Medical Image Analysis 102 (2025) 103512

Contents lists available at ScienceDirect

Medical Image Analysis

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

Domain knowledge based comprehensive segmentation of Type-A aortic dissection with clinically-oriented evaluation

Shanshan Song [a], Hailong Qiu [b], Meiping Huang [c], Jian Zhuang [b], Qing Lu [d], Yiyu Shi [d], Xiaomeng Li [a,*], Wen Xie [b,*], Guang Tong [b,*], Xiaowei Xu [b,*]

[a] Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong
[b] Department of Cardiovascular Surgery, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China
[c] Department of Catheterization Lab, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China
[d] Department of Computer Science and Engineering, University of Notre Dame, IN, 46656, USA

ARTICLE INFO

Keywords:
Medical Image Segmentation
Type-A Aortic Dissection
Benchmark
Domain Knowledge
Clinically-oriented Evaluation

ABSTRACT

Type-A aortic dissection (TAAD) is a cardiac emergency in which rapid diagnosis, prognosis prediction, and surgical planning are critical for patient survival. A comprehensive understanding of the anatomic structures and related features of TAAD patients is the key to completing these tasks. However, due to the emergent nature of this disease and requirement of advanced expertise, manual segmentation of these anatomic structures is not routinely available in clinical practice. Currently, automatic segmentation of TAAD is a focus of the cardiovascular imaging research. However, existing works have two limitations: no comprehensive public dataset and lack of clinically-oriented evaluation. To address these limitations, in this paper we propose imageTAAD, the first comprehensive segmentation dataset of TAAD with clinically-oriented evaluation. The dataset is comprised of 120 cases, and each case is annotated by medical experts with 35 foreground classes reflecting the clinical needs for diagnosis, prognosis prediction and surgical planning for TAAD. In addition, we have identified four key clinical features for clinically-oriented evaluation. We also propose SegTAAD, a baseline method for comprehensive segmentation of TAAD. SegTAAD utilizes two pieces of domain knowledge: (1) the boundaries play a key role in the evaluation of clinical features, and can enhance the segmentation performance, and (2) the tear is located between TL and FL. We have conducted intensive experiments with a variety of state-of-the-art (SOTA) methods, and experimental results have shown that our method achieves SOTA performance on the ImageTAAD dataset in terms of overall DSC score, 95% Hausdorff distance, and four clinical features. In our study, we also found an interesting phenomenon that a higher DSC score does not necessarily indicate better accuracy in clinical feature extraction. All the dataset, code and trained models have been published (Xiaowei, 2024).

1. Introduction

Type-A aortic dissection (TAAD) is a medical emergency, characterized by an abrupt tear in the inner lining of the aorta, which leads to blood forcefully entering the middle layer of the aortic wall (Criado, 2011; Zhu et al., 2020). The annual incidence rate of TAAD is roughly 3/100,000 (Harris et al., 2011). Without timely surgical treatment, the prognosis of TAAD is grim, with a mortality rate of approximately 1% per hour after onset and 50% within three days without treatment. Despite the continuous advancement in surgical techniques, the operative mortality rate still remains as high as 12% today (Criado, 2011).

Rapid diagnosis, prognosis prediction, and surgical planning are vital for the survival of TAAD patients (Yuan et al., 2022). A comprehensive understanding of the anatomic structures and related features of the patients is the key to completing these tasks (Nienaber et al., 2016). However, the anatomic structures and related features are quite complex. As shown in Fig. 1, in a typical TAAD, the singular lumen of the aorta is split into two distinct parts: a true lumen (TL) and a false lumen (FL) that starts from the ascending aorta (Daily et al., 1970). Analyzing the vessels connected to FL, the tears between TL and FL are located as shown in Fig. 1(b) and (c). In current clinical practice,

* Corresponding authors.
E-mail addresses: eexmli@ust.hk (X. Li), wen.xie.cn@gmail.com (W. Xie), tongguang@gdph.org.cn (G. Tong), xiao.wei.xu@foxmail.com (X. Xu).

https://doi.org/10.1016/j.media.2025.103512
Received 3 August 2024; Received in revised form 12 February 2025; Accepted 15 February 2025
Available online 26 February 2025

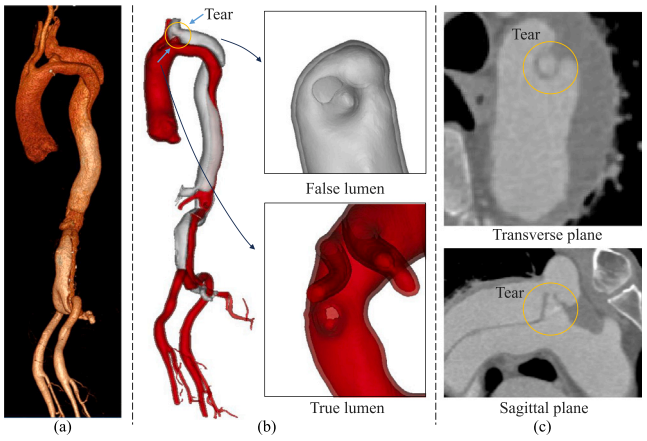

Fig. 1. 3D and 2D views of TAAD from CT images. (a) The 3D aortic dissection extracted from the Philips Vue picture archiving and communication system.[1] (b) Detailed 3D visualizations of the tear, false lumen, and true lumen structures. The top inset zooms in on the false lumen, and the bottom inset on the true lumen. (c) Cross-sectional images showing the tear in both transverse and sagittal planes, circled for clarity.

enhanced computed tomography (CT) is the most commonly used imaging method for both diagnosis and surgical planning for TAAD (Erbel et al., 2001). The CT images are extensively evaluated by the radiologist and aortic surgeon so anatomic features such as **locations of tears (LOT)** (Furui et al., 2024), **branch vessels involvement (BVI)** (Wang et al., 2023), **false lumen area ratio (FLAR)** (Igarashi et al., 2022; Kim et al., 2022) and **true-lumen collapse (TLC)** (Chung et al., 2000) by the dissection (Fig. 1(a) and (c)) are identified and considered for surgical planning. Recently, manual segmentation of related anatomic structures, as shown in Fig. 1(b), has been used in clinical practice to provide clinicians with a vivid visualization of the anatomic structures and related features (Mastrodicasa et al., 2022). However, due to the demanding time and expertise required, manual segmentation of the dissected aorta is not routinely available in clinical practice.

Currently, automatic segmentation of TAAD has been a hot topic in the research community of Medical AI. For comprehensiveness of discussion, we also include segmentation methods of type-B aortic dissection here as it shares a similar anatomic structure with TAAD. Note that according to Stanford classification (Daily et al., 1970), type-A means the dissection includes the ascending aorta, and type-B means the dissection does not. As TAAD segmentation involves identifying complex and variable structures such as the true lumen, false lumen, and tear sites, existing works tailor their implementation to suit specific application characteristics. Some works incorporate morphological features specific to the aorta's structure to enhance segmentation performance (Chen et al., 2021; Lyu et al., 2021; Zhao et al., 2022), while others segment the entire aorta first and then refine the local classification such as the true and false lumens (Cao et al., 2019; Yu et al., 2021; Feng et al., 2023). In addition, Zhang et al. (2023b) utilized semi-supervised learning to leverage both the labeled data and unlabeled data to achieve improved segmentation performance at a lower annotation cost.

Generally, existing works have two limitations. **(1) No comprehensive public dataset**. Previous research has not considered the relationship between the aorta, its branch arteries, and surrounding anatomical structures, but focused primarily on distinguishing between the true and false lumens of the aorta. However, delineating features such as LOT, BVI, FLAR, and TLC are vital for guiding clinical practice. Simply identifying different parts of aortic lumens is insufficient. Furthermore, there is still a lack of publicly available and high-quality datasets for TAAD, leading to limited attention from AI researchers. **(2) No clinically-oriented evaluation**. The general metrics used to evaluate segmentation results are limited, focusing only on pixel-level accuracy but not considering clinical features. This gap has led to a discrepancy between the evaluated segmentation performance and actual clinical applicability. In summary, there is lack of a comprehensive yet generic benchmark to evaluate various segmentation models effectively.

To address the above limitations and advance this field, we propose imageTAAD, the first comprehensive segmentation dataset of TAAD featuring clinically-oriented evaluation. The dataset is comprised of 120 cases, and each case is annotated by medical experts with 35 foreground classes including aorta, associated branch arteries, and primary organs, reflecting the clinical needs for diagnosis, prognosis prediction, and surgical planning of TAAD. In addition, based on clinical practice, we have identified four key clinical features for TAAD: LOT (Weiss et al., 2012; Furui et al., 2024), BVI (Wang et al., 2023), FLAR (Jiang et al., 2023), and TLC (Chung et al., 2000). To assess these features, we propose a new set of segmentation evaluation standards, including feature extraction methods from the segmentation results and calculation metrics. These standards are aimed at assessing whether segmentation results accurately reflect key clinical features rather than merely focusing on pixel-level accuracy. Our proposed evaluation standards are more aligned with actual clinical applicability. Afterall, clinicians are more concerned with whether the segmentation results can accurately reflect relevant clinical features rather than minor improvements in overall DSC score. Based on the imageTAAD dataset, we benchmarked consistent comparisons and evaluations of various advanced medical segmentation methods, including both generic segmentation metrics and our newly proposed clinical feature metrics, allowing for more effective assessment of the algorithms against standardized criteria.

In addition, we propose SegTAAD, a baseline method for comprehensive segmentation of TAAD. SegTAAD utilizes two pieces of domain knowledge: (1) the boundaries of different anatomical parts play a key role in the evaluation of clinical features, and can enhance the segmentation performance, and (2) the tear is located between TL and FL. Consequently, two decoders are used. Particularly, one is for segmentation of all classes, and another one is for segmentation of edges and tears. The decoder for tear and edge segmentation is fed into the decoder for all class segmentation to enhance performance. We conduct extensive experiments with a variety of deep neural network architectures including CNN-based networks such as nnUNet (Isensee et al., 2021), MedNeXt (Roy et al., 2023), 3DUXNet (Lee et al., 2022), and BANet (Hu et al., 2022), Transformer-based networks, namely TransBTS (Wenxuan et al., 2021), nnformer (Zhou et al., 2021), UNETR (Hatamizadeh et al., 2022), UTNet (Gao et al., 2021), TransFuse (Zhang et al., 2021), and SwinUNETR (Hatamizadeh et al., 2021), and a Mamba-based network named Umamba (Ma et al., 2024). Experimental results show that our method achieves SOTA performance on the ImageTAAD dataset in terms of overall DSC score, 95% Hausdorff distance, and four clinical features. We also find an interesting phenomenon that a higher DSC score does not necessarily indicate better accuracy in clinical feature extraction. However, both the DSC score (64.01%) and the accuracy of clinical features (precision, 0.62–0.69; recall, 0.63–0.80) are still too low for reliable diagnostic application.

The main contributions of this paper are as follows:

- We present ImageTAAD, the first comprehensive segmentation dataset of TAAD with clinically-oriented evaluation. It consists of 120 cases with annotations for 35 foreground classes. This dataset provides a solid foundation for developing and evaluating advanced segmentation algorithms towards TAAD diagnosis and treatment.
- Based on the proposed dataset, we introduce four segmentation evaluation metrics to better assess the clinical relevance. This evaluation criterion can accurately reflect the practical capability of a method for extracting four key clinical features.

[1] https://www.philips.com.hk/healthcare/solutions/diagnostic-informatics/enterprise-imaging-pacs.

Table 1
Overview of TAAD segmentation works and used datasets. For comprehensiveness of discussion, segmentation methods of type-B aortic dissection are also included here as they share a similar anatomic structure with TAAD. TL: true lumen; FL: false lumen; FLT: false lumen thrombus.

Dataset	Year	Scans_num	Class_num	Foreground classes	Train_num	Val_num	Test_num	Type	Open-resource
Cao et al., 2019	2019	276	4	Whole Aorta, TL, and FL	246	–	30	Type-B	×
Cheng et al., 2020	2020	20	2	TL	12	4	4	–	×
Lyu et al., 2021	2021	42	2	Aorta	35	–	7	–	×
Yu et al., 2021	2021	139	3	Whole Aorta, TL, and FL	99	15	25	Type-B	×
Yao et al., 2021	2021	100	4	TL, FL and FLT	67	–	33	Type-B	✓
Chen et al., 2021	2021	120	4	TL, FL and Branch trunk	80	20	20	Type-B	×
Zhou et al., 2022	2022	35	10	Ascending Aorta (AAO), Descending Aorta (DAO), Aortic Arch (AA), Right Lung (RL), Left Lung (LL), Pulmonary Artery (PA), TL, FL, and Flap (IF)	29	–	6	–	×
Xiang et al., 2023	2023	108	4	TL, FL, and Branch trunk	68	–	40	Type-B	×
Feng et al., 2023	2023	463	5	Whole Aorta, TL, PFL, and TFL	309	77	77	Type-A	×
Zhang et al., 2023b	2023	306	5	TL, FL, TH, and BV (Consisting of aortic arch branches, abdominal branches, and iliac arteries)	191	38	77	Type-B	×
Jung et al., 2024	2024	253	4	TL, FL, and TH.	173	–	80	Type-A&Type-B	×
Ours	2024	120	36	Refer to Table 2	80	20	20	Type-A	✓

- We propose SegTAAD, a baseline method for comprehensive segmentation of TAAD. SegTAAD leverages domain knowledge by utilizing two specialized decoders to optimize its performance: one is dedicated to segmenting all classes, while the other specifically targets the segmentation of edges and tears. Additionally, the output from the tear and edge segmentation decoder is integrated into the all-class segmentation decoder.
- Experimental results show that our method achieve SOTA performance on the ImageTAAD dataset in terms of DSC score, 95% Hausdorff distance, and four clinical features. We also find an interesting phenomenon that a higher DSC score does not necessarily indicate high accuracy of clinical feature extraction.

2. Related work

2.1. TAAD segmentation

Table 1 summarizes the existing TAAD segmentation works and their related datasets. Note that for comprehensiveness of discussion, segmentation methods of type-B aortic dissection are also included and discussed here as it shares a similar anatomic structure with TAAD. As their approaches for TAAD segmentation can be categorized into two different types: one-stage approach and multi-stage approach, we review these works accordingly.

In the one-stage approach, end-to-end frameworks with the one-stage training strategy are used. Cao et al. (2019) compared four different CNN models (single one-task, single multi-task, and serial multi-task) on type-B aortic dissection segmentation. The experiments showed that the serial multi-task model performed the best. Lyu et al. (2021) designed an algorithm based on both 3D and 2D CNN networks. The 3D model identified the proximal and distal regions of the dissected aorta data and the 2D model was incorporated to extract the boundary information. A multi-label segmentation network, MOLS-Net (Zhou et al., 2022), was proposed for aortic dissection segmentation. MOLS-Net utilized the sequence feature pyramid attention module to capture sequence features across different scales, employing the attention mechanism to enhance the model's accuracy in targeting specific areas and improving feature utilization. Zhao et al. (2022) developed a morphology-constrained stepwise deep mesh regression for the true lumen segmentation, which improved the efficiency of the deep network and the uniformity of the mesh points. Zhang et al. (2023a) proposed an innovative semi-supervised segmentation framework for aortic dissections. This framework, which utilized a time-dependent weighted feedback fusion approach, effectively leveraged unlabeled data for improved segmentation accuracy. The feedback network could encode the predicted output from the backbone network into high-level feature space, and then fuse it with the original image features to correct previous mistakes.

In the multi-stage approach, multi-stage or multi-model segmentation is used to more effectively recognize different classes and locations of aortic dissection. Xu et al. (2019) firstly adopted Mask-RCNN (He et al., 2017) to detect and segment the aorta and then extracted the edges by Canny edge detector, after which two ResNets were employed for aortic dissection detection using the processed aorta images. Chen et al. (2021) employed an aortic straightening method to better identify both the global volumetric features of the aorta and the local characteristics of the primary tear. Prior to this, they utilized two cascaded neural networks to segment the aortic trunk and branches to separate the dual lumen, respectively. Feng et al. (2023) employed a two-stage approach, first segmenting the whole aorta and then the thrombus, to effectively highlight the aorta within a complex background. They also designed skip connection attention refinement modules to improve the segmentation of the thrombus details. ADSeg (Xiang et al., 2023) proposed a cascaded network structure with feature reuse, a novel flap attention module, and a two-step strategy for segmenting type-B aortic dissection. They focused on the intimal flap structure that separates the true and false lumens, a signal neglected by previous works. Zhang et al. (2023b) also presented a two-stage method to segment different categories. To enhance the segmentation of thrombus and branch vessels, a global-local fusion learning mechanism was designed to compensate for the missing contextual features in the cropped images from the first stage. ZOZI-Seg (Jung et al., 2024) utilized a cascade strategy that captured both the global context (anatomical structure) and the local detail texture based on the dynamic patch size with Zoom-Out and Zoom-In schemes. The process consisted of two stages, with a 3D transformer employed for panoptic context-awareness and a 3D UNet used for localized texture refinement.

Of all the aforementioned studies, only one published their dataset and code (Yao et al., 2021). All the previous work just performed comparison with some general segmentation methods like 3D U-Net (Çiçek et al., 2016). In addition, existing works mainly focused on segmentation of several classes like TL, FL and TFL. However, identification of both branch vessels originated from the FL and tears are vital for diagnosis, prognosis prediction, and surgical planning of TAAD. In this paper, we fist propose ImageTAAD with 35 annotated classes, with a set of evaluation criteria including general segmentation evaluation metrics and four clinical features. With the ImageTAAD dataset, we performed comparison with a variety of popular networks including nnUNet, MedNeXt, 3DUXNet, BANet, TransBTS, nnformer, UNETR,

UTNet, TransFuse, SwinUNETR, and Umamba. Existing TAAD segmentation methods were not covered for comparison and the reason is threefold. First, no codes were publicly available. Second, it would be difficult to re-implement previous code with complex training hyper parameters. Third, our tasks were quite different from published ones in terms of related classes (35 v.s. 3–5) and evaluation metrics (segmentation metrics and clinical features v.s. segmentation metrics).

2.2. Clinical feature extraction of segmentation results

Some existing works performed clinical feature extraction based on segmentation results, mainly using two approaches: post-processing based extraction, and end-to-end extraction. In the post-processing based extraction approach, most of the works focused on quantification of the true and false lumen areas. Yu et al. (2021) introduced an automatic method to measure the diameter of the type-B aortic dissection segmentation. This approach involved defining eight measurement positions and utilizing both deep learning (DL) and manual methods to calculate the maximum and minimum diameters, demonstrating excellent consistency between the reference measurements and those obtained from manual and DL methods. Sieren et al. (2022) quantified the physiological and diseased aorta to evaluate the segmentation of the aorta, using the method proposed by Selle et al. (2002) to calculate the maximum diameter, effective diameter, and area at eight anatomical landmarks, the maximum area of an aneurysm. However, their method requires pre-defining positions for calculation, which can vary and raise uncertainty across different cases.

In the end-to-end extraction, there is no explicit post-processing. Zhao and Feng (2021) proposed a fully automatic algorithm to extract the centerline based on a convolutional regression network and the morphological properties of aortic dissection. This method has been evaluated on two datasets and achieved high overlapping ratios. Pepe et al. (2023) presented four methods based on convolutional neural networks and uncertainty quantification methods to determine the orientation of cross-sectional planes of the aorta. Their trained model provided faster and more reproducible results than previous methods. However, their approaches lack reliability in evaluating segmentation results for our task due to the uninterpretability of the CNN-based networks.

In conclusion, all the existing works have limited consideration of clinical features during segmentation. Most of the works only use post-processing techniques to extract clinical features, and no clinically-oriented evaluation is performed to assess the used methods. In this paper, we take clinical features including LOT (Weiss et al., 2012; Furui et al., 2024), BVI (Wang et al., 2023), FLAR (Jiang et al., 2023), and TLC (Chung et al., 2000) into consideration, and propose a new set of segmentation evaluation standards. These standards aim to assess whether the segmentation results accurately reflect the key clinical features rather than merely focus on pixel-level accuracy. Our proposed evaluation standards are more aligned with actual clinical applicability.

3. ImageTAAD benchmark

3.1. Basic information of ImageTAAD dataset

Our proposed ImageTAAD dataset is a collection of computed tomography (CT) scans, comprising 120 individual cases. Each case in this dataset originates from a unique patient, ensuring a diverse and representative sample. CT scans are collected at Guangdong Provincial People's Hospital from May 16, 2019 to March 18, 2023. The examination time for each CT scan is 10–15 min. Patients are those who have been diagnosed with TAAD for the first time and have not undergone heart or aortic surgeries. The contrast agent protocol is as follows. The contrast agent was injected through the right cubital vein using a high-pressure injector at a flow rate of 4.0 ml/s, with a total injection volume of 80 ml. After contrast injection, 30 ml of saline was injected at the same flow rate. Bolus tracking was used, and monitoring started 10 s after injection. The region of interest was placed in the descending aorta with a trigger threshold of 150 HU, and scanning commenced 6 s after triggering. The scan parameters are set as follows: (1) detector configuration 128×0.625 mm, (2) pitch: 1.19, (3) rotation time: 0.27 s, (4) tube voltage: 120 kV, (5) automatic tube current modulation and (6) reconstruction slice thickness: 1.5 mm. The through-plane resolution of the CT scans is 1.5 mm for most cases. The spatial resolution of the CT scans is 0.75 mm in the x-y planes. Most of the CT scans are obtained by Philips Brilliance iCT 256 slice CT scanner, and the rest by Siemens SOMATOM definition flash CT scanner. These high-resolution images provide an in-depth view of the anatomical structures, which is critical for precise segmentation tasks. These high-resolution images provide an in-depth view of the anatomical structures, critical for precise segmentation tasks.

Table 2
35 annotated foreground categories in the ImageTAAD dataset. Each label is listed alongside its corresponding short form and full name.

Label index	Short form	Full name
0	BG	Background
1	ICA	Intercostal artery
2	Tear	Tear
3	AS	Aortic sinus
4	AA	Ascending aorta
5	ARCH	Aortic arch
6	IMA	Innominate artery
7	LCCA	Left common carotid artery
8	LSA	Left subclavian artery
9	DAO	Descending aorta
10	AA1	Abdominal aorta
11	SA	Splenic artery
12	Spleen	Spleen
13	AG	Gastric artery
14	Stomach	Stomach
15	HA	Hepatic artery
16	Liver	Liver
17	SMA	Superior mesenteric artery
18	RRA	Right renal artery
19	Kidney	Kidney
20	LRA	Left renal artery
21	Bone	Pelvis, sternum, vertebra, rib and a part of femur
22	RCIA	Right iliac artery
23	LCIA	Left iliac artery
24	RFA	Right femoral artery
25	LFA	Left femoral artery
26	ASFL	Aortic sinus false lumen
27	AAFL	Ascending aorta false lumen
28	ARCHFL	Aortic arch false lumen
29	IMAFL	Innominate artery false lumen
30	LCCAFL	Left common carotid artery false lumen
31	LSAFL	Left subclavian artery false lumen
32	DAOFL	Descending aorta false lumen
33	AA1FL	Abdominal aorta false lumen
34	RCIAFL	Right iliac artery false lumen
35	LCIAFL	Left iliac artery false lumen

We have meticulously cataloged the clinical characteristics of the ImageTAAD dataset, detailed in Table 2. There are 35 foreground categories that can be divided into four main groups as follows:

- Organs and ribs. Organs include spleen, stomach, liver, kidneys, pelvis, sternum, vertebras, and a part of both femurs, which are involved with organ perfusion (Jiang et al., 2023). The rib is involved with surgical planning (Sherif et al., 2017).
- The true lumen of the aorta located in the aortic sinus, ascending aorta, aortic arch, descending aorta, and abdominal aorta, which are included by extension (Wang et al., 2023).
- The branch arteries, including the intercostal arteries, innominate artery, left common carotid artery, left subclavian artery, splenic artery, gastric artery, hepatic artery, superior mesenteric artery, right renal artery, left renal artery, right iliac artery, left iliac

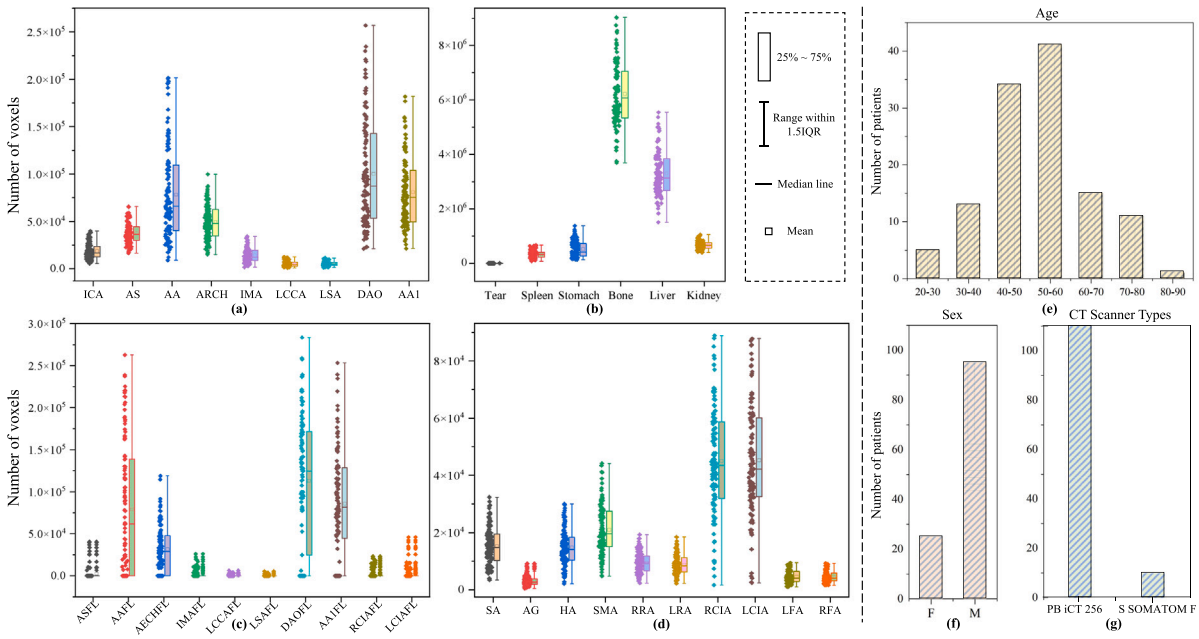

Fig. 2. Voxels and basic information distribution of ImageTAAD. The vertical scales vary across the sub-figures, highlighting the highly unbalanced distribution among different categories. (a) Voxel distribution across aortic sections. (b) Voxel distribution of organs, tears, and bone. (c) Voxel distribution across false lumen categories. (d) Voxel distribution of branch arteries. IQR: interquartile range. (e) Age distribution. (f) Gender distribution. (g) CT scanner types distribution. PB iCT 256: Philips Brilliance iCT 256; S SOMATOM F: Siemens SOMATOM Force.

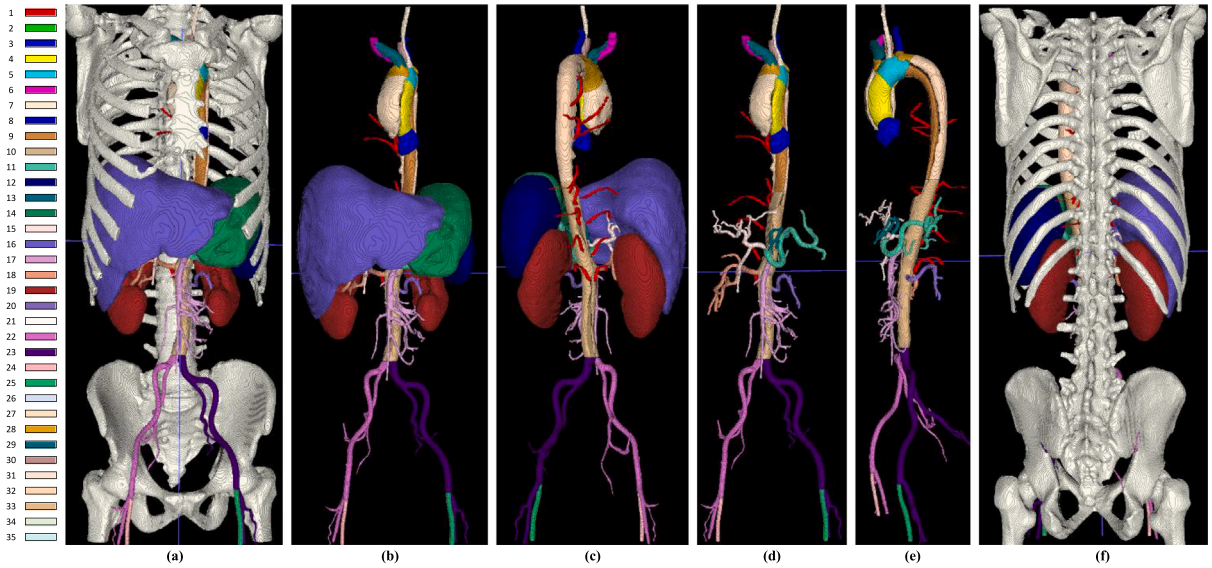

Fig. 3. Example of 3D rendering labels in ImageTAAD, with 35 annotated categories listed on the left. For the specific names corresponding to each class label number, please refer to Table 2. (a) and (f) display all annotated types, while (b), (c), and (d), (e) provide visualizations with bones and organs removed for clear internal visualization.

artery, right femoral artery, and left femoral artery. The origination of branch arteries are vital for organ perfusion (Jiang et al., 2023).

- The false lumen and associated tears located in the aortic sinus, ascending aorta, aortic arch, innominate artery, left common carotid artery, left subclavian artery, descending aorta, abdominal aorta, right iliac artery, and left iliac artery, which are included by extension (Wang et al., 2023), and LOT (Weiss et al., 2012; Furui et al., 2024).

Our classification scheme provides valuable insights into the dataset's composition and potential applications in aortic dissection image analysis. Furthermore, Fig. 2 presents the voxel distributions of all categories within the dataset. Notably, it highlights the challenges posed by the extremely unbalanced distribution among different classes, a critical consideration for developing effective aortic dissections segmentation models. To ensure the consistency in data distribution, we have split the dataset based on the proportion of specific categories, such as tears and different false lumen, since these categories are not present in every case. Finally, we divide the dataset into 80 scans for training, 20 scans for validation, and 20 scans for testing.

3.2. Annotation

Overall, our annotation process is completed by two processes: segmentation annotation and clinical feature annotation. The first

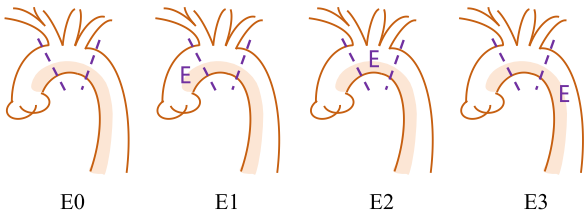

Fig. 4. LOT and associated label categories. "E" signifies the entry point of the tear, with 4 distinct scenarios each corresponding to a specific label: "E0" denotes the tear is unclear or invisible; "E1" denotes the tear is located in the ascending aorta; "E2" denotes the tear is located in the aortic arch; and "E3" denotes the tear is located in the descending aorta or below.

process involves manual annotation of the segmentation masks, while the second process aims at automatically extracting the labels of four clinical features from the segmentation masks.

3.2.1. Segmentation annotation

The annotation process is a rigorous and collaborative effort. Initially, a team of 3 experienced radiologists/aortic surgeons attentively annotated each scan, ensuring that a high level of detail and precision are maintained in the annotations. Following this initial phase, another group of 3 clinicians conducted a thorough review of these annotations. They provided correction and verification to ensure the utmost precision and reliability of the annotated masks. A semi-automatic tool, named Materialise Mimics[1] is used to execute the annotation. It takes approximately 1–1.5 h to annotate all 35 foreground categories for each case, followed by an additional 5–10 min to review, discuss, and refine the annotations. The ImageTAAD dataset was compiled, annotated, and reviewed over a period of approximately four months. An illustrative example of these detailed annotations is showcased in Fig. 3.

3.2.2. Clinical feature annotation

We have summarized four main clinical features based on clinical practice, which are used to assess clinical applicability of the segmentation results and serve as additional evaluation criteria for the proposed benchmark.

Location of tears (LOT). As shown in Fig. 4, LOT falls into four categories: "E0" (the tear is unclear or invisible), "E1" (the tear is located in the ascending aorta), "E2" (the tear is located in the aortic arch), or "E3" (the tear is located in the descending aorta or below). More than one tear may be present in different sections of the aorta. This variability is often discerned through differences in radiographic imaging, highlighting the complex nature of aortic dissections.

Identifying the location of the tears is crucial because the location directly influences the choice of treatment methods and prognosis (Weiss et al., 2012; Furui et al., 2024). After initial surgical management, follow-up monitoring of residual tears is also vital for assessing treatment effectiveness, prognosis, and further treatment planning.

Branch vessels Involvement (BVI). As shown in Fig. Fig. 5, this characteristic primarily evaluates the dissection's impact on branch arteries (Wang et al., 2023). In our dataset, we focus on the following 14 branch arteries: intercostal artery, innominate artery, left common carotid artery, left subclavian artery, splenic artery, gastric artery, hepatic artery, superior mesenteric artery, right renal artery, left renal artery, right iliac artery, left iliac artery, right femoral artery, and left femoral artery. For each branch, a binary label is employed to denote involvement: '0' signifies 'no involvement', while '1' indicates 'involvement'. Therefore, each case is assigned a binary label consisting of 14 elements corresponding to 14 branch arteries (see Fig. 5).

[1] https://www.materialise.com/en/healthcare/mimics-innovation-suite/mimics.

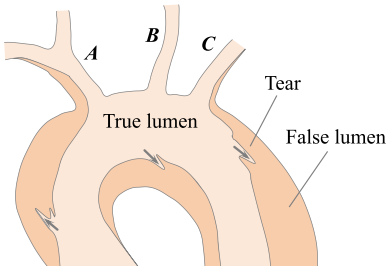

Fig. 5. Example of BVI in TAAD. Branch arteries A and C are involved by the dissection, while branch artery B remains unaffected by the dissection.

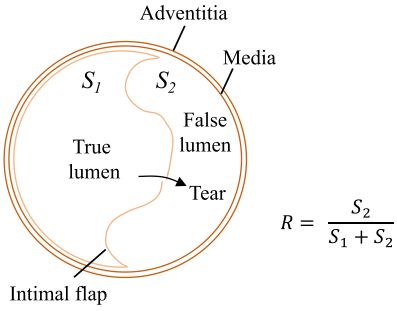

Fig. 6. Example of FLAR in TAAD. The image presents a cross-section along the direction of blood flow.

Note that if the false lumen extends to critical branches of the aorta, it may lead to interruption or significant reduction in blood supply to these branches. For example, involvement of the coronary arteries could result in acute myocardial infarction, while involvement of the renal arteries could lead to acute renal failure. Furthermore, if the false lumen affects arteries in supplying vital organs, prioritized surgical intervention may be necessary to restore normal blood flow. Understanding the extent of branch artery involvement is crucial for assessing the urgency, risks, and complexities of surgical interventions for treating aortic dissection.

False Lumen Area Ratio (FLAR). For parts affected by the dissection, measuring the FLAR can help analyze the severity of aortic dissection (Igarashi et al., 2022; Kim et al., 2022). An expanded false lumen accompanied by a collapsed true lumen may indicate organ malperfusion and high structural instability of the dissection. This also correlates with an increased risk of complications, potentially necessitating more urgent treatment (Immer et al., 2005; Kim et al., 2022). Delineating FLAR is crucial for assessing the urgency, risks, and complexities of surgical interventions for treating aortic dissection. For example, if the expanded false lumen affects blood supply to vital organs, prioritized surgical intervention may be necessary to restore normal blood flow. Certain treatments, such as stent placement, can be guided by LOT and FLAR (Evangelista et al., 2012; Li et al., 2020). Additionally, understanding this ratio before surgical intervention is crucial for assessing the risks and likelihood of surgical success (Igarashi et al., 2022; Fattouch et al., 2009; Kim et al., 2022).

In our work, FLARs at 10 aortic parts including aortic sinus, ascending aorta, aortic arch, innominate artery, left common carotid artery, left subclavian artery, descending aorta, abdominal aorta, right iliac artery, and left iliac artery are involved. As shown in Fig. 6, we could calculate the FLAR R of each position. Then, for each part of the aorta, we calculate the maximum and average FLAR. Particularly, the maximum FLAR refers to the highest value of FLARs recorded at any point along the centerline of each involved part, and average FLAR is the mean of them. Consequently, the quantified label for this feature consists of 10×2 numerical values, with those parts not accumulating

false lumens being assigned a zero value.

True-lumen collapse (TLC). TLC is classified based on FLAR, and has three distinct categories: no dissection, no true lumen collapse, and true lumen collapse (Chung et al., 2000). The evaluation criteria are derived from the above two ratios (R_{max} and R_{avg}). Taking R_{max} as an example, it is defined as follows:

1. False lumen/ (true + false lumen area ratio (R_{max})) = 0 → No dissection.
2. False lumen/ (true + false lumen area ratio (R_{max})) <50% → No true lumen collapse.
3. False lumen/ (true + false lumen area ratio (R_{max})) >50% → True lumen collapse.

The 10 aorta related classes including AS, AA, ARCH, IMA, LCCA, LSA, DAO, AA1, RCIA, and LCIA are involved with TLC.

3.3. Evaluate metric

General segmentation evaluation. To evaluate the general segmentation performance, we choose the popular Dice Similarity Coefficient (DSC) and 95% Hausdorff distance (HD_{95}) as the metrics. Denoting the ground truth by $truth$ and the predictions by $pred$, the mean DSC is computed as:

$$DSC = \frac{2|pred \cap truth|}{|pred| + |truth|} \tag{1}$$

As a surface measurement, we compute the mean Hausdorff distance over all vertebrae as:

$$HD_{95} = \frac{P_{95}[D(pred, truth)] + P_{95}[D(truth, pred)]}{2} \tag{2}$$

where $pred$ and $truth$ represent the prediction and corresponding ground truth respectively. $D(pred, truth)$ is the set of distances from boundary voxels of $pred$ to the nearest boundary voxel of $truth$. $P_{95}[D]$ means the 95 percentile of D.

Clinical feature evaluation. To evaluate the extracted clinical features from the segmentation mask, we employ precision, recall, and F1 scores ($F1_{micro}$ and $F1_{macro}$), which are fundamental metrics for classification tasks including LOT, BVI, and TLC. Mean absolute error (MAE), and mean squared error (MSE) are used for evaluation of FLAR regression results.

Specifically, precision (P) and recall (R) are defined as follows,

$$P = \frac{TP}{TP + FP}, \tag{3}$$

$$R = \frac{TP}{TP + FN}, \tag{4}$$

where TP represents the number of true positives, FP is the number of false positives, FN is the number of false negatives.

The $F1_{micro}$ is defined as follows:

$$Precision = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FP_i)}, \tag{5}$$

$$Recall = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FN_i)}, \tag{6}$$

$$F1_{micro} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \tag{7}$$

where C is the total number of classes. TP_i, FP_i, and FN_i are the TP, FP, and FN of the ith class.

The $F1_{macro}$ is defined as follows:

$$F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i}, \tag{8}$$

$$F1_{macro} = \frac{1}{C} \sum_{i=1}^{C} F1_i, \tag{9}$$

where $Precision_i$ and $Recall_i$ are the precision and recall of the ith class.

Additionally, common regression evaluation metrics MAE and MSE are used to evaluate the accuracy of FLAR regression results. Their formulas are given as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|, \tag{10}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \tag{11}$$

where N is the number of ratios for every case, y_i is the ith ground truth value, and $\hat{y}_i$ is the ith predicted value.

3.4. Characteristics of the benchmark

Our task presents the following characteristics:

- **High region and category diversity.** The anatomical regions span extensively from the neck to the femoral artery, encompassing 35 diverse categories. As shown in Fig. 2, the distribution across these categories is highly unbalanced;
- **Variable morphology and indistinct boundaries.** The shapes of various aortic sections and the false lumen are highly variable, and their boundaries are ambiguous, complicating precise identification efforts;
- **Inconsistent occurrence of abnormal features.** Not all samples exhibit tears and specific false lumens, and these categories vary in location. Images with such categories often lack sufficient data, which poses challenges for effective extraction;
- **Challenging tear locations.** The category of tears presents unique challenges due to the small size and unpredictable positioning. Tears can occur at any location between the true and false lumens of the aorta, contributing to difficulties in segmentation.

4. Methods

4.1. SegTAAD

In this section, we introduce the proposed SegTAAD for comprehensive segmentation of TAAD. Note that our primary objective is to design a baseline network specifically tailored to the aortic dissection segmentation task, aiming to enhance model performance for this particular application rather than to develop a unified segmentation framework capable of handling both 2D and 3D tasks. Furthermore, as demonstrated in Section 5.2, experimental results from prior mainstream advanced frameworks reveal that 2D segmentation struggles to capture all categories effectively, particularly smaller yet clinically critical features, such as tears, which are vital for prognosis. Based on these observations, we concluded that 3D segmentation networks are better suited to address the unique challenges of TAAD segmentation. Consequently, our proposed SegTAAD is built upon a 3D segmentation framework. As shown in Fig. 7, SegTAAD consists of three major components: visual encoder, tear&boundary decoder, and enhanced segmentation decoder. SegTAAD utilizes two pieces of domain knowledge: (1) the boundaries plays a key role in the evaluation of clinical features and can enhance the segmentation performance, and (2) the tear locates between TL and FL. Accordingly, two decoders are used for all class segmentation and edge and tear segmentation, respectively. This approach may helps mitigate the issue of imbalanced pixel distribution between the tear and other regions.

Visual encoder. Based on the comparison of current advanced medical segmentation methods in Section 5.3, we contend that transformer-based segmentation models are not yet robust enough to replace CNN-based methods. You et al. (2022) Therefore, we choose the CNN-based modules to obtain multi-scale variations. Our visual encoder is the same as nnUNet (Isensee et al., 2021) that could effectively extract features of different resolutions. It has five encoder stages each of

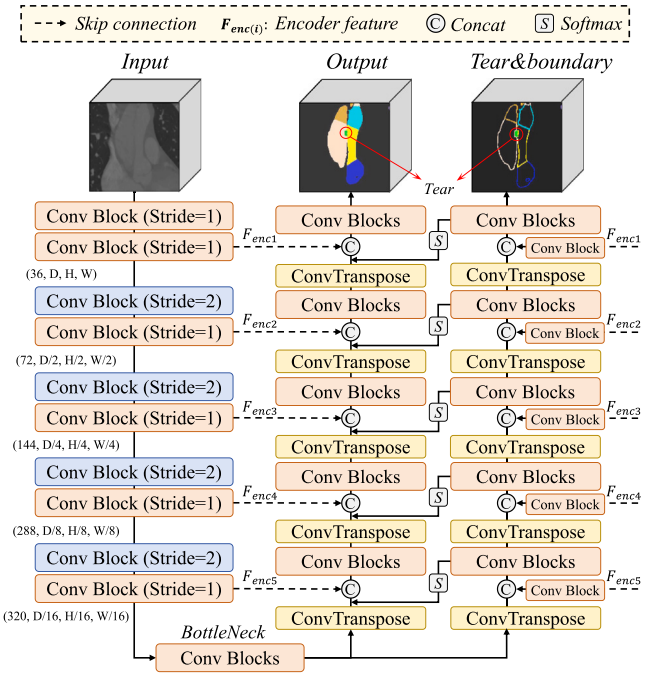

Fig. 7. Overview of the proposed SegTAAD architecture. The input image is initially processed by a CNN encoder to generate multi-stage feature maps, and then the output of the last stage is separately fed into two decoder branches. The tear&boundary decoder uses convolutional blocks to merge skip features from the encoder. The enhanced segmentation decoder is optimized with outputs from each stage of the tear&boundary decoder and encoder feature maps. Finally, it outputs the predicted segmentation result.

which consists of two conv-blocks including a convolutional layer, instance normalization (Ulyanov et al., 2016) and LeakyReLU activation function (Maas et al., 2013). Downsampling within the encoder is accomplished through strided convolutions. Different from the nnUNet default settings, in our dataset, the number of convolutional filters begins with 36 (30 on nnUNet default settings of our dataset) in the initial layer and doubles at each subsequent block. This is because the number of convolutional filters reflects the feature representation capacity of a layer (Khan et al., 2020). Compared to most medical segmentation tasks, our task involves a larger number of categories and covers a broader range of anatomical regions. To effectively capture the distinctions among these categories, we utilize an encoder with enhanced feature representation capacity. This progression continues until the filter count exceeds 288, at which point it is capped at 320 for all the following blocks.

Tear&boundary decoder. Inspired by previous works (Jia et al., 2019; Lee et al., 2020; Hu et al., 2022), it is effective to utilize the boundary of each connection to enhance the segmentation of medical images. Thus, we followed the BANet (Hu et al., 2022) by using the boundary probability maps produced by the tear&boundary decoder at each scale. Different from the original work, our boundary probability maps contain all the tears and the boundary of other categories. We argue that tears, inherently situated at the position between the true and false lumen, should be collectively considered as the boundary type. Additionally, by focusing on the entire tear and boundaries of other classes, we may alleviate the issue of extreme imbalance in the pixel distribution across different classes. This perspective can enhance the network's ability to perceive LOT more accurately. Moreover, instead of directly utilizing skip features from encoder, we add a projection block prior to concatenating each of them with the output generated by the tear&boundary decoder at each scale. We believe that the skip connection information provided by the encoder should be differentiated for each decoder. This projection can clearly distinguish the encoder features required by two different decoders, thus ensuring that each encoder receives the most relevant information and avoiding potential confusion.

Specifically, the visual feature map $X_{\text{enc}} \in \mathbb{R}^{C\times D\times H\times W}$ from the last stage of encoder is fed into the tear&boundary decoder. First, X_{enc} goes through every stage and produces boundary maps $P_{\text{bdr}(i)} \in \mathbb{R}^{C_i\times \frac{D}{k}\times \frac{H}{k}\times \frac{W}{k}}$ $(k = 1, 2, 4, 6, 8)$ from the ith stage. Each stage consists of a transposed convolution with stride 2 for upsampling and two conv-blocks that is similar to encoder conv-block structure. Every output after upsampling is concatenated with the corresponding encoder feature after projection block. Therefore, every decoder stage processes the input feature as follows,

$$F^{tbd}_{\text{enc}(i)} = Convblocks\left(F_{\text{enc}(i)}\right), \tag{12}$$

$$F^{tbd}_{\text{dec}(i)} = Convblocks\left([TransConv(F^{tbd}_{\text{dec}(i-1)}); F^{tbd}_{\text{enc}}(i)]\right), \tag{13}$$

$$P_{\text{bdr}(i)} = Softmax\left(F^{tbd}_{\text{dec}(i)}\right), \tag{14}$$

where $F^{tbd}_{\text{dec}(i-1)}$ and $F^{tbd}_{\text{dec}(i)}$ denote the input and output decoder feature of the ith stage, $F_{\text{enc}(i)}$ indicates the ith encoder feature, $F^{tbd}_{\text{enc}}(i)$ is the ith encoder feature after projection block, and $P_{\text{bdr}(i)}$ indicates the ith boundary probability maps, [;] is the concatenation operation.

Enhanced segmentation decoder. The enhanced segmentation decoder is similar to the tear&boundary decoder. As shown in Fig. 7, each boundary probability map is injected to enhance segmentation feature before it is concatenated with corresponding encoder feature map. To be more specific, every enhanced segmentation decoder stage is formulated as

$$F_{\text{ups}(i)} = TransConv\left(F_{\text{dec}(i-1)}\right), \tag{15}$$

$$F_{\text{enh}(i)} = F_{\text{ups}(i)} + P_{\text{bdr}(i)} \circ F_{\text{ups}(i)}, \tag{16}$$

$$F_{\text{dec}(i)} = Convblocks\left([F_{\text{enh}(i)}; F_{\text{enc}(i)}]\right), \tag{17}$$

where $F_{\text{enh}(i)}$ is the upsampled feature, $F_{\text{enh}(i)}$ indicates the enhancement feature with boundary probability maps, and $F_{\text{dec}(i-1)}$ and $F_{\text{dec}(i)}$ denote the input and output feature maps of the ith stage.

Training objective. Dice loss and cross-entropy loss are used as the loss functions, and we jointly consider minimizing both output segmentation loss and tear&boundary segmentation loss as our training objective. To ensure consistency, all methods did not employ deep supervision. Hence, the training process of SegTAAD can be formulated as

$$\mathcal{L} = 1 - \frac{2\sum_{v=1}^{V} p_v y_v}{\sum_{v=1}^{V}(p_v + y_v + \epsilon)} - \sum_{v=1}^{V}\left(y_v \log p_v + (1-y_v)\log(1-p_v)\right), \tag{18}$$

$$\mathcal{L} = \mathcal{L}^{bdr} + \mathcal{L}^{Seg}. \tag{19}$$

Here, the ground truth and prediction of the vth pixel in the output segmentation map are represented by y_v and p_v, respectively. The total number of voxels is denoted by V, and ϵ serves as a smoothing factor to avoid division by zero.

4.2. Clinical feature extraction

Clinical feature extraction of LOT, BVI, FLAR, and TLC are presented as follows. Note that for the definition of our clinical features, please refer to Section 3.2.

LOT. For each connected component of tears, we first perform dilation on the tear region, and then overlapped pixels between the dilated tear region and other parts are counted. The category of the aorta with the predominant pixel count is identified as the tear's location. Consequently, each case is assigned a one-hot classification label [E0, E1, E2, E3] corresponding to LOT as shown in Fig. 4. For instance, a

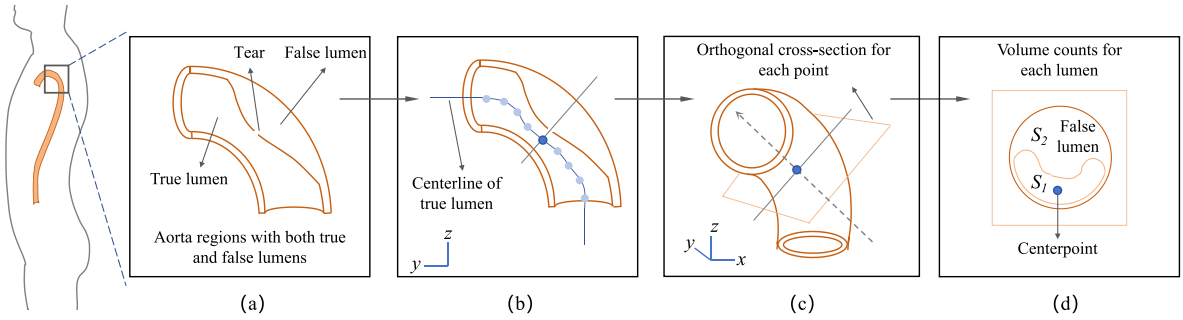

Fig. 8. Assessment pipeline of FLAR in the aorta: (a) Extraction of aortic regions containing both the true and false lumens. (b) Identification of the true lumen's centerline by calculating its skeleton. (c) Generation of orthogonal cross-sections at each point along the centerline. (d) Volume calculation for each lumen within the orthogonal cross-sections, distinguishing between areas S_1 (true lumen) and S_2 (false lumen). The section exhibiting the largest proportion of the false lumen is used to determine $FLAR_{max}$, while all the sections are considered to calculate $FLAR_{avg}$.

label of [1, 0, 0, 0] indicates that the tear is unclear or invisible, while [0, 1, 0, 0] denotes that the tear is located in the ascending aorta.

BVI. 14 branch arteries are considered for BVI. At the origination of each affected branch artery, we quantify the pixels within a region that expands one pixel width outward which is similar to determine LOT. Particularly, we quantitatively assess the pixel count within the true and false lumens among the specified pixels. A branch artery is considered to be originated from false lumen if the pixel count within the false lumen exceeds that of the true lumen.

FLAR. The assessment pipeline is illustrated in Fig. 8. First, the binary mask of the true lumen from the segmentation result I_{mask} for the specified class is extracted. Second, morphological opening and closing operations are applied to smoothing the binary mask, resulting in I'_{mask}. Third, skeletonization (Lee et al., 1994) is employed to derive the 3D skeleton line of the true lumen. The primary skeleton line is then identified by selecting points on the 3D skeleton line that have exactly two adjacent pixels. Forth, the centerline cross-section of the true lumen for each point is calculated based on its adjacent elements. The centerline cross-section of each point can be expressed as follows,

$$(a, b, c) = (x_{i+1} - x_{i-1}, y_{i+1} - y_{i-1}, z_{i+1} - z_{i-1}), \tag{20}$$

$$a(x - x_i) + b(y - y_i) + c(z - z_i) = 0, \tag{21}$$

$$d = \frac{|a(x - x_i) + b(y - y_i) + c(z - z_i)|}{\sqrt{a^2 + b^2 + c^2}} < 0.5, \tag{22}$$

where (x_i, y_i, z_i) represents the current center point under consideration. The points $(x_{i-1}, y_{i-1}, z_{i-1})$ and $(x_{i+1}, y_{i+1}, z_{i+1})$ are the two adjacent points along the skeleton line relative to the center point. The vector (a, b, c) denotes the normal vector of the skeleton line at the point (x_i, y_i, z_i). Eq. (21) describes the cross-section passing through this center point. The variable d represents the distance from a given point (x, y, z) to this cross-section. All voxels with a distance less than 0.5 are considered to be on this cross-section, followed by the calculation of the ratio of pixels between the true and false lumens. This ratio r_i is utilized to assess the perfusion status at ith center point

$$r_i = \frac{S_{2i}}{S_{1i} + S_{2i}}, \tag{23}$$

where S_{1i} is the volume count for true lumen and S_{2i} is the volume count for false lumen.

Finally, we calculate r_i for each point on the centerline, and then get the maximum FLAR, $FLAR_{max}$, and the average FLAR, $FLAR_{avg}$, for each section of the aorta. The two FLAR ratios representing the final perfusion extent are calculated as,

$$FLAR_{max} = \max_{i \in (1,N)} (r_i), \tag{24}$$

$$FLAR_{avg} = \frac{\sum_{i=1}^{N} r_i}{N}. \tag{25}$$

TLC. For each part of the aorta, we calculate TLC_{max} and TLC_{avg}, the maximum and average value of TLC, based on maximum FLAR and average FLAR, respectively. After calculating the regression results, we classify each category of aorta parts based on the commonly used threshold of 50%. Each part is categorized into one of three states: no dissection, no true lumen collapse, or true lumen collapse, thereby obtaining a multi-class label. The results of this metric is influenced by the chosen threshold (50% in our experiment), especially the FLAR value near the predefined threshold significantly determines its reliability. Therefore, this metric should be analyzed based on the specific context and real-world requirements.

5. Experiments and results

5.1. Experimental setup

The experiment was conducted using the Pytorch framework (Paszke et al., 2019) on a single NVIDIA RTX A40 GPU. Several existing SOTA deep network architectures for medical segmentation including the famous CNN-based networks such as nnUNet (Isensee et al., 2021), MedNeXt (Roy et al., 2023), 3DUXNet (Lee et al., 2022), and BANet (Hu et al., 2022), Transformer-based networks, namely TransBTS (Wenxuan et al., 2021), nnformer (Zhou et al., 2021), UNETR (Hatamizadeh et al., 2022), UTNet (Gao et al., 2021), TransFuse (Zhang et al., 2021), and SwinUNETR (Hatamizadeh et al., 2021), and the Mamba-based network named Umamba (Ma et al., 2024) are adopted for comprehensive evaluation. These methods have covered most, if not all, mainstream 2D and 3D segmentation models.

Regarding optimization recipe, nnUNet, BANet (Hu et al., 2022), and SegTAAD employed the stochastic gradient descent algorithm with a momentum of 0.99, whereas other methods were optimized using AdamW (Loshchilov and Hutter, 2017). The initial learning rate was set at 0.01 for SegTAAD model, and other baseline methods aligned with the learning rate settings used in the MedNeXt experiments (Roy et al., 2023). For 3D networks, the input patch size was 128 × 128 × 128 and the batch size was 2, while the 2D networks utilized an input patch size of 512 × 512 and a batch size of 14. Additionally, the 2D network employed sampling of 2D slices from the training subjects and processing in a shuffled manner to ensure that the model learns from a diverse set of examples. The data augmentation techniques was the same as nnUNet: rotations, scalling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring. The data was resampled to 1.0 mm × 1.0 mm × 1.0 mm spacing during both training and inference phases. The training process spanned 1000 epochs without deep supervision. In the testing stage, all experiments were performed using 50% patch overlap without any post-processing. As to the loss function and training configurations, we adopt the settings used in Hu et al. (2022). To fairly compare with existing medical segmentation methods, all baseline methods and our

Table 3
General segmentation performance comparison on ImageTAAD in DSC (%) and HD_{95} (mm). O&B means organs and bone, TLA represents the true lumen of the aorta, BA means branch arteries, FL represents the false lumen. The best scores are marked in bold, and the second-best scores are underlined.

Methods	TransBTS	nnFormer	UNETR	Umamba[a]	UTNet[a]	TransFuse[a]	nnUNet[a]	3DUXNet	SwinUNETR	MedNeXt	BANet	nnUNet	Umamba	SegTAAD
DSC (%)														
Tear	0	0	0	0	0	0	0	17.84	5.36	23.09	16.44	19.87	23.30	**28.62**
O&B	91.804	90.796	89.946	92.064	90.948	90.680	91.510	91.850	91.922	92.994	93.078	93.468	93.348	**93.664**
BA	30.294	48.950	52.634	57.766	59.512	63.505	**66.572**	62.424	62.129	63.874	63.864	65.158	66.226	65.723
TLA	71.990	76.264	75.472	78.752	77.472	77.620	77.338	81.242	80.276	82.184	**82.494**	81.740	82.106	82.410
FL	21.519	22.734	21.628	23.785	30.620	28.692	27.942	32.252	36.250	38.509	37.816	40.446	37.951	**41.137**
Mean	41.659	49.941	50.864	54.304	56.613	57.643	58.733	59.421	59.962	62.237	61.902	63.217	63.121	**64.014**
HD_{95} (mm)														
Tear	–	–	–	–	–	–	–	**17.838**	31.006	48.479	28.035	15.56	39.081	20.281
O&B	23.1278	11.904	33.7364	11.5412	16.687	18.5264	11.4802	25.1888	12.972	10.6098	**9.538**	11.1282	9.5514	10.8266
TLA	27.5658	12.1172	17.884	12.435	11.7506	12.7542	14.5764	11.943	12.4202	16.2838	11.848	**10.1926**	11.8422	10.9166
FL	–	–	–	–	–	–	–	–	30.7123	28.7802	**22.5596**	25.9207	29.8056	22.8318
BA	–	–	101.0074	–	–	50.7848	–	108.4149	96.7007	95.2411	95.8021	101.623	102.4069	**74.9311**
Mean	–	–	–	–	–	–	–	–	51.969	51.546	48.623	51.546	53.651	**40.181**

[a] Denotes 2D model.

proposed SegTAAD model were implemented in the nnUNet framework (Isensee et al., 2021). Such a uniform framework can serve as a neutral testbed for all models, ensuring no bias toward any network in terms of patch size, spacing, augmentations, training, or evaluation.

5.2. General segmentation results

The quantitative segmentation results, measured in terms of DSC and HD_{95}, are presented in Table 3. We have presented the performance results based on the category groupings defined in Section 3.1. Detailed results for each category can be found in Appendix A.1 for further reference.

We can easily notice that our proposed SegTAAD surpasses existing SOTA approaches in the comprehensive segmentation task of TAADs, achieving a DSC of 64.014% and HD_{95} of just 40.181 mm. Notably, in the challenging and critical category of tear, SegTAAD achieves a score of 28.62% and an absolute improvement of 4.69% in DSC, significantly outperforming other techniques. Our method also outperforms existing approaches on the crucial branch artery classes (i.e. ICA 60.80%, IMA 69.43%, LCCA 70.07%, and LSA 71.24% in Appendix A.1). These classes are characterized by smaller sizes, indistinct boundaries, and variable shapes, which pose significant challenges in obtaining accurate segmentation masks. The improvement may due to the fact that we employ the Tear&Boundary decoder branch to enhance boundary learning and reduce volume distribution disparities, effectively aiding the model in focusing on learning these categories. Additionally, to mitigate the impact of training and validation set distributions on model performance, we conducted five-fold cross-validation for both nnUNet and our proposed SegTAAD. The comparative results of this cross-validation can be found in Appendix A.2. We have also calculated the p-value to compare our SegTAAD with the SOTA nnUNet on the official split. The result, $p = 0.04184$, demonstrates the statistical significance of the performance improvement of our method. All experiment results convincingly prove the effectiveness of our approach in addressing the comprehensive segmentation of aortic dissections.

By detailed result analysis of existing methods, it seems that existing segmentation methods do not effectively solve the comprehensive segmentation task of aortic dissection. While these methods achieve high accuracy for classes with large volumes and distinct geometric features, such as the liver, kidneys, and bone, they performed poorly in segmenting various arterial parts and false lumens, which have variable shapes and indistinct boundary. Even in certain classes, such as tears and several branch false lumens (ASFL, IMAFL, LCCAFL, LSAFL), some networks fail to capture relevant features of these classes, resulting in a DSC score of zero during testing. As shown in the HD_{95} part of Table 3, it also can be observed that several existing methods fail to predict any class correctly (indicated by '-'). Only a subset of 3D networks can predict all categories and obtain the overall HD_{95} score.

Based on comparisons among existing methods, we can also find that under the current constraints of limited datasets, CNN-based segmentation methods, particularly nnUNet, demonstrate greater robustness and generally outperform Transformer-based approaches. Transformer-based segmentation models still face some challenges. For instance, the visual tokenization might not effectively capture the intrinsic structures of objects or the detailed spatial information required for the dense prediction task. CNN-based method are more effective at modeling multi-scale variations by learning feature maps at various resolutions, which is crucial for the comprehensive TAAD segmentation. Although it is undisputed that transformers can better model long-range dependencies compared to CNNs, the existing hybrid networks have not demonstrated a clear advantage. Additionally, compared to transformer-based models, Umamba has shown a significant advantage, highlighting the potential of the Mamba network (Gu and Dao, 2023) in medical image segmentation tasks.

Furthermore, it can be observed that 3D models are capable of capturing the information of various classes more comprehensively than 2D models. This deficiency in 2D segmentation models stems from a lack of spatial information. Therefore, for multi-category medical image segmentation tasks covering extensive bodily regions, 3D models are more suitable. However, for categories located near the lower edge of the image (RCIA, LCIA, RFA, LFA), 2D segmentation models outperform the 3D models. As shown in Fig. 12, the left and right arteries of iliac and femoral feature similarly shaped branches that are spatially distant from each other. Patches of $128 \times 128 \times 128$ in 3D models are unable to simultaneously encompass both the left and right arteries. Consequently, following random data augmentation during training, 3D models struggle to differentiate between the left and right information, which can lead to confusion during sliding window predictions. In contrast, 2D model slices can cover both branches simultaneously, facilitating more accurate categorization of each side.

5.3. Clinical feature extraction results

The quantitative results of clinical feature extraction on our proposed dataset are shown in Tables 4–7. For detailed methods of clinical feature extraction and evaluation, readers can refer to Section 4.2.

LOT. Table 4 presents the results of LOT extraction. Note that networks failing to learn the characteristics of the tear will predict its nonexistence, resulting in a label of [1, 0, 0, 0]. Among the 20 test cases, the ground truth for five cases is either no tear or indistinct tears with the same label as no tear. Despite that, these networks still achieve a precision of 0.250, a recall of 0.250, an $F1_{macro}$ score of 0.625, and an $F1_{macro}$ score of 0.500. We have additionally annotated

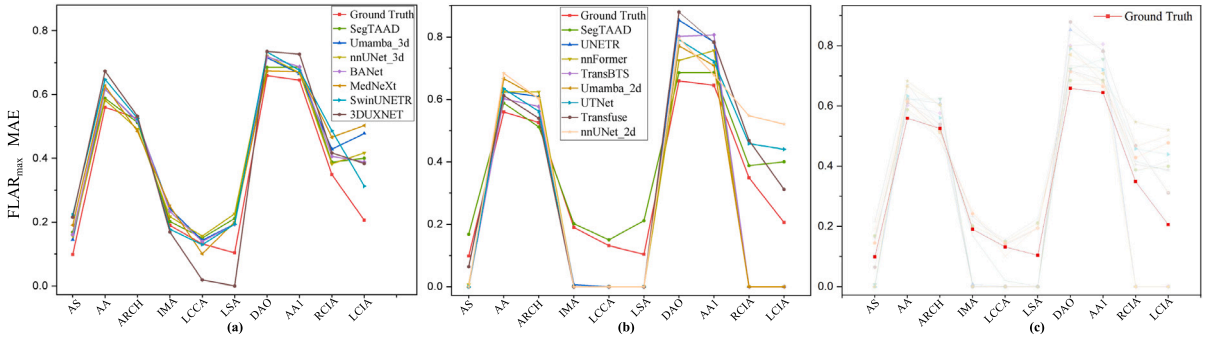

Fig. 9. Comparison of FLAR$_{max}$ for each category. (a) and (b) show the prediction results of different methods compared to the ground truth and SegTAAD. (c) highlights the distribution of the ground truth.

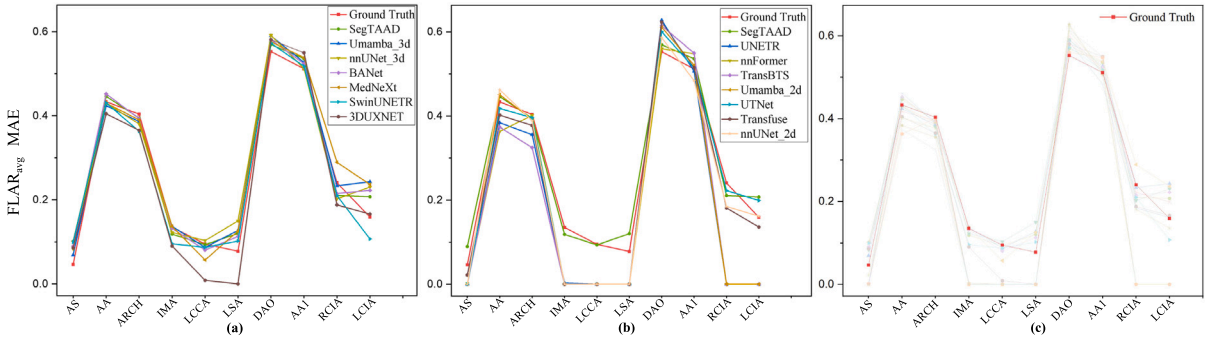

Fig. 10. Comparison of FLAR$_{avg}$ for each category: (a) and (b) show the prediction results of different methods compared to the ground truth and SegTAAD. (c) highlights the distribution of the ground truth.

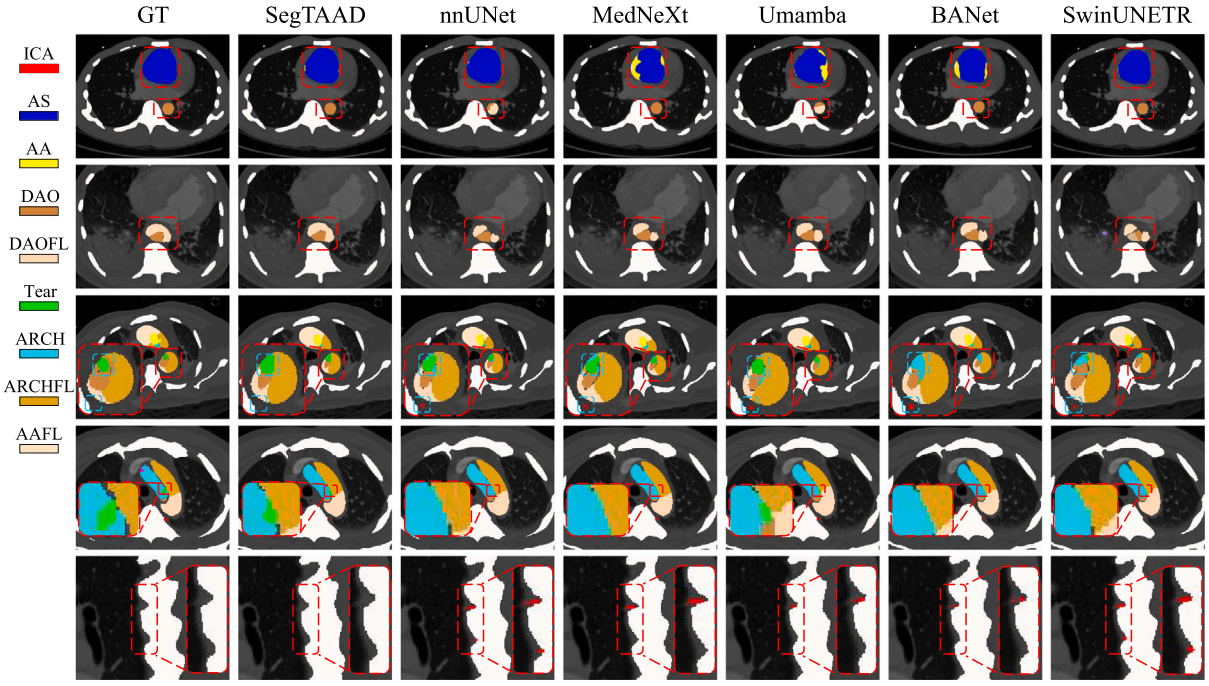

Fig. 11. 2D visualization comparison of important classes regarding TAAD including ICA, AS, AA, AAFL, DAO, DAOFL, Tear, ARCH, and ARCHFL. The first and second rows illustrate the segmentation details of the true and false lumens. The third and fourth rows display the segmentation of tears, while the fifth row shows the segmentation of the intercostal arteries.

the numerical differences from the results that assumed no tears were present in parentheses. This highlights the networks' ability to learn features associated with tears. It can be observed that our SegTAAD outperforms the previous best model by a large margin and achieves an absolute improvement of **6.03**% and **4.8**% in $F1_{micro}$ and $F1_{macro}$, respectively. The improvement may owe to the fact that by conceptualizing tears as boundary features, we can significantly improve the network's performance in identifying these challenging regions, thus effectively addressing the difficulties of tear detection. Moreover, 3D segmentation significantly outperforms 2D segmentation for feature

Table 4
Clinical feature extraction performance comparison on LOT in precision, recall, $F1_{micro}$, and $F1_{macro}$. The highest scores are highlighted in bold, and the second-highest scores are underlined.

Methods	TransBTS	nnFormer	UNETR	Umamba[a]	UTNet[a]	TransFuse[a]	nnUNet[a]	3DUXNet	SwinUNETR	MedNeXt	BANet	nnUNet	Umamba	SegTAAD
Precision	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.424	0.435	0.433	0.500	0.519	0.533	**0.615**
recall	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.700	0.500	0.650	0.750	0.700	**0.800**	**0.800**
$F1_{micro}$	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.688	0.713	0.700	0.750	0.762	0.755	**0.825**
$F1_{macro}$	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.647	0.634	0.651	0.709	0.714	0.738	**0.786**

[a] Denotes 2D models.

Table 5
Clinical feature extraction performance comparison on BVI in precision, recall, $F1_{micro}$, and $F1_{macro}$. SegTAAD is our proposed method for this task. The highest scores are highlighted in bold, and the second highest scores are underlined.

Methods	TransBTS	nnFormer	UNETR	Umamba[a]	UTNet[a]	TransFuse[a]	nnUNet[a]	3DUXNet	SwinUNETR	MedNeXt	BANet	nnUNet	Umamba	SegTAAD
Precision	0.267	0.686	0.455	0.739	0.667	0.679	0.643	**0.750**	0.667	0.692	0.628	0.694	0.674	0.690
Recall	0.087	0.552	0.109	0.370	0.348	0.413	0.391	0.261	0.478	0.587	0.587	0.543	**0.630**	**0.630**
$F1_{micro}$	0.735	0.835	0.765	0.825	0.810	0.820	0.810	0.810	0.825	0.845	0.825	0.840	0.845	**0.850**
$F1_{macro}$	0.487	0.745	0.519	0.694	0.671	0.702	0.685	0.637	0.724	0.768	0.747	0.755	0.776	**0.781**

[a] Denotes 2D models.

Table 6
Clinical feature extraction performance comparison on $FLAR_{max}$ and $FLAR_{avg}$ in both mean absolute error (MAE) and mean squared error (MSE). The lowest scores are highlighted in bold, and the second-lowest scores are underlined.

Methods	TransBTS	nnFormer	UNETR	Umamba[a]	UTNet[a]	TransFuse[a]	nnUNet[a]	3DUXNet	SwinUNETR	MedNeXt	BANet	nnUNet	Umamba	SegTAAD
$FLAR_{max}$														
MAE	0.1730	0.1580	0.1834	0.1587	0.1440	0.1698	0.1714	0.1362	0.1218	0.1337	0.1144	0.1133	0.1253	**0.1055**
MSE	0.1097	0.1020	0.1209	0.1058	0.0874	0.1005	0.1071	0.0630	0.0514	0.0640	0.0507	0.0497	0.0570	**0.0463**
$FLAR_{avg}$														
MAE	0.1195	0.1113	0.1210	0.1089	0.1006	0.1051	0.1087	0.0976	0.0873	0.0960	0.0852	0.0842	0.0862	**0.0836**
MSE	0.0565	0.0547	0.0585	0.0536	0.0399	0.0416	0.0412	0.0305	0.0272	0.0325	**0.0256**	0.0265	0.0266	**0.0256**

[a] Denotes 2D models.

Table 7
Clinical feature extraction performance comparison on TLC_{max} and TLC_{avg} in precision, recall, $F1_{micro}$, and $F1_{macro}$. The highest scores are highlighted in bold, and the second highest scores are underlined.

Methods	TransBTS	nnFormer	UNETR	Umamba[a]	UTNet[a]	TransFuse[a]	nnUNet[a]	3DUXNet	SwinUNETR	MedNeXt	BANet	nnUNet	Umamba	SegTAAD
TLC_{max}														
Precision	0.621	0.566	0.576	0.598	0.601	0.593	0.554	0.592	**0.636**	0.614	0.635	0.610	0.627	0.633
Recall	0.633	0.541	0.571	0.584	0.632	0.632	0.543	0.556	0.635	0.612	0.620	0.631	0.634	**0.641**
$F1_{micro}$	0.720	0.740	0.740	0.745	0.659	0.730	0.725	0.690	0.710	0.680	0.615	**0.770**	0.710	0.760
$F1_{macro}$	0.587	0.547	0.567	0.585	0.601	0.594	0.544	0.548	0.581	0.558	0.592	**0.616**	0.578	0.610
TLC_{avg}														
Precision	0.653	0.650	0.637	0.612	0.631	0.644	0.606	0.655	0.681	0.676	0.674	0.675	0.682	**0.694**
Recall	0.614	0.615	0.607	0.579	0.612	0.615	0.564	0.615	0.640	0.663	0.638	0.671	0.677	**0.690**
$F1_{micro}$	0.705	0.645	0.705	**0.710**	0.690	0.705	0.630	0.610	0.610	0.640	0.630	0.695	0.650	0.655
$F1_{macro}$	0.613	0.615	0.608	0.588	0.605	0.618	0.562	0.578	0.587	0.614	0.604	**0.639**	0.618	0.632

[a] Denotes 2D models.

extraction of tear, which is due to the fact that 3D modeling improves the context of spatial information.

BVI. Table 5 presents the results of BVI classification. Our method achieves a recall of 0.630, an $F1_{micro}$ score of 0.850, and an $F1_{macro}$ score of 0.781, which outperform others significantly. Umamba2d achieves the second highest $F1_{macro}$ score of 0.776, while 3DUXNet obtains the highest precision at 0.750, followed closely by Umamba2d with a precision of 0.739. It is noteworthy that networks with suboptimal *DSC* score are able to extract relatively better clinical features. For instance, nnFormer, despite a total *DSC* score of only 49.941%, achieves an $F1_{macro}$ of 0.745 for BVI extraction. Although UNETR, Umamba2d, 3DUXNet, and so on have higher *DSC* scores than nnFormer, UNETR has an $F1_{macro}$ of only 0.519, and others also underperform relative to nnFormer. Consequently, for clinical practitioners assessing the extent of involvement, the effectiveness of nnFormer is higher, even though other methods have a higher *DSC* score.

FLAR. The quantitative results of FLAR extraction are presented in Table 6. We have presented the overall performance results of $FLAR_{max}$ and $FLAR_{avg}$. Detailed FLAR results for each category can

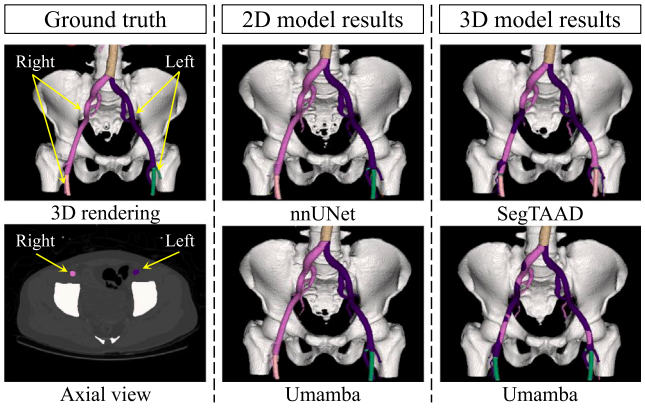

Fig. 12. Visual comparison of segmentation results for the left and right iliac and femoral arteries between ground truth, 2D models, and 3D models.

Table 8
Ablation study on ImageTAAD dataset. CF36 means that the number of convolutional filters is 36. TBD stands for using Tear&boundary decoder. P stands for adding the projection blocks prior to concatenating the encoder features with the output features of tear&boundary decoder.

Settings	CF36	TBD	P	General metrics		Clinical metrics			
				DSC↑	HD95↓	LOT↑	BVI↑	$FLAR_{MAE}$↓	$FLAR_{MSE}$↓
nnUNet	✗	✗	✗	63.217	51.546	0.714	0.755	0.0842	0.0265
a	✓	✗	✗	63.302	50.085	0.715	0.724	0.1989	0.1257
b	✓	✓	✗	63.887	46.878	0.768	0.778	0.1698	0.0998
Ours	✓	✓	✓	**64.014**	**40.181**	**0.786**	**0.781**	**0.0836**	**0.0256**

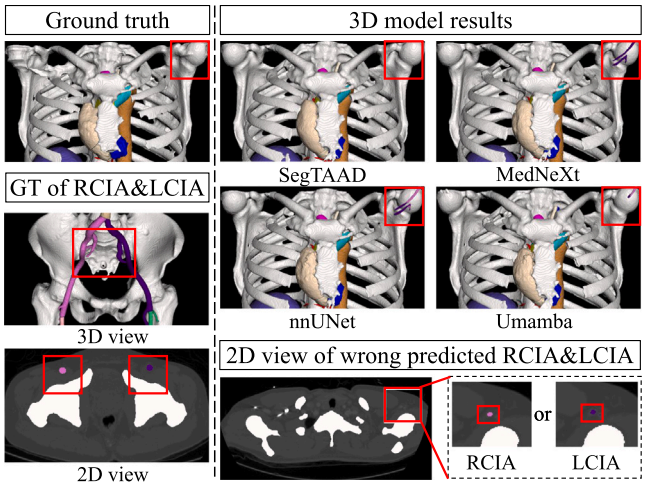

Fig. 13. Examples of the incorrectly predicted left and right iliac arteries (RCIA and LCIA) in 3D and 2D views. Additional examples of the ground truth (RCIA and LCIA) in 3D and 2D views are included for comparison.

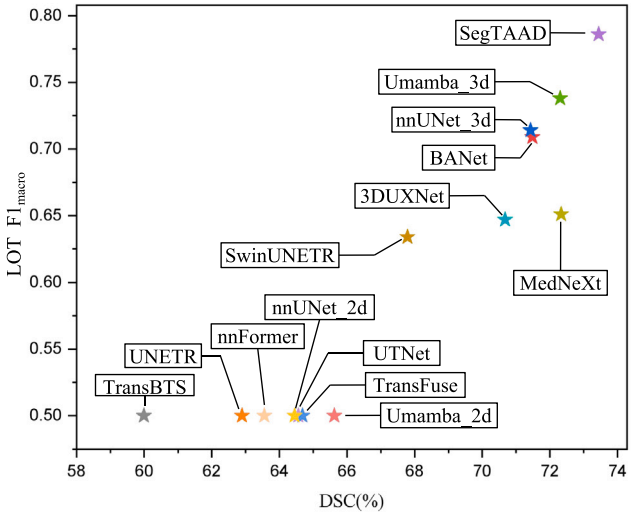

Fig. 14. The correlation between the DSC score and the $F1_{macro}$ score for LOT. The x-axis represents the DSC score, while the y-axis represents the $F1_{macro}$ score.

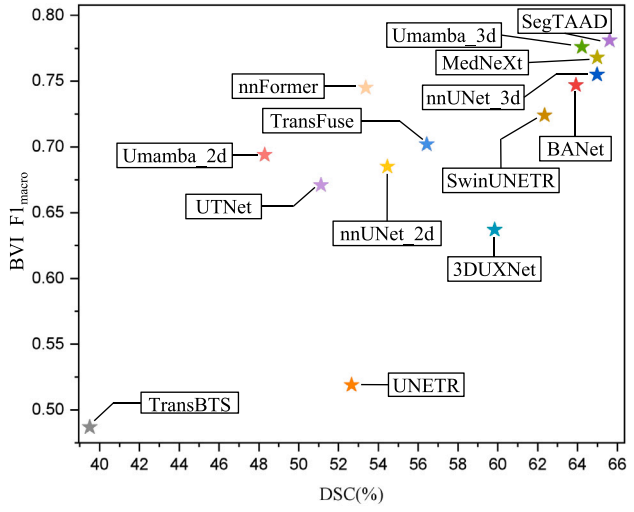

Fig. 15. The correlation between the DSC score and the $F1_{macro}$ score for BVI. The x-axis represents the DSC score, while the y-axis represents the $F1_{macro}$ score.

be found in Appendix A.3 for further reference. If a false lumen was not predicted in a segment result, the false lumen perfusion area ratio is recorded as zero. For $FLAR_{max}$, our method achieves the lowest MAE (R_{max}) of 0.1055 and an MSE (R_{max}) of 0.0463. nnUNet followed closely with an MAE (R_{max}) of 0.1133 and an MSE (R_{max}) of 0.0497. SwinUNETR and MedNext also show high performances in several classes. Regarding $FLAR_{avg}$, our method again achieves the lowest MAE (R_{avg}) of 0.0836 and MSE (R_{avg}) of 0.0256. BANet similarly achieves the lowest MSE (R_{avg}), while nnUNet obtains the second lowest MAE (R_{avg}). Different methods exhibit substantial variation in effectiveness across various classes. For example, Transfuse significantly outperforms others in assessing the perfusion ratio for category 'AS', though its overall performance is comparatively weaker. Figs. 9 and 10 show the results of $FLAR_{max}$ and $FLAR_{avg}$, respectively. It is evident that our method's overall distribution more closely approximates the ground truth compared to other approaches. Furthermore, as illustrated in Fig. 9(c) and Fig. 10(c), except for the instances marked as zero (where the false lumen is not recognized), all methods exhibit various degrees of overprediction of the false lumen area, resulting in larger predicted false lumen proportion than the ground truth.

TLC. The quantitative results of TLC classification are presented in Table 7 for TLC_{max} and TLC_{avg}. For TLC_{max}, nnUNet_3d achieves the best overall performance with an $F1_{micro}$ of 0.770 and an $F1_{macro}$ of 0.770. Following that, SegTAAD achieves an $F1_{micro}$ of 0.760 and an $F1_{macro}$ of 0.610. However, our method achieves the highest recall of 0.641. SwinUNETR records the highest precision of 0.636. For TLC_{avg}, nnUNet_3d also achieves the best overall performance with an $F1_{macro}$ of 0.639. Our method obtains the highest precision (0.694) and recall (0.690).

5.4. Ablation study

As shown in Table 8, we conducted ablation studies on the ImageTAAD dataset, leveraging nnU-Net as a robust baseline to assess the impact of three key components of our proposed SegTAAD: the configuration of convolutional filters (CF36), the tear&boundary decoder (TBD), and the projection blocks (P), which are applied before concatenating encoder features with the output features of the tear&boundary decoder.

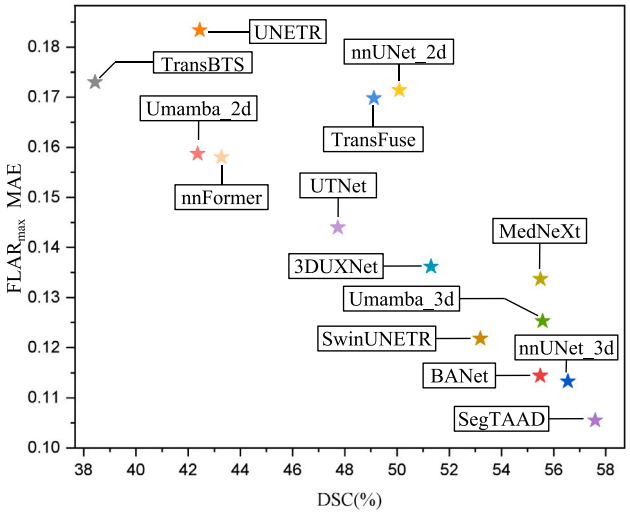

Fig. 16. The correlation between the *DSC* score and the MAE for $FLAR_{max}$. The *x*-axis represents the *DSC* score, while the *y*-axis represents the MAE score.

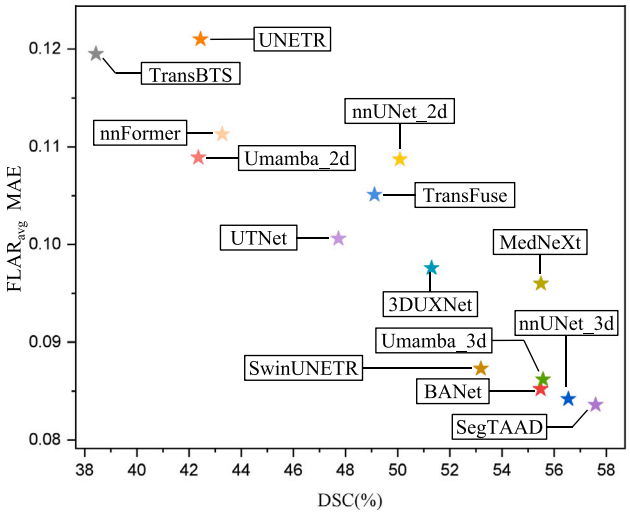

Fig. 17. The correlation between the *DSC* score and the MAE for $FLAR_{avg}$. The *x*-axis represents the *DSC* score, while the *y*-axis represents the MAE score.

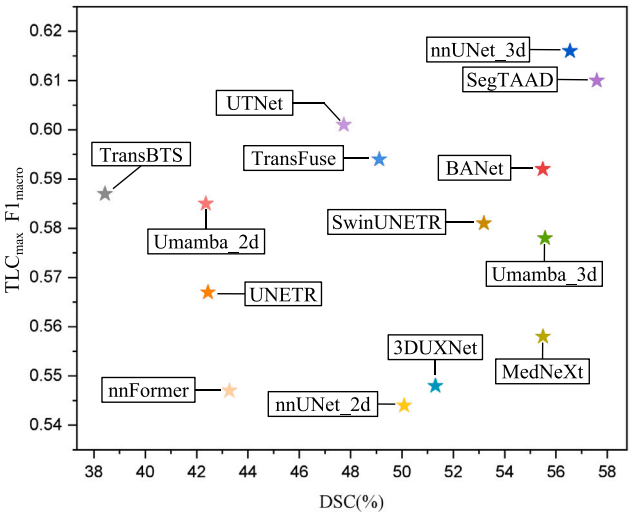

Fig. 18. The correlation between the *DSC* score and the $F1_{macro}$ for TLC_{max}. The *x*-axis represents the DSC score, while the *y*-axis represents the $F1_{macro}$ score.

First, when the convolutional filters are set to 36 without TBD and P (as shown in setting "a"), the results show a slight improvement over the baseline nnU-Net in general metrics. This indicates that increasing the number of convolutional filters enhances the model's feature extraction capability. However, clinical metrics such as average $FLAR_{MAE}$ and average $FLAR_{MSE}$ still exhibit significant room for improvement.

When TBD is enabled without projection blocks (setting "b"), the results improve further, particularly in clinical metrics like LOT and BVI. This suggests that the tear&boundary decoder contributes to a better understanding of structural and boundary information, which is crucial for clinically meaningful segmentation. Nevertheless, the average $FLAR_{MSE}$, which measures errors in false lumen prediction, remains relatively high, indicating limitations in effectively integrating detailed tear and boundary information into the segmentation output.

Finally, when all components are enabled (ours), the model achieves the best performance across all metrics. Notably, significant improvements are observed in HD95 and clinical metrics such as average $FLAR_{MAE}$ and average $FLAR_{MSE}$. This demonstrates the effectiveness of the projection blocks in aligning encoder features with tear and boundary features, enabling the model to produce more accurate and clinically relevant segmentations. Additionally, the DSC metric also shows a modest yet meaningful improvement, reflecting overall consistency in segmentation performance.

5.5. Qualitative results

Qualitative results are shown in Fig. 11. We primarily compare the results of those methods capable of predicting all categories. To more clearly observe the details, a 2D perspective is utilized to display the segmentation results. The first and second row illustrate the segmentation details of the true and false lumens. Our method distinguishes between these parts more effectively, exhibiting fewer instances of confusion and more accurate contour determination. The third and fourth rows display the segmentation of tears, where our approach is shown to more precisely identify tears and their exact locations. The fifth row shows the segmentation of the intercostal arteries. Compared with existing approaches, our method tend to have less false positive cases in the above nine classes. This discussion primarily highlights classes crucial for diagnosing TAAD, where 2D visual clarity is superior. Additional 3D visualization results of more classes are presented in Appendix A.4.

Furthermore, as shown in Fig. 13, we observed that without post-processing, 3D segmentation models tend to erroneously predict categories (RCIA, LCIA, RFA, LFA) in the area in front of the shoulders. As show in the 2D view for GT of RCIA and LCIA and the 2D view of wrong predicted RCIA and LCIA in Fig. 13, this misclassification arises because the patch features of these two specific areas are very similar, misleading the model to make incorrect judgments without a global view of the entire CT scan. Therefore, another potential direction for this task involves acquiring global spatial information from whole CT scans and integrating it with high-resolution local features for segmentation.

5.6. Performance correlation between general segmentation and clinical feature extraction

Results of performance correlation analyses on clinical features and segmentation results are shown in Figs. 14–19 for LOT, BVI, $FLAR_{avg}$, $FLAR_{max}$, respectively. For ease of discussion, commonly used metrics including *DSC* and $F1_{macro}$ are considered. Overall, there exists a clear correlation between *DSC* and clinical features, i.e. a higher *DSC* generally corresponds to higher feature extraction performance. However, there exist quite a lot of outliers. As shown in Fig. 14, we can notice that while MedNeXt achieves a higher Dice score compared to nnUNet_3d and BANet, it significantly underperforms in terms of the $F1_{macro}$ score for detecting LOT. We can also find similar phenomena in Figs. 15–19.

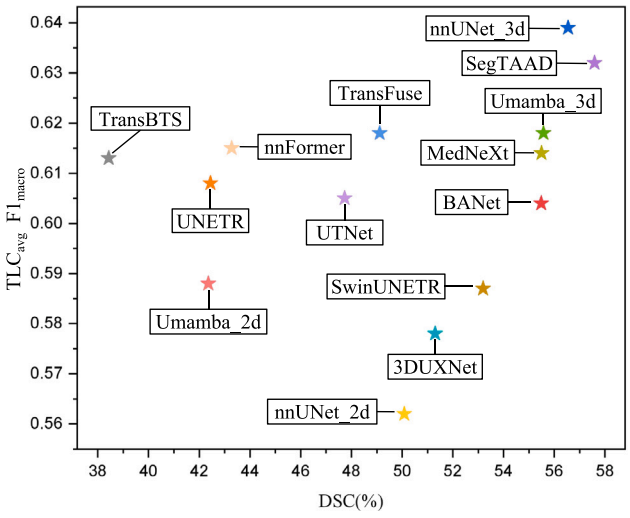

Fig. 19. The correlation between the DSC score and the $F1_{macro}$ for TLC_{avg}. The x-axis represents the DSC score, while the y-axis represents the $F1_{macro}$ score.

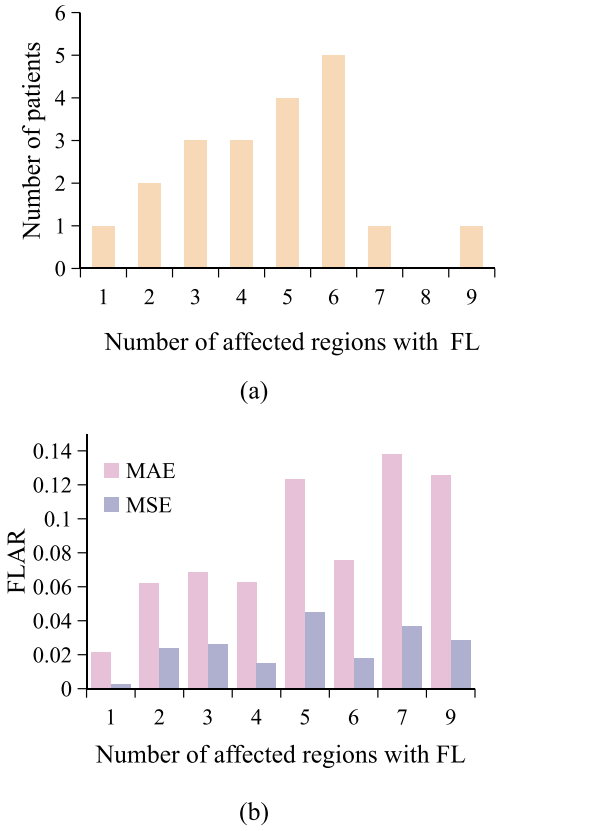

Fig. 20. Impact of disease severity on the performance of clinical feature extraction. (a) Distribution of aortic dissection extent in the test set. (b) Correlation between FLAR and disease severity.

In Fig. 15, although UNETR achieves higher DSC score compared to Umamba_2d and UTNet, its $F1_{macro}$ score for the feature of involvement extent is significantly lower than that of Umamba2d and UTNet. Although nnformer has relatively lower DSC scores, its $F1_{macro}$ score is comparatively higher. This indicates that relying solely on DSC score does not accurately assess the clinical effectiveness of segmentation results. Figs. 16 and 17, such trend is mitigated without as many outliers as UNETR in Fig. 15. As for Figs. 18 and 19, the correlation between the DSC score and clinical features appears less pronounced. For instance, the DSC scores of models such as MedNeXt, Umamba_3d, and BANet are very similar, yet there is an obvious difference in their F1 scores. Notably, TransBTS exhibits the lowest DSC score but surpasses over 50% of the methods in terms of F1 score. This phenomenon primarily stems from the fact that the TLC focuses only on distinguishing effects at a given threshold, hence reflecting the accuracy for samples close to this threshold. By looking at the details of the methods with bad performance in Fig. 14, we can discover that there seems to exist a DSC threshold for good $F1_{macro}$. For example, considering the group of methods including TransBTS, UNETR, nnFormer, nnUnet_2d, UTNet, TransFuse and Umamba_2d, their $F1_{macro}$ are almost the same around 0.50 though their DSC are quite different. However, when the DSC of SwinUNETR reaches beyond for about 0.65, its $F1_{macro}$ obtains a large improvement. We would like to highlight that our SegTAAD achieves the highest performance in all the figures. The main reason is possibly that we emphasize boundaries and tears in the network design, which is quite important for clinical feature extraction.

We also analyzed the number of regions affected by the aortic false lumen in each patient from the testset across 10 specific regions: the aortic sinus, ascending aorta, aortic arch, innominate artery, left common carotid artery, left subclavian artery, descending aorta, abdominal aorta, right iliac artery, and left iliac artery. A higher number of affected regions suggests a more severe aortic dissection in the patient. Fig. 20(a) displays the distribution of affected regions within the testset, with the horizontal axis representing the number of affected regions and the vertical axis showing the patient count. Based on this, we compared the average FLAR ($FLAR_avg$) results with the number of affected regions, which serves as an indicator of disease severity. As shown in Fig. 20(b), a greater extent of false lumen involvement correlates with a higher potential error in the FLAR results. However, when the number of affected regions is 6, the error appears comparable to that observed in cases with fewer affected regions, and it is obviously smaller than the error when 5 regions are involved.

This suggests that although more severe cases of aortic dissection, which involve a larger tear and a broader area of false lumen, may present additional challenges for segmentation, this does not always result in worse segmentation outcomes. The segmentation performance also depends on whether the morphological characteristics of the affected aortic regions are easily distinguishable. Considering that LOT is solely determined by the location of the tear, BVI focuses only on the impact of the aortic arch and abdominal aortic dissection on the branches of the arteries, and TLC is a further manifestation of FLAR features, a deeper analysis of the relationship between these metrics and disease severity may be not particularly valuable. Therefore, we have not performed additional comparisons of these metrics.

6. Discussion and conclusion

In this study, we have established a new benchmark for comprehensive TAAD segmentation. This includes releasing the ImageTAAD dataset to the public, proposing a series of innovative clinical feature metrics for evaluating this task, and inventing a solid baseline segmentation framework named SegTAAD. The ImageTAAD dataset comprises 120 cases, with 35 foreground categories annotated, essential for clinical and prognostic purposes. The publication of this dataset will fill existing gaps in fundamental resources and encourage researchers to further explore in this field, thereby fostering the development of relevant techniques. Besides the dataset, we propose the clinical feature metrics mainly focusing on four clinical features which are highly aligned with clinical application. We evaluate high-level features from the segmentation results to compare the clinical efficiency between different methods. Finally, we propose a novel network architecture named SegTAAD which is suitable for the TAAD segmentation task and

has surpassed all previous models in both general segmentation metrics and clinical feature metrics. The key insight is to integrate the tear and the boundary to capture the local features of comprehensive categories to help enhance the final segmentation results.

Furthermore, we have compared existing SOTA methods using the SegTAAD dataset and identified several unresolved or challenging issues. Although frameworks such as nnUNet are robust and generalizable, effectively tackling most medical segmentation tasks and outperforming existing methods, they still fall short in addressing the specific problems presented in our proposed dataset and tasks. Even our baseline method, which achieved SOTA results, exhibits limitations in accuracy and offers room for improvement. The comprehensive segmentation task of TAAD remains a challenging problem that requires further exploration, given its significance in clinical practice. Future work could focus on a more in-depth analysis under a cascaded framework and the development of more complex, targeted technical innovations tailored to the characteristics of aortic dissection tasks hold significant potential. Additionally, exploring how to leverage advanced large pre-trained models and multimodal information to enhance TAAD segmentation represents another promising direction.

CRediT authorship contribution statement

Shanshan Song: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis. **Hailong Qiu:** Resources, Formal analysis, Data curation, Conceptualization. **Meiping Huang:** Resources, Formal analysis, Data curation, Conceptualization. **Jian Zhuang:** Resources, Formal analysis, Data curation, Conceptualization. **Qing Lu:** Writing – review & editing, Software, Formal analysis, Conceptualization. **Yiyu Shi:** Software, Resources, Formal analysis, Conceptualization. **Xiaomeng Li:** Supervision, Project administration, Formal analysis, Conceptualization. **Wen Xie:** Writing – review & editing, Resources, Formal analysis, Data curation. **Guang Tong:** Supervision, Project administration, Formal analysis, Conceptualization. **Xiaowei Xu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work and the collection of data of retrospective data on implied consent received Research Ethics Committee (REC) approval from Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences under Protocol No. 2022-048-01. It complies with all relevant ethical regulations. Deindentification was performed in which all CT files are transformed into NIfTI format, and sensitive information of the patients including name, birther day, admission year, admission number, and CT number is removed. Only de-identified retrospective data were used for research, without the active involvement of patients.

This work was supported by the Natural Science Foundation of China (No. 62276071), Guangdong Special Support Program-Science and Technology Innovation Talent Project, China (No. 0620220211), the Science and Technology Planning Project of Guangdong Province, China (No. 2019B020230003), Guangdong Peak Project, China (No. DFJH201802), Guangzhou Science and Technology Planning Project, China (No. 202206010049), the National Guangdong Basic and Applied Basic Research Foundation, China (No. 2022A1515010157, 2022A1515011650), 2024A1515010105), Guangzhou Science and Technology Planning Project, China (No. 202102080188), the National Natural Science Foundation of China (No. 62306254, the National Natural Science Foundation of China (NSFC) and the Research Grants Council (RGC) of Hong Kong under the Joint Research Scheme (JRS) (No. N_HKUST654).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2025.103512.

Data availability

All the dataset, code and trained models have been published (Xiaowei, 2024).

References

Cao, L., Shi, R., Ge, Y., Xing, L., Zuo, P., Jia, Y., Liu, J., He, Y., Wang, X., Luan, S., et al., 2019. Fully automatic segmentation of type B aortic dissection from cta images enabled by deep learning. Eur. J. Radiol. 121, 108713.

Chen, D., Zhang, X., Mei, Y., Liao, F., Xu, H., Li, Z., Xiao, Q., Guo, W., Zhang, H., Yan, T., et al., 2021. Multi-stage learning for segmentation of aortic dissections using a prior aortic anatomy simplification. Med. Image Anal. 69, 101931.

Cheng, J., Tian, S., Yu, L., Ma, X., Xing, Y., 2020. A deep learning algorithm using contrast-enhanced computed tomography (CT) images for segmentation and rapid automatic detection of aortic dissection. Biomed. Signal Process. Control. 62, 102145.

Chung, J.W., Elkins, C., Sakai, T., Kato, N., Vestring, T., Semba, C.P., Slonim, S.M., Dake, M.D., 2000. True-lumen collapse in aortic dissection 1: Part I. Evaluation of causative factors in phantoms with pulsatile flow. Radiology 214 (1), 87–98.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer, pp. 424–432.

Criado, F.J., 2011. Aortic dissection: A 250-year perspective. Tex. Heart Inst. J. 38 (6), 694.

Daily, P.O., Trueblood, H.W., Stinson, E.B., Wuerflein, R.D., Shumway, N.E., 1970. Management of acute aortic dissections. Ann. Thorac. Surg. 10 (3), 237–247.

Erbel, R., Alfonso, F., Boileau, C., Dirsch, O., Eber, B., Haverich, A., Rakowski, H., Struyven, J., Radegran, K., Sechtem, U., et al., 2001. Diagnosis and management of aortic dissection: Task force on aortic dissection, European society of cardiology. Eur. Heart J. 22 (18), 1642–1681.

Evangelista, A., Salas, A., Ribera, A., Ferreira-González, I., Cuellar, H., Pineda, V., González-Alujas, T., Bijnens, B., Permanyer-Miralda, G., Garcia-Dorado, D., 2012. Long-term outcome of aortic dissection with patent false lumen: Predictive role of entry tear size and location. Circulation 125 (25), 3133–3141.

Fattouch, K., Sampognaro, R., Navarra, E., Caruso, M., Pisano, C., Coppola, G., Speziale, G., Ruvolo, G., 2009. Long-term results after repair of type a acute aortic dissection according to false lumen patency. Ann. Thorac. Surg. 88 (4), 1244–1250.

Feng, H., Fu, Z., Wang, Y., Zhang, P., Lai, H., Zhao, J., 2023. Automatic segmentation of thrombosed aortic dissection in post-operative CT-angiography images. Med. Phys. 50 (6), 3538–3548.

Furui, M., Uesugi, N., Matsumura, H., Hayashida, Y., Kuwahara, G., Fujii, M., Shimizu, M., Morita, Y., Ito, C., Hayama, M., et al., 2024. Relationship between false lumen morphology and entry tear in acute type a aortic dissection. Eur. J. Cardiothorac Surg. 65 (2), ezad389.

Gao, Y., Zhou, M., Metaxas, D.N., 2021. UTNet: A hybrid transformer architecture for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24. Springer, pp. 61–71.

Gu, A., Dao, T., 2023. Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752.

Harris, K.M., Strauss, C.E., Eagle, K.A., Hirsch, A.T., Isselbacher, E.M., Tsai, T.T., Shiran, H., Fattori, R., Evangelista, A., Cooper, J.V., et al., 2011. Correlates of delayed recognition and treatment of acute type A aortic dissection: The international registry of acute aortic dissection (IRAD). Circulation 124 (18), 1911–1918.

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D., 2021. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. Springer, pp. 272–284.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 574–584.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

Hu, S., Liao, Z., Xia, Y., 2022. Boundary-aware network for abdominal multi-organ segmentation. arXiv preprint arXiv:2208.13774.

Igarashi, T., Sato, Y., Satokawa, H., Takase, S., Iwai-Takano, M., Seto, Y., Yokoyama, H., 2022. Ratio of the false lumen to the true lumen is associated with long-term prognosis after surgical repair of acute type A aortic dissection. JTCVS Open 10, 75–84.

Immer, F.F., Kráhenbuúhl, E., Hagen, U., Stalder, M., Berdat, P.A., Eckstein, F.S., Schmidli, J., Carrel, T.P., 2005. Large area of the false lumen favors secondary dilatation of the aorta after acute type A aortic dissection. Circulation 112 (9_supplement), I–249.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211.

Jia, H., Song, Y., Huang, H., Cai, W., Xia, Y., 2019. HD-Net: Hybrid discriminative network for prostate segmentation in MR images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. Springer, pp. 110–118.

Jiang, Q., Huang, K., Wang, D., Xia, J., Yu, T., Hu, S., 2023. A comparison of bilateral and unilateral cerebral perfusion for total arch replacement surgery for non-marfan, type A aortic dissection. Perfusion 02676591231161919.

Jung, J.-H., Oh, H.M., Jeong, G.-J., Kim, T.-W., Koo, H.J., Lee, J.-G., Yang, D.H., 2024. ZOZI-seg: A transformer and unet cascade network with zoom-out and zoom-in scheme for aortic dissection segmentation in enhanced CT images. Comput. Biol. Med. 175, 108494.

Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53, 5455–5516.

Kim, J.-H., Lee, S.H., Lee, S., Youn, Y.-N., Yoo, K.-J., Joo, H.-C., 2022. Role of false lumen area ratio in late aortic events after acute type I aortic dissection repair. Ann. Thorac. Surg. 114 (6), 2217–2224.

Lee, H.H., Bao, S., Huo, Y., Landman, B.A., 2022. 3D ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. arXiv preprint arXiv:2209.15076.

Lee, T.-C., Kashyap, R.L., Chu, C.-N., 1994. Building skeleton models via 3-D medial surface axis thinning algorithms. CVGIP, Graph. Models Image Process. 56 (6), 462–478.

Lee, H.J., Kim, J.U., Lee, S., Kim, H.G., Ro, Y.M., 2020. Structure boundary preserving segmentation for medical image with ambiguous boundary. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4817–4826.

Li, X., Qiao, H., Shi, Y., Xue, J., Bai, T., Liu, Y., Sun, L., 2020. Role of proximal and distal tear size ratio in hemodynamic change of acute type A aortic dissection. J. Thorac. Dis. 12 (6), 3200.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Lyu, T., Yang, G., Zhao, X., Shu, H., Luo, L., Chen, D., Xiong, J., Yang, J., Li, S., Coatrieux, J.-L., et al., 2021. Dissected aorta segmentation using convolutional neural networks. Comput. Methods Programs Biomed. 211, 106417.

Ma, J., Li, F., Wang, B., 2024. U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722.

Maas, A.L., Hannun, A.Y., Ng, A.Y., et al., 2013. Rectifier nonlinearities improve neural network acoustic models. In: Proc. Icml. vol. 30, Atlanta, GA, p. 3, 1.

Mastrodicasa, D., Codari, M., Bäumler, K., Sandfort, V., Shen, J., Mistelbauer, G., Hahn, L.D., Turner, V.L., Desjardins, B., Willemink, M.J., et al., 2022. Artificial intelligence applications in aortic dissection imaging. In: Seminars in Roentgenology. vol. 57, Elsevier, pp. 357–363, 4.

Nienaber, C.A., Clough, R.E., Sakalihasan, N., Suzuki, T., Gibbs, R., Mussa, F., Jenkins, M.P., Thompson, M.M., Evangelista, A., Yeh, J.S., et al., 2016. Aortic dissection. Nat. Rev. Dis. Prim. 2 (1), 1–18.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Pepe, A., Egger, J., Codari, M., Willemink, M.J., Gsaxner, C., Li, J., Roth, P.M., Schmalstieg, D., Mistelbauer, G., Fleischmann, D., 2023. Automated cross-sectional view selection in CT angiography of aortic dissections with uncertainty awareness and retrospective clinical annotations. Comput. Biol. Med. 165, 107365.

Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H., 2023. Mednext: Transformer-driven scaling of convnets for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 405–415.

Selle, D., Preim, B., Schenk, A., Peitgen, H.-O., 2002. Analysis of vasculature for liver surgical planning. IEEE Trans. Med. Imaging 21 (11), 1344–1357.

Sherif, M., Podila, R.S., Von Oppell, U., 2017. Acute aortic dissections initially incorrectly managed as acute coronary syndromes prior to surgery-5 years review: 0287. Int. J. Surg. 47, S23.

Sieren, M.M., Widmann, C., Weiss, N., Moltz, J.H., Link, F., Wegner, F., Stahlberg, E., Horn, M., Oecherting, T.H., Goltz, J.P., et al., 2022. Automated segmentation and quantification of the healthy and diseased aorta in CT angiographies using a dedicated deep learning approach. Eur. Radiol. 32 (1), 690–701.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2016. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.

Wang, W., Wang, L., Chai, C., Sun, Q., Yuan, Y., Wang, T., Wu, L., Tang, Z., 2023. Prognostic impact of branch vessel involvement on organ malperfusion and mid-term survival in patients with acute type A aortic dissection. Int. J. Cardiol. 381, 81–87.

Weiss, G., Wolner, I., Folkmann, S., Sodeck, G., Schmidli, J., Grabenwöger, M., Carrel, T., Czerny, M., 2012. The location of the primary entry tear in acute type B aortic dissection affects early outcome. Eur. J. Cardiothorac Surg. 42 (3), 571–576.

Wenxuan, W., Chen, C., Meng, D., Hong, Y., Sen, Z., Jiangyun, L., 2021. Transbts: Multimodal brain tumor segmentation using transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 109–119.

Xiang, D., Qi, J., Wen, Y., Zhao, H., Zhang, X., Qin, J., Ma, X., Ren, Y., Hu, H., Liu, W., et al., 2023. ADSeg: A flap-attention-based deep learning approach for aortic dissection segmentation. Patterns 4 (5).

Xiaowei, X., 2024. Dataset. https://github.com/XiaoweiXu/Comprehensive-Segmentation-of-Type-A-Aortic-Dissection-with-Clinically-Oriented-Evaluation.

Xu, X., He, Z., Niu, K., Zhang, Y., Tang, H., Tan, L., 2019. An automatic detection scheme of acute stanford type a aortic dissection based on DCNNs in CTA images. In: Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing. pp. 16–20.

Yao, Z., Xie, W., Zhang, J., Dong, Y., Qiu, H., Yuan, H., Jia, Q., Wang, T., Shi, Y., Zhuang, J., et al., 2021. Imagetbad: A 3d computed tomography angiography image dataset for automatic segmentation of type-b aortic dissection. Front. Physiol. 12, 732711.

You, C., Zhao, R., Liu, F., Dong, S., Chinchali, S., Topcu, U., Staib, L., Duncan, J., 2022. Class-aware adversarial transformers for medical image segmentation. Adv. Neural Inf. Process. Syst. 35, 29582–29596.

Yu, Y., Gao, Y., Wei, J., Liao, F., Xiao, Q., Zhang, J., Yin, W., Lu, B., 2021. A three-dimensional deep convolutional neural network for automatic segmentation and diameter measurement of type B aortic dissection. Korean J. Radiol. 22 (2), 168.

Yuan, X., Mitsis, A., Nienaber, C.A., 2022. Current understanding of aortic dissection. Life 12 (10), 1606.

Zhang, X., Cheng, G., Han, X., Li, S., Xiong, J., Wu, Z., Zhang, H., Chen, D., 2023b. Deep learning-based multi-stage postoperative type-b aortic dissection segmentation using global-local fusion learning. Phys. Med. Biol. 68 (23), 235011.

Zhang, Y., Liu, H., Hu, Q., 2021. Transfuse: Fusing transformers and cnns for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer, pp. 14–24.

Zhang, J., Liu, J., Wei, S., Chen, D., Xiong, J., Gao, F., 2023a. Semi-supervised aortic dissections segmentation: A time-dependent weighted feedback fusion framework. Comput. Med. Imaging Graph. 106, 102219.

Zhao, J., Feng, Q., 2021. Automatic aortic dissection centerline extraction via morphology-guided CRN tracker. IEEE J. Biomed. Heal. Inform. 25 (9), 3473–3485.

Zhao, J., Zhao, J., Pang, S., Feng, Q., 2022. Segmentation of the true lumen of aorta dissection via morphology-constrained stepwise deep mesh regression. IEEE Trans. Med. Imaging 41 (7), 1826–1836.

Zhou, H.-Y., Guo, J., Zhang, Y., Yu, L., Wang, L., Yu, Y., 2021. Nnformer: Interleaved transformer for volumetric segmentation. arXiv preprint arXiv:2109.03201.

Zhou, Q., Qin, J., Xiang, X., Tan, Y., Ren, Y., 2022. MOLS-Net: Multi-organ and lesion segmentation network based on sequence feature pyramid and attention mechanism for aortic dissection diagnosis. Knowl.-Based Syst. 239, 107853.

Zhu, Y., Lingala, B., Baiocchi, M., Tao, J.J., Toro Arana, V., Khoo, J.W., Williams, K.M., Traboulsi, A.A.-R., Hammond, H.C., Lee, A.M., et al., 2020. Type a aortic dissection—experience over 5 decades: JACC historical breakthroughs in perspective. J. Am. Coll. Cardiol. 76 (14), 1703–1713.