A Deformable Constraint Transport Network for Optimal Aortic Segmentation from CT Images

Weiyuan Lin, Zhifan Gao, Hui Liu, Heye Zhang

Abstract— Aortic segmentation from computed tomography (CT) is crucial for facilitating aortic intervention, as it enables clinicians to visualize aortic anatomy for diagnosis and measurement. However, aortic segmentation faces the challenge of variable geometry in space, as the geometric diversity of different diseases and the geometric transformations that occur between raw and measured images. Existing constraint-based methods can potentially solve the challenge, but they are hindered by two key issues: inaccurate definition of properties and inappropriate topology of transformation in space. In this paper, we propose a deformable constraint transport network (DCTN). The DCTN adaptively extracts aortic features to define intra-image constrained properties and guides topological implementation in space to constrain inter-image geometric transformation between raw and curved planar reformation (CPR) images. The DCTN contains a deformable attention extractor, a geometry-aware decoder and an optimal transport guider. The extractor generates variable patches that preserve semantic integrity and long-range dependency in long-sequence images. The decoder enhances the perception of geometric texture and semantic features, particularly for low-intensity aortic coarctation and false lumen, which removes background interference. The guider explores the geometric discrepancies between raw and CPR images, constructs probability distributions of discrepancies, and matches them with inter-image transformation to guide geometric topology in space. Experimental studies on 267 aortic subjects and four public datasets show the superiority of our DCTN over 23 methods. The results demonstrate DCTN's advantages in aortic segmentation for different types of aortic disease, for different aortic segments, and in the measurement of clinical indexes.

Index Terms— Computed Tomography, Aortic Segmentation, Geometric Constraint, Optimal Transport, Transformer

I. INTRODUCTION

Weiyuan Lin, Zhifan Gao and Heye Zhang are with School of Biomedical Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China; (e-mail: linwy37@mail2.sysu.edu.cn; gaozhifan@mail.sysu.edu.cn; zhangheye@mail.sysu.edu.cn).
Hui Liu is with Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China (e-mail: liuhuijiujiu@gmail.com).
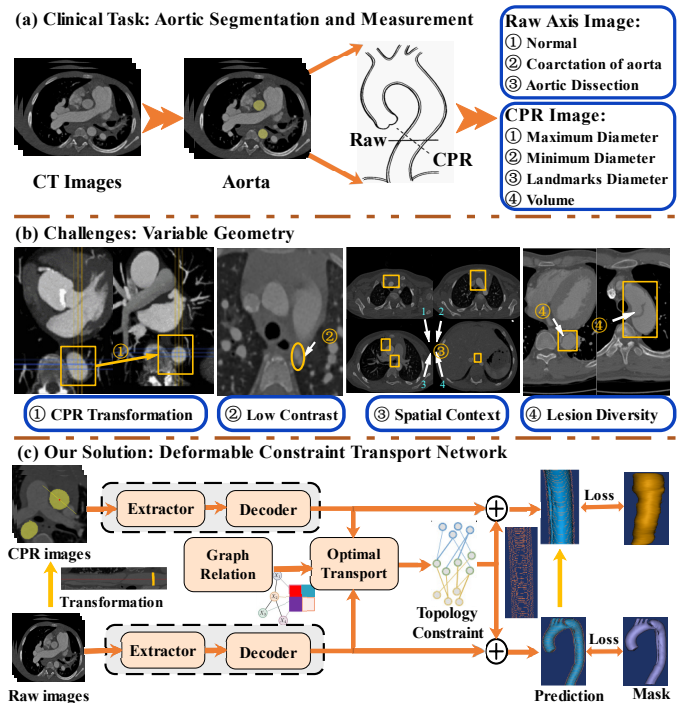
Fig. 1. (a) Aortic segmentation from CT images is important for aortic intervention; (b) Challenges of variable geometry: CPR transformation, small stenosis, diverse morphology, irregular lesions; (c) Our solution: deformable attention extractor, hierarchical geometry-aware decoder and optimal transport guider are included in segmentation.

AORTIC segmentation from computed tomography (CT) images is crucial for carrying out aortic intervention. Aortic intervention is a common treatment for aortic disease. On average, more than 27,000 patients undergo aortic intervention each year according to the Centers for Disease Control and Prevention [1]. Aortic intervention requires anatomical information of the aorta [1], [2], such as morphological visualization and measurement. CT is an important imaging technique that provides anatomical information [3]. Aortic segmentation from CT images visualizes the whole morphological structure of the aorta. Clinicians can then better diagnose lesions, such as coarctation and dissection [4], [5]. Clinicians can also better anchor the stent-graft in the uninvolved aortic wall above and below the lesion [1]. Clinical guidelines recommend morphological indexes and measurement methods [2], [3]. For that, clinicians require volume indexes to assess curative efficacy [6]. The size of stent-grafts should match the index of lesions,

such as maximum and minimum diameters [7]. The anterior-posterior diameter should be measured in the longitudinal view, with the diameter perpendicular to the longitudinal axis of the aorta [8]. However, raw axial view of CT images does not meet such requirements. Curved planar reformation (CPR) of the axial view source data can create aortic images in a plane perpendicular to the longitudinal axis [8], [9], as shown in Fig. 1(a). Clinicians rely on CPR for aortic measurements. CPR requires clinicians to manually select an anatomical landmark of the aorta and the viewpoint of its center point. They have to repeatedly adjust the angle of viewpoint until it is perpendicular. There are ten key anatomical landmarks of the aorta, and clinicians need to repeat such operations several times. Hence, aortic segmentation of raw and CPR images in the CT workstation is time-consuming [10]. The result of segmentation needs manual correction. The subjective corrections between observers is inaccurate.

However, aortic segmentation confronts the challenge of variable geometry in space, as shown in Fig.1(b), including intra-image and inter-image challenges. The intra-image challenge refers to geometric relationship between the variable aortic morphology and its spatial position in the long CT sequence of images. The aorta spans from the thorax to the abdomen, resulting in a long CT sequence with many images. Aortic recognizability and morphological regularity gradually decrease in long-sequence images. For example, the lumen is large in the ascending aorta, but it is small in the descending aorta. The ascending aorta is shaped like an ellipse in the axial view, but the aortic arch is irregular. This phenomenon occurs due to the interaction between the anatomical structure and imaging technology. Aortic disease increase the difficulty of segmentation, including low-contrast regions and morphological abnormalities. For example, coarctation of the aorta (CoA) is the low-contrast stenosis in the aortic isthmus on CT images [11]. Type B aortic dissection (AD) has irregular beak signs and a small true lumen compressed by the false lumen (due to tear in the inner wall causing blood-filled channel in the media) [12]. The inter-image challenge refers to geometric relationship between raw and CPR images. This geometric relationship is the geometric transformation produced by different viewing angles. Deep learning (DL) can fit a nonlinear relationship to segment two images. However, DL is hard to learn their geometric transformation due to its "black box" nature. The geometric relationship of images stays independent. DL segmentation is not always based on relevant clinical operation by clinicians, which restricts the diffusion process in the medical domain. DL segmentation should match the high level clinical knowledge to improve its credibility.

The research on geometric constraint has the potential to solve the aforementioned challenges. Geometric constraint defines the properties of target objects relative to other objects [13], including shape, orientation, distance, and perpendicularity, to establish the relationship among different geometries. Constrained properties increase the sensitivity of the model to aortic morphology. Despite these advantages, the existing geometry-constrained methods ignore two crucial issues for aortic segmentation. Different types of aortic diseases, complex anatomy and diverse morphology are averse to the definition of representative properties (geometric features). Inaccurate definitions weaken the constraint on the intra-image aorta. The geometric transformation between raw and CPR images needs Inter-image constraint, which is a spatial topology representing the relationship between different geometries in space. Topology requires that the initial and target object are homeomorphic to preserve the invariance of the geometric properties. The invariance is beneficial to the robustness of constraints in space. Aortic segmentation requires accurate geometric features and suitable spatial topology.

In this paper, we have developed the Deformable Constraint Transport Network (DCTN) for aortic segmentation, as shown in Fig. 1(c). As illustrated in the figure, the DCTN comprises of the extractor, decoder, and guider. DCTN adaptively extracts aortic features to define constrained properties. The DCTN guides the topology in space to constrain the aortic variable geometry. For feature extraction, we design an extractor consisting of deformable self-attention modules. The extractor generates patches of variable sizes and variable positions to adaptively capture the diverse morphology. The variable patches preserve the aortic semantic integrity. The self-attention mechanism of the extractor is beneficial to capture the long-range dependency in long-sequence CT images. Semantic integrity and long-range dependency contribute to the robustness of constraints. The decoder combines local and global features at different scales, which enhances foreground texture information and suppresses low-contrast background information. For topological implementation, we design a geometry-constrained optimal transport (OT) to match CPR images with raw images. The OT explores the graph structure of two images, compares geometric structure, and constructs probability distributions to describe the geometric difference. The guider learns the distance discrepancy of probability distributions to map the geometric transformation between two images. The guider models spatial geometric relationship, which adaptively guides the topology of constraints in space. Finally, DCTN is sensitive to the variable aortic morphology, as DCTN imposes strong constraints on the aortic contour.

Our main contributions are summarized as follows:

1. We develop a clinical tool for aortic segmentation and quantitative measurement from long-sequence CT images to improve the work efficiency of the aortic intervention.

2. DCTN adaptively extracts geometric features and accurately guides spatial topology between raw and CPR images. It constrains the intra-image and inter-image aortic geometry to address the challenge of variable geometry in space.

3. Experiments on the in-house and public CT dataset show that our DCTN outperforms 23 state-of-the-art methods.

This study advances our work presentation in MICCAI2022 [14]: (1) While our MICCAI paper focuses on segmenting the aorta in only one view (the raw view), our current paper goes a step further by simultaneously segmenting the aorta in two views: the axis image of the raw view and the CPR image of the short-axis view. (2) Our current paper focuses on addressing the challenge of geometric transformation and presents the solution. The guider establishes the spatial topology between the two image to constrain the morphological affine relationship. Besides, our current paper further considers

the challenge of semantic confusion and spatial continuity. The decoder removes background interference and refines the foreground feature in low-contrast regions. The extractor further captures the features of adjacent slices to maintain the local spatial continuity in long CT sequences. (3) We add more than 30% cases of aortic subjects on in-house dataset. We also add four public datasets for validation. We discuss our advantages in deciphering different types of aortic diseases and different aortic segments, and measuring clinical indexes.

II. RELATED WORKS

A. Segmentation Methods Based on Deep Learning

DL methods employ neural networks to automatically extract feature from images with minimal manual intervention. DL have always been widely applied in the segmentation of medical images for many years [15]. For example, Lei designs SGU-Net, an ultralight shape-guided network for the segmentation of abdominal medical images [16]. Xing proposes an end-to-end diff-Unet [17], which integrates the denoising diffusion model into a U-shaped architecture for medical volumetric segmentation. Over the past few years, researchers have developed convolution neural networks (CNN)-based methods for aortic segmentation from CT images [18]–[20]. For example, cascaded and coarse-to-fine segmentation methods combine multiple stages to jointly enhance the aortic features, but they rely on the cooperation between stages, thereby resulting in error accumulation. The number of excellent works is increasing. Xiong makes significant progress by applying adversarial learning to enhance CT features in non-enhanced CT images [21]. Feiger proposes a novel variant network, SU-net, to solve some of the limitations of Unet [19]. 3D U-Net, which retains spatial contextual information between slices [22]–[24], improves spatial continuity. However, 3D models require a large memory for long-sequence images [25]. The multi-stage straightening methods simplify the shape of aortic dissection [26], although the back-restoring process is computational expensive, especially for the curved aortic arch. Multi-task methods can simultaneously accomplish various clinical requirements [27], [28], including the aortic segmentation and quantification. Nevertheless, effective coordination of the implicit relationship between tasks is crucial for better results. Researchers have also combined CNN with traditional rule-based methods to improve segmentation performance. For example, multi-scale wavelet analysis was introduced into network to detect aorta through the density distribution within the aortic boundary. However, most CNN-based frameworks only consider the correlation between short-range pixels [29].

B. Segmentation Methods Based on Constraints

Constraint-based methods are an effective approaches for segmentation. These methods artificially set up certain conditions, such as curvature, shape, and orientation, to guide the segmentation process. One popular constraint-based method is the hypothetical prior algorithm which artificially generates a shape-based prior knowledge for segmentation. For example, the aortic segmentation can utilizes a circular shape as a prior knowledge via deformable Hough [30]. Semi-automatic

interactive segmentation methods involve manual intervention to help the model identify regions that may be difficult to distinguish automatically. During training, the model gets corrected and constraints based on new sampling points and reference contours provided by the user [31]. Nonetheless, the hypothetical prior algorithm may not be suitable for complex shapes and cannot effectively segment structures with complex curves, planes, or variation in shapes. On the other hand, Graph-cut constraint method is another viable method that introduces discriminative features and random forests to generate probability maps of self-similarity descriptor [32]. The maps can adaptively smooth the constraints. The above methods are pixel-based local methods, which might lead to false-positive results due to a lack of spatial continuity. In other words, they might fail to catch subtle variations in shape or curvature that occur across multiple pixels. Global models based on curves and meshes are preferred over pixel-based local models in several cases as they offer better spatial continuity and accuracy. These models use discrete meshes or curves to provide a global constraint on the segmentation process. The morphology-guided method is an example of the global constraint. Zhao employs the centerline of the aorta as a morphological guide and proposes novel CRN regression to constrain the boundary distance [33], [34]. Another such approach is mesh-based method. Cascaded network predicts 2D curves among adjacent cross-sectional images, and then integrates them into a 3D mesh [35]. The network regularizes the mesh to ensure the continuity of the curve. However, 2D curves may lose spatial information between slices, making them less effective in cases with large curvatures. Similarly, graph network, an approach to establish graphic constraints between mesh and pixels [36], deforms the nodes of the mesh through a stepwise regression strategy. Overall, global models utilize curves or meshes as constraints, allowing for greater adaptability in complex cases.

III. METHODOLOGY

A. Problem Formulation

Our method is aimed to learn the feature representation of the aorta and constrain the topology between axis image in raw view and CPR image in short-axis view. Raw axis image and its ground truth (GT) are defined as x and y. The CPR image and its GT are defined as x' and y'. Our method (DCTN) needs to solve the problem as follows:

$$F_{DCTN} = \min_{\Omega^*} \frac{1}{M} \sum_{i=1}^{M} L \left(\overset{J}{\underset{j=1}{\boldsymbol{R}}} \left(F_j; \boldsymbol{\Omega_j} \right) \left(\boldsymbol{x}_i, x_i' \right), \left(\boldsymbol{y}_i, y_i' \right) \right) \tag{1}$$

where x and y are raw axis image and its ground truth; x' and y' are CPR image and its ground truth; M is the number of training samples; J is the number of constraint levels and set to 3; L is the loss function; $\boldsymbol{R}$ is the abbreviation of the compound function; Ω is the set of model parameters to be learned. The Eqs.(1) is a holistic problem, which can be viewed as a combination of three problem, namely the segmentation problem of raw axis images (F_1), segmentation problem of curved planar reformation (CPR) images (F_2), and

morphological transformation constraint problem between raw axis image and CPR image (F_3).

For F_1 and F_2, our aim is to find optimal mapping f_1 maps x to y and f_2 maps x' to y', as follows:

$$mapping \quad f_1 : x \xrightarrow{\Omega_1} y \quad ; \quad f_2 : x' \xrightarrow{\Omega_2} y'$$

$$\min_{\Omega} \frac{1}{M} \sum_{i=1}^{M} L_1(f_1(x_i; \Omega_1)), y_i) + L_2(f_2(x'_i; \Omega_1)), y'_i) \quad (2)$$

where M is the number of training samples, $L_{1,2}$ is the loss function, and Ω is the set of model parameters to be learned.

For F_3, our aim is to find optimal mapping f_3 to constrain geometric transformation T between raw axis image and CPR image, which can be expressed as the optimal transmission problem to seek a optimal plan π between raw axis space X and CPR space X' [13], [37], as follow:

$$\min \left\{ \mathcal{P}(\Gamma) + \alpha \int_{\Gamma} |\nabla f_3|^2 dx + \beta \int_{\omega} |f_3 - T|^2 dx \right\}$$
$$= \min \iint_{X \times X'} L_3(x, y; x', y') \, d\pi(x, y) d\pi(x', y') \quad (3)$$

where ω is a connected open set with boundary, T is true constrained rules, α and β are positive coefficients, and $\mathcal{P}(\Gamma)$ is the compact spatial information based on the centerline Γ.

B. Deformable Attention Extractor

The aorta comprised of different CT slices in a long sequence exhibits a variable geometry in space. The transformers are suitable for capturing the long-range context information. The previous developed transformers divide the image into non-overlapping fixed patches [38]. However, the fixed patches destroy the geometric integrity of the aorta, because a complete structure may be divided into different patches. It is difficult for the fixed patches to capture aortic morphology and handle the geometric variations in space. We improve the deformable attention extractor to extract intra-image complete features and inter-slice spatial correlation [39], as shown in Fig. 2(a) .

Given an input axis image $x \in R^{H \times W \times C}$, it is divided into a finite number of fixed rectangle patches. The coordinate points of vertex on each patch and its rectangle size are defined as $p = (p_x, p_y)$ and $s \times s$. From the left-top corner to the right-bottom corner of the image, the coordinates points of all patches can be expressed as $\{(0,0), ..., (H\text{-}1, W\text{-}1)\}$. We normalize them to the range $[(\text{-}1,\text{-}1), ..., (+1,+1)]$. In order to obtain deformable patch, we add an offset $\Delta p = (o_x, o_y)$ and a scale $\Delta s = (s_x, s_y)$, which allow the patch to move around the original center and change the patch size. To obtain the parameter matrix $(\Delta p, \Delta s)$, the input x and all patches are fed into the dense dilated block M_{DDB} for prediction. The M_{DDB} module that includes a dense block and a dilated block, enlarge the receptive field of the image and make dense predictions for all patches. Offsets and scales are predicted as follows:

$$\Delta p = w_p \cdot M_{DDB}(x \cdot W_q), \Delta s = w_s \cdot M_{DDB}(x \cdot W_q) + b \quad (4)$$

where w_p and w_s are learned weights, W_q is the weight matrix of query tokens q, and b is a bias to prevents the size of new patch from being less than 0. The coordinate and size of the

new patch are updated to $p' = p + \Delta p = (p'_x, p'_y)$ and $s' = (s + s_x) \times (s + s_y)$. However, the new patches with different sizes are not conducive to batches of input in the training. To solve this problem, we employ bilinear interpolation to resample the deformable regions [40]. Assuming the deformable attention module (DAM) extracts feature at (p'_x, p'_y), the corresponding feature is as follows:

$$f(x; (p'_x, p'_y)) = \sum_{q_x, q_y} B(p'_x, p'_y; q_x, q_y) \cdot x[q_x, q_y] \quad (5)$$

$$with \ B(p'_x, p'_y; q_x, q_y) = B(p'_x, q_x) B(p'_y, q_y)$$
$$= \max(0, 1 - |p'_x - q_x|) \max(0, 1 - |p'_y - q_y|) \quad (6)$$

where B is the bilinear interpolation function, and $x[q_x, q_y]$ indexes all the locations on x. Since $B(p_x, p_y; q_x, q_y)$ is non-zero only at four integral locations near (p_x, p_y), it can be computed with few weighted average. Besides, in order to capture correlation between adjacent slices, we develop the inter-slice correlation module (ICM). ICM extracts the features of $N/2$ slices before and after the current image x, and then calculate the correlation between the current image and them. Here we take $N = 8$. Each slice is a patch and downsampled 4 times. All slices as patches are converted into tokens to calculate the local spatial correlation. We define multi-head self-attention (MSA) as $MSA(q, k, v)_{module} = softmax(qk^T/\sqrt{dim}) \cdot v$ in the module. Hence, the final output f_E of the extractor is:

$$f_E = FFN(MSA(q, k, v)_{DAM} \oplus MSA(q, k, v)_{ICM}) \quad (7)$$

where FFN is Feed Forward Network.

C. Geometry-Constrained OT Guider

The curved planar reformation (CPR) reorganizes pixel data of raw axis image to create aortic images perpendicular to the longitudinal axis. The geometry-constrained OT guider is designed to learn the morphological transformation between raw axis image and curved planar reformation (CPR) image.

The CPR images hardly change the inherent amount of CT data (mass conservation), as it is obtained from the surface reconstruction of raw axis image. There exists a transformation between the two images. It can be expressed as $x' = T(x; C; \Omega)$, where T represents the morphological transformation; Ω is the learned model parameter; C is the reconstructed matrix about location of the center points, tangential vector of the center points, and distance of the center points to skeleton, respectively. The guider is aimed to constrain the topology between the raw axis images and the CPR images.

The optimal transport (OT) is a great way to solve problems about constraints for convex geometry and learn geometric affine transformations. OT can distinguish geometric discrepancies in space and construct probability distributions of their discrepancies, even when there are non-overlapping supports. OT aims to seek a optimal plan π to constrain the transformations T and map the measure μ in raw space X onto another measure ν in CPR space X', during which the law of mass conservation needs to be satisfied (the transformation is based on the centerline). Let $c : X \times X'$ be a cost function where $c(X, X')$ measures the cost of transporting from point $x_i \in X$
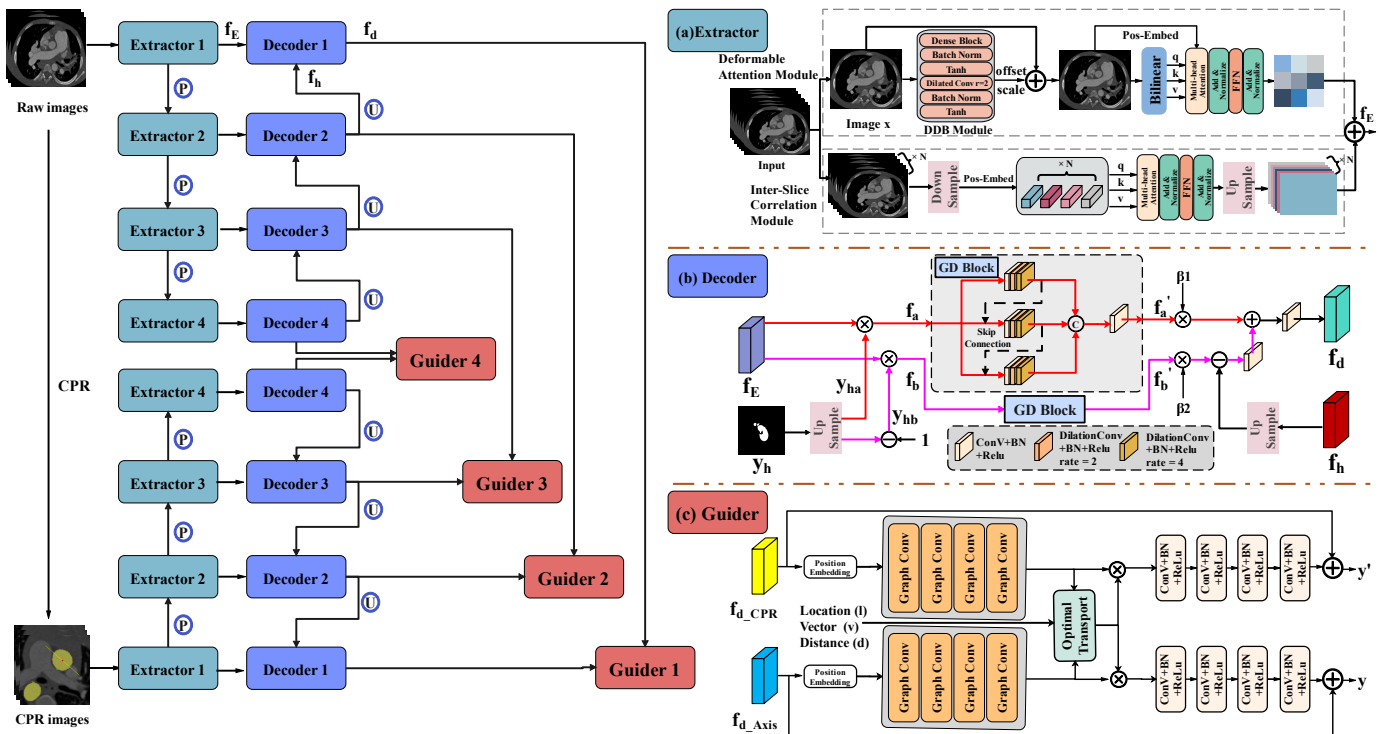
Fig. 2. The network details of the DCTN, including the deformable attention extractor (Extractor), the hierarchical geometry-aware decoder (Decoder), and the geometry-constrained OT Guider (guider).

to point $x_i' \in X'$. The problem is to transport μ to ν whilst minimizing the cost c [41], as follows:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{X \times X'} c(x_i, x_i') \mathrm{d}\pi(x_i, x_i') \qquad (8)$$

where Π is the convex set of all probability measures. The model measure the probability distribution between approximate geometric domains (spaces). The formula can be simplified to Wasserstein distance d_W by Kantorovich Duality [42]:

$$T \to d_W = \min_{(\varphi, \psi)} \int_X \varphi(x_i) \mathrm{d}\mu(x_i) + \int_{X'} \psi(x_i') \mathrm{d}\mu(x_i') \quad (9)$$

To calculate the d_W, We use graph convolutions to explore the discrepancy of graphic structure between the raw and the CPR image. CPR reconstructs the vessel along center points, which gets short-axis images perpendicular to the centerline. We calculate the aortic skeleton from the mask by skeletonization algorithm, then extract the longest path of the skeleton, and set it as the centerline of the whole aorta. The reconstructed image matrix can be represented as $C = \{C_m | m = 1, 2, ..., r\}, C_m = (l_m, v_m, d_m)$, where r is the number of center points; and l_m, v_m and d_m are respectively location, tangential vector, and distance to skeleton of rth center point. The graph convolution extracts the characteristic information of each adjacent center point, which effectively analyzes the relationship between two kinds of imaging data and non-linearly models the relationship between the internal elements in sequence. Graph convolution is defined as [43]:

$$G^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} G^{(l)} W^{(l)} \right) \qquad (10)$$

where G and $G^{(l+1)}$ are the feature of input and output, respectively. σ is activation function. $\tilde{A} = A + I$ is the

adjacent matrix of CPR with position embedding. I is the identity matrix. $\tilde{D}$ is the degree matrix of $\tilde{A}$. The output is the weighted sum of the features at C. We utilize graph convolution to capture the node relationship of raw image and CPR image, i.e. $\varphi(x_i)$ and $\psi(x_i')$. Geometric information is combined with images for position embedding. All the features are input into the OT guider, as shown in Fig. 2(c). The error of the output continuously optimizes parameters through the back propagation of the loss function:

$$L_{guider} = \arg\min_C \frac{1}{M} \sum_{i=1}^{M} \frac{1}{r} \sum_{m=1}^{r} ||C_{Seg}'^{(i,m)} - C_{Gt}'^{(i,m)}|| \quad (11)$$

where C' is matrix C with the minimum Wasserstein distance and $C' = C \mid d_W^{min}(G(x), G(x'))$; $||C'||$ is the Frobenius norm of matrix C'; $C_{Seg}'^{(i,r)}$ and $C_{Gt}'^{(i,r)}$ are the predicted and real labels of mth point in ith sample, respectively; M is the number of samples. The guider learns the inter-image geometric transformation to realize the variation reasoning of the topology on the aorta. The topology develops the aortic structure with similarity and homogeneity in space, while the non-aortic region has greater anisotropy. The guider imposes strong constraints on the aortic contour via marginal probability, and corrects the semantic features.

D. Hierarchical Geometry-Aware Decoder

The hierarchical geometry-aware decoder (HGAD) aims to solve the semantic confusion of the low-contrast foreground regions. The low-contrast foreground regions are common in the cases of CoA and AD (false lumen). Low-contrast foreground regions are often similar to the background, so

aortic segmentation suffers from false positive and false negative predictions. Our HGAD perceives the foreground and background, then removes these false predictions and background interference, and finally refines the feature maps.

The framework of the HGAD is shown in Fig. 2(b). First, HGAD upsamples the higher-level foreground prediction y_h. Then HGAD multiplies y_{ha} by the current-level features f_E of the deformable attention extractor. Finally, HGAD generates the foreground-attentive features f_a. Similarly, HGAD multiplies the reverse version of y_{ha} by f_E to generate the background-attentive features f_b. The two types of features are fed into the geometry distraction (GD) block in parallel to detect the foreground false-positive distractions f'_a and the background false-negative distractions f'_b [44], respectively. The GD block contains three parallel branches for contextual semantic exploration. Each branch includes a convolution and two dilated convolution with a dilation rate of 2 and 4. The outputs of the first and second branches are respectively fed to the second and third branches via skip connections. All outputs of the three branches are finally concatenated and fused. This process is expressed as the following formula:

$$f'_a = GD((f_E \otimes y_{ha})), \ f'_b = GD((f_E \otimes (1 - y_{ha}))) \quad (12)$$

The parallel multi-scale structure allows HGAD to analyze features and perceive interference at different scales. Multiple dilated convolution operations obtain large ranges of receptive fields to adapt to scale variations and explore distracting features in the foreground and background. The skip connections cascade different branches to enhance the depth of exploration and improve the strength of perception. In total, the structure of the GD block are beneficial to perceive contexts in a wide range of receptive field, which is used for contextual distraction reasoning and interference detection.

HGAD needs to remove the interference after interference detection. HGAD utilizes element-wise subtraction to suppresses background interference (false-positive predictions), and then utilizes element-wise addition of the upsampled high-level features f_h to enhance foreground feature (false-negative predictions). The refined output feature f_d of HGAD is expressed as the following formula:

$$f_d = ConV(\beta_1 f'_a + ConV(Up(f_h) - \beta_2 f'_b)) \quad (13)$$

where $ConV$ represents the convolution, batch normalization and ReLU function, Up is the upsample operation, and β_1 and β_2 are the learnable scale parameters to achieve the suppression and enhancement. HGAD does not directly supervise interference, as the appearance of interference is a dynamic process in the imaging setting. Therefore, we incorporate ground truth of background and foreground as a supervision constraint. This constraint helps the GD block to learn diverse representation of interference, and help HGAD to remove interference and refine features in an implicit manner.

IV. EXPERIMENTAL STUDIES

A. Dataset

This study conducted a retrospective search for aortic patients referred for CT from October 2016 to June 2022 in the Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences). A total of 267 patients were enrolled, including 73 normal aorta, 74 CoA, and 120 AD. Our work was exempted from formal ethical approval by the Medical Research Ethics Committees of the hospital.

Aortic examinations were performed with 64-slice (America, GE LightSpeed VCT) and 256-slice (America, Philips Brilliance iCT) wide CT scanner. An average of 100ml of contrast agents (China, iopamidol 370, BRACCO) was injected into the antecubital vein, followed by 30ml of saline solution. The bolus tracking technique (BolusPro, Philips Healthcare) was simultaneously used for scanning. The scanning ranged from the thoracic inlet to the beginning of the bilateral femoral artery. The volume size of images in our dataset is $512 \times 512 \times S(300 \leq S \leq 800)$. The CT imaging parameters are as follows:(1) Siemens: tube voltage is 70 kV; slice increment is 0.50mm; slice thickness is 0.75mm; (2) 256-slice Philips: tube voltage is 120 kV; slice increment is 0.75mm; slice thickness is 1.5mm; (3) 256-slice Philips: tube voltage is 120 kV; slice increment is 1mm; slice thickness is 1mm.

All imaging data were recorded in the picture archiving and communication system (PACS). The aortic images were processed by the radiologists with 6-year experience in CT imaging. We implement quality control to ensure the quality of annotations. First, a radiologist (annotator) manually annotates the images as ground truth (GT) in a voxel-wise manner. The average annotation time is approximately 4 seconds per image (20 minutes per case). Then, another radiologist (checker) reviews the work for inaccuracies. Finally, if the checker confirms any incorrect annotations in an image, it is sent back to the annotator for re-annotation. All images have passed the quality verification. We also measured the aorta along centerline. All monitored diameters are the outer diameter, and the results are the average of two measurements.

B. Implementation Details

Our framework is implemented based on PyTorch 1.7. All data are trained and tested on two NVIDIA RTX A6000 GPUs with 48GB memory. We count the window level and window width of the aorta in the dataset and readjusted window settings of 500 Hounsfield units (HU) for window width and 850 HU for window level. The images are rescaled to a size of $256 \times 256 \times 512$ to reduce the pressure of calculation and memory. The pixel intensities are normalized to [0, 1]. We do not use data augmentation techniques on the dataset. During the training, we employ the Adam optimizer to minimize the loss function, with a batch of 16 images per step. The initial learning rate is 0.01. It gradually decay in a step-wise manner to 0.1 times of the previous every 10 epochs. The momentum is 0.9 and the weight decay is 0.005 in a total of 200 epochs. we apply an early stopping strategy with a patience of 10 epochs. It is a common approach to set a large epoch in this strategy, as the model needs enough epochs to complete training [45]. If the validation loss error does not improve for 10 consecutive epochs, the training process will be halted [25]. This strategy allows us to terminate the training early. We use the stratified ten-fold cross-validation respecting independent subjects with
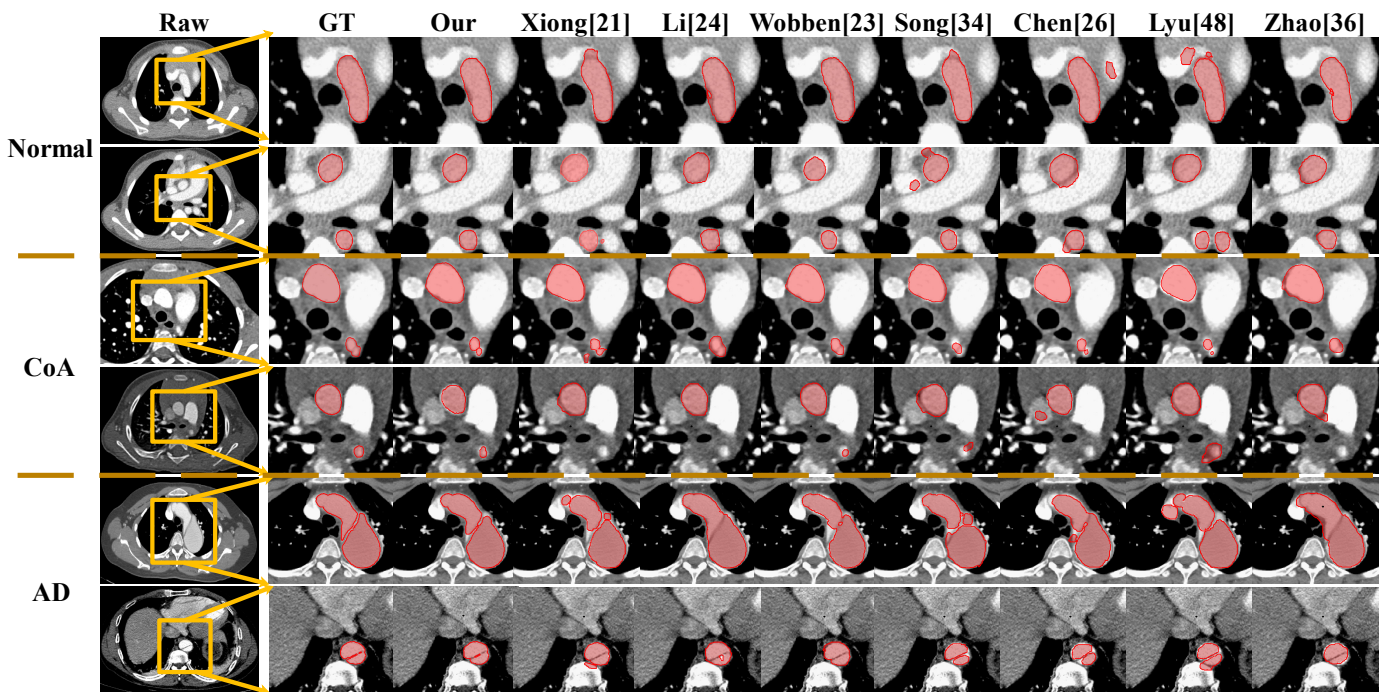
Fig. 3. Representative performances of aortic segmentation in our DCTN and other state-of-the-art seven methods. Types of the aorta studied include the normal aorta, coarctation of the aorta (CoA), and aortic dissection (AD).

eight folds as the training set, one fold as the validation set, and one fold as the testing set. Besides, we perform the grid search for the hyperparameter optimality of our method. Based on the optimal hyperparameters, we obtain performance on the testing set as the final experimental results. Our code can be available at https://github.com/HICL-SYSU/DCTN.git.

C. Evaluation Metric

To quantitatively evaluate the performance, we employ metrics as follows: dice similarity score (DSC), Hausdorff distance (HD), precision (Pre), recall, root mean square distance (RMSE) and average symmetric surface distance (ASSD) [46].

D. Comparison of DCTN with the State-of-the-Art Aortic Segmentation Methods

To demonstrate the superiority of our DCTN, we have compared it with the existing methods of aortic segmentation, including Xiong [21], Sieren [27], Li [24], Feiger [19], Wobben [23], Song [34], Hahn [28], Abdolmanafi [47], Chen [26], Lyu [48], Zhao [36], Deng [32], Yu [49], Cheng [50], and Cao [22]. We visualize the segmentation results of DCTN and other top seven comparison methods in Fig. 3. The results clearly show that DCTN can better segment the aorta, including small and irregular vessels. In the case of coarctation of the aorta (CoA), small vessels with low contrast values are non-significant, but DCTN still learns non-significant features. In the case of aortic dissection (AD), the aortic arch and descending aorta are irregular and complex, but DCTN still captures their anatomy. The segmentation is highly similar to the GT, which proves the potential of our DCTN.

The upper half of Table I records the quantitative results of these above mentioned methods for objective evaluation. The

results show that our DCTN gets the largest DSC of 0.949 on the testing set. The DCTN outperforms other methods by more than 1.50% DSC. The DCTN also gets the smallest RMSD and HD, which are 3.51 mm and 3.45 mm, respectively. To better analyze the aortic segmentation performance, the whole aorta is divided into ascending aorta, aortic arch and descending aorta according to the anatomy in the clinical guidelines [2], [8]. The DSC of the ascending aorta is 0.951, outperforming other methods by over 1.93%. The average DSC of the aortic arch is 0.941, outperforming other methods by over 0.86%. The DSC of the descending aorta is 0.950, also outperforming other methods by over 1.28%. Overall, our DCTN has more advantages in aortic segmentation, especially for the variable morphology and complex structure. It is worth mentioning that our model performances better than others when faced with irregular aortic arch and low-contrast descending aorta.

We verify the model performances of DCTN and other methods in different types of aortic diseases. The quantitative results are recorded in the upper half of Table II. The DCTN achieves the best performance in all evaluation metrics, including DCS, and ASSD. The DSC of the normal aorta is 0.952, outperforming other methods by more than 0.63%. The DSC of CoA is 0.944. The segmentation performance is slightly less than of the normal aorta. The reason is that the intensity of the contrast agent in the stenosis is too low. However, DCTN still shows excellent effectiveness on low-contrast stenosis. The DSC of our model outperform other methods by more than 1.83%, with the ASSD less than 2.17%. The DSC and ASSD of AD are 0.936 and 3.22; the segmentation performance is greatly improved by more than 0.97% and 3.31% compared with the other methods. For the true and false lumen in the case of AD, the dice coefficients are 0.938 and 0.930, with

TABLE I

COMPARISON OF OUR MODEL WITH OTHER METHODS FOR AORTIC SEGMENTATION, ACCORDING TO ANATOMY. THE UPPER PART IS THE SOTA AORTIC SEGMENTATION METHODS. THE MIDDLE IS THE MEDICAL SEGMENTATION MODEL. THE LOWER PART IS THE COMPUTER VISION SEGMENTATION BENCHMARKS

Segments / Methods	Whole Aorta			Ascending aorta			Aortic arch			Descending aorta		
	DSC ↑	RMSD ↓	HD ↓	DSC ↑	RMSD ↓	HD ↓	DSC ↑	RMSD ↓	HD ↓	DSC ↑	RMSD ↓	HD ↓
Xiong [21]	0.925	3.77	3.66	0.930	3.70	3.61	0.914	3.83	3.71	0.926	3.74	3.64
Sieren [27]	0.913	3.84	3.72	0.915	3.83	3.72	0.915	3.83	3.71	0.910	3.84	3.73
Li [24]	0.935	3.64	3.56	0.931	3.67	3.58	0.933	3.66	3.60	0.938	3.62	3.54
Zhao [33]	0.914	3.86	3.71	0.919	3.82	3.73	0.906	4.11	4.03	0.918	3.85	3.71
Feiger [19]	0.918	3.82	3.69	0.924	3.76	3.64	0.906	4.02	3.96	0.920	3.79	3.68
Wobben [23]	0.931	3.67	3.59	0.933	3.68	3.61	0.928	3.72	3.63	0.933	3.64	3.55
Song [34]	0.927	3.72	3.62	0.932	3.68	3.59	0.926	3.74	3.61	0.925	3.76	3.62
Hahn [28]	0.906	4.02	3.96	0.912	3.84	3.72	0.898	4.13	4.08	0.914	3.83	3.70
Abdolmanafi [47]	0.917	3.88	3.71	0.920	3.84	3.69	0.907	4.01	3.89	0.920	3.82	3.70
Chen [26]	0.921	3.78	3.67	0.923	3.77	3.63	0.913	3.83	3.75	0.919	3.81	3.69
Lyu [48]	0.918	3.83	3.68	0.922	3.80	3.67	0.906	4.04	3.95	0.920	3.81	3.67
Zhao [36]	0.927	3.77	3.65	0.931	3.67	3.60	0.914	3.84	3.72	0.928	3.74	3.60
Deng [32]	0.911	3.85	3.76	0.907	4.04	3.97	0.912	3.85	3.75	0.916	3.81	3.71
Yu [49]	0.904	4.05	3.97	0.917	3.83	3.69	0.897	4.12	4.08	0.904	4.04	3.95
Cheng [50]	0.897	4.14	4.10	0.903	4.06	3.98	0.901	4.07	4.00	0.893	4.18	4.15
Cao [22]	0.895	4.17	4.13	0.896	4.13	4.12	0.895	4.15	4.12	0.894	4.17	4.16
Unet [51]	0.893	4.18	4.15	0.910	3.85	3.74	0.886	4.33	4.28	0.905	4.02	3.93
TransUnet [52]	0.920	3.77	3.69	0.926	3.78	3.62	0.914	3.87	3.68	0.917	3.79	3.68
PSPNet [53]	0.896	4.15	4.11	0.903	4.05	3.95	0.892	4.21	4.20	0.902	4.05	3.97
SegNet [54]	0.888	4.33	4.29	0.892	4.23	4.21	0.885	4.34	4.29	0.887	4.35	4.30
ViT [55]	0.902	4.09	4.02	0.904	4.02	3.92	0.891	4.19	4.17	0.903	4.06	3.91
SwinTrans [56]	0.918	3.81	3.74	0.911	3.82	3.75	0.916	3.82	3.67	0.923	3.76	3.70
SegFormer [57]	0.913	3.84	3.81	0.908	4.06	3.96	0.910	3.84	3.74	0.926	3.75	3.59
Ours	**0.949**	**3.51**	**3.45**	**0.951**	**3.47**	**3.34**	**0.941**	**3.61**	**3.53**	**0.950**	**3.48**	**3.39**

TABLE II

COMPARISON OF OUR MODEL WITH OTHER METHODS FOR AORTIC SEGMENTATION, ACCORDING TO TYPES OF DISEASES. THE UPPER PART IS THE SOTA AORTIC SEGMENTATION METHODS SUGGESTED BY THE REVIEWER. THE MIDDLE IS THE MEDICAL SEGMENTATION BENCHMARKS. THE LOWER PART IS THE COMPUTER VISION SEGMENTATION BENCHMARKS. UNIT ABBREVIATIONS, M:MILLION; MS:MILLISECOND.

Diseases / Methods	Normal Aorta		CoA		AD		True Lumen		False Lumen		Parameter Size (M)	Inference Time (ms)
	DSC ↑	ASSD ↓	DSC ↑	ASSD ↓	DSC ↑	ASSD ↓	DSC ↑	ASSD ↓	DSC ↑	ASSD ↓		
Xiong [21]	0.928	3.33	0.924	3.41	0.917	3.58	0.904	3.92	0.922	3.46	75.4	63 ± 3.7
Sieren [27]	0.915	3.65	0.918	3.57	0.908	3.84	0.914	3.69	0.898	4.07	47.2	39 ± 2.9
Li [24]	0.946	3.13	0.930	3.23	0.922	3.35	0.923	3.32	0.921	3.50	51.6	44 ± 3.1
Zhao [33]	0.920	3.49	0.914	3.69	0.912	3.68	0.916	3.55	0.910	3.72	40.1	41 ± 2.3
Feiger [19]	0.920	3.51	0.922	3.44	0.910	3.81	0.915	3.64	0.901	3.99	63.4	56 ± 2.5
Wobben [23]	0.931	3.30	0.935	3.29	0.924	3.42	0.927	3.36	0.919	3.54	55.8	48 ± 1.8
Song [34]	0.923	3.43	0.925	3.38	0.918	3.54	0.918	3.57	0.916	3.59	56.4	47 ± 3.0
Hahn [28]	0.912	3.76	0.906	3.89	0.900	4.04	0.903	3.95	0.894	4.17	33.7	37 ± 1.6
Abdolmanafi [47]	0.921	3.52	0.918	3.54	0.914	3.63	0.916	3.62	0.911	3.75	53.4	50 ± 2.2
Chen [26]	0.924	3.41	0.919	3.55	0.917	3.59	0.923	3.44	0.915	3.63	52.2	46 ± 4.2
Lyu [48]	0.916	3.60	0.924	3.41	0.915	3.62	0.917	3.58	0.912	3.76	57.5	53 ± 2.3
Zhao [36]	0.931	3.25	0.925	3.45	0.927	3.48	0.928	3.45	0.919	3.55	46.8	47 ± 3.7
Deng [32]	0.919	3.54	0.913	3.72	0.904	3.91	0.910	3.80	0.900	4.03	40.3	42 ± 2.8
Yu [49]	0.908	3.84	0.905	3.89	0.895	4.15	0.907	3.88	0.893	4.22	34.6	40 ± 1.9
Cheng [50]	0.905	3.88	0.903	3.93	0.896	4.12	0.896	4.13	0.894	4.18	88.2	105 ± 3.5
Cao [22]	0.907	3.86	0.901	4.00	0.893	4.20	0.901	4.02	0.892	4.24	42.1	40 ± 3.3
Unet [51]	0.906	3.87	0.901	4.02	0.887	4.37	0.893	4.21	0.885	4.41	28.9	38 ± 2.1
TransUnet [52]	0.923	3.44	0.917	3.55	0.919	3.52	0.925	3.38	0.912	3.77	126.2	206 ± 6.8
PSPNet [53]	0.906	3.88	0.906	3.87	0.890	4.29	0.901	3.98	0.889	4.28	48.8	53 ± 2.6
SegNet [54]	0.901	3.99	0.898	4.08	0.883	4.46	0.887	4.39	0.880	4.52	29.5	35 ± 2.4
ViT [55]	0.908	3.85	0.902	3.96	0.896	4.12	0.903	3.94	0.893	4.21	85.6	101 ± 5.6
SwinTrans [56]	0.921	3.50	0.916	3.58	0.924	3.48	0.922	3.47	0.923	3.42	88.3	95 ± 4.1
SegFormer [57]	0.918	3.55	0.912	3.76	0.906	3.89	0.913	3.72	0.903	3.92	64.1	58 ± 2.6
Ours	**0.952**	**3.07**	**0.944**	**3.16**	**0.936**	**3.22**	**0.938**	**3.21**	**0.930**	**3.26**	73.2	61 ± 2.4

the ASSD of 3.22 and 3.92. Our model has a strong ability to capture morphological information to still maintain smaller errors. DCTN has advantages in many types of aortic diseases.

E. Comparison of DCTN with Medical Benchmarks and Computer Vision Segmentation Benchmarks

We compare DCTN with models commonly used for computer vision in recent years, mainly including Unet [51], TransUNet [52], PSPNet [53], SegNet [54], Vision Transformer (ViT) [55], Swim Transformer (SwinTrans) [56] and SegFormer [57]. The lower half of Table I and Table II both record the evaluation results of these benchmark models. SegNet achieve relatively inferior segmentation performance, with a DSC of 0.888 for the whole aorta. The results of Unet and PSPNet are slightly better than SegNet, with the DSC of 0.893 and 0.896, indicating that the skip connection and multi-

scale pyramid structure is beneficial to strengthen learning for small features in the descending aorta and polymorphic features in the aortic arch, but the effect is not obvious. Despite the face that U-structure is a robust structure for feature representation of medical images [58], U-Net relies heavily on convolutional layers and down-sampling to handle CT images. Its structure maybe lead to the loss of spatial resolution and crucial aortic details, especially in areas such as aortic intima, break of aortic dissection, and coarctation. U-Net's encoder-decoder architecture with skip connections may limit the model's ability to capture long-range dependencies and maintain spatial information. Overall, U-Net primarily focuses on local feature representations, which may struggle to capture fine-grained details, various morphology, geometric relationships between distant regions, and a large receptive field. The experimental results show that CNN-based models with short-range dependency are hard to learn long-range dependencies, so the CNN models are not suitable for capturing diverse shape information of the aorta in the long CT sequence. The local receptive field of the models based on short-range dependence limits the perception range and ignores some global information.

We further compare our DCTN with seven 3D segmentation methods. These methods include nn-Unet [59], 2.5D Unet [60], 3D Unet [19], 3D Res-Unet [23],3D Attention-Unet [34], Vnet [19] and SU-Net [19]. The results are shown in the Table III. Our network achieves a DSC of 0.949, while the other best-performing method gets a Dice coefficient of 0.935. Our segmentation results also achieve the highest performance on different aortic diseases (normal, CoA, AD, true lumen and false lumen) and different anatomical segments (ascending aorta, aortic arch and descending aorta). The results indicate that our method DCTN outperforms these 3D networks. The extractor includes inter-slice correlation module (ICM). The ICM learns the spatial correlation between the z-axis slices. Therefore, our DCTN can learn spatial correlation in the depth dimension to obtain contextual information. Our DCTN approximates the ability of learning contextual information in 3D networks [61]. We also evaluate other model based on self-attention mechanism. The performance of ViT is a little better, with the DSC of 0.908. Models based on the self-attention mechanism have better performance due to their advantages in learning long-range features of long-sequence images. Multi-scale feature fusion of SegFormer improves the segmentation performance, which is similar to the network structure of Unet and PSPNet. TransUnet combines the advantages of Unet and transformer, and achieves a DSC of 0.923, but the parameter size and inference time are relatively large. It can be found that transformer is still not good enough in the case of CoA and AD. Overall, these transformers divide the complete aortic structure into different patches. Traditional patches with fixed size and position seriously destroy the integrity of semantic features, so diverse morphology (dissection and arch) and small vessels (stenosis) produce false predictions with high confidence. These benchmarks demonstrate that the potential of DCTN employment in aortic segmentation.

We employ the Wilcoxon Signed-Rank test to evaluate the statistical significance of the scores in TABLE I and TABLE II. The test includes the whole aorta, different aortic segments and different types of aortic disease. The p-values in this test are all less than the predetermined significance threshold of 0.05 ($p < 0.05$). The scores of our method is better than other segmentation methods. Therefore, there is a significant improvement in our method compared to other methods.

F. Quantitative Analyses of Clinical Indexes

According to multiple guidelines prescribed by the European Society of Cardiology, American Heart Association and European Association of Cardiovascular Imaging [1], [2], [8], we have validated the recommended clinical indexes of the aorta, including volume, maximum and minimum diameter, the diameter at the coarctation, the diameter at the tear, and the diameter at ten key landmarks : sinuses of Valsalva, sinotubular junction, mid ascending aorta, proximal aortic arch (aorta at the origin of the brachiocephalic trunk), mid aortic arch (between left common carotid and subclavian arteries), proximal descending thoracic aorta ($20mm$ distal to left subclavian artery), mid descending aorta (level of the pulmonary arteries), diaphragm, celiac axis origin, and bifurcation.
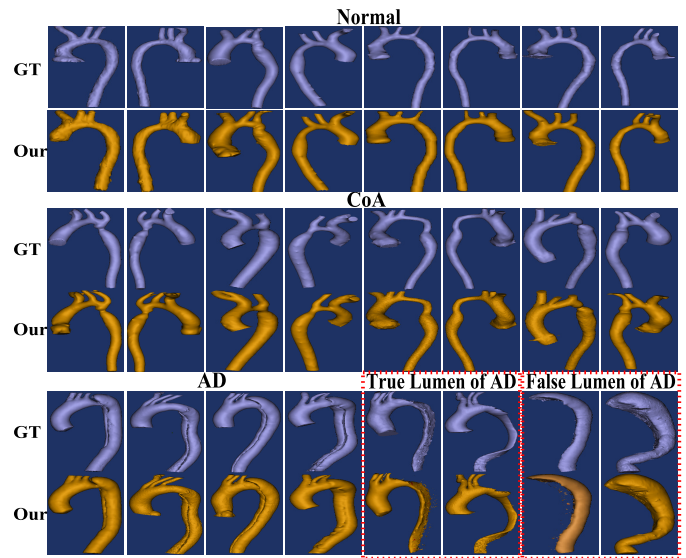

Fig. 4. 3D visualization segmented by our DCTN in patients with normal aorta, CoA and AD (including true lumen and false lumen).

1) Volume: We reconstruct the 3D aorta to intuitively visualize the overall anatomy, as shown in Fig. 4. The segmentation results are seen to be consistent with the aortic structure and our results are close to the GT in different viewpoints. In the case of the normal aorta, main and branch vessels are well segmented, including the ascending, arch, descending aorta. In the case of CoA, stenosis is segmented well without severe under-segmentation or fracture surface. In the case of AD, our DCTN can segment irregular and complex morphology, especially at the tear between true lumen and false lumen. To fully display the segmentation result in the lumen, we further divide AD into the true lumen and the false lumen. It is found that DCTN can well preserve the shape of the true and false lumen, but there are some adhesion between the two lumen. The reason is that the rupture of the intima

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2023.3339142

10   IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2022

TABLE III
COMPARISON OF OUR METHOD WITH OTHER 3D NETWORKS BASED ON U-NET.

| Methods	Whole Aorta	Ascending aorta	Aortic arch	Descending aorta	Normal	CoA	AD	True Lumen	False Lumen
nnU-Net [59]	0.933	0.934	0.931	0.935	0.942	0.936	9.926	0.930	0.924
2.5D U-Net [60]	0.908	0.913	0.905	0.912	0.914	0.907	0.905	0.903	0.908
3D U-Net [19]	0.904	0.911	0.897	0.909	0.908	0.904	0.902	0.904	0.901
3D Res-Unet [23]	0.931	0.931	0.928	0.933	0.931	0.930	0.924	0.927	0.919
3D AttentionUnet [34]	0.927	0.932	0.926	0.925	0.923	0.925	0.918	0.918	0.916
V-Net [19]	0.912	0.922	0.904	0.910	0.916	0.915	0.906	0.910	0.904
SU-Net [19]	0.925	0.931	0.918	0.923	0.930	0.927	0.918	0.922	0.915
Ours	**0.949**	**0.951**	**0.941**	**0.950**	**0.952**	**0.944**	**0.936**	**0.938**	**0.930**
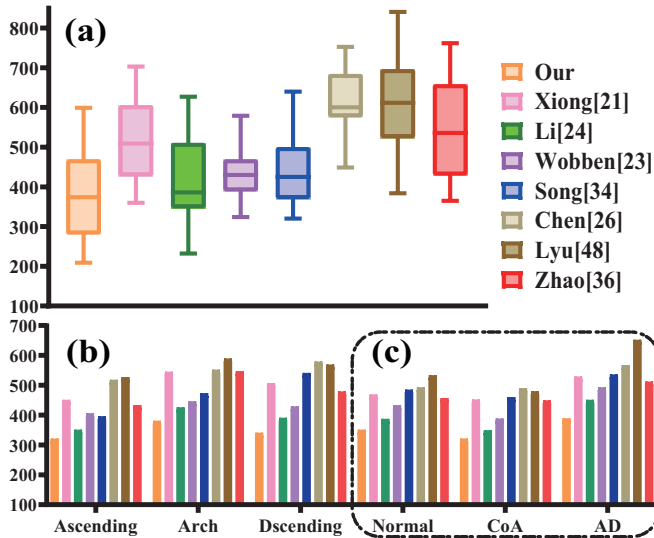
Fig. 5.   Comparison of volume error between our DCTN and other advanced methods. (a) is for the whole aorta, (b) is for different segments and (c) is for different diseases. The unit in the y-axis is mm^3.

Fig. 6.   Bland-Altman analysis and Pearson correlation are used for statistics. High consistent between our DCTN and manual GT.

does not completely separate the two lumen. The physiological structure determines that the distance between the two lumen is very close, which has a negligible effect on volume and diameter measurements. There are a few protrusions and roughness on the aortic surface. However, they do not affect clinical diagnosis, because the aorta is not a smooth tubular structure. We quantify aortic volume, including the whole aorta, ascending, arch, descending, normal aorta, CoA and AD. The volume errors are shown in Fig. 5. The mean error of volume is 374 mm^3 in the whole aorta, 323 mm^3 in the ascending, 381 mm^3 in the arch, 342 mm^3 in the descending, 350 mm^3 in the normal aorta, 322 mm^3 in the CoA , and 386 mm^3 in the AD. Among all methods of aortic segmentation, our model gets the smallest error. Our DCTN is thereby shown to be suitable for volume estimation in clinical practice.

2) Diameter Measurement: We automatically calculate aortic diameters and compare them with their actual values measured by the clinician. The measurements include maximum diameter, minimum diameter, diameter at the coarctation and diameter at the tear. Bland-Altman analysis and Pearson correlation analysis are used for evaluation, as shown in Fig. 6. The average bias and correlation coefficient R of the maximum diameter are $1.2mm$ and 0.92, while those of the
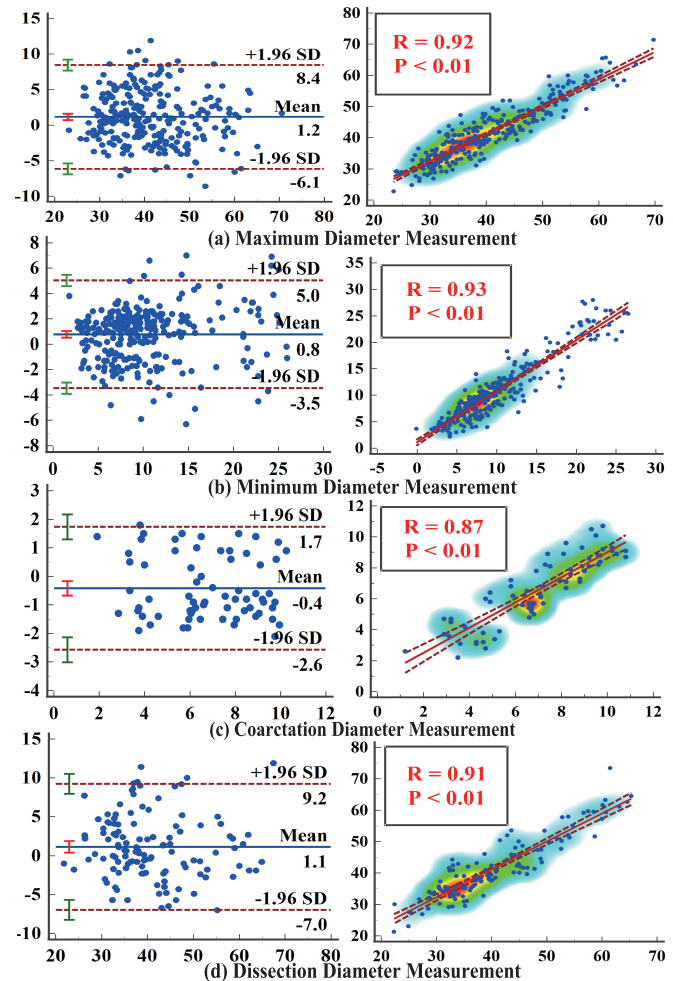
minimum diameter are $0.8mm$ and 0.93. The average bias and correlation coefficient of the stenosis are $-0.4mm$ and 0.87. Stenosis with low contrast intensity is under-segmented. The fluctuation range of the error is relatively small, as can be seen from the confidence interval. The average bias and correlation coefficient of the tear are $1.1mm$ and 0.91. The DCTN only produces a small amount of over-segmentation in the edge area, and so the bias is relatively small. All measurements are statistically significant with p-values less than 0.01. These results indicate that there is no significant difference between
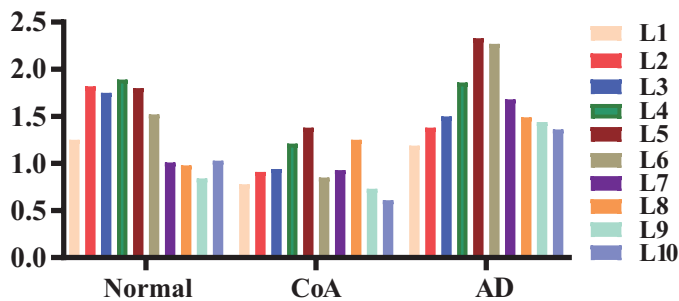
Fig. 7. Diameter errors of our DCTN and GT at different aortic landmarks. Sinuses of Valsalva (L1), sinotubular junction (L2), mid ascending aorta (L3), proximal aortic arch (L4), mid aortic arch (L5), proximal descending thoracic aorta (L6), mid descending aorta (L7), diaphragm (L8), celiac axis origin (L9), and aortic bifurcation (L10).

our DCTN and manual measurement. Besides, we also measure the diameter of the ten key landmarks and record them as L1 to L10, as shown in Fig. 7. The results indicate that there is no significant difference between our DCTN and manual measurements. The DCTN performs well on the measurement.

G. Ablation Study

1) Effectiveness of Deformable Attention Extractor: We decouple each key component from the DAM to verify the effectiveness of the offset, scale and position embedding, as shown in Table IV. Offset improves our model performance the most when one of the components is used alone, with a DSC of 0.912, followed by scale with a DSC of 0.907, and finally embedding with a DSC of 0.901. When two components are employed together, the combination of offsets and scales improves model performance by 2.79%, with a DSC of 0.920. Embedding works better in combination with offset and scale. All components increase the DSC by 0.031, and improve the performance by more than 3.46%. It is seen that offset, scale and embedding are vital components of our DAM. The offset moves more patches to aortic areas, the scale enables better preservation of the complete semantic features, and the embedding coordinates these components.

TABLE IV
DECOUPLE OF OFFSET,SCALE AND EMBEDDING IN DAM. "$\sqrt{}$" MEANS TO PERFORM THE OPERATION IN THE MODULE, WHILE "$\times$" MEANS NOT.

Offset	Scale	Eembedding	DSC ↑	Pre ↑	Recall ↑
$\times$	$\times$	$\times$	0.895	0.898	0.887
$\sqrt{}$	$\times$	$\times$	0.912	0.915	0.906
$\times$	$\sqrt{}$	$\times$	0.907	0.903	0.908
$\times$	$\times$	$\sqrt{}$	0.901	0.896	0.904
$\sqrt{}$	$\sqrt{}$	$\times$	0.920	0.924	0.918
$\sqrt{}$	$\times$	$\sqrt{}$	0.916	0.917	0.913
$\times$	$\sqrt{}$	$\sqrt{}$	0.913	0.908	0.915
$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	**0.926**	**0.928**	**0.922**

We conduct experiments to evaluate at which stage the DAM should be integrated into the extractor. We replace the original self-attention block with the DAM at different stages. Besides, we add an extra stage, Stage 5, to illustrate the impact of the number of stages on aortic segmentation performance. According to the results shown in Table V, DAM is minimally effective when applied in Stage 1. The reason may be that the

raw image contains more messy geometric features and fewer high-level semantic features in the early stage. Segmentation performance with a DSC of 0.926 is highest at Stage 4, while it begins to decline after Stage 5. Hence, integrating the DAM before Stage 4 can yield the optimal results.

TABLE V
ABLATION STUDY OF DAM IN DIFFERENT STAGES. "$\sqrt{}$" MEANS DAM IS USED IN THIS STAGE OF OUR DCTN.

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	DSC ↑	Pre ↑	Recall ↑
					0.895	0.898	0.887
$\sqrt{}$					0.901	0.906	0.895
$\sqrt{}$	$\sqrt{}$				0.908	0.912	0.904
$\sqrt{}$	$\sqrt{}$	$\sqrt{}$			0.916	0.919	0.913
$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$		**0.926**	**0.928**	**0.922**
$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	0.920	0.924	0.918

We verify the ability of the ICM to extract contextual information in space. We calculate the feature correlation between the current slice and the two slices before and after it, where the current slice is marked as CS, the two slices before current slice are marked as $CS - 2$ and $CS - 1$, and the two slices after current slice are marked as $CS + 1$ and $CS + 2$. We visualize feature correlations between slices through heat maps. The results are shown in Fig. 8(a). The results show good spatial correlation between slices, with the correlations of 0.78-0.93. The correlation between adjacent slices is more than 0.86. The results prove that our ICM can better capture the sequence relationship between slices in space, extract the local feature correlation of morphology, and promote the DCTN learn spatial context information. We further compare our ICM and the 3D convolution to illustrate the performance of both in extracting spatial correlation. We calculate the spatial correlation in the depth dimension to compare the differences of the contextual information. The results are shown in Fig. 8(b). The results show that our ICM approximates the ability of learning contextual information in 3D convolution, with the correlations of 0.80-0.93.
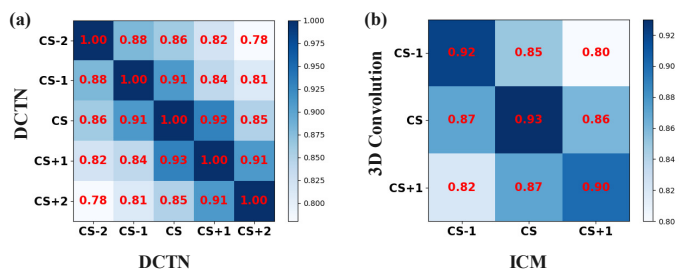

Fig. 8. (a) The heat map shows that our DCTN has high feature correlations between slices. (b) The heat map show that our ICM approximates the ability of learning contextual information in 3D convolution

2) Effectiveness of Geometry-constrained OT Guider: We visualize the topological process of constraints, as shown in Fig. 9. For that, we employ the Gaussian distribution to display that OT matches the source probability vector of raw images into the target probability vector of CPR images. The OT guider is beneficial for the accuracy of the geometric constraint, because it learns inter-image geometric transformation and ensures the sensitivity to the aortic boundary.
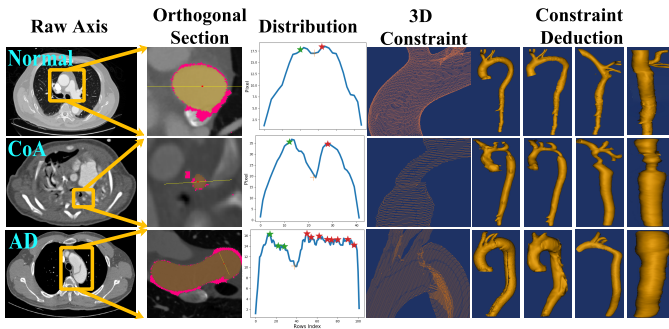
Fig. 9. Visualization of OT process in the guider. The role of the guider is shown from the orthogonal section, curves of Gaussian distribution, results of 3D constraint and constraints deduction, respectively. The optimal points in the curve are indicated by the stars for ascending aorta (green), aortic arch (orange) and descending aorta (red).

In the case of the normal aorta, the distribution curve is seen to be relatively smooth; the curve is more consistent with the Gaussian distribution, which helps the network to learn the morphological correlation between graphs. In the case of CoA, the curve has a fluctuation in the middle; the reason is anatomical mutations caused by aortic stenosis, but nevertheless DCTN remains sensitive to stenotic vessels. In the case of AD, the curve is more jagged; AD is irregular, the false lumen is beak-shaped and the true lumen is small. The variational inference via OT is not as smooth. However, the DCTN still has high matching and good robustness. As can be seen from the 3D constraints, our DCTN decodes the constraints of spatial topology quite well. The results conform to the anatomical structure of the aorta.

TABLE VI
ABLATION STUDY ON ALL MODULES.

Extractor	Decoder	Guider	DSC ↑	Pre ↑	Recall ↑
×	×	×	0.895	0.898	0.887
√	√	×	0.936	0.938	0.923
×	×	√	0.912	0.907	0.915
√	√	√	**0.949**	**0.952**	**0.933**

3) Effectiveness of all Modules: We conduct experiments on all modules to verify their effectiveness, including the extractor, decoder and guider. The results are shown in Table VI. The model achieves an average DSC of 0.895 for aortic segmentation without any additional modules. However, the model that includes the extractor and the decoder achieves a higher DSC of 0.936. The extractor module performs well in extracting semantic features of the aorta and capturing intra-image and inter-slice correlations. The decoder enables the model to perceive geometric contours of the aorta at multiple scales, and enhance the foreground information of the aorta with a low-contrast agent intensity. The combination of the two modules leads to a substantial increase in the model's learning ability. When solely the guider module is employed, the model achieves a DSC of 0.912. The guider can effectively learn geometric features between raw and CPR images, enforce morphological constraints, and facilitate the learning of the aortic features. All modules in the DCTN cooperate and complement each other to achieve the best performance, with

a DSC of 0.949, precision of 0.952 and recall of 0.933. The convergence curve of our model is shown in Fig.10.
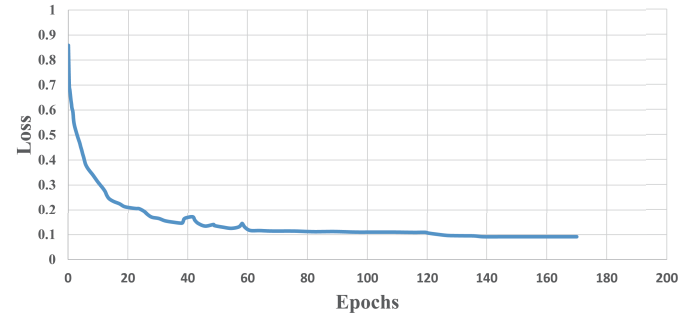

Fig. 10. The convergence curve of our model during the training.

H. Applicability and Superiority on the Public Datasets

We conduct experiments on four public CT datasets (AVT [62], TBAD [63], AAPM [64], and SegTHOR [65]) for better evaluation of the applicability and superiority. These datasets contain the raw CT images and the corresponding organ labels. Table VII shows the comparison between the DCTN and the fourteen aortic segmentation methods on four public datasets. Our method achieves a mean DSC of 0.903, RMSD of 4.74 and HD of 6.21 on AVT; a DSC of 0.922, RMSD of 3.33 and HD of 3.43 on TBAD; a DSC of 0.893, RMSD of 5.52 and HD of 7.54 on AAPM; and a DSC of 0.941, RMSD of 3.06 and HD of 3.12 on SegTHOR. The values of these results show that our method is highly consistent and reliable across different datasets and close to our private dataset (DSC:0.949, RMSD: 3.51, HD:3.45). These results also indicate that our method is applicable on the separate, independent, and public datasets. The performance of our method on these four public datasets is superior to that of other state-of-the-art methods, as shown by its higher DSC and lower HD values in comparison to the others. These results provide evidence that the effectiveness of our method extends beyond the separate dataset and is superior to other methods in diverse datasets.

V. CONCLUSION

In this paper, we have developed a novel deformable constraint transport newtwork (DCTN) for aortic segmentation. The DCTN adaptively extracts aortic features to define intra-image geometric properties; it guides topological implementation in space to constrain inter-image morphological transformation. Our method conduct on 267 in-house subjects, including normal aorta, CoA, and AD. Our method also conduct on four extra public datasets. The results show the effectiveness of DCTN for more precision segmentation in all these cases, and its superior segmentation accuracy compared with other advanced segmentation methods. This contributes to distinctive employment of DCTN for aortic segmentation, for measurement of clinical indexes and for carrying out appropriate aortic interventions.

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2023.3339142

AUTHOR *et al.*: PREPARATION OF PAPERS FOR IEEE TRANSACTIONS ON MEDICAL IMAGING 13

TABLE VII

THE RESULTS OF COMPARISON BETWEEN OUR MODEL AND OTHER SEGMENTATION METHODS ON FOUR PUBLIC DATASETS.

Method	AVT [62]			TBAD [63]			AAPM [64]			SegTHOR [65]		
	DSC ↑	RMSD ↓	HD ↓	DSC ↑	RMSD ↓	HD ↓	DSC ↑	RMSD ↓	HD ↓	DSC ↑	RMSD ↓	HD ↓
Xiong [21]	0.886	5.03	7.72	0.907	4.51	5.31	0.890	5.56	8.22	0.931	3.76	3.34
Sieren [27]	0.879	6.05	10.01	0.894	5.01	5.22	0.883	6.19	9.13	0.927	3.86	4.02
Li [24]	0.885	5.03	7.43	0.918	4.35	3.41	0.887	5.82	8.75	0.938	3.84	3.67
Zhao [33]	0.881	5.23	9.45	0.903	4.55	5.43	0.880	6.32	9.51	0.924	4.12	4.93
Feiger [19]	0.883	5.06	8.76	0.902	4.55	5.02	0.880	6.44	10.08	0.929	4.06	3.75
Wobben [23]	0.887	4.95	7.54	0.906	4.50	4.57	0.886	5.86	9.04	0.935	3.47	3.16
Song [34]	0.884	5.08	7.86	0.907	4.52	6.06	0.884	6.13	8.32	0.931	3.78	4.13
Hahn [28]	0.880	6.12	12.35	0.885	6.06	8.65	0.875	6.62	11.67	0.922	4.32	6.31
Abdolmanafi [47]	0.882	5.08	9.12	0.903	4.58	5.77	0.880	6.31	9.45	0.922	4.25	5.67
Chen [26]	0.889	5.02	8.36	0.905	4.53	5.63	0.886	5.73	8.45	0.925	4.16	5.24
Lyu [48]	0.888	4.96	8.41	0.902	4.52	6.53	0.882	6.41	9.43	0.927	4.14	4.89
Zhao [36]	0.898	4.82	7.13	0.912	4.38	4.02	0.884	6.23	9.34	0.930	3.95	4.03
Deng [32]	0.883	6.06	10.36	0.894	4.71	5.88	0.880	6.42	10.02	0.918	5.01	6.28
Yu [49]	0.882	5.07	9.33	0.897	5.22	6.26	0.878	6.48	9.97	0.915	4.89	6.35
Cheng [50]	0.880	5.18	9.88	0.901	4.59	4.97	0.875	6.56	9.83	0.913	4.92	6.56
Cao [22]	0.881	5.11	9.86	0.896	5.18	6.31	0.879	6.51	11.23	0.912	4.87	5.82
Unet [51]	0.881	5.19	9.91	0.893	5.33	6.30	0.877	6.54	11.18	0.902	5.12	5.38
TransUnet [52]	0.886	5.05	7.65	0.912	4.41	4.06	0.885	5.83	8.77	0.910	4.92	5.57
PSPNet [53]	0.880	6.08	10.11	0.893	5.32	6.29	0.880	6.33	9.92	0.898	5.16	6.01
SegNet [54]	0.877	6.06	9.97	0.891	5.37	6.28	0.878	6.53	11.25	0.893	5.34	6.03
ViT [55]	0.882	5.30	8.92	0.908	4.52	5.58	0.881	6.25	9.73	0.905	5.07	5.39
SwinTrans [56]	0.888	4.92	7.41	0.911	4.08	4.68	0.883	5.90	8.89	0.913	4.37	4.92
SegFormer [57]	0.884	5.01	7.56	0.906	4.54	5.89	0.881	6.27	9.61	0.908	5.23	5.45
Ours	**0.903**	**4.74**	**6.21**	**0.922**	**3.33**	**3.43**	**0.894**	**5.52**	**7.53**	**0.941**	**3.06**	**3.12**

REFERENCES

[1] L. F. Hiratzka, G. Bakris, J. A. Beckman, R. M. Bersin *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, pp. 27–129, 2010.

[2] R. Erbel, V. Aboyans, C. Boileau, E. Bossone *et al.*, "2014 esc guidelines on the diagnosis and treatment of aortic diseases," *Russian Journal of Cardiology*, vol. 123, no. 7, pp. 7–72, 2015.

[3] N. Manghat, G. Morgan-Hughes, and C. Roobottom, "Multi-detector row computed tomography: imaging in acute aortic syndrome," *Clinical radiology*, vol. 60, no. 12, pp. 1256–1267, 2005.

[4] Q. Zhao, K. Shi, Z.-g. Yang, K.-y. Diao, H.-y. Xu, X. Liu, and Y.-k. Guo, "Predictors of aortic dilation in patients with coarctation of the aorta: evaluation with dual-source computed tomography," *BMC cardiovascular disorders*, vol. 18, no. 1, pp. 1–7, 2018.

[5] C. S. Roberts and W. C. Roberts, "Aortic dissection with the entrance tear in the descending thoracic aorta. analysis of 40 necropsy patients." *Annals of surgery*, vol. 213, no. 4, pp. 356–368, 1991.

[6] D. Spinelli, F. Benedetto, R. Donato, G. Piffaretti, M. M. Marrocco-Trischitta, H. J. Patel, K. A. Eagle, and S. Trimarchi, "Current evidence in predictors of aortic growth and events in acute type b aortic dissection," *Journal of vascular surgery*, vol. 68, no. 6, pp. 1925–1935, 2018.

[7] G. Garzón, M. Fernández-Velilla, M. Martí, I. Acitores, F. Ybáñez, and L. Riera, "Endovascular stent-graft treatment of thoracic aortic disease," *Radiographics*, vol. 25, no. suppl_1, pp. S229–S244, 2005.

[8] S. A. Goldstein, A. Evangelista, S. Abbara, A. Arai *et al.*, "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.

[9] R. Raman, S. Napel, C. F. Beaulieu, E. S. Bain *et al.*, "Automated generation of curved planar reformations from volume data: method and evaluation," *Radiology*, vol. 223, no. 1, pp. 275–280, 2002.

[10] J. Rueckel, P. Reidler, N. Fink, J. Sperl, T. Geyer, M. Fabritius, J. Ricke, M. Ingrisch, and B. Sabel, "Artificial intelligence assistance improves reporting efficiency of thoracic aortic aneurysm ct follow-up," *European Journal of Radiology*, vol. 134, p. 109424, 2021.

[11] J. W. Nance, R. E. Ringel, and E. K. Fishman, "Coarctation of the aorta in adolescents and adults: a review of clinical features and ct imaging," *Journal of cardiovascular computed tomography*, vol. 10, no. 1, pp. 1–12, 2016.

[12] D. Fleischmann, R. O. Afifi, A. I. Casanegra, J. A. Elefteriades, T. G. Gleason *et al.*, "Imaging and surveillance of chronic aortic dissection: A scientific statement from the american heart association," *Circulation: Cardiovascular Imaging*, vol. 15, no. 3, p. e000075, 2022.

[13] K. Chen, N. Debroux, and C. Le Guyader, "A survey of topology and geometry-constrained segmentation methods in weakly supervised settings," *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pp. 1–46, 2021.

[14] W. Lin, H. Liu, L. Gu, and Z. Gao, "A geometry-constrained deformable attention network for aortic segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, 2022, pp. 287–296.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[16] T. Lei, R. Sun, X. Du, H. Fu, C. Zhang, and A. K. Nandi, "Sgu-net: Shape-guided ultralight network for abdominal image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1431–1442, 2023.

[17] Z. Xing, L. Wan, H. Fu, G. Yang, and L. Zhu, "Diff-unet: A diffusion embedded network for volumetric segmentation," *arXiv preprint arXiv:2303.10326*, 2023.

[18] A. Pepe, J. Li, M. Rolf-Pissarczyk, C. Gsaxner, X. Chen *et al.*, "Detection, segmentation, simulation and visualization of aortic dissections: A review," *Medical Image Analysis*, vol. 65, p. 101773, 2020.

[19] B. Feiger, E. Lorenzana-Saldivar, C. Cooke, R. Horstmeyer, M. Bishawi *et al.*, "Evaluation of u-net based architectures for automatic aortic dissection segmentation," *ACM Transactions on Computing for Healthcare*, vol. 3, no. 1, pp. 1–16, 2021.

[20] V. V. Danilov, K. Y. Klyshnikov, O. M. Gerget, I. P. Skirnevsky *et al.*, "Aortography keypoint tracking for transcatheter aortic valve implantation based on multi-task learning," *Frontiers in Cardiovascular Medicine*, vol. 8, p. 699, 2021.

[21] X. Xiong, Y. Ding, C. Sun, Z. Zhang, X. Guan *et al.*, "A cascaded multitask generative framework for detecting aortic dissection on 3-d non-contrast-enhanced computed tomography," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 5177–5188, 2022.

[22] 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. 108713, 2019.

[23] L. D. Wobben, M. Codari, G. Mistelbauer, A. Pepe, K. Higashigaito *et al.*, "Deep learning-based 3d segmentation of true lumen, false lumen, and false lumen thrombosis in type-b aortic dissection," in *2021 43rd*

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2023.3339142

14   IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2022

Annual International Conference of the IEEE Engineering in Medicine & Biology Society. IEEE, 2021, pp. 3912–3915.

[24] F. Li, L. Sun, K.-Y. Lam, S. Zhang, Z. Sun *et al.*, "Segmentation of human aorta using 3d nnu-net-oriented deep learning," *Review of Scientific Instruments*, vol. 93, no. 11, p. 114103, 2022.

[25] A. Fantazzini, M. Esposito, A. Finotello, F. Auricchio, B. Pane *et al.*, "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.

[26] D. Chen, X. Zhang, Y. Mei, F. Liao, H. Xu *et al.*, "Multi-stage learning for segmentation of aortic dissections using a prior aortic anatomy simplification," *Medical Image Analysis*, vol. 69, p. 101931, 2021.

[27] M. M. Sieren, C. Widmann, N. Weiss, J. H. Moltz, F. Link *et al.*, "Automated segmentation and quantification of the healthy and diseased aorta in ct angiographies using a dedicated deep learning approach," *European Radiology*, vol. 32, no. 1, pp. 690–701, 2022.

[28] L. D. Hahn, G. Mistelbauer, K. Higashigaito, M. Koci *et al.*, "Ct-based true-and false-lumen segmentation in type b aortic dissection using machine learning," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 3, p. e190179, 2020.

[29] N. Lee, H. Tek, and A. F. Laine, "True-false lumen segmentation of aortic dissection using multi-scale wavelet analysis and generative-discriminative model matching," in *Medical Imaging 2008: Computer-Aided Diagnosis*, vol. 6915, 2008, pp. 878 – 888.

[30] T. Kovács, P. Cattin, H. Alkadhi, S. Wildermuth, and G. Székely, "Automatic segmentation of the aortic dissection membrane from 3d cta images," in *Medical Imaging and Virtual Reality*, 2006, pp. 317–324.

[31] M. De Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen, "Interactive segmentation of abdominal aortic aneurysms in cta images," *Medical Image Analysis*, vol. 8, no. 2, pp. 127–138, 2004.

[32] 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.

[33] J. Zhao and Q. Feng, "Automatic aortic dissection centerline extraction via morphology-guided crn tracker," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3473–3485, 2021.

[34] Z. Song, S. Chai, and E. Zhu, "Segmentation of aorta with aortic dissection based on centerline and boundary distance," in *2022 41st Chinese Control Conference.* IEEE, 2022, pp. 7292–7297.

[35] Z. Li, J. Feng, Z. Feng, Y. An, Y. Gao, B. Lu, and J. Zhou, "Lumen segmentation of aortic dissection with cascaded convolutional network," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 122–130.

[36] J. Zhao, J. Zhao, S. Pang, and Q. Feng, "Segmentation of the true lumen of aorta dissection via morphology-constrained stepwise deep mesh regression," *IEEE Transactions on Medical Imaging*, 2022.

[37] C. Liu, M. K.-P. Ng, and T. Zeng, "Weighted variational model for selective image segmentation with application to medical images," *Pattern Recognition*, vol. 76, pp. 367–379, 2018.

[38] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo *et al.*, "A survey on vision transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2022.

[39] Z. Xia, X. Pan, S. Song, L. E. Li, and G. Huang, "Vision transformer with deformable attention," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 4794–4803.

[40] E. J. Kirkland, "Bilinear interpolation," in *Advanced Computing in Electron Microscopy*, 2010, pp. 261–263.

[41] L. Caicedo Torres, L. Manella Pereira, and M. Hadi Amini, "A survey on optimal transport for machine learning: Theory and applications," *arXiv e-prints*, pp. arXiv–2106, 2021.

[42] C. Villani, "Topics in optimal transportation: American mathematical society," *Graduate studies in mathematics*, vol. 58, 2003.

[43] F. Manessi, A. Rozza, and M. Manzo, "Dynamic graph convolutional networks," *Pattern Recognition*, vol. 97, p. 107000, 2020.

[44] H. Mei, G.-P. Ji, Z. Wei, X. Yang, X. Wei, and D.-P. Fan, "Camouflaged object segmentation with distraction mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8772–8781.

[45] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural networks*, vol. 11, no. 4, pp. 761–767, 1998.

[46] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, "Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on fcns," *Artificial intelligence in medicine*, vol. 83, pp. 58–66, 2017.

[47] A. Abdolmanafi, A. Forneris, R. D. Moore, and E. S. Di Martino, "Deep-learning method for fully automatic segmentation of the abdominal aortic aneurysm from computed tomography imaging," *Frontiers in Cardiovascular Medicine*, vol. 9, 2022.

[48] T. Lyu, G. Yang, X. Zhao, H. Shu, L. Luo, D. Chen, J. Xiong, J. Yang, S. Li, J.-L. Coatrieux, and Y. Chen, "Dissected aorta segmentation using convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 211, p. 106417, 2021.

[49] Y. Yu, Y. Gao, J. Wei, F. Liao, Q. Xiao, J. Zhang, W. Yin, and B. Lu, "A three-dimensional deep convolutional neural network for automatic segmentation and diameter measurement of type b aortic dissection," *Korean journal of radiology*, vol. 22, no. 2, pp. 168–178, 2021.

[50] J. Cheng, S. Tian, L. Yu, X. Ma, and Y. Xing, "A deep learning algorithm using contrast-enhanced computed tomography (ct) images for segmentation and rapid automatic detection of aortic dissection," *Biomedical Signal Processing and Control*, vol. 62, p. 102145, 2020.

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

[52] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[53] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[54] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[55] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[56] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.

[57] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.

[58] Z. Gao, Y. Guo, J. Zhang, T. Zeng, and G. Yang, "Hierarchical perception adversarial learning framework for compressed sensing mri," *IEEE Transactions on Medical Imaging*, vol. 42, no. 6, pp. 1859–1874, 2023.

[59] 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.

[60] L. Gu and X.-C. Cai, "Fusing 2d and 3d convolutional neural networks for the segmentation of aorta and coronary arteries from ct images," *Artificial Intelligence in Medicine*, vol. 121, p. 102189, 2021.

[61] S. Guo, X. Liu, H. Zhang, Q. Lin, L. Xu, C. Shi, Z. Gao *et al.*, "Causal knowledge fusion for 3d cross-modality cardiac image segmentation," *Information Fusion*, vol. 99, p. 101864, 2023.

[62] L. Radl, Y. Jin, A. Pepe, J. Li, C. Gsaxner, F. hua Zhao, and J. Egger, "Aortic Vessel Tree (AVT) CTA Datasets and Segmentations," 1 2022. [Online]. Available: https://doi.org/10.6084/m9.figshare.14806362.v1

[63] Z. Yao, W. Xie, J. Zhang, Y. Dong, H. Qiu, H. Yuan *et al.*, "Imagetbad: A 3d computed tomography angiography image dataset for automatic segmentation of type-b aortic dissection," *Frontiers in Physiology*, p. 1611, 2021. [Online]. Available: https://github.com/XiaoweiXu/Dataset_Type-B-Aortic-Dissection

[64] J. Yang, H. Veeraraghavan, S. G. Armato III, K. Farahani, J. S. Kirby *et al.*, "Auto-segmentation for thoracic radiation treatment planning: a grand challenge at aapm 2017," *Medical physics*, vol. 45, no. 10, pp. 4568–4581, 2018. [Online]. Available: http://aapmchallenges.cloudapp.net/competitions/3

[65] Z. Lambert, C. Petitjean, B. Dubray, and S. Kuan, "Segthor: Segmentation of thoracic organs at risk in ct images," in *2020 Tenth International Conference on Image Processing Theory, Tools and Applications*, 2020, pp. 1–6. [Online]. Available: https://competitions.codalab.org/competitions/21145