exacerbation in patients with COPD. With the segmentation and diameter measured methods presented in this thesis, the full 3D shape, volume, and diameters of the aorta and pulmonary artery can be considered to extract novel imaging biomarkers for COPD and cardiovascular

disease. This can be a good direction for future research.

In the non-contrast CT scans, the relatively high noise level and the motion artifacts caused by the non-ECG-gated data intensifies the difficulty of vessel segmentation. For improving the segmentation quality in such scans, one approach can be to denoise the scans and improve the image quality with techniques such as presented in [146–148], prior to the segmentation method. Another approach can be to reduce the aortic motion artifacts [149, 150]. In this thesis, noise and artifact reduction techniques are not considered, whereas it would be interesting to apply noise and motion reduction techniques as a pre-processing step and thereafter investigate their effect on the proposed segmentation methods.

A limitation of the studies proposed in this thesis is that the methods are evaluated on data from a single scan protocol (**Chapter 2** and **Chapter 4**) and a relatively healthy screening population(**Chapter 2**, **Chapter 4**, and **Chapter 5**). For application in data from different imaging and scanning protocols or in scans with abnormalities in the vessel shapes, parameters may need to be adjusted, or the network may need to be retrained. However, the optimal surface graph cut method is successfully applied, with identical parameter settings, on other data including CTA scans of patients with abnormal aortic shape due to Turner syndrome. A sample of this segmentation result is shown in the appendix of **Chapter 4**.

6.6 Conclusion

In conclusion, in this thesis, fully automatic segmentation and diameter measurement techniques are developed and validated to quantify the shape and size of the aorta and pulmonary artery in non-ECG-gated, non-contrast CT scans. The methods presented robust and reproducible results and are proven to be of sufficient accuracy and reliability for use in the clinical study. I hope this work will be continued in further studies, thereby contributing to future incorporation into clinical practice and facilitating clinical and epidemiological research, for the final goal of accurate diagnosis of silent disease in an early stage contributing to improved everyday healthcare.

Bibliography

[1] H. Ritchie and M. Roser, "Causes of death," *Our World in Data*, 2018.

[2] S. Roversi, L. M. Fabbri, D. D. Sin, N. M. Hawkins, and A. Agusti, "Chronic obstructive pulmonary disease and cardiac diseases. An urgent need for integrated care," *American journal of respiratory and critical care medicine*, vol. 194, no. 11, pp. 1319–1336, 2016.

[3] M. Ruparel, S. L. Quaife, J. L. Dickson, C. Horst, S. Burke, M. Taylor, A. Ahmed, P. Shaw, M.-J. Soo, A. Nair, *et al.*, "Evaluation of cardiovascular risk in a lung cancer screening cohort," *Thorax*, vol. 74, no. 12, pp. 1140–1146, 2019.

[4] S. André, B. Conde, E. Fragoso, J. Boléo-Tomé, V. Areias, J. Cardoso, *et al.*, "COPD and cardiovascular disease," *Pulmonology*, vol. 25, no. 3, pp. 168–176, 2019.

[5] P. Carter, J. Lagan, C. Fortune, D. L. Bhatt, J. Vestbo, R. Niven, N. Chaudhuri, E. B. Schelbert, R. Potluri, and C. A. Miller, "Association of cardiovascular disease with respiratory disease," *Journal of the American College of Cardiology*, vol. 73, no. 17, pp. 2166–2177, 2019.

[6] J. H. Pedersen, H. Ashraf, A. Dirksen, K. Bach, H. Hansen, P. Toennesen, H. Thorsen, J. Brodersen, B. G. Skov, M. Døssing, *et al.*, "The Danish randomized lung cancer CT screening trial—overall design and results of the prevalence round," *Journal of Thoracic Oncology*, vol. 4, no. 5, pp. 608–614, 2009.

[7] W. W. Labaki, C. H. Martinez, F. J. Martinez, C. J. Galbán, B. D. Ross, G. R. Washko, R. G. Barr, E. A. Regan, H. O. Coxson, E. A. Hoffman, *et al.*, "The role of chest computed tomography in the evaluation and management of the patient with chronic obstructive pulmonary disease," *American journal of respiratory and critical care medicine*, vol. 196, no. 11, pp. 1372–1379, 2017.

[8] O. Mets, P. De Jong, B. Van Ginneken, H. Gietema, and J. Lammers, "Quantitative computed tomography in COPD: Possibilities and limitations," *Lung*, vol. 190, no. 2, pp. 133–145, 2012.

[9] J. M. Wells, G. R. Washko, M. K. Han, N. Abbas, H. Nath, A. J. Mamary, E. Regan, W. C. Bailey, F. J. Martinez, E. Westfall, *et al.*, "Pulmonary arterial enlargement and acute exacerbations of COPD," *New England Journal of Medicine*, vol. 367, no. 10, pp. 913–921, 2012.

[10] O. M. Mets, C. F. Buckens, P. Zanen, I. Isgum, B. van Ginneken, M. Prokop, H. A. Gietema, J.-W. J. Lammers, R. Vliegenthart, M. Oudkerk, *et al.*, "Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans," *Jama*, vol. 306, no. 16, pp. 1775–1781, 2011.

[11] H. O. Coxson, J. Mayo, S. Lam, G. Santyr, G. Parraga, and D. D. Sin, "New and current clinical imaging techniques to study chronic obstructive pulmonary disease," *American journal of respiratory and critical care medicine*, vol. 180, no. 7, pp. 588–597, 2009.

[12] O. M. Mets, P. A. de Jong, and M. Prokop, "Computed tomographic screening for lung cancer: An opportunity to evaluate other diseases," *JAMA*, vol. 308, no. 14, pp. 1433–1434, 2012.

[13] D. Singhvi and J. Bon, "Computed Tomography Imaging and Comorbidities in Chronic Obstructive Pulmonary Disease: Beyond Lung Cancer Screening," *Chest*, 2020.

[14] M. A. Heuvelmans, M. Vonder, M. Rook, H. J. Groen, G. H. De Bock, X. Xie, M. J. Ijzerman, R. Vliegenthart, and M. Oudkerk, "Screening for early lung cancer, chronic obstructive pulmonary disease, and cardiovascular disease (the Big-3) using low-dose chest computed tomography," *Journal of thoracic imaging*, vol. 34, no. 3, pp. 160–169, 2019.

[15] Z. Li, Y. Xia, Y. Fang, Y. Guan, Y. Wang, S. Liu, and L. Fan, "The importance of CT quantitative evaluation of emphysema in lung cancer screening cohort with negative findings by visual evaluation," *The clinical respiratory journal*, vol. 13, no. 12, pp. 741–750, 2019.

[16] N. L. S. T. R. Team, "The national lung screening trial: Overview and study design," *Radiology*, vol. 258, no. 1, pp. 243–253, 2011.

[17] Y. J. Suh, J. W. Lee, S. Y. Shin, J. M. Goo, Y. Kim, and H. S. Yong, "Coronary artery calcium severity grading on non-ECG-gated low-dose chest computed tomography: A multiple-observer study in a nationwide lung cancer screening registry," *European radiology*, pp. 1–8, 2020.

[18] M. Oudkerk, A. Devaraj, R. Vliegenthart, T. Henzler, H. Prosch, C. P. Heussel, G. Bastarrika, N. Sverzellati, M. Mascalchi, S. Delorme, *et al.*, "European position statement on lung cancer screening," *The Lancet Oncology*, vol. 18, no. 12, e754–e766, 2017.

[19] T. E. Raymond, J. E. Khabbaza, R. Yadav, and A. R. Tonelli, "Significance of main pulmonary artery dilation on imaging studies," *Annals of the American Thoracic Society*, vol. 11, no. 10, pp. 1623–1632, 2014.

[20] Q. A. Truong, H. S. Bhatia, J. Szymonifka, Q. Zhou, Z. Lavender, A. B. Waxman, M. J. Semigran, and R. Malhotra, "A four-tier classification system of pulmonary artery metrics on computed tomography for the diagnosis and prognosis of pulmonary hypertension," *Journal of cardiovascular computed tomography*, vol. 12, no. 1, pp. 60–66, 2018.

[21] A. C. of Cardiology Foundation, A. H. A. T. F. on Practice Guidelines, A. A. for Thoracic Surgery, A. C. of Radiology, A. S. Association, S. of Cardiovascular Anesthesiologists, S. for Cardiovascular Angiography, Interventions, S. of Interventional Radiology, S. of Thoracic Surgeons, S. for Vascular Medicine, *et al.*, "2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with thoracic aortic disease," *Journal of the American College of Cardiology*, vol. 55, no. 14, e27–e129, 2010.

[22] A. Wolak, H. Gransar, L. E. Thomson, J. D. Friedman, R. Hachamovitch, A. Gutstein, L. J. Shaw, D. Polk, N. D. Wong, R. Saouaf, *et al.*, "Aortic size assessment by noncontrast cardiac computed tomography: Normal limits by age, gender, and body surface area," *JACC: Cardiovascular Imaging*, vol. 1, no. 2, pp. 200–209, 2008.

[23] E. L. Frandsen, L. J. Burchill, A. M. Khan, and C. S. Broberg, "Ascending aortic size in aortic coarctation depends on aortic valve morphology: Understanding the bicuspid valve phenotype," *International journal of cardiology*, vol. 250, pp. 106–109, 2018.

[24] Y. von Kodolitsch, M. A. Aydin, D. H. Koschyk, R. Loose, I. Schalwat, M. Karck, J. Cremer, A. Haverich, J. Berger, T. Meinertz, *et al.*, "Predictors of aneurysmal formation after surgical correction of aortic coarctation," *Journal of the American College of Cardiology*, vol. 39, no. 4, pp. 617–624, 2002.

[25] J. Y. Rho, D. A. Lynch, Y. J. Suh, J. W. Nah, J. A. Zach, J. D. Schroeder, C. W. Cox, R. P. Bowler, B. E. Fenster, M. T. Dransfield, *et al.*, "CT measurements of central pulmonary vasculature as predictors of severe exacerbation in COPD," *Medicine*, vol. 97, no. 3, 2018.

[26] N. Terzikhan, D. Bos, L. Lahousse, L. Wolff, K. M. Verhamme, M. J. Leening, J. F. Felix, H. Gall, H. A. Ghofrani, O. H. Franco, *et al.*, "Pulmonary artery to aorta ratio and risk of all-cause mortality in the general population: The Rotterdam Study," *European Respiratory Journal*, vol. 49, no. 6, 2017.

[27] K. F. Rabe, J. R. Hurst, and S. Suissa, "Cardiovascular disease and COPD: Dangerous liaisons?" *European Respiratory Review*, vol. 27, no. 149, 2018.

[28] R. R. Davies, A. Gallo, M. A. Coady, G. Tellides, D. M. Botta, B. Burke, M. P. Coe, G. S. Kopf, and J. A. Elefteriades, "Novel measurement of relative aortic size predicts rupture of thoracic aortic aneurysms," *The Annals of thoracic surgery*, vol. 81, no. 1, pp. 169–177, 2006.

[29] J. B. Kim, K. Kim, M. E. Lindsay, T. MacGillivray, E. M. Isselbacher, R. P. Cambria, and T. M. Sundt III, "Risk of rupture or dissection in descending thoracic aortic aneurysm," *Circulation*, vol. 132, no. 17, pp. 1620–1629, 2015.

[30] I. H. Melvinsdottir, S. H. Lund, B. A. Agnarsson, K. Sigvaldason, T. Gudbjartsson, and A. Geirsson, "The incidence and mortality of acute thoracic aortic dissection: Results from a whole nation study," *European Journal of Cardio-Thoracic Surgery*, vol. 50, no. 6, pp. 1111–1117, 2016.

[31] A. Ruff, K. Patel, J. R. Joyce, H. L. Gornik, and M. B. Rothberg, "The use of pre-existing CT imaging in screening for abdominal aortic aneurysms," *Vascular Medicine*, vol. 21, no. 6, pp. 515–519, 2016.

[32] H. Teimouri, F. Sabzi, and S. Dabiri, "Congenital saccular aneurysm of coarctation of aorta: A case report," *The Journal of Tehran University Heart Center*, vol. 8, no. 4, p. 210, 2013.

[33] L. R. Cangussú, M. R. Lopes, and R. H. d. A. Barbosa, "The importance of the early diagnosis of aorta coarctation," *Revista da Associação Médica Brasileira*, vol. 65, no. 2, pp. 240–245, 2019.

[34] A. Chaouat, R. Naeije, and E. Weitzenblum, "Pulmonary hypertension in COPD," *European Respiratory Journal*, vol. 32, no. 5, pp. 1371–1385, 2008.

[35] F. Aluja Jaramillo, F. R. Gutierrez, F. G. Díaz Telli, S. Yevenes Aravena, C. Javidan-Nejad, and S. Bhalla, "Approach to pulmonary hypertension: From CT to clinical diagnosis," *Radiographics*, vol. 38, no. 2, pp. 357–373, 2018.

[36] M. Gupta, A. Agrawal, A. Iakovou, S. Cohen, R. Shah, and A. Talwar, "Pulmonary artery aneurysm: A review," *Pulmonary circulation*, vol. 10, no. 1, p. 2 045 894 020 908 780, 2020.

[37] H. S. Park, M. R. Chamarthy, D. Lamus, S. S. Saboo, P. D. Sutphin, and S. P. Kalva, "Pulmonary artery aneurysms: Diagnosis & endovascular therapy," *Cardiovascular diagnosis and therapy*, vol. 8, no. 3, p. 350, 2018.

[38] T. I. Murray, L. M. Boxt, J. Katz, K. Reagan, and R. J. Barst, "Estimation of pulmonary artery pressure in patients with primary pulmonary hypertension by quantitative analysis of magnetic resonance images.," *Journal of thoracic imaging*, vol. 9, no. 3, pp. 198–204, 1994.

[39] S. P. Raman, M. Mahesh, R. V. Blasko, and E. K. Fishman, "CT scan parameters and radiation dose: Practical advice for radiologists," *Journal of the American College of Radiology*, vol. 10, no. 11, pp. 840–846, 2013.

[40] S. P. Raman, P. T. Johnson, S. Deshmukh, M. Mahesh, K. L. Grant, and E. K. Fishman, "CT dose reduction applications: Available tools on the latest generation of CT scanners," *Journal of the American College of Radiology*, vol. 10, no. 1, pp. 37–41, 2013.

[41] H. Berhane, M. Scott, M. Elbaz, K. Jarvis, P. McCarthy, J. Carr, C. Malaisrie, R. Avery, A. J. Barker, J. D. Robinson, *et al.*, "Fully automated 3D aortic segmentation of 4D flow MRI for hemodynamic analysis using deep learning," *Magnetic resonance in medicine*, vol. 84, no. 4, pp. 2204–2218, 2020.

[42] O. Ecabert, J. Peters, M. J. Walker, T. Ivanc, C. Lorenz, J. von Berg, J. Lessick, M. Vembar, and J. Weese, "Segmentation of the heart and great vessels in CT images using a model-based adaptation framework," *Medical image analysis*, vol. 15, no. 6, pp. 863–876, 2011.

[43] A. Biesdorf, K. Rohr, D. Feng, H. von Tengg-Kobligk, F. Rengier, D. Böckler, H.-U. Kauczor, and S. Wörz, "Segmentation and quantification of the aortic arch using joint 3D model-based segmentation and elastic image registration," *Medical image analysis*, vol. 16, no. 6, pp. 1187–1201, 2012.

[44] P. Entezari, A. R. Honarmand, M. S. Galizia, Y. Yang, J. Collins, V. Yaghmai, J. C. Carr, *et al.*, "Analysis of the thoracic aorta using a semi-automated post processing tool," *European journal of radiology*, vol. 82, no. 9, pp. 1558–1564, 2013.

[45] M. A. Elattar, E. Wiegerinck, R. Planken, H. van Assen, J. Baan, H. Marquering, *et al.*, "Automatic segmentation of the aortic root in CT angiography of candidate patients for transcatheter aortic valve implantation," *Medical & biological engineering & computing*, vol. 52, no. 7, pp. 611–618, 2014.

[46] X. Gao, P. H. Kitslaar, R. P. Budde, S. Tu, M. A. de Graaf, L. Xu, B. Xu, A. J. Scholte, J. Dijkstra, and J. H. Reiber, "Automatic detection of aortofemoral vessel trajectory from whole-body computed tomography angiography data sets," *The international journal of cardiovascular imaging*, vol. 32, no. 8, pp. 1311–1322, 2016.

[47] X. Gao, S. Boccalini, P. H. Kitslaar, R. P. Budde, M. Attrach, S. Tu, M. A. de Graaf, T. Ondrus, M. Penicka, A. J. Scholte, *et al.*, "Quantification of aortic annulus in computed tomography angiography: Validation of a fully automatic methodology," *European journal of radiology*, vol. 93, pp. 1–8, 2017.

[48] L. Cao, R. Shi, Y. Ge, L. Xing, P. Zuo, Y. Jia, J. Liu, Y. He, X. Wang, S. Luan, *et al.*, "Fully automatic segmentation of type B aortic dissection from CTA images enabled by deep learning," *European journal of radiology*, vol. 121, p. 108 713, 2019.

[49] M. G. Linguraru, J. A. Pura, R. L. Van Uitert, N. Mukherjee, R. M. Summers, C. Minniti, M. T. Gladwin, G. Kato, R. F. Machado, and B. J. Wood, "Segmentation and quantification of pulmonary artery for noninvasive CT assessment of sickle cell secondary pulmonary hypertension," *Medical physics*, vol. 37, no. 4, pp. 1522–1532, 2010.

[50] D. Moses, C. Sammut, and T. Zrimec, "Automatic segmentation and analysis of the main pulmonary artery on standard post-contrast CT studies using iterative erosion and dilation," *International journal of computer assisted radiology and surgery*, vol. 11, no. 3, pp. 381–395, 2016.

[51] K. L.-L. Román, I. de La Bruere, J. Onieva, L. Andresen, J. Q. Holsting, F. N. Rahaghi, I. Macía, M. A. G. Ballester, and R. S. J. Estepar, "3D pulmonary artery segmentation from CTA scans using deep learning with realistic data augmentation," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, Springer, 2018, pp. 225–237.

[52] L. Baskaran, S. J. Al'Aref, G. Maliakal, B. C. Lee, Z. Xu, J. W. Choi, S.-E. Lee, J. M. Sung, F. Y. Lin, S. Dunham, *et al.*, "Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning," *PloS one*, vol. 15, no. 5, e0232573, 2020.

[53] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. Van Ginneken, "Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans," *IEEE transactions on medical imaging*, vol. 28, no. 7, pp. 1000–1010, 2009.

[54] O. C. Avila-Montes, U. Kurkure, R. Nakazato, D. S. Berman, D. Dey, and I. A. Kakadiaris, "Segmentation of the thoracic aorta in noncontrast cardiac CT images," *IEEE journal of biomedical and health informatics*, vol. 17, no. 5, pp. 936–949, 2013.

[55] A. Dasgupta, S. Mukhopadhyay, S. A. Mehre, and P. Bhattacharyya, "Morphological geodesic active contour based automatic aorta segmentation in thoracic CT images," in *Proceedings of International Conference on Computer Vision and Image Processing*, Springer, 2017, pp. 187–195.

[56] P. G. Tahoces, L. Alvarez, E. González, C. Cuenca, A. Trujillo, D. Santana-Cedrés, J. Esclarín, L. Gomez, L. Mazorra, M. Alemán-Flores, *et al.*, "Automatic estimation of the aortic lumen geometry by ellipse tracking," *International journal of computer assisted radiology and surgery*, vol. 14, no. 2, pp. 345–355, 2019.

[57] T. Kitasaka, K. Mori, J.-i. Hasegawa, J.-i. Toriwaki, and K. Katada, "Automated extraction of aorta and pulmonary artery in mediastinum from 3D chest X-ray ct images without contrast medium," in *Medical Imaging 2002: Image Processing*, International Society for Optics and Photonics, vol. 4684, 2002, pp. 1496–1507.

[58] M. Feuerstein, T. Kitasaka, and K. Mori, "Automated anatomical likelihood driven extraction and branching detection of aortic arch in 3-D chest CT," in *Second international workshop on pulmonary image analysis*, Med. Image Comput. Comput. Assist. Interv.(MICCAI), 2009, pp. 49–60.

[59] Y. Xie, J. Padgett, A. M. Biancardi, and A. P. Reeves, "Automated aorta segmentation in low-dose chest CT images," *International journal of computer assisted radiology and surgery*, vol. 9, no. 2, pp. 211–219, 2014.

[60] S. Kurugol, C. E. Come, A. A. Diaz, J. C. Ross, G. L. Kinney, J. L. Black-Shinn, J. E. Hokanson, M. J. Budoff, G. R. Washko, and R. San Jose Estepar, "Automated quantitative 3D analysis of aorta size, morphology, and mural calcification distributions," *Medical physics*, vol. 42, no. 9, pp. 5467–5478, 2015.

[61] J. M. Noothout, B. D. De Vos, J. M. Wolterink, and I. Išgum, "Automatic segmentation of thoracic aorta segments in low-dose chest CT," in *Medical Imaging 2018: Image Processing*, International Society for Optics and Photonics, vol. 10574, 2018, 105741S.

[62] R. Trullo, C. Petitjean, D. Nie, D. Shen, and S. Ruan, "Joint segmentation of multiple thoracic organs in CT images with two collaborative deep architectures," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 21–29.

[63] T. He, J. Hu, Y. Song, J. Guo, and Z. Yi, "Multi-task learning for the segmentation of organs at risk with label dependence," *Medical image analysis*, vol. 61, p. 101 666, 2020.

[64] M. Feuerstein, T. Kitasaka, and K. Mori, "Adaptive model based pulmonary artery segmentation in 3D chest CT," in *Medical Imaging 2010: Image Processing*, International Society for Optics and Photonics, vol. 7623, 2010, 76234S.

[65] Y. Xie, M. Liang, D. F. Yankelevitz, C. I. Henschke, and A. P. Reeves, "Automated measurement of pulmonary artery in low-dose non-contrast chest CT images," in *Medical Imaging 2015: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 9414, 2015, 94141G.

[66] S. Bruns, J. M. Wolterink, R. W. van Hamersvelt, T. Leiner, and I. Išgum, "CNN-Based segmentation of the cardiac chambers and great vessels in non-contrast-enhanced cardiac CT," *arXiv preprint arXiv:1908.07727*, 2019.

[67] X. Deng, Y. Zheng, Y. Xu, X. Xi, N. Li, and Y. Yin, "Graph cut based automatic aorta segmentation with an adaptive smoothness constraint in 3D abdominal CT images," *Neurocomputing*, vol. 310, pp. 46–58, 2018.

[68] A. M. Arias-Lorza, J. Petersen, A. van Engelen, M. Selwaness, A. van der Lugt, W. J. Niessen, and M. de Bruijne, "Carotid artery wall segmentation in multispectral MRI by coupled optimal surface graph cuts," *IEEE transactions on medical imaging*, vol. 35, no. 3, pp. 901–911, 2015.

[69] J. Petersen, M. Nielsen, P. Lo, L. H. Nordenmark, J. H. Pedersen, M. M. W. Wille, A. Dirksen, and M. de Bruijne, "Optimal surface segmentation using flow lines to quantify airway abnormalities in chronic obstructive pulmonary disease," *Medical image analysis*, vol. 18, no. 3, pp. 531–541, 2014.

[70] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, "Deep learning for cardiac image segmentation: A review," *Frontiers in Cardiovascular Medicine*, vol. 7, p. 25, 2020.

[71] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images," *Medical image analysis*, vol. 41, pp. 40–54, 2017.

[72] A. Fantazzini, M. Esposito, A. Finotello, F. Auricchio, B. Pane, C. Basso, G. Spinella, and M. Conti, "3D automatic segmentation of aortic computed tomography angiography combining multi-view 2D convolutional neural networks," *Cardiovascular engineering and technology*, vol. 11, no. 5, pp. 576–586, 2020.

[73] G. A. Roth, M. D. Huffman, A. E. Moran, V. Feigin, G. A. Mensah, M. Naghavi, and C. J. Murray, "Global and regional patterns in cardiovascular mortality from 1990 to 2013," *Circulation*, vol. 132, no. 17, pp. 1667–1678, 2015.

[74] Y. Itani, S. Watanabe, Y. Masuda, K. Hanamura, K. Asakura, S. Sone, Y. Sunami, and T. Miyamoto, "Measurement of aortic diameters and detection of asymptomatic aortic aneurysms in a mass screening program using a mobile helical computed tomography unit," *Heart and vessels*, vol. 16, no. 2, pp. 42–45, 2002.

[75] H. Kälsch, N. Lehmann, S. Möhlenkamp, A. Becker, S. Moebus, A. Schmermund, A. Stang, A. A. Mahabadi, K. Mann, K.-H. Jöckel, *et al.*, "Body-surface adjusted aortic reference diameters for improved identification of patients with thoracic aortic aneurysms: Results from the population-based Heinz Nixdorf Recall study," *International journal of cardiology*, vol. 163, no. 1, pp. 72–78, 2013.

[76] A. F. members, R. Erbel, V. Aboyans, C. Boileau, E. Bossone, R. D. Bartolomeo, H. Eggebrecht, A. Evangelista, V. Falk, H. Frank, G. Oliver, G. Martin, H. Axel, I. Bernard, M. Athanasios John, M. Folkert, N. Christoph A, R. Marco, R. Hervé, S. Udo, S. Per Anton, and v. A. Regula S, "2014 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)," *European heart journal*, vol. 35, no. 41, pp. 2873–2926, 2014.

[77] M. M. Wille, A. Dirksen, H. Ashraf, Z. Saghir, K. S. Bach, J. Brodersen, P. F. Clementsen, H. Hansen, K. R. Larsen, J. Mortensen, *et al.*, "Results of the randomized Danish lung cancer screening trial with focus on high-risk profiling," *American journal of respiratory and critical care medicine*, vol. 193, no. 5, pp. 542–551, 2016.

[78] K. Hameeteman, M. A. Zuluaga, M. Freiman, L. Joskowicz, O. Cuisenaire, L. F. Valencia, M. A. Gülsün, K. Krissian, J. Mille, W. C. Wong, *et al.*, "Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading," *Medical image analysis*, vol. 15, no. 4, pp. 477–488, 2011.

[79] F. Heckel, O. Konrad, H. K. Hahn, and H.-O. Peitgen, "Interactive 3D medical image segmentation with energy-minimizing implicit functions," *Computers & Graphics*, vol. 35, no. 2, pp. 275–287, 2011.

[80] Z. Sedghi Gamechi, A. M. Arias-Lorza, J. H. Pedersen, and M. de Bruijne, "Aorta and pulmonary artery segmentation using optimal surface graph cuts in non-contrast CT," in *SPIE*, vol. 10574, 2018, p. 105742D.

[81] H. Kirişli, M. Schaap, S. Klein, S.-L. Papadopoulou, M. Bonardi, C.-H. Chen, A. C. Weustink, N. R. Mollet, E.-J. Vonken, R. J. van der Geest, *et al.*, "Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: A large-scale, multicenter, and multivendor study," *Medical physics*, vol. 37, no. 12, pp. 6279–6291, 2010.

[82] H. Tang, T. van Walsum, R. S. van Onkelen, R. Hameeteman, S. Klein, M. Schaap, F. L. Tori, Q. J. van den Bouwhuijsen, J. C. Witteman, A. van der Lugt, *et al.*, "Semiautomatic carotid lumen segmentation for quantification of lumen geometry in multispectral MRI," *Medical image analysis*, vol. 16, no. 6, pp. 1202–1215, 2012.

[83] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[84] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol. 15, no. 2, pp. 155–163, 2016.

[85] O. Vriz, C. Driussi, M. Bettio, F. Ferrara, A. D'Andrea, and E. Bossone, "Aortic root dimensions and stiffness in healthy subjects," *The American journal of cardiology*, vol. 112, no. 8, pp. 1224–1229, 2013.

[86] L. E. Quint, P. S. Liu, A. M. Booher, K. Watcharotone, and J. D. Myles, "Proximal thoracic aortic diameter measurements at CT: Repeatability and reproducibility according to measurement method," *The international journal of cardiovascular imaging*, vol. 29, no. 2, pp. 479–488, 2013.

[87] R. R. Davies, L. J. Goldstein, M. A. Coady, S. L. Tittle, J. A. Rizzo, G. S. Kopf, and J. A. Elefteriades, "Yearly rupture or dissection rates for thoracic aortic aneurysms: Simple prediction based on size," *The Annals of thoracic surgery*, vol. 73, no. 1, pp. 17–28, 2002.

[88] A. Hager, H. Kaemmerer, U. Rapp-Bernhardt, S. Blücher, K. Rapp, T. M. Bernhardt, M. Galanski, and J. Hess, "Diameters of the thoracic aorta throughout life as measured with helical computed tomography," *The Journal of thoracic and cardiovascular surgery*, vol. 123, no. 6, pp. 1060–1066, 2002.

[89] S. S. Mao, N. Ahmadi, B. Shah, D. Beckmann, A. Chen, L. Ngo, F. R. Flores, Y. lin Gao, and M. J. Budoff, "Normal thoracic aorta diameter on cardiac computed tomography in healthy asymptomatic adults: Impact of age and gender," *Academic radiology*, vol. 15, no. 7, pp. 827–834, 2008.

[90] F. Y. Lin, R. B. Devereux, M. J. Roman, J. Meng, V. M. Jow, A. Jacobs, J. W. Weinsaft, L. J. Shaw, D. S. Berman, A. Gilmore, *et al.*, "Assessment of the thoracic aorta by multidetector computed tomography: Age-and sex-specific reference values in adults without evident cardiovascular disease," *Journal of cardiovascular computed tomography*, vol. 2, no. 5, pp. 298–308, 2008.

[91] T. Rasmussen, L. Køber, J. H. Pedersen, A. Dirksen, L. H. Thomsen, S. Stender, J. Brodersen, J. Groen, H. Ashraf, and K. F. Kofoed, "Relationship between chronic obstructive pulmonary disease and subclinical coronary artery disease in long-term smokers," *European Heart Journal–Cardiovascular Imaging*, vol. 14, no. 12, pp. 1159–1166, 2013.

[92] S. A. Goldstein, A. Evangelista, S. Abbara, A. Arai, F. M. Asch, L. P. Badano, M. A. Bolen, H. M. Connolly, H. Cuéllar-Calàbria, M. Czerny, R. B. Devereux, R. A. Erbel, R. Fattori, E. M. Isselbacher, J. M. Lindsay, M. McCulloch, H. I. Michelena, C. A. Nienaber, J. K. Oh, M. Pepi, A. J. Taylor, J. W. Weinsaft, J. L. Zamorano, H. Dietz, K. Eagle, J. Elefteriades, G. Jondeau, H. Rousseau, and M. Schepens, "Multimodality imaging of diseases of the thoracic aorta in adults: From the american society of echocardiography and the european association of cardiovascular imaging: Endorsed by the society of cardiovascular computed tomography and society for cardiovascular magnetic resonance," *Journal of the American Society of Echocardiography*, vol. 28, no. 2, pp. 119–182, 2015.

[93] D. Detaint, H. I. Michelena, V. T. Nkomo, A. Vahanian, G. Jondeau, and M. E. Sarano, "Aortic dilatation patterns and rates in adults with bicuspid aortic valves: A comparative study with marfan syndrome and degenerative aortopathy," *Heart*, vol. 100, no. 2, pp. 126–134, 2014.

[94] Y. Agmon, B. K. Khandheria, I. Meissner, G. L. Schwartz, J. D. Sicks, A. J. Fought, W. M. O'Fallon, D. O. Wiebers, and A. J. Tajik, "Is aortic dilatation an atherosclerosis-related process? Clinical, laboratory, and transesophageal echocardiographic correlates of thoracic aortic dimensions in the population with implications for thoracic aortic aneurysm formation," *Journal of the American College of Cardiology*, vol. 42, no. 6, pp. 1076–1083, 2003.

[95] B. Mensel, L. Heßelbarth, M. Wenzel, J.-P. Kühn, M. Dörr, H. Völzke, W. Lieb, K. Hegenscheid, and R. Lorbeer, "Thoracic and abdominal aortic diameters in a general population: MRI-based reference values and association with age and cardiovascular risk factors," *European radiology*, vol. 26, no. 4, pp. 969–978, 2016.

[96] A. R. Brady, S. G. Thompson, F. G. R. Fowkes, R. M. Greenhalgh, and J. T. Powell, "Abdominal aortic aneurysm expansion: Risk factors and time intervals for surveillance," *Circulation*, vol. 110, no. 1, pp. 16–21, 2004.

[97] Z. Sedghi Gamechi, L. R. Bons, M. Giordano, D. Bos, R. P. Budde, K. F. Kofoed, J. H. Pedersen, J. W. Roos-Hesselink, and M. de Bruijne, "Automated 3D segmentation and diameter measurement of the thoracic aorta on non-contrast enhanced CT," *European radiology*, vol. 29, no. 9, pp. 4613–4623, 2019.

[98] C. S. Lam, V. Xanthakis, L. M. Sullivan, W. Lieb, J. Aragam, M. M. Redfield, G. F. Mitchell, E. J. Benjamin, and R. S. Vasan, "Aortic root remodeling over the adult life course: Longitudinal data from the framingham heart study," *Circulation*, vol. 122, no. 9, pp. 884–890, 2010.

[99] J. F. Rodriguez-Palomares, G. Teixido-Tura, V. Galuppo, H. Cuellar, A. Laynez, L. Gutierrez, M. T. González-Alujas, D. Garcia-Dorado, and A. Evangelista, "Multimodality assessment of ascending aortic diameters: Comparison of different measurement methods," *Journal of the American Society of Echocardiography*, vol. 29, no. 9, pp. 819–826, 2016.

[100] L. R. Bons, A. L. Duijnhouwer, S. Boccalini, A. T. van den Hoven, M. J. van der Vlugt, R. G. Chelu, J. S. McGhie, I. Kardys, A. E. Van Den Bosch, H.-M. J. Siebelink, K. Nieman, A. Hirsch, C. S. Broberg, R. P. J. Budde, and J. W. Roos-Hesselink, "Intermodality variation of aortic dimensions: How, where and when to measure the ascending aorta," *International journal of cardiology*, vol. 276, pp. 230–235, 2019.

[101] M. A. Zafar, Y. Li, J. A. Rizzo, P. Charilaou, A. Saeyeldin, C. A. Velasquez, A. M. Mansour, S. U. B. Mahmood, W.-G. Ma, A. J. Brownstein, M. Tranquilli, J. Dumfarth, P. Theodoropoulos, K. Thombre, M. Tanweer, Y. Erben, S. Peterss, B. A. Ziganshin, and J. A. Elefteriades, "Height alone, rather than body surface area, suffices for risk estimation in ascending aortic aneurysm," *The Journal of thoracic and cardiovascular surgery*, vol. 155, no. 5, pp. 1938–1950, 2018.

[102] C. H. Gravholt, N. H. Andersen, G. S. Conway, O. M. Dekkers, M. E. Geffner, K. O. Klein, A. E. Lin, N. Mauras, C. A. Quigley, K. Rubin, D. E. Sandberg, T. C. J. Sas, M. Silberbach, V. Söderström-Anttila, K. Stochholm, J. A. van Alfen-van derVelden, J. Woelfle, and P. F. Backeljauw, "Clinical practice guidelines for the care of girls and women with Turner syndrome: Proceedings

from the 2016 Cincinnati International Turner Syndrome Meeting," *European journal of endocrinology*, vol. 177, no. 3, G1–G70, 2017.

[103] A. Redheuil, W.-C. Yu, C. O. Wu, E. Mousseaux, A. De Cesare, R. Yan, N. Kachenoura, D. Bluemke, and J. A. Lima, "Reduced ascending aortic strain and distensibility: Earliest manifestations of vascular aging in humans," *Hypertension*, vol. 55, no. 2, pp. 319–326, 2010.

[104] D. Craiem, G. Chironi, A. Redheuil, M. Casciaro, E. Mousseaux, A. Simon, and R. L. Armentano, "Aging impact on thoracic aorta 3D morphometry in intermediate-risk subjects: Looking beyond coronary arteries with non-contrast cardiac CT," *Annals of biomedical engineering*, vol. 40, no. 5, pp. 1028–1038, 2012.

[105] D. Oladokun, B. Patterson, J. Sobocinski, A. Karthikesalingam, I. Loftus, M. Thompson, and P. Holt, "Systematic review of the growth rates and influencing factors in thoracic aortic aneurysms," *European Journal of Vascular and Endovascular Surgery*, vol. 51, no. 5, pp. 674–681, 2016.

[106] J. E. Roos, J. K. Willmann, D. Weishaupt, M. Lachat, B. Marincek, and P. R. Hilfiker, "Thoracic aorta: Motion artifact reduction with retrospective and prospective electrocardiography-assisted multi–detector row CT," *Radiology*, vol. 222, no. 1, pp. 271–277, 2002.

[107] S. Dou, C. Zheng, X. Ji, W. Wang, M. Xie, L. Cui, and W. Xiao, "Co-existence of COPD and bronchiectasis: A risk factor for a high ratio of main pulmonary artery to aorta diameter (PA: A) from computed tomography in COPD patients," *International journal of chronic obstructive pulmonary disease*, vol. 13, p. 675, 2018.

[108] Z. Sedghi Gamechi, A. M. Arias-Lorza, J. H. Pedersen, and M. de Bruijne, "Aorta and pulmonary artery segmentation using optimal surface graph cuts in non-contrast CT," in *Medical Imaging 2018: Image Processing*, International Society for Optics and Photonics, vol. 10574, 2018, p. 105742D.

[109] P. Lo, J. Sporring, H. Ashraf, J. J. Pedersen, and M. de Bruijne, "Vessel-guided airway tree segmentation: A voxel classification approach," *Medical image analysis*, vol. 14, no. 4, pp. 527–538, 2010.

[110] M. A. Gülsün and H. Tek, "Robust vessel tree modeling," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2008, pp. 602–611.

[111] U. Bozlar, F. Ors, O. Deniz, M. Uzun, S. Gumus, M. Ugurel, F. Yazar, and C. Tayfun, "Pulmonary artery diameters measured by multidetector-row computed tomography in healthy adults," *Acta Radiologica*, vol. 48, no. 10, pp. 1086–1091, 2007.

[112] A. R. Tonelli, S. Johnson, L. Alkukhun, R. Yadav, and R. A. Dweik, "Changes in main pulmonary artery diameter during follow-up have prognostic implications in pulmonary arterial hypertension," *Respirology*, vol. 22, no. 8, pp. 1649–1655, 2017.

[113] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[114] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[115] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[116] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[117] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 424–432.

[118] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[119] L. Yu, X. Yang, J. Qin, and P.-A. Heng, "3D FractalNet: Dense volumetric segmentation for cardiovascular MRI volumes," in *Reconstruction, segmentation, and analysis of medical images*, Springer, 2016, pp. 103–110.

[120] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[121] H. J. Kuijf, J. M. Biesbroek, J. De Bresser, R. Heinen, S. Andermatt, M. Bento, M. Berseth, M. Belyaev, M. J. Cardoso, A. Casamitjana, *et al.*, "Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge," *IEEE transactions on medical imaging*, vol. 38, no. 11, pp. 2556–2568, 2019.

[122] O. Maier, M. Wilms, J. von der Gablentz, U. M. Krämer, T. F. Münte, and H. Handels, "Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences," *Journal of neuroscience methods*, vol. 240, pp. 89–100, 2015.

[123] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with gaussian edge potentials," *arXiv preprint arXiv:1210.5644*, 2012.

[124] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[125] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1529–1537.

[126] A. G. Schwing and R. Urtasun, "Fully connected deep structured networks," *arXiv preprint arXiv:1503.02351*, 2015.

[127] S. Chen and M. de Bruijne, "An end-to-end approach to semantic segmentation with 3D CNN and Posterior-CRF in medical images," *arXiv preprint arXiv:1811.03549*, 2018.

[128] R. Vemulapalli, O. Tuzel, M.-Y. Liu, and R. Chellapa, "Gaussian conditional random field network for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3224–3233.

[129] G. Lin, C. Shen, A. Van Den Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3194–3203.

[130] Y. Li and W. Ping, "Cancer metastasis detection with neural conditional random field," *arXiv preprint arXiv:1806.07064*, 2018.

[131] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

[132] R. Selvan, M. Welling, J. H. Pedersen, J. Petersen, and M. de Bruijne, "Mean field network based graph refinement with application to airway tree extraction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 750–758.

[133] Z. Wang, N. Zou, D. Shen, and S. Ji, "Non-local U-nets for biomedical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 6315–6322.

[134] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[135] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[136] H. Yuan, N. Zou, S. Zhang, H. Peng, and S. Ji, "Learning hierarchical and shared features for improving 3D neuron reconstruction," in *2019 IEEE International Conference on Data Mining (ICDM)*, IEEE, 2019, pp. 806–815.

[137] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[138] M. Monteiro, M. A. Figueiredo, and A. L. Oliveira, "Conditional random fields as recurrent neural networks for 3D medical imaging segmentation," *arXiv preprint arXiv:1807.07464*, 2018.

[139] A. Adams, J. Baek, and M. A. Davis, "Fast high-dimensional filtering using the permutohedral lattice," in *Computer Graphics Forum*, Wiley Online Library, vol. 29, 2010, pp. 753–762.

[140] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, *et al.*, "ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI," *Medical image analysis*, vol. 35, pp. 250–269, 2017.

[141] Y. Li and R. Zemel, "Mean-field networks," *arXiv preprint arXiv:1410.5884*, 2014.

[142] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[143] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[144] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[145] A. M. Arias-Lorza, D. Bos, A. van der Lugt, and M. de Bruijne, "Cooperative carotid artery centerline extraction in MRI," *Plos one*, vol. 13, no. 5, e0197180, 2018.

[146] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT denoising with convolutional neural network," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, IEEE, 2017, pp. 143–146.

[147] T. Higaki, Y. Nakamura, F. Tatsugami, T. Nakaura, and K. Awai, "Improvement of image quality at CT and MRI using deep learning," *Japanese journal of radiology*, vol. 37, no. 1, pp. 73–80, 2019.

[148] Y. Nakamura, T. Higaki, F. Tatsugami, Y. Honda, K. Narita, M. Akagi, and K. Awai, "Possibility of deep learning in medical imaging focusing improvement of computed tomography image quality," *Journal of computer assisted tomography*, vol. 44, no. 2, pp. 161–167, 2020.

[149] D. Tamada, "Noise and artifact reduction for MRI using deep learning," *arXiv preprint arXiv:2002.12889*, 2020.

[150] T. Lossau, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass, "Motion estimation and correction in cardiac CT angiography images using convolutional neural networks," *Computerized Medical Imaging and Graphics*, vol. 76, p. 101 640, 2019.

Acronyms

AA Ascending Aorta.

ASI Aortic Size Index.

AVD Average volume difference.

BAV Bicuspid Aortic Valve.

BMI Body Mass Index.

CI Confidence Interval.

CMPR Curved Multi-Planar Reformatting.

CNN Convolutional Neural Networks.

COPD Chronic Obstructive Pulmonary Disease.

CRF Conditional Random Field.

CT Computed Tomography.

CTA Computed Tomography Angiography.

DA Descending Aortic.

DLCST Danish Lung Cancer Screening Trial.

DSC Dice Similarity Coefficient.

ECG Electrocardiography.

EMC Erasmus MC, University Medical Centre Rotterdam.

FLAIR T2-weighted Fluid Attenuated Inversion Recovery.

GBM Glioblastoma Multiforme.

GLCM Grey Level Co-occurence Matrix.

GLRLM Grey Level Run Length Matrix.

GLSZM Grey Level Size Zone Matrix.

GNN Graph Neural Networks.

GPU Graphics Processing Unit.

H95 95^{th} percentile Hausdorff Distance.

HU Hounsfield Unities.

ICC Intra-Class Correlation.

IQR Interquartile Range.

ISLES Ischemic Stroke Lesions Segmentation.

LME Linear Mixed Effects.

LPA Left Pulmonary Artery.

MAP Maximum Posteriori Probability.

MCD Mean Centerline Distance.

MDCT Multidetector Computed Tomography.

MFA Mean-Field Approximation.

MFN Mean-Field Network.

MRI Magnetic Response Imaging.

MSD Mean Surface Distance.

PA Pulmonary Artery.

PABIFL Pulmonary Artery Bifurcation Level.

PAT Pulmonary Artery Trunk.

RCT Randomized Control Trial.

RNN Recurrent Neural Network.

ROI Region Of Interest.

RPA Right Pulmonary Artery.

SD Standard Deviation.

SHAP SHapley Additive exPlanations.

WMH White Matter Hyperintensities.

Lay Summary (English Summary)

Cardiovascular diseases and Chronic Obstructive Pulmonary Disease (COPD) are among the major leading causes of death globally. In the search for early identification of individuals at risk of cardiovascular disease in COPD, imaging-based assessments of the shape and size of the aorta and pulmonary artery have rapidly gained interest. Changes in these two large arteries may indicate cardiovascular diseases such as pulmonary hypertension and aortic aneurysm. Furthermore, the ratio of the diameter of the pulmonary artery to ascending aorta at the level of pulmonary artery bifurcation is shown to be associated with an increased risk of severe exacerbations and increased mortality in patients with COPD. Therefore, it is essential to have an accurate delineation and quantification of the aorta and pulmonary artery anatomy. With the growing use of low-dose non-contrast thoracic CT scans for lung cancer screening, there is an opportunity to measure the aorta and pulmonary artery in these scans. However, performing diameter measurements manually is labor-intensive and time-consuming; therefore, automatic 3D segmentation and measurement techniques are desirable.

This thesis aims to develop and validate fully automatic segmentation and diameter measurement techniques to quantify the shape and size of the aorta and pulmonary arteries in CT scans.

Chapter 1 provides a background on the aorta and pulmonary artery anatomy and thoracic imaging and introduces the diseases associated with these vessels and the challenges associated with automatic vessel segmentation. The main challenges for the segmentation of the aorta and pulmonary artery in non-contrast CT scans are the unclear vessel boundary and intensity similarities with adjacent vessels. Additionally, in non-ECG-gated CT scans, the existence of motion artifacts caused by the motion of the heart during the cardiac cycle increases the ambiguity of the vessel boundaries and makes segmentation a more challenging task.

Chapter 2 develops and validates a fully automatic aorta segmentation and diameter measurement technique based on an optimal surface graph cuts method. This 3D segmentation algorithm is evaluated on 100 non-ECG-gated, non-contrast CT scans and performs well with high overlap between the full 3D manual and automatic segmentations. Besides the 3D segmentation, the aortic centerline and a landmark for the level of the pulmonary artery bifurcation are extracted, which made the diameter

measurements at cross-sections perpendicular to the vessel centerline at a standard landmark level possible. Subsequently, the aortic diameters are assessed at multiple, fixed levels relative to the extracted landmark level, and a high agreement between the manual and automatic diameters is achieved. This study shows that the proposed automatic method is a promising technique to accurately and robustly assess aorta diameters and is a valuable method for both clinical practice and study purposes.

The automatic aorta segmentation and diameter measurement method evaluated in **Chapter 2** is then applied in **Chapter 3** to study the aortic growth rate in a large cohort. This cohort consists of current and former smokers of 50- to 70-year-old adults from the Danish Lung Cancer Screening Trial. In this subgroup of the general population (smokers), the ascending and descending aorta diameters are measured at the level of the pulmonary artery bifurcation of almost 2000 participants. An aortic growth rate of approximately 0.1 mm/year for both males and females is found, consistent with numbers reported for growth in cross-sectional studies of the general population. In addition, it is found that larger changes of aortic diameters in time are associated with lower age, increased height, absence of medication for hypertension or hypercholesterolemia, lower Agatston score, and a large thoracic aortic diameter.

Since it is shown that the ratio of the diameter of the pulmonary artery to the diameter of the ascending aorta at the level of the pulmonary artery bifurcation (PA:AA) is associated with increased risk of severe exacerbations and increased mortality in patients with COPD, measuring the PA:AA ratio is important for patient management. In **Chapter 4**, a fully automatic optimal surface graph cut based method is applied to segment and measure the diameters of the pulmonary artery and aorta on non-ECG-gated, non-contrast CT scans. The segmentation method is validated by comparing the automatic segmentations with full 3D manual annotations and achieved high accuracy. Diameters and the PA:AA ratio measured in the axial plane at the level of the pulmonary artery bifurcation show a high agreement with manual diameters measured in the same slice. **Chapter 4** also presents a 3D volumetric diameter measurement technique that is less subjective and is a more robust and reproducible technique than diameter measurement in 2D axial slices. The 3D volumetric diameters of the aorta and pulmonary artery and the PA:AA ratio are measured in 10 mm segments along the vessel centerlines around the level of the pulmonary artery bifurcation and show high scan-rescan repeatability.

The methods in **Chapter 2** and **4** segment the aorta and the pulmonary artery separately. In **Chapter 5**, an end-to-end deep learning method based on the combination of a Convolutional Neural Network (CNN) with a Conditional Random Field (CRF), named Posterior-CRF, is proposed to jointly segment the aorta and pulmonary artery. Posterior-CRF refines the voxel-level predictions made by CNN by encouraging spatial coherence. In non-contrast CT, intensity-based information alone provides a low-quality feature space for the CRF due to the intensity similarity between vessels and surrounding structures. Therefore, the Posterior-CRF presented in **Chapter 5** uses adaptive features that are learned by CNN. The high accuracy achieved by the Posterior-CRF network compared with other networks in **Chapter 5**

and the qualitative comparisons demonstrate that the method has good performance on segmenting unclear vessel boundaries.

Finally, in **Chapter 6**, the contributions of this thesis towards the development and evaluation of fully automatic segmentation and diameter measurement techniques for the aorta and pulmonary artery are discussed.

The methods proposed in this thesis presented robust and reproducible results of sufficient accuracy and reliability for use in the clinical study. I hope this work will be continued in further studies, thereby contributing to future incorporation into clinical practice and facilitating clinical and epidemiological research, for the final goal of accurate diagnosis of silent disease in an early stage contributing to improved everyday healthcare.

Samenvatting (Dutch Summary)

Hart- en vaatziekten en chronische obstructieve longziekte (COPD) behoren wereldwijd tot de belangrijkste doodsoorzaken. Ten behoeve van een vroege identificatie van COPD-patiënten met een risico op hart- en vaatziekten is er toenemende interesse voor een op beeldvorming gebaseerde beoordeling van de vorm en diameters van de longslagaders en de thoracale aorta. Veranderingen in deze twee grote slagaders kunnen wijzen op hart- en vaatziekten zoals pulmonale hypertensie en een aneurysma van de thoracale aorta. De verhouding tussen de diameter van de longslagader en de opgaande aorta ter hoogte van de vertakking van de longslagader blijkt geassocieerd te zijn met een verhoogd risico op ernstige exacerbaties en verhoogde mortaliteit bij patiënten met COPD. Daarom is het essentieel om een nauwkeurige afbakening en kwantificering van de anatomie van de aorta en longslagader te hebben. Het toenemende gebruik van CT-thorax scans voor longkankerscreening biedt de mogelijkheid om de aorta en longslagaders in deze scans te meten. Omdat het arbeidsintensief en tijdrovend is om deze metingen handmatig te verrichten, zijn automatische 3D-segmentatie en meettechnieken gewenst.

Dit proefschrift richt zich op de ontwikkeling en validatie van volautomatische segmentatie- en meettechnieken waarmee de vorm en grootte van de aorta en longslagaders kunnen worden gekwantificeerd in CT-scans.

Hoofdstuk 1 geeft achtergrondinformatie over de anatomie van de aorta en longslagaders, thoracale beeldvorming, de aandoeningen die met deze vaten geassocieerd zijn en over de uitdagingen van automatische vaatsegmentatie. De belangrijkste moeilijkheden voor de segmentatie van de aorta en longslagader zijn de onduidelijke bloedvatgrenzen die worden veroorzaakt door het gebrek aan contrast tussen bloedvaten en hun omgeving en door de bewegingsartefacten in niet-ECG-getriggerde CT-scans zonder contrastmiddel.

Hoofdstuk 2 ontwikkelt en valideert een automatische techniek voor het segmenteren en meten van de aorta, op basis van "optimal surface graph cuts". Dit 3D-segmentatie-algoritme is geëvalueerd op 100 niet-ECG-getriggerde CT-scans zonder contrastmiddel, waarbij een grote overeenkomst werd gemeten tussen de automatische en handmatige 3D-segmentaties. Naast de 3D-segmentatie bepaalt het algoritme het niveau van de vertakking van de longslagader en de aortamiddellijn, waardoor het mogelijk is om de diameter loodrecht op de middellijn van het vat te meten. Hierop volgend wordt de aortadiameter gemeten op een aantal niveaus ten opzichte

van het vertakkingsniveau van de longslagader. De automatisch gemeten diameters kwamen op alle 13 meetniveaus in hoge mate overeen met de handmatige metingen. De voorgestelde automatische methode is een veelbelovende techniek waarmee aortadiameters nauwkeurig kunnen worden bepaald en is zodoende waardevol voor zowel de klinische praktijk als voor onderzoeksdoeleinden.

De methode uit **Hoofdstuk 2** wordt vervolgens toegepast in **Hoofdstuk 3** om de groeisnelheid van de aorta te bestuderen in een groot cohort van huidige of voormalige rokers uit de groep van 50- tot 70-jarigen in het Deense onderzoek naar longkankerscreening (DLCST). In deze subgroep van de algemene populatie (rokers) hebben we de aortadiameters op het niveau van de longslagadervertakking van alle bijna 2000 deelnemers van DLCST gemeten. We vonden een aorta-groeisnelheid van ongeveer 0,1 mm per jaar voor zowel mannen als vrouwen, wat overeenkomt met de cijfers die eerder zijn gerapporteerd voor aortagroei in de algemene bevolking op basis van transversale onderzoeken. Grotere veranderingen van aortadiameter in de tijd waren geassocieerd met een lagere leeftijd, hogere lichaamslengte, afwezigheid van medicatie voor hypertensie of hypercholesterolemie, lagere Agatston-score en een grotere thoracale aortadiameter.

Omdat is aangetoond dat de verhouding van de diameter van de longslagader tot de diameter van de het stijgende deel van de aorta (PA:AA ratio) geassocieerd is met een verhoogd risico op ernstige verslechtering en een verhoogde mortaliteit bij patiënten met COPD, is het meten van deze ratio belangrijk. **Hoofdstuk 4** presenteert een volautomatische methode op basis van optimal graph cuts voor het segmenteren en meten van de diameter van longslagaders en de aorta in niet-ECG-getriggerde CT-scans zonder contrast. Deze segmentatiemethode is gevalideerd door de automatische segmentaties te vergelijken met volledig handmatige 3D-segmentaties en geeft een hoge nauwkeurigheid. In dit hoofdstuk worden de diameters en de PA:AA ratio gemeten in het axiale vlak op het niveau van de longslagadervertakking, waarbij een grote overeenkomst wordt bereikt met handmatig gemeten diameters en PA:AA-verhouding in hetzelfde vlak. Het is lastig om de meting van de longslagaderdiameter klinisch te standaardiseren. **Hoofdstuk 4** presenteert daarom een 3D-volumetrische techniek die minder subjectief is en een robuustere en reproduceerbare techniek is dan diametermeting in axiale coupes. De 3D-volumetrische diameters van de aorta en longslagader en de PA:AA-verhouding zijn gemeten in segmenten van 10 mm langs de middellijnen van het vat rond het niveau van de longslagadervertakking en deze vertoonden een hoge reproduceerbaarheid.

In de methodes uit **Hoofdstuk 2** en **Hoofdstuk 4** worden de aorta en de longslagaders afzonderlijk gesegmenteerd. In **Hoofdstuk 5** wordt een nieuwe methode voorgesteld, gebaseerd op de combinatie van een Convolutional Neural Network (CNN) met een Conditional Random Field (CRF), de zogeheten “Posterior-CRF”, die de aorta en de longslagaders gelijktijdig segmenteert. Posterior-CRF geeft een beter resultaat dan een CNN door in de voorspelling van de CNN de ruimtelijke coherentie te verbeteren met behulp van een CRF. Vanwege de vergelijkbare intensiteiten van vaten en omliggende structuren in CT zonder contrastmiddel, geeft intensiteit

onvoldoende informatie voor de CRF. De Posterior-CRF die wordt gepresenteerd in **Hoofdstuk 5** gebruikt daarom beeldeigenschappen die adaptief worden geleerd door de CNN als input voor de CRF. Zowel de kwantitatieve als de kwalitatieve vergelijking van de resultaten van het Posterior-CRF-netwerk met die van andere netwerken in **Hoofdstuk 5**, tonen aan dat de methode goede prestaties levert bij het segmenteren van onduidelijke vaatgrenzen.

Ten slotte worden in **Hoofdstuk 6** de bijdragen van dit proefschrift aan de ontwikkeling en evaluatie van volautomatische technieken voor segmentatie en diametermetermetingen voor de aorta en longslagader besproken.

Onze methoden geven nauwkeurige en reproduceerbare resultaten en zijn daarmee bewezen voldoende betrouwbaar voor gebruik in klinische studies. Ik hoop dat dit werk in verdere studies zal worden voortgezet en in de toekomst kan worden geïntegreerd in de klinische praktijk, zodat klinisch en epidemiologisch onderzoek kan worden vergemakkelijkt, met als einddoel een nauwkeurige diagnose van stille ziekte in een vroeg stadium en een verbeterde dagelijkse gezondheidszorg.

Acknowledgement

It is time to close this chapter of my life! No individual is alone responsible for reaching a scientific achievement. The impact and support of many are required. A PhD thesis is no exception. In my PhD journey, I met many brilliant, talented, and wonderful people, of whom I learned valuable lessons. I am deeply grateful for the professional and social support of my family, friends, and colleagues towards my goal and would like to express my sincere appreciation to them.

First and foremost, I would like to express my gratitude to my promoters Prof. dr. Marleen de Bruijne, and Prof. dr. Wiro Niessen. I am grateful to be part of your research lab.

Dear Marleen, I would like to thank you for giving me the freedom to explore different research directions, present new ideas, and acquire knowledge in related fields. Simultaneous to the freedom, you provided sufficient guidance and directed me to the right track when I hit a dead-end. I am impressed with your knowledge and insight into the medical image analysis field and your rigorous attitude towards science. You taught me how to think profoundly and provide out-of-the-box ideas to conduct an appropriate research study. You are an example of a brilliant scientist to me. I especially appreciate your availability and your management style in providing one-hour weekly meetings for each of your PhD students and your availability. I could step into your office for a quick question and you were available, even late in the evening or on the weekends, via email. You listened carefully to my ideas and my plans - my "Big Dreams" - and helped me to make them realistic. Thank you for guiding me through my PhD path. PhD time was not always easy for me, but in the end, it worked out well.

Dear Wiro, I would like to thank you for your supervision. Although we did not have regular meetings, you guided me with your constructive feedback when required. I admire your enthusiasm and passion for science and your ability to lead a diverse and inclusive research group successfully. It has been a pleasure working with you.

I would like to extend my gratitude to my committee members for their time and comments. Members of my inner committee, Dr. Raúl San José Estépar, Prof. dr. Ivana Išgum and Prof. dr. Aad van der Lugt, thank you for reading and approving my thesis. Prof. dr. Jolien Roos-Hesselink, Dr. Ricardo P.J. Budde, and Dr. ir. Rob J. van der Geest, thank you for your willingness to be part of my committee.

Special regards go to Dr. Mehdi Chehel Amirani, my Master's thesis supervisor at Urmia University. Thank you for supporting my decision and for encouraging me to pursue my PhD abroad.

I am also grateful to my co-authors for their contribution to this thesis. Special thanks to Jolien Roos and Daniel Bos for providing valuable clinical insights. Lidia, Andries, and Shuai, thank you for your hard work and your insights.

I would like to express my thank to all my current and former colleagues at BIGR for their willingness to help, their kindness, the chats and fun times during movie nights, board game nights, outings, drinks, and lunch and coffee breaks. I consider myself fortunate to have the opportunity to work in an international group filled with amazing, talented, and friendly researchers.

In particular, I would like to thank Carolyn and Annegreet that accepted to be my paranymphs. Carolyn, I'm so grateful for all your kindness, sharing experiences, and assistance in everything that matters in life. You had a significant influence on my life, especially after the fourth year of my PhD. Annegreet, thanks for the coffee break chats and stories we shared during lunch breaks. Your interesting cross-fit stories motivated me to work out more. Thank you both for encouraging me during the preparation of my thesis.

My special thanks go to my colleagues and friends Yao and Gokhan. Yao, I am delighted to have made such a great friend. Thank you for all the dinners and the Christmas trees we decorated together with Wenhao and Ethen. Being with you made me feel like I'm at home. I still remember when we went to Efteling with Gokhan, and you surprised Mojtaba and me for our birthday. Gokhan, thanks for inviting us to explore Schiedam together; it was amazing. You are a good friend to Mojtaba and me.

I would like to extend my thanks to the BIGirl ladies, Esther, Veronika, Wyke, Emilie, Annegreet, Carolyn, and Yao. Thanks for all the events and good times together. It is so lovely that this group of friends is extending with the new family members.

To the former and current members of the model-based medical image analysis group (MBM), Adria, Veronika, Andriés, Annegreet, Gijs, Florian, Gerda, Arno, Kim, Antonio, Shuai, Deep, Oliver, Robin and Raghav, it was a pleasure working and learning together. Gijs, thank you for being an amazing colleague. You were always volunteering to review drafts of our papers. Thank you for your comments and feedback and for your help in translating my Dutch summary. Andirés, thank you for the brainstorms on graph cuts. I learned a lot from your experiences and your knowledge. Florian and Gerda, I was introduced more to the exciting world of deep learning when you joined MBM. Thanks to your brainstorms and discussions, I got more enthusiastic about this field. Arno, I enjoyed working with you on your Master's thesis. Your project gave me the opportunity to examine my knowledge and my supervision capabilities. You are a self-motivated person with new ideas, and I am happy that you continued in

academia.

My office mates, Martijn, Sebastian and Gerda, thanks for making our working environment so friendly and full of happiness. I enjoyed my working time with you. Florian, Gijs, and Jose, thank you for joining most of our talks and making it even more fun. Martijn, thank you for answering my questions related to XNAT and Python. You made it look way too simple. I learned a lot about Radiomics from you and Sebastian. Gerda, I am so happy having you seated beside me in our office. You are a very kind, strong, and brilliant girl. Sebastian, thanks for having cool ideas for gatherings and for organizing joyful events. The board game nights, movie nights, stair climbing, and many other activities brought fun to BIGR and made it even more friendly and socially enjoyable. Florian, you made it difficult to understand when you are serious and when you are joking. You always had a fixed amount of questions, 2 or 3! Thanks for bringing lots of laughter to the group.

I would like to thank the infrastructure group, Marcel, Hakim, Thomas, and Mattias, for being helpful with the BIGR-cluster. I benefited from the high speed of the cluster in the parameter optimization process of my methods. Additionally, I would like to thank Mart for the IT support.

I am also grateful for the assistance given to me by Petra, Desiree, Tineke, and Andreas for the administration tasks. Desiree thank you for your patience and your help in administration tasks. Petra, thank you for being kind and friendly. You made me feel welcome in the Netherlands and at Erasmus MC.

I will take the opportunity to thank the other members of BIGR and colleagues I met from other departments during my PhD. Jose, Hua, Gennedy, Karin, Yuanyuan, Jiahang, Choeping, Vikram, Ghassan, Kasper, Jean-Marie, Pierre, Bo, Muhammad, Nora, Roman, Stefan, Dirk, Theo, Erik, Jifke, Henri, Raluca, and Natalie, thank you for sharing your knowledge and experiences.

I would like to extend my deepest gratitude to Leila Karimi from the Medical Informatics department. Dear Leila, you were the first Iranian person I met at Erasmus MC, and I am delighted to have this nice friendly relationship with you and your dear family, Nima and Ali. Thank you for our long Persian chats and all the laughter. You eagerly shared your life experience in the Netherlands and helped me understand the Dutch culture better and integrate easier into society.

I would also like to extend my thank you to the great colleagues I met in my new work environment since my PhD contract ended. Dear colleagues at Almende (and daughter companies), thank you for sharing your knowledge and all the projects and fun events we have at the office. In the past few years, I have learned a lot from you. I am very grateful to be working among such talented people.

The completion of my thesis would not have been possible without the support and nurturing of my family. I express my deep appreciation and gratitude to my parents. My dear mother and father, thank you so much for your consistent support, your guidance and, words of wisdom. I would not be where I am today without your care and trust. Dad, living as a researcher, you always showed me new perspectives on things that matter and you guided me to the right path. Your determination and persistence at work encouraged me to work hard. Mom, you made many sacrifices for me. You always have a smile on your face, no matter what the circumstances are. This melts my heart and gives me strength. You are the greatest source of inspiration for me. Words can not express how grateful I am for your unconditional love and lifelong support.

Special thanks go to my sister Leila and her family (Ramin, Helma, and Hirad), my younger sister Maryam, and my brother, Mohammadreza. Thank you for your encouragement and for supporting me with your generous love despite the long distance between us. After moving to the Netherlands, I miss you and all the fun we used to have together. Mohammadreza, I am so happy that you have been inspired to continue your academic life in a similar field as mine. I enjoyed our brainstorms and discussions related to research over these years.

I would like to extend my gratitude to my in-laws. My dear parents-in-law, thank you for welcoming me into your family and for encouraging me all these years. Somayeh and Morteza, my sister- and brother-in-law, thank you for your heart-warming kindness. Morteza, it was a pleasure working with you at Urmia University of Technology, before moving to the Netherlands.

I want to finalize this acknowledgment by thanking the most valued person in my life, my beloved husband. Dear Mojtaba, thank you for accompanying me on this journey, and holding my hands tight through the roller coaster of my PhD life. You helped me in everything, from reading early drafts of my papers to giving me advice on the cover and the layout. Your patience and endurance helped me pass through the hard days and calmed me down in stressful times. You had faith in me when I did not think I could do it. You are as important to this thesis getting done as I am. I am greatly indebted to you and truly appreciate you, my soul mate, my love, and my best friend.

Publications

Journal Papers

Z. Sedghi Gamechi[*], L. R. Bons[*], M. Giordano, D. Bos, R. P. Budde, K. F. Kofoed, J. H. Pedersen, J. W. Roos-Hesselink, and M. de Bruijne, "Automated 3D segmentation and diameter measurement of the thoracic aorta on non-contrast enhanced CT," *European radiology*, vol. 29, no. 9, pp. 4613–4623, 2019.

L. R. Bons[*], **Z. Sedghi Gamechi**[*], C. G. Thijssen, K. F. Kofoed, J. H. Pedersen, Z. Saghir, J. J. Takkenberg, I. Kardys, R. P. Budde, M. de Bruijne, and J. W. Roos-Hesselink, "Growth of the thoracic aorta in the smoking population: The Danish Lung Cancer Screening Trial," *International journal of cardiology*, vol. 299, pp. 276–281, 2020.

Z. Sedghi Gamechi, A. M. Arias-Lorza, Z. Saghir, D. Bos, and M. de Bruijne, "Optimal surface graph cuts to segment the pulmonary artery and aorta on non-contrast CT," [**under review**].

S. Chen, **Z. Sedghi Gamechi**, F. Dubost, G. van Tulder, and M. de Bruijne, "An end-to-end approach to segmentation in medical images with CNN and Posterior-CRF," [**under review**].

Conference Papers

Z. Sedghi Gamechi, A. M. Arias-Lorza, J. Holst Pedersen, and M. de Bruijne, "Aorta and pulmonary artery segmentation using optimal surface graph cuts in non-contrast CT," in *Medical Imaging 2018: Image Processing*, vol. 10574, SPIE, 2018, pp. 616–622.

Conference Abstracts

Z. Sedghi Gamechi, A. M. Arias-Lorza, J. Holst Pedersen, K. F. Kofoed, D. Bos, and M. de Bruijne, "3D quantitative analysis of the aorta and pulmonary artery

on non-contrast CT," in *Insights into Imaging*, ser. ECR 2019: Book of Abstracts, vol. 10, 2019, B0497.

L. R. Bons, **Z. Sedghi Gamechi**, K. F. Kofoed, J. H. Pederson, R. PJ, M. de Bruijne, and J. W. Roos-Hesselink, "Imaging before and after aortic valve repair," in *Insights into Imaging*, ser. ECR 2019: Book of Abstracts, vol. 10, 2019, SS203.

* indicates equal contributions

PhD Portfolio

PhD portfolio

Name	Zahra Sedghi Gamechi
Department	Radiology & Nuclear Medicine
Research School	Cardiovascular Research School (COEUR) - ASCI
PhD Period	Nov 2014 - Sep 2021
Thesis Title	Automatic quantification of the aorta and pulmonary artery in chest CT; methods and validation in lung screening
1st Promotor	Prof. dr. Marleen de Bruijne
2nd Promotor	Prof. dr. Wiro J. Niessen

PhD Training (24.1 ECTS)

		ECTS
2014	Knowledge driven Image Segmentation (ASCI) *Leiden University Medical Center (LUMC)*	4.0
2015	Front-End Vision and Multi-Scale Image Analysis (ASCI) *Eindhoven University of Technology (TU/e)*	4.0
2015	Advanced Pattern Recognition (ASCI) *Delft University of Technology (TU Delft)*	4.0
2015	Presentation Course *Erasmus MC - BIGR*	1.0
2016	NFBIA Summer School *Radboud University Medical Center (Radboudumc)*	2.0
2016	Biomedical English Writing and Communication *Erasmus MC*	3.0
2017	Scientific Integrity *Erasmus MC - Department Medical Ethics and Philosophy of Medicine*	0.3
2017	Cardiovascular Imaging and Diagnostics Part I *Cardiovascular School Erasmus Rotterdam (COEUR)*	0.5
2017	MRI Based Assessment of Biomechanical Stress and Atherosclerosis in Carotid Arteries *Cardiovascular School Erasmus Rotterdam (COEUR)*	0.4

2018	Cardiovascular Imaging and Diagnostics Part II: Clinical non-invasive Cardiac Imaging *Cardiovascular School Erasmus Rotterdam (COEUR)*	0.5
2018	Cardiovascular Imaging and Diagnostics Part III: Atherosclerotic Plaque Imaging *Cardiovascular School Erasmus Rotterdam (COEUR)*	0.5
2018	Bayesian Statistics and JASP *Erasmus Postgraduate school Molecular Medicine (MolMed)*	0.3
2018	Statistical Reasoning for Public Health 1: Estimation, Inference, & Interpretation *Johns Hopkins University - School of Public Health (Online coursey)*	1.5
2018	Statistical Reasoning for Public Health 2: Regression Methods *Johns Hopkins University - School of Public Health (Online coursey)*	1.5
2019	Software Carpentry workshop on Python programming *Amsterdam Since Park Study*	0.6

Teaching Activities (6.5 ECTS)

Lectures and Teacher Assistance

2015 - 2016	Imaging - Skills Training / MeVisLab Practical *Master Technical Medicine*	1.25
2016 - 2018	Image Processing - Practical Session /MeVisLab *Bachelor Clinical Technology*	1.25
2016 - 2018	Image Processing - Practical Session /MeVisLab *Master Technical Medicine*	0.75
2018 - 2019	Advanced Image Processing -Practical Session /MeVisLab *Master Technical Medicine*	0.75
2019	Advanced Image Processing - Lecturer *Master Technical Medicine*	1.0

Supervision

2016	Supervision Internship - Arno van Hilten *Machine learning for Carotid artery Plaque Segmentation*	0.5
2016 - 2018	Supervision of Master Thesis - Arno van Hilten *Carotid Plaque Component Segmentation in MRI*	1.0

Seminar and Research Meetings (6.4 ECTS)

Posters

2018	SPIE Medical Imaging Conference - Houston, USA	0.3
2017	Imaging Research on the Move - Rotterdam, The Netherlands	0.3

Oral Presentations

2016	NVPHBV symposium - Eindhoven	0.3
2017	Plenary Meeting of the Department of Radiology & Nuclear Medicine	0.3
2019	European Congress of Radiology (ECR) - Vienna, Austria	1.0

Seminars

2014 - 2019	Biomedical Imaging Group Seminars (biweekly)	1.0
2014 - 2019	Medical Informatics Research Lunch Meeting (biweekly)	1.0
2014 - 2019	Biomedical Imaging Group Literature Review (weekly)	1.0
2014 - 2019	Model Based Meetings (MBM) (weekly)	–
2014 - 2015	Software Development Seminar (MeVisLab)	0.3
2015 - 2018	COEUR PhD Day	0.3
2017	COEUR Symposium	0.3
2018 - 2019	Deep Learning Study Group	0.3

Grants

2019	Stichting Erasmus Trustfonds - Conference Participation Grant

Committees

2015 - 2019	Organizer - Biomedical Imaging Group Literature Review
2018	Organizer - Medical Imaging Symposium for PhD Students (MISP)
2019	Organizer - Healthy PhD life
2019	Leader of the section “Better Medicine Through Machine Learning: What’s Real, And What’s Artificial?” - Health(Y) Sciences Seminar

About the author

Zahra Sedghi Gamechi was born on the 5th of May 1987 in Urmia, Iran. She grow up in a city famous for the tasty grapes and a large salty lack in the west of Iran. At the age of 4, she moved to Australia with her parents, where she attended preschool and part of the elementary school in a multicultural environment. After returning to Iran, she enjoyed performing scientific experiments in her father's (Prof. Hassan Sedghi) physics laboratory and got very interested in physics and mathematics. These experiments led her to select physics as a major in high school and publish papers. She got rewarded for the best student paper in the physics conference for high school students in 2004.

In 2006 she started her bachelors in the field of Electrical Engineering at Urmia University to explore the field of electrical circuits. During her Bachelor's, besides the courses related to electronics, she gained research experience in several subjects including, signal and digital image processing. She graduated from her Bachelor's with first-class honors in 2010.

Selected as an "Exceptional Talented Student", she followed her Master's studies in Communication Engineering at Urmia University, where she focused on stochastic processes, pattern recognition, and image processing with a great interest in medical image analysis. She successfully defended her Master's thesis on an automated feature extraction technique based on spectral correlation functions for texture analysis in brain MRI scans, and received her diploma in February 2013.

During and after her Master's studies, she worked as the director of the Electrical Laboratory and a lecturer for Physics and Electrical Circuits in Urmia University of Technology for about four years.

In November 2014, Zahra moved to the Netherlands to pursue her PhD under the supervision of Prof. Marleen de Bruijne and Prof. Wiro Niessen. She joined the Biomedical Imaging Group, Rotterdam, to work on model-based medical image analysis. Her topic was 3D segmentation and quantification of the aorta and pulmonary artery in non-contrast CT, the results of which are presented in this thesis.

Upon completing her PhD contract, Zahra began working as data analytic/scientist at Almedne B.V. She applies machine learning and analytical methods to various healthcare, energy, and security problems to improve the quality of human life.

No Regrets in Life.
Just Lessons Learned!

ISBN 978-94-6423-355-1