Validation of a Robust Method for Quantification of Three-Dimensional Growth of the Thoracic Aorta Using Deformable Image Registration

Zhangxing Bian[1,2], Jiayang Zhong[1,2], Jeffrey Dominic[1,2], Gary E. Christensen[3], Charles R. Hatt[1,4], and Nicholas S. Burris[1,5].

1. Department of Radiology, University of Michigan
2. Department of Electrical Engineering and Computer Science, University of Michigan
3. Department of Electrical and Computer Engineering, University of Iowa
4. Imbio, LLC
5. Department of Biomedical Engineering, University of Michigan

Version typeset May 18, 2021

Correspondence author: Nicholas S. Burris. Email: nburris@med.umich.edu

Abstract

Purpose: Accurate assessment of thoracic aortic aneurysm (TAA) growth is important for appropriate clinical management. Maximal aortic diameter is the primary metric that is used to assess growth, but it suffers from substantial measurement variability. A recently proposed technique, termed Vascular Deformation Mapping (VDM), is able to quantify three-dimensional aortic growth using clinical computed tomography angiography (CTA) data using an approach based on deformable image registration (DIR). However, the accuracy and robustness of VDM remains undefined given the lack of a ground truth from clinical CTA data, and furthermore the performance of VDM relative to standard manual diameter measurements is unknown.

Methods: To evaluate the performance of the VDM pipeline for quantifying aortic growth we developed a novel and systematic evaluation process to generate 31 unique synthetic CTA growth phantoms with variable degrees and locations of aortic wall deformation. Aortic deformation was quantified using two metrics: Area Ratio (AR), defined as the ratio of surface area in triangular mesh elements, and the magnitude of deformation in the normal direction (DiN) relative to the aortic surface. Using these phantoms, we further investigated the effects on VDM's measurement accuracy resulting from factors that influence quality of clinical CTA data such as respiratory translations, slice thickness and image noise. Lastly, we compare the measurement error of VDM TAA growth assessments against two expert raters performing standard diameter measurements of synthetic phantom images.

Results: Across our population of 31 synthetic growth phantoms, the median absolute error was 0.048 (IQR: 0.077-0.037) for AR and 0.138mm (IQR: 0.227-0.107mm) for DiN. Median relative error was 1.9% for AR and $< 6.4\%$ for DiN at the highest tested noise level (CNR = 2.66). Error in VDM output increased with slice thickness,

with highest median relative error of 1.4% for AR and 6.3% for DiN at slice thickness of 2.0 mm. Respiratory motion of the aorta resulted in maximal absolute error of 3% AR and 0.6 mm in DiN, but bulk translations in aortic position had a very small effect on measured AR and DiN values (relative errors $< 1\%$). VDM-derived measurements of magnitude and location of maximal diameter change demonstrated significantly high accuracy and lower variability compared to two expert manual raters ($p < 0.03$ across all comparisons).

Conclusions: VDM yields accurate, three-dimensional assessment of aortic growth in TAA patients and is robust to factors such as image noise, respiration-induced translations and differences in patient position. Further, VDM significantly outperformed two expert manual raters in assessing the magnitude and location of aortic growth despite optimized experimental measurement conditions. These results support validation of the VDM technique for accurate quantification of aortic growth in patients and highlight important several advantages over current measurement techniques.

Contents

I. Introduction

The thoracic aorta is the largest artery in the body, carrying blood from the heart to the rest of the systemic circulation. A variety of degenerative and inflammatory processes cause the degradation of the structural integrity of the normally elastic aortic wall resulting in thoracic aortic aneurysm (TAA). Aneurysms of the thoracic aorta are often asymptomatic and indolent, either remaining stable or growing slowly over a period of years or decades; however, a small fraction of patients experience acute complications[1] such as rapid growth, aortic dissection or aortic rupture, all of which necessitate urgent surgical repair and are potentially fatal. Current clinical guidelines recommend routine imaging surveillance of TAA and surveillance regimens typically consist of annual or biannual computed tomography angiography (CTA) examinations to assess for interval growth for other aortic complications. Maximal aortic diameter is the primary metric that is used to assess growth and determine candidacy for surgical repair, with measurements typically performed either manually or in a semi-automated fashion using analysis software that allows for multi-planer or centerline-based measurements in planes orthogonal to the aortic axis.

Despite optimal measurement technique and operator experience, current diameter measurement techniques are associated with substantial measurement variability – on the order of ±2-5 mm – often limiting confident assessment of aortic growth at typical aortic growth rates (<1mm per year)[2,3]. There are many potential sources of error/variability with diameter measurements. Common issues involve differences in measurement location along the length of the aorta, differences in angulation of the 2-dimensional measurement planes, and differences in radial position of the diameter calipers (especially when the aortic cross-section is non-circular/elliptical). Without improved methods to measure aortic growth, confident determination of disease progression, accurate assessment of patient risk and fully informed treatment decisions will not be possible.

To address this problem, our group has recently proposed a method, termed Vascular Deformation Mapping (VDM)[4], to quantify aortic growth in a more accurate and comprehensive fashion. This approach employs deformable image registration (DIR) to quantify three-dimensional changes in the aortic wall morphology using high-resolution volumetric computed tomography angiography (CTA) data. Preliminary reports show that the VDM technique may be useful for more complete depiction of the extent of TAA growth to inform surgical planning and for the assessment of growth during imaging surveillance. However, the VDM approach and key algorithms have not yet been validated in a manner that supports the improved accuracy of VDM-derived measurements compared to standard diameter assessments. B-spline based techniques for deformable image registration are well-validated, and can achieve sub-millimeter registration accuracy using clinical CT data. However, a variety of factors related to physiologic motion and image reconstruction may influence the accuracy of registration results between serial aortic CTA examinations, and thus a comprehensive evaluation of the influence of these factors is warranted.

The objectives of this study were threefold: 1) To determine the accuracy of our VDM pipeline for measuring deformation of the aortic wall in TAA using using a representative

sample of synthetically generated CTA phantom pairs; 2) examine the influence of a variety of variables that influence clinical CTA data (e.g., respiratory motion, slice thickness and image noise) on the accuracy of VDM-derived deformation assessment; and 3) compare the accuracy of growth measurements between VDM and experienced manual raters using synthetic phantoms to better quantify the potential benefit on clinical growth assessments.

II. Methods

This section describes the VDM registration pipeline and the procedure to create the synthetically deformed images used in this study. The validation procedure for assessing the accuracy of VDM-based maximal diameter change measurements compared with ground truth is also described.

II.A. VDM Registration

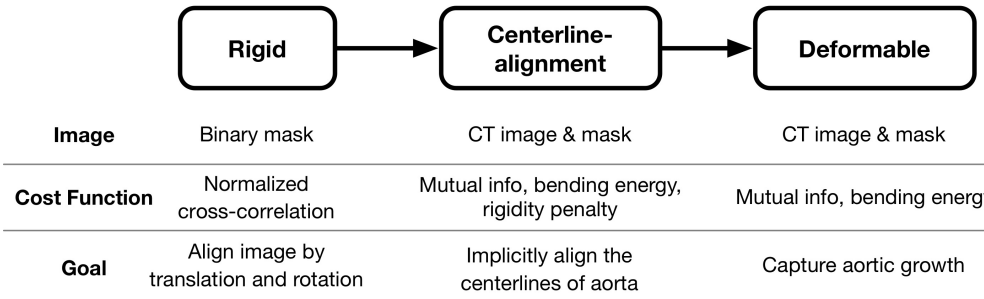


Figure 1: The registration pipeline.

Given two serial CT images with corresponding aortic segmentation masks, we use the VDM pipeline, as shown in Fig.1, to measure growth of the aortic wall.

The registration consists of three main steps: rigid registration, aortic centerline alignment, and deformable registration. The rigid registration uses segmentations of the aorta to rigidly align the images based on the normalized cross-correlation metric, while the centerline alignment and deformable registration steps both use a multi-image, multi-cost function strategy, with each pair of images focusing on a different cost.

Centerline alignment is a DIR step that is highly regularized by bending energy and aortic rigidity penalties, which registers the aortic centerlines by allowing non-rigid movement of the tissues adjacent to the aorta but rigid movement of the aorta itself. This allows the final DIR step to 1) focus primarily on aortic growth via measurement of wall deformation and 2) reduces the need for a large capture range.

The final DIR step performs B-spline-based registration on a finer grid ($0.48 \times 0.48 \times 0.625$) and with a smaller bending energy term than the centerline alignment step to align the aortic wall between the baseline and follow-up images. The displacement field used for

further steps is generated from the final deformable registration step.

The centerline alignment and deformable registration steps utilize one similarity metric (mutual information, MI), and two regularization penalties (bending energy and rigidity). MI[5] is a widely used metric that had originally been developed for multi-modality registration[6]. In our initial experiments[4], we found MI to produce the most accurate results in comparison to other metrics such as normalized cross correlation and sum of squared differences, presumably because MI implicitly focuses on the alignment of boundaries as well as the fact that the intensity of the intraluminal iodine contrast agent can vary between CTA scans. A bending energy penalty[7] is used to regularize DIR by penalizing the high-frequency changes in the deformation field, while rigidity penalty[8] can be used enforce rigidity of the deformation field by penalizing local compression/expansion and deviations from linearity and orthonormality of the Jacobian of of the deformation field. Our workflow is implemented in Elastix[9].

II.B. Generation of Synthetically Deformed CTA images

Step 1: Manually Deformed Aortic Mesh Modeling

A 3D surface was built using the Marching Cubes algorithm[10] applied to an aortic segmentation of the fixed CT image. We used an open-source 3D modeling software (Blender, www.blender.org) to perform deformation of the aortic surface and create synthetic aortic growth phantoms. Each mesh was defined as a set of vertices $\mathcal{V} = v_1, v_2, ..., v_N$, and each face, $f_{\{v_i,v_j,v_k\}}$, was constructed by grouping three neighboring vertices. Each vertex v_i has a position (x_i, y_i, z_i) in the 3D space. We denote the vertices in deformed surface as $\tilde{\mathcal{V}}$; vertex-wise correspondence is maintained during the manual-deform process, i.e. $v_i \leftrightarrow \tilde{v_i}, v_i \in \mathcal{V}, \tilde{v_i} \in \tilde{\mathcal{V}}$.

A total of 31 unique synthetic growth phantoms were derived from a from a single high-quality, electrocardiogram-gated CTA scan of the thoracic aorta acquired at 100 kVP, tube current 340-480 mA, using 95-mL iopamidol 370 mg I/mL (Isovue 370, Bracco Diagnostics, Inc., Princeton, NJ) injected at 4 mL/s, followed by a 100-ml saline chaser at 4 mL/s with axial reconstructions at 0.625-mm section thickness and 0.625-mm intervals at 75% of the R–R cycle. Manual deformations were created with variable locations along the aorta and magnitudes (ranging from 0 to 7 mm) under the guidance of an experienced cardiothoracic radiologist (N.S.B), and were designed to simulate clinically observed aortic shapes and growth patterns. Three primary modes of growth were utilized to create growth phantoms (as depicted in Fig. 2, Step 1):

- Outward radial deformation along the circumference of an aortic cross-section, which mimics typical fusiform growth.
- Sculpting, which mimics an irregular regions of eccentric/saccular buldging often seen in association with atherosclerotic plaque.

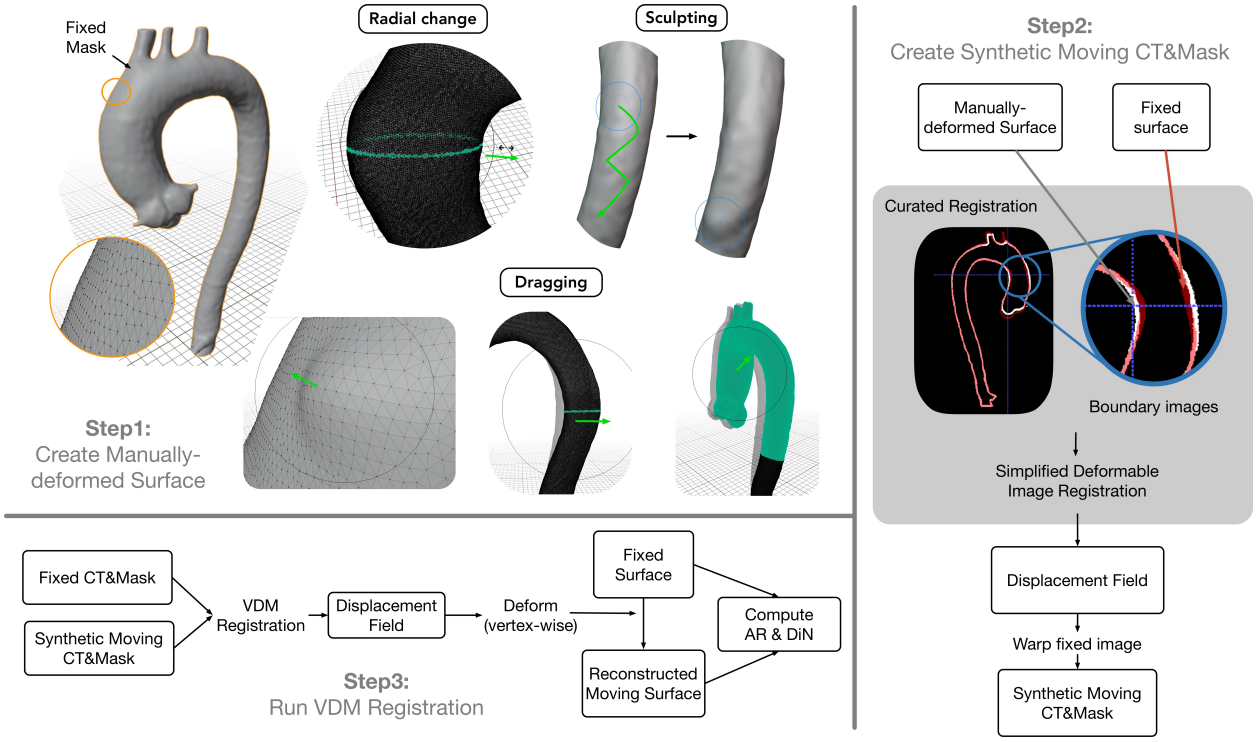


Figure 2: Pipeline for creating synthetic images and validation process. Step 1: three techniques are used to create the deformations on 3D meshes: radial change, sculpting, and dragging. Step 2: a single-step curated DIR registration is used to align the fixed and manually-deformed surfaces. Subsequently, the resulting transformation is used to warp the fixed CT and mask to obtain the synthetic moving images. The displacement field is used to deform the fixed surface to create the synthetic moving surface. Step 3: VDM is used to register the synthetic moving images with fixed images, and compute the metrics – AR and DiN.

- Dragging a group of vertices to simulate bending and/or stretching. Specifically, we used this operation to simulate respiratory related aortic translations.

Step 2: Synthetic Moving Image & Mesh Creation

Following creation of the original and deformed meshes (defined by $\mathcal{V}$ and $\tilde{\mathcal{V}}$), synthetically deformed CT and aortic segmentation masks are generated. This is done by using $\mathcal{V}$ and $\tilde{\mathcal{V}}$ to create "boundary" images ($\mathcal{B}$ and $\tilde{\mathcal{B}}$) , which are then registered to create a deformation field, and we consider this deformation field as the ground truth for all further experiments. Specifically, in the "boundary image", voxels that occupy any vertex are set to one and are otherwise zero. Then we register these two boundary images with $\mathcal{B}$ the moving image and $\tilde{\mathcal{B}}$ as the fixed image, using a simplified (single-step) B-spline based deformable registration. The resulting deformation fields are used to create a deformed CTA aorta mask, $\hat{M}$, and a new set of vertices defining a 3rd mesh $\hat{\mathcal{V}}$. Note that $\hat{\mathcal{V}}$ rather than $\tilde{\mathcal{V}}$ represents an aortic surface that is perfectly concordant with the anatomy shown in the synthetic moving $\hat{\mathcal{I}}$ and the simulated deformation field. A schematic depiction of this workflow is shown in Fig.2, Step 2.

Step 3: Run VDM Registration

Given the synthetic moving image and mask from Step 2, we register it with fixed image through VDM and deform the fixed surface using the deformation field (resulting from the VDM). Then we compute AR and DiN based on the deformation field and deformed mesh. To visualize the results, we interpolate the AR and DiN onto the vertices of fixed surface; one example can be found in Fig.3. The computation of AR and DiN are explained in Section.II.C.1..

II.C. Validation

II.C.1. Quantitative Growth Metrics

We define two mesh-based metrics for measuring aortic growth: Area ratio (AR) and deformation in the normal direction to the aortic mesh surface (DiN). AR is defined as the ratio of the area of a face in one mesh (e.g., moving surface) to that of the corresponding face in another mesh (e.g., moving surface).

$$AR_f = \frac{S(f_{\{\hat{v_i},\hat{v_j},\hat{v_k}\}})}{S(f_{\{v_i,v_j,v_k\}})}, \tag{1}$$

where the $S(\cdot)$ computes the area for a given face.

The DiN metric, which is computed at each mesh vertex and defined in Eq.2, is computed by projecting registration-derived displacement vectors between two corresponding vertices (one on the fixed surface and another on moving surface) onto the corresponding normal vector on the fixed surface mesh. This metric reflects the magnitude of deformation (in millimeters) perpendicular to the aortic surface at each vertex.

$$DIN_{v_i} = \vec{n}_{v_i} \cdot (v_i - \hat{v_i}). \tag{2}$$

II.C.2. Validation of Quantitative Measurement Robustness

The robustness of VDM growth quantification using AR and DiN metrics was assessed for a variety of factors that may affect registration accuracy including slice thickness, image noise, and bulk patient motion. The effect of image noise was tested by adding add various magnitudes of Gaussian noise (50 HU, 100 HU, and 150 HU) to the CT images before performing registration, corresponding to contrast-to-noise ratios (CNRs) of 6.84 3.88, 2.66, respectively. CNR was computed using the following equation:

$$\text{CNR} = \frac{\text{Contrast}}{\text{Noise}} = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \tag{3}$$

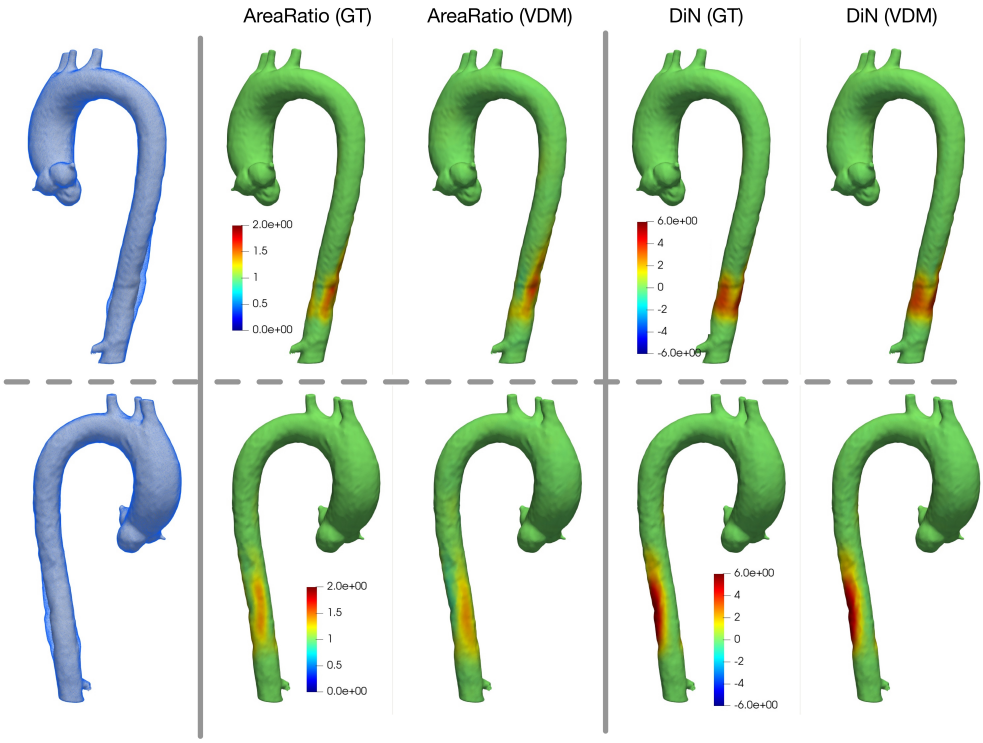


Figure 3: Examples of the ground-truth and VDM-based AreaRatio and DiN metrics for growth quantification shown for a representative synthetic phantom case. The white solid surface is the fixed surface and the blue semi-transparent surface is the synthetic moving surface.

where $\mu_1, \mu_2, \sigma_1\sigma_2$ are the means and standard deviations of the HU values in regions of interest in the aorta and myocardial fat, respectively. More details regarding the CNR calculation can be found in Appendix.VI.

The effect of CT slice thickness on AR and DiN was also tested at three different slice thicknesses representative of a range typically used for clinical CTA: 1.0, 1.5, and 2.0 (mm). We tested the effect of patient bulk motion by randomly rotating (according to a uniform distribution $\{+5, -5\}$ degrees) and translating the image by {20,40,60} (mm) along three axes. For each level of these factors (i.e., noise, slice thickness and bulk motion), a pair of perturbed fixed and moving synthetic images were created. The full VDM analysis pipeline was performed, and the resulting AR and DiN values were compared to unperturbed results by calculation of absolute and relative errors. A schematic depicting this workflow is shown in Fig.4.

Finally, while clinical CTA is most often acquired during inspiration, we tested the effect of respiratory motion of the aorta and how serial CTA scans acquired at different phases of the respiration would affect the accuracy of VDM growth measurements. To do this, an additional 6 synthetic moving images were created that had a combination of localized deformation of the aortic wall in addition to differences in respiratory position of the aorta based on published values[11]. Specifically, we selected 6 cases from the 31 synthetic phantoms with varying degrees of growth and used Blender's dragging tool (Fig.2) to translate the

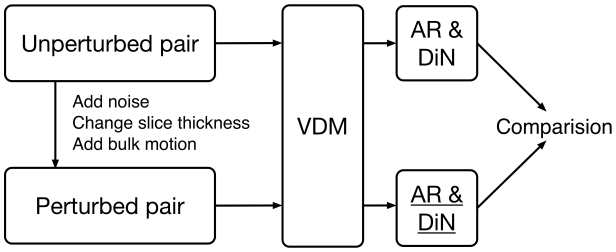


Figure 4: Workflow of robustness test.

ascending aortic, arch and proximal descending aorta in a physiologically realistic manner.

II.C.3. Maximal Diameter Measurement: Expert Manual Measurement vs. VDM

In this section, we focus on the typical clinical task, i.e., measuring the maximal aortic diameter change (i.e., growth), and describe the procedure used to compare VDM-based growth measurements against manual measurements.

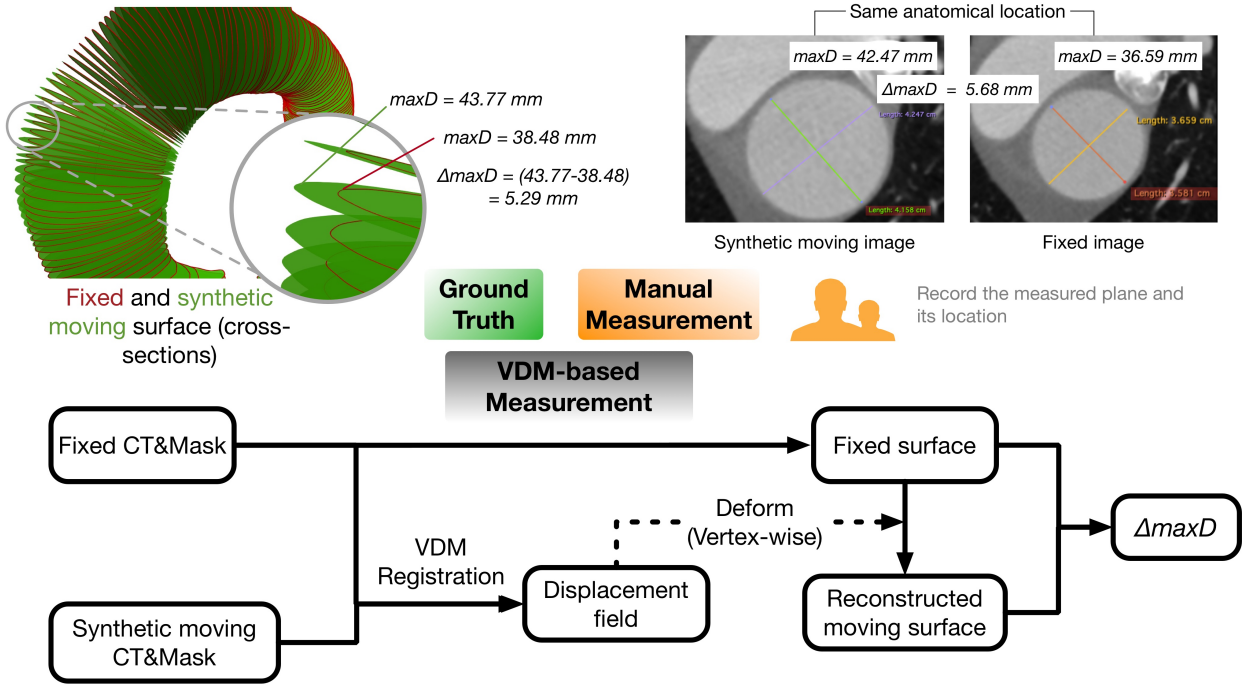


Figure 5: Validation process on maximal diameter change.

Two independent, expert raters (advanced image analysis technologists) with 5 and 15 years of experience with aortic measurements, respectively (Rater 1 and Rater 2) identified the location where the maximum diameter change happens and measured the change accord-

ing to a standard workflow: each rater viewed the synthetically deformed and original CTA images side-by-side and attempted to locate the position where the maximum deformation occurred.

Given that the deformed moving image was synthetically created from the original image, the anatomy was intrinsically registered except at the local region of deformation, which made this task much easier than in a real-life clinical scenario where changes in patient positioning and the positioning of adjacent organs makes visual comparison of side-by-side images much more difficult. Thus, the rater's performance on the synthetic cases was considered the best case scenario for what can be achieved with routine manual measurements.

The ground truth maximal diameter change was measured by first extracting the aortic centerline of the fixed image then sampling the centerline at points every 0.5mm. The maximum diameter of each cross-section (orthogonal to the centerline) was then computed by the open-source Vascular Modeling Toolkit (VMTK, www.vmtk.org)[12]. We denote the results as two one-dimensional arrays $\mathbf{d}_{\mathcal{V}_{fixed}}$ and $\mathbf{d}_{\mathcal{V}_{smoving}}$, with each having the length equal to the number of point samples on the center-line. Then we take $max(|\mathbf{d}_{\mathcal{V}_{fixed}} - \mathbf{d}_{\mathcal{V}_{smoving}}|)$ as the ground-truth maximal diameter change and record the location of the maximal diameter change along the centerline.

In the VDM-based diameter measurement, we obtained the reconstructed moving surface by deforming the fixed surface using the displacement field resulting from registration step. Similarly, we take the same sampled centerline and measure the maximum diameter at each centerline point for both reconstructed moving surface and fixed surface, and record the magnitude and location of the largest change in diameter.

Statistical Analysis. We performed a piori sample size estimates for our manual rater experiments using an F-test of variances and assuming a conservative standard deviation of measurement error of ± 0.3 mm for VDM (based on preliminary experiments) and standard deviation of manual aortic diameter measurements of ± 1 mm from prior literature[13]. This calculation showed a 99% power to detect a difference between groups with a sample size of n=30 synthetic phantoms. Levene's test was used to examine differences in variance of errors and Wilcoxon Test was used to examine group differences in absolute errors. A p-value of < 0.05 was considered significant for all statistical tests. Statistical analyses were performed using Stata 14.0 (StataCorp LP, College Station, TX).

III. Results

III.A. Comparison Between VDM and Ground Truth Growth Metrics

Across our population of 31 synthetic growth phantoms, the median absolute error was 0.048 (IQR: 0.077-0.037) for AR and 0.138mm (IQR: 0.227-0.107mm) for DiN. Absolute error for AR and DiN showed moderate correlation with the degree of maximal aortic deformation

(AR: R=0.52; DiN: R=0.57). There was not a statistically significant difference in the median absolute error between cases of ascending vs. descending TAA for AR (median ascending 0.040, IQR: 0.036-0.047 vs. median descending 0.079, IQR: 0.063-0.118; p=0.079) or DiN (median ascending 0.12, IQR: 0.100-0.217 vs. median descending 0.170, IQR: 0.135-0.213; p=0.477).

A summary of the robustness of the AR and DiN measurements to noise, variable slice thicknesses, and bulk motion is shown in Fig.6. In the case of image noise, the 99^{th} percentile error of AR and DiN measurements increased with increasing degrees of image noise, however, median relative error of 1.9% for AR and $< 6.4\%$ for DiN at the highest tested noise level (Noise-150, CNR = 2.66). Considering the effects of slice thickness variations, generally speaking, the error similarly increased with thicker slices and was highest at slice thickness of 2.0 mm, with highest median relative error of 1.4% for AR and $< 6.3\%$ for DiN. Bulk motion had a very small effect on measured AR and DiN values with relative errors $< 1\%$ at all degrees of translation.

Results of the 6 synthetic phantoms combining growth and respiratory motion are shown in Fig.7. Among synthetic phantoms with growth of the ascending and descending aorta ranging in magnitude from 1.5 mm to 6.5 mm the absolute and relative errors associated with respiratory motion were small for AreaRatio (maximally 0.031 and 2.2% respectively). For these same 6 phantoms, the mean absolute error was 0.23 mm (range: 0.055-0.458 mm).

III.B. Comparison Between VDM and Manual Raters

Following the procedure described in Fig.5, we compared VDM-based measurements with the manual measurements from two expert raters. Fig.8 shows that the VDM-based measurements had significantly less variability (i.e., were more precise) than that of the two manual raters and also was significantly more accurate in regard to localization of the area of maximal diameter change. Rater 1 (more experienced) did demonstrate significantly higher accuracy compared to Rater 2 (less experienced), but there were no significant differences between raters for localization of maximal diameter or varaice of diameter measurement error.

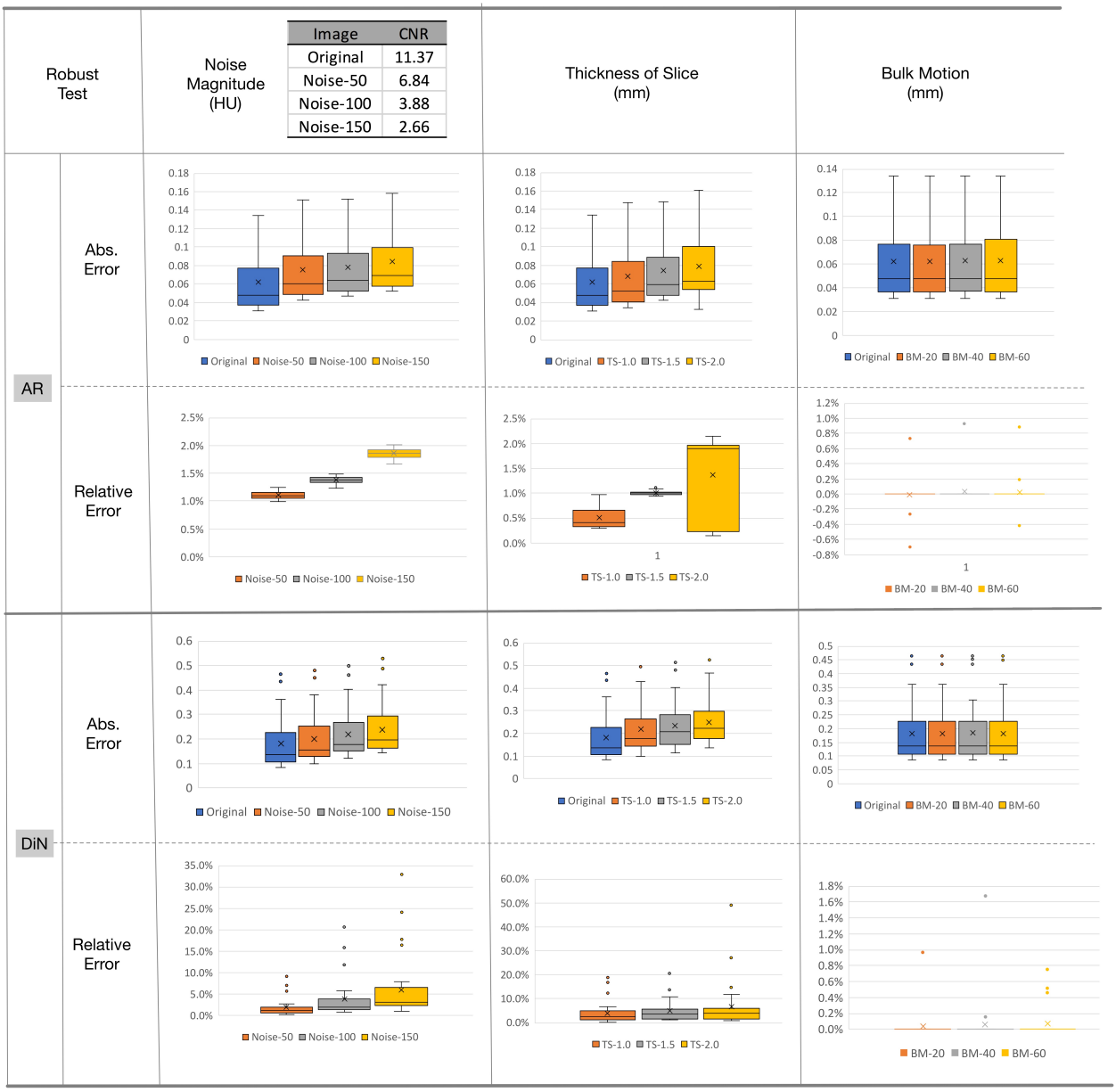


Figure 6: Absolute and relative errors in VDM metrics of aortic growth. Original indicates the VDM result without any perturbations. The reminder of the test reflect the effects of one graded perturbations applied to the fixed and moving images on the VDM outputs. The 99-th percentile errors for both AR and DiN is reported, e.g., $err = |GT_{AreaRatio,DiN} - VDM_{AreaRatio,DiN}|^{99th}$. The relative error is computed by $(err_{perturbed} - err_{original})/GT^{99th}$. In the box plots, the "x" in the box indicates the mean and line indicates median value.

Resp. Case ID	Location of Anuesysm	Anuesysm Growth (mm)	AreaRatio GT (99-th)	Error of AreaRatio (99-th)				DiN GT (99-th)	Error of DiN (99-th)			
				w/o resp.	w/ resp.	Abs. Inc.	Rel. Inc.		w/o resp.	w/ resp.	Abs. Inc.	Rel. Inc.
A		6.5	1.179	0.069	0.074	0.005	+0.4%	4.042	0.156	0.331	0.175	+4.3%
B	Ascending	3.5	1.120	0.033	0.043	0.01	+0.9%	1.610	0.214	0.269	0.055	+3.4%
C		1.5	1.056	0.031	0.035	0.004	+0.4%	0.680	0.103	0.329	0.226	+33.2%
D		6.5	1.420	0.123	0.154	0.031	+2.2%	4.043	0.247	0.434	0.187	+4.6%
E	Descending	4.1	1.222	0.106	0.092	-0.014	-1.1%	2.210	0.194	0.652	0.458	+20.7%
F		2.3	1.243	0.092	0.093	0.001	+0.1%	1.224	0.126	0.378	0.252	+20.6%

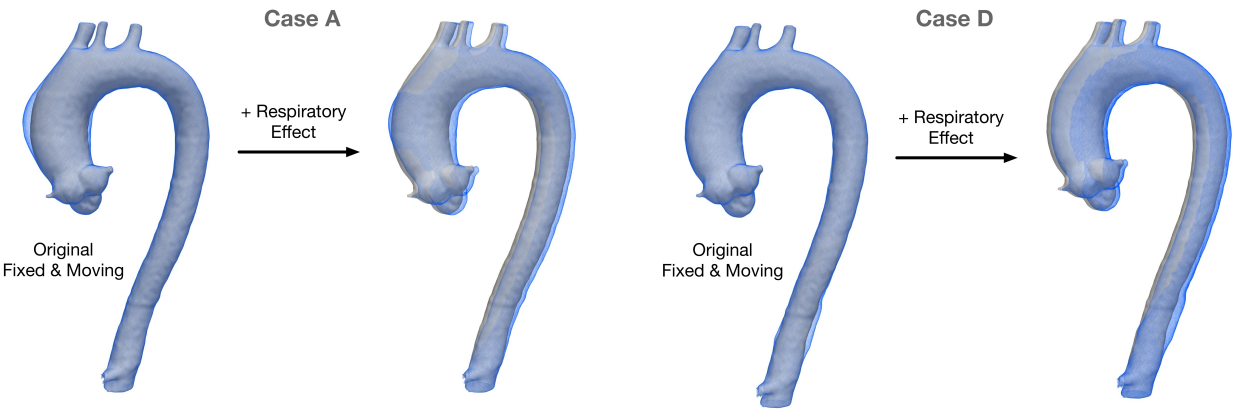
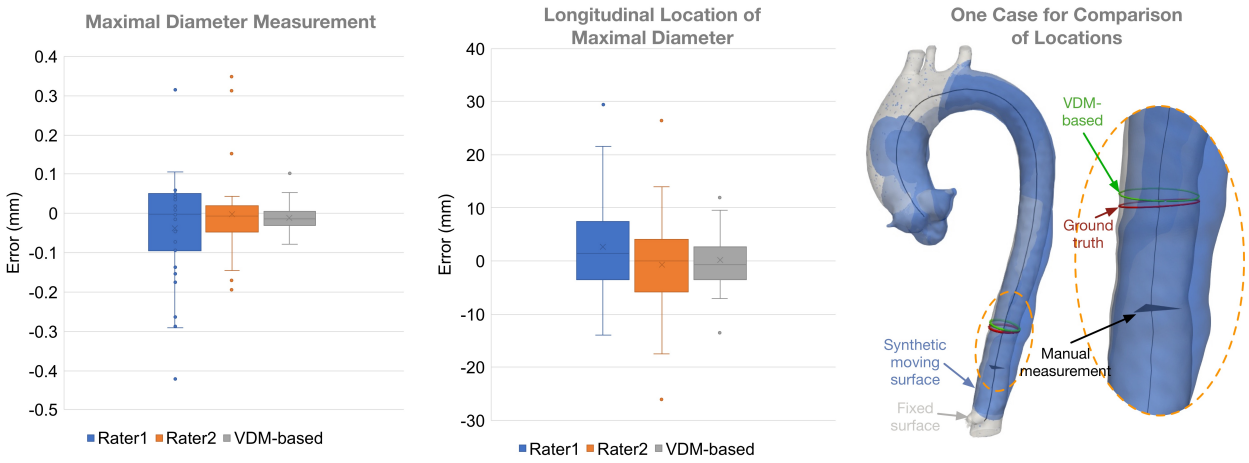


Figure 7: Error in VDM output with in growing TAA with respiratory translations. The white surface is the fixed surface, while the blue surface is the synthetic moving surface. We reported the 99^{th} percentile of the absolute error on AreaRatio and DiN. The relative error is computed by $(Resp - NonResp)/GT^{99th}$. Zoom in to have a better view.



	Purpose of Test	R1 v.s. R2	R1 v.s. VDM	R2 v.s. VDM
Maximal Diameter Measurement Error	Difference in variance	0.114	<0.001	0.010
	Difference in accuracy	0.033	<0.001	0.026
Location Measurement Error	Difference in variance	0.759	0.002	0.005
	Difference in accuracy	0.397	0.004	0.017

Figure 8: Measurement error of VDM versus manual raters. Two box-plots on the left show the error in maximal diameter measurements and longitudinal localization by two raters (R1 and R2) and the VDM-based method. The right figure gives an example of the three locations along the centerline of maximal growth: ground-truth location, Rater1 location (manual), and VDM-based location. The table below shows the p-values corresponding to comparisions between raters and VDM for testing differences in variance (Levene's test) and accuracy (Wilcoxon test). Statistically significant values (< 0.05) are underlined.

IV. Discussion

Accurate measurement of aortic growth remains an important challenge in the management of patients with TAA. A technique such as VDM that more fully utilizes the three-dimensional nature of aortic CTA data may improve aortic growth assessment by avoiding the variability in plane placement in angulation encountered with 2-dimensional diameter measurements. In this study we investigated how the measurement accuracy of VDM compares with manual diameter measurements performed by expert readers, and quantify the effects of physiologic and image quality parameters on the measurement performance of VDM. In summary, we found that the deformable image registration based VDM-pipeline was quite robust to Gaussian image noise and variations in slice thickness within the typical range encountered in clinical CTA examinations. Furthermore, we found that VDM derived Area Ratio measurements were highly robust to physiologic motion of the thoracic aorta due to respiration, although measurement of deformation magnitude in the normal direction demonstrated higher sensitivity to respiratory motion effects. Lastly and perhaps most importantly, we demonstrated that VDM-derived diameter measurements demonstrated significantly higher accuracy and lower variability in aortic growth measurements compared to manual assessments by expert raters and was more accurate in identifying the location of maximal aortic growth along the aortic length.

Few prior studies have attempted to quantify aortic growth in a 3-dimensional fashion using deformable image registration. Gao et al employed a deformable registration-based analysis technique which used a centerline to generate semi-automatic aortic diameter measurements at several discrete locations along the aortic length, and compared the reliability of these measurements with manual raters[14]. However, this study did not attempt to map localized deformation along the surface of the aortic wall and did not employ synthetic phantoms to assess the accuracy of either the semi-automated or manual measurement techniques against a reference configuration. As demonstrated from the data in this paper, manual diameter measurements can be significantly variable and inaccurate despite expert raters and an optimal measurement scenario. Specifically we identified instances where measurement error was up to 3 mm on synthetic phantoms despite excellent image quality, identical CT datasets outside of area of growth, and no differences in patient positioning or physiologic motion. Further, Subramaniam et al. described an approach for quantification of longitudinal aortic growth using contrast-enhanced magnetic resonance angiography (MRA) in patients with Turner syndrome[15]. Their technique involved measurement of the Euclidean distance between aortic centerline points and the aortic segmentation boundary along the length of the aorta, with aortic growth quantified as the differences in these Euclidean distance values between two MRA studies after rigid registration using an iterative closest point algorithm. Similar to Gao et al, Subramaniam et al reported the agreement of their investigational measurements with standard manual diameter measurements, but did not examine the accuracy or robustness of their approach using phantoms, and the accuracy of their approach may be degraded by inaccuracy in segmentation at the aortic boundary and of their point-cloud based rigid registration. Assessment of measurement accuracy against a reference standard aortic growth/deformation, as performed in this study, is an important step

in understanding the real-world clinical utility of such novel measurement techniques considering the small magnitudes of aortic growth typically encountered in clinical practice (often < 2 mm). Similar to previously described techniques, our approach uses aortic segmentation and centerline generation, however, unlike other studies VDM uses of the displacement field (calculated from deformable registration) to deform an aortic mesh. This approach offers several unique advantages including: ability to quantify localized aortic surface area changes, allows point-to-point correspondence between baseline and follow-up aortic geometries, and quantification of aortic wall deformation does not rely on 2D geometric properties such as diameter or Euclidean distance.

Using a group of synthetic growth phantoms with realistic shapes, magnitudes and distributions of growth, we found that VDM measurement of Area Ratio and DefromOnNormal were robust to a variety of image characteristics including image noise and slice thickness with median increases in relative error being $< 2\%$ for Area Ration and $< 5\%$ for DiN at maximal values for Gaussian noise intensity (150) and slice thicknesses (2.0 mm). While medial relative errors were higher with DiN, the absolute magnitude of errors with this metric was still < 0.5 mm. We believe the errors encountered in these synthetic experiments are acceptable for routine clinical scenarios given that ECG-gated CT angiography examinations are commonly reconstructed at slice thicknesses $<$ 2mm and that clinical CT scanners employ dose modulation techniques (e.g., noise index, quality reference mA) to maintain image noise within reasonable limits[16]. While we acknowledge that Gaussian noise is not a true representation of CT image noise, synthetically generating realistic CT image noise can be a challenging procedure, and we believe that Gaussian noise still allows us to examine the effect on registration accuracy attributable to blurring at aortic boundary.

Furthermore, we found minimal error associated with bulk translations/rotations of synthetic CTA pairs ($< 2\%$ relative error), simulating differences in patient position in the CT scanner between examination, but this is an unsurprising result given that rigid registration techniques are commonly used technique to account for such positional differences. Finally, we found that the errors in Area Ratio and DiN values associated with positional changes of the thoracic aorta with respiration (inspiration to expiration), were overall small at physiologic magnitudes[11], and while relative errors for DiN attributable to respiratory motion reached 67% maximally, absolute errors were less than 0.46 mm. In clinical practice we expect these respiratory effects to be even smaller given that our synthetic phantoms simulated the motion associated with peak inspiration to expiration, whereas smaller differences in breath-hold position would be expected based on standard inspiratory CTA acquisition procedures. Of note, we chose not to systematically evaluate the effects differing phases of image reconstruction throughout the cardiac cycle (i.e., % R-R interval), as varying the cardiac phase would instead quantify the effects of pulsatile aortic strain rather than longitudinal aortic wall growth, however, this does assume that the two CTAs used for VDM analysis are reconstructed at the same phase of the cardiac cycle (typically and mid-late diastole in clinical practice).

A unique contribution of this paper is the systematic evaluation of measurement accuracy between VDM and manual expert raters of using synthetic phantoms with defined degrees of growth. Multiple prior papers have examined interrater variability of aortic diam-

eter measurements or have compared novel measurement techniques with standard manual measurements, however neither of these approaches which utilize only clinical data allows for assessment of measurement error. In an attempt to isolate the effects of measurement error attributable to variability in the location and angulation of measurement planes, we designed our aortic phantom experiment to optimize manual raters ability to produce accurate measurements. Specifically, for these experiments the baseline and follow-up (deformed) CTAs were identical outside of the area of synthetic deformation eliminating any possibility for differences in contrast timing or extra-aortic image artifacts, manual raters told the region (e.g., ascending, descending or arch) in which the deformation was created, and no bulk translations or rotations were assigned between baseline and follow-up CTs. Nonetheless we found that VDM had a significantly lower error in determining maximal aortic diameter change and the location of maximal growth compared to experience manual raters with 5 and 16 years of aortic measurement experience respectively. While this highly constrained experiment is not realistic a realistic representation of the routine clinical task of aortic diameter measurements, we believe this experimental design highlights the fundamental limitations in two-dimensional diameter measurements for assessing complex three-dimensional aortic anatomy and emphasizes the advantage of a technique such as VDM that more fully utilizes the volumetric CTA data. The measurement errors with manual raters in our study were lower than the typical degrees of measurement variability reported in the literature (+/- 2mm on average)[3,13,17], which probably reflect the highly constrained nature of our experiment.

This study has several limitations. First, our population of synthetic aortic phantoms was created manually using mesh editing software and thus there may be minor geometric differences in patterns and shapes of growth between these phantoms and the morphologies of TAA seen in patients. However, we made substantial effort to generate synthetic growth in realistic locations, patterns and magnitudes based on prior experience with VDM analysis in a clinical TAA population, and all synthetic phantoms were reviewed by an experienced cardiovascular imager prior to evaluation to confirm only realistic geometries were used. Secondly, our synthetic CT phantoms did not include atherosclerotic plaque (calcified or nonnoncalcified) which can be present in clinical aortic CTA - most often in the descending aorta - and thus the effects of such luminal irregularities were not assessed. However, theoretically features of the aortic wall with unique shapes and intensities would be expected to improve registration accuracy rather than inhibit it, and the relatively featureless luminal boundary in our synthetic phantoms would be expected to be a more challenging registration task. Thirdly, rather than calculating a displacement field directly from the edited mesh vertices, we employed a simplified b-spline deformable image registration between boundary images to generate the displacement field from which reference values for Area Ration and DiN were determined. We believe this approach is valid given that we found very small registration errors at this step, and such small errors would have the equal effects on measurement errors for both VDM and manual measurements. Lastly, we did not aim to evaluate aortic growth in the root (i.e., sinuses of Valsalva) given that the irregular and non-cylindrical geometry makes measurement of maximal aortic diameter challenging in this segment.

V. Conclusion

Our results confirm that Vascular Deformation Mapping is an accurate technique for three-dimensional assessment of aortic growth in patients with thoracic aortic aneurysm, and is robust to a variety of factors related to image quality and physiologic motion which are present in clinical CTA examinations. Using a group of realistic TAA growth phantoms, we were able to investigate the error of growth assessments in a fashion that is not possible using clinical data, and overall we observed that absolute errors in VDM-derived measurements of magnitude of normal deformation and surface area change were less than 0.6 mm and 16% respectively across all phantoms. Furthermore, we found that VDM significantly outperformed experienced manual raters in head-to-head measurements of the magnitude and location of aortic growth, suggesting that this technique could significantly improve the accuracy and reliability of aortic measurements compared to standard-of-care measurement techniques. Further work will be needed to validate the VDM technique in a clinical setting, but these synthetic experiments support both validity of this technique in a controlled setting and provide guidance as to the image and physiologic characteristics that can be tolerated in clinical practice.

Appendix

VI. Noise

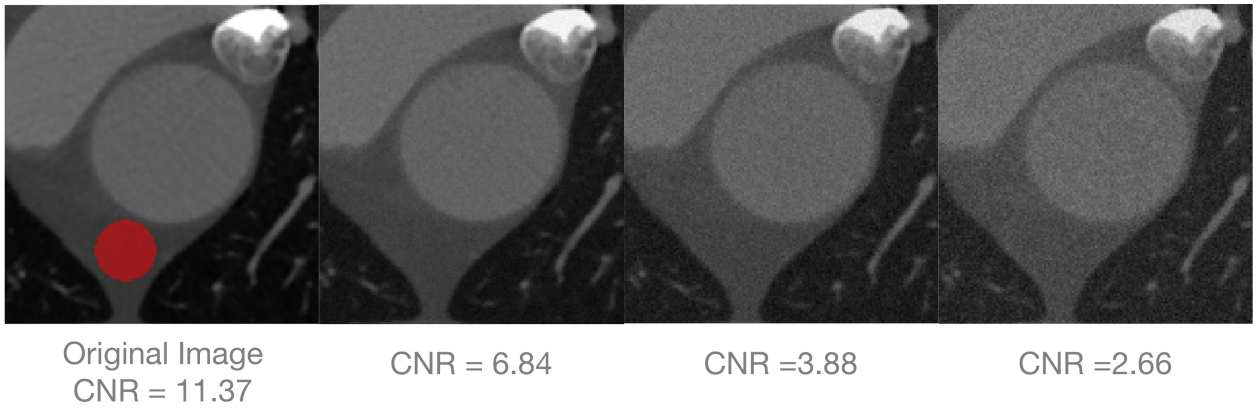


Figure 9: We manually select three slices around the background region (labelled in red) to compute μ_2, σ_2 (mean and standard deviation values) for background in Eq.3. We erode the aorta mask by 3 voxels to obtain the region of interest, and use the voxel in the mask to compute the μ_1, σ_1 for aorta.

References

[1] L. F. Hiratzka et al., ACCF, Circulation. **121**, 1544 (2010).

[2] M. H. Guo, J. J. Appoo, R. Saczkowski, H. N. Smith, M. Ouzounian, A. J. Gregory, E. J. Herget, and M. Boodhwani, Association of mortality and acute aortic events with ascending aortic aneurysm: a systematic review and meta-analysis, JAMA network open **1**, e181281–e181281 (2018).

[3] 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 **29**, 479–488 (2013).

[4] N. S. Burris, B. A. Hoff, H. J. Patel, E. A. Kazerooni, and B. D. Ross, Three-Dimensional Growth Analysis of Thoracic Aortic Aneurysm With Vascular Deformation Mapping, Circulation: Cardiovascular Imaging **11**, e008045 (2018).

[5] P. Thévenaz and M. Unser, Optimization of mutual information for multiresolution image registration, IEEE transactions on image processing **9**, 2083–2099 (2000).

[6] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, Multimodality image registration by maximization of mutual information, IEEE transactions on Medical Imaging **16**, 187–198 (1997).

[7] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, Non-rigid registration using free-form deformations: application to breast MR images, IEEE transactions on medical imaging **18**, 712–721 (1999).

[8] M. Staring, S. Klein, and J. P. Pluim, A rigidity penalty term for nonrigid registration, Medical physics **34**, 4098–4108 (2007).

[9] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, Elastix: a toolbox for intensity-based medical image registration, IEEE transactions on medical imaging **29**, 196–205 (2009).

[10] W. E. Lorensen and H. E. Cline, Marching cubes: A high resolution 3D surface construction algorithm, ACM siggraph computer graphics **21**, 163–169 (1987).

[11] G.-Y. Suh, R. E. Beygui, D. Fleischmann, and C. P. Cheng, Aortic arch vessel geometries and deformations in patients with thoracic aortic aneurysms and dissections, Journal of Vascular and Interventional Radiology **25**, 1903–1911 (2014).

[12] R. Izzo, D. Steinman, S. Manini, and L. Antiga, The vascular modeling toolkit: a Python library for the analysis of tubular structures in medical images, Journal of Open Source Software **3**, 745 (2018).

[13] T.-L. C. Lu, E. Rizzo, P. M. Marques-Vidal, L. K. v. Segesser, J. Dehmeshki, and S. D. Qanadli, Variability of ascending aorta diameter measurements as assessed with electrocardiography-gated multidetector computerized tomography and computer assisted diagnosis software, Interactive cardiovascular and thoracic surgery **10**, 217–221 (2010).

[14] X. Gao, S. Boccalini, P. H. Kitslaar, R. P. Budde, S. Tu, B. P. Lelieveldt, J. Dijkstra, and J. H. Reiber, A novel software tool for semi-automatic quantification of thoracic aorta dilatation on baseline and follow-up computed tomography angiography, The international journal of cardiovascular imaging **35**, 711–723 (2019).

[15] D. R. Subramaniam, W. A. Stoddard, K. H. Mortensen, S. Ringgaard, C. Trolle, C. H. Gravholt, E. J. Gutmark, G. Mylavarapu, P. F. Backeljauw, and I. Gutmark-Little, Continuous measurement of aortic dimensions in Turner syndrome: a cardiovascular magnetic resonance study, Journal of Cardiovascular Magnetic Resonance **19**, 1–17 (2017).

[16] M. K. Kalra, M. M. Maher, T. L. Toth, B. Schmidt, B. L. Westerman, H. T. Morgan, and S. Saini, Techniques and applications of automatic tube current modulation for CT, Radiology **233**, 649–657 (2004).

[17] F. M. Asch et al., The need for standardized methods for measuring the aorta: multimodality core lab experience from the GenTAC registry, JACC: Cardiovascular Imaging **9**, 219–226 (2016).