Medical Image Analysis 101 (2025) 103432

Contents lists available at ScienceDirect

Medical Image Analysis

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

AutoFOX: An automated cross-modal 3D fusion framework of coronary X-ray angiography and OCT

Chunming Li [a], Yuchuan Qiao [b], Wei Yu [a], Yingguang Li [c], Yankai Chen [a], Zehao Fan [a], Runguo Wei [a], Botao Yang [a], Zhiqing Wang [d], Xuesong Lu [e], Lianglong Chen [d], Carlos Collet [f], Miao Chu [a,g,*], Shengxian Tu [a,g,*]

[a] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China
[b] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 201200, China
[c] International Smart Medical Devices Innovation Center, Kunshan Industrial Technology Research Institute, Suzhou, China
[d] Department of Cardiology, Fujian Medical University Union Hospital, Fuzhou, China
[e] School of Biomedical Engineering, South-Central Minzu University, Wuhan 430074, Hubei, China
[f] Cardiovascular Center Aalst, OLV Clinic, Aalst, Belgium
[g] Department of Cardiovascular Medicine, University of Oxford, OX39DU, UK

ARTICLE INFO

Keywords:
Coronary artery disease
X-ray angiography
OCT
Deep learning-based alignment
3D fusion

ABSTRACT

Coronary artery disease (CAD) is the leading cause of death globally. The 3D fusion of coronary X-ray angiography (XA) and optical coherence tomography (OCT) provides complementary information to appreciate coronary anatomy and plaque morphology. This significantly improve CAD diagnosis and prognosis by enabling precise hemodynamic and computational physiology assessments. The challenges of fusion lie in the potential misalignment caused by the foreshortening effect in XA and non-uniform acquisition of OCT pullback. Moreover, the need for reconstructions of major bifurcations is technically demanding. This paper proposed an automated 3D fusion framework AutoFOX, which consists of deep learning model TransCAN for 3D vessel alignment. The 3D vessel contours are processed as sequential data, whose features are extracted and integrated with bifurcation information to enhance alignment via a multi-task fashion. TransCAN shows the highest alignment accuracy among all methods with a mean alignment error of 0.99 ± 0.81 mm along the vascular sequence, and only 0.82 ± 0.69 mm at key anatomical positions. The proposed AutoFOX framework uniquely employs an advanced side branch lumen reconstruction algorithm to enhance the assessment of bifurcation lesions. A multi-center dataset is utilized for independent external validation, using the paired 3D coronary computer tomography angiography (CTA) as the reference standard. Novel morphological metrics are proposed to evaluate the fusion accuracy. Our experiments show that the fusion model generated by AutoFOX exhibits high morphological consistency with CTA. AutoFOX framework enables automatic and comprehensive assessment of CAD, especially for the accurate assessment of bifurcation stenosis, which is of clinical value to guiding procedure and optimization.

1. Introduction

Coronary artery diseases (CAD) remains the leading cause of death worldwide (Martin et al., 2024). Percutaneous coronary intervention (PCI) is the primary diagnostic and treatment option for CAD, where coronary X-ray angiography (XA) and optical coherence tomography (OCT) are utilized to provide complementary information to guide PCI (Räber et al., 2018). As illustrated in Fig. 1, XA enables an intuitive assessment of overall coronary anatomy, while OCT provides a detailed assessment of internal morphology with the super-resolution imaging of lumen and plaque compositions (e.g., calcification, lipid, and fibrous) within vessel wall (Bezerra et al., 2009). The fusion of XA with OCT provides complementary information that enhances the understanding of coronary anatomy and plaque morphology, playing a significant role in CAD diagnosis and prognosis. An ideal 3D fusion model should accurately provide the morphological details of the main vessel (MV) as well as the anatomical structure of side branch (SB) ostia, as this enables accurate hemodynamic (e.g., endothelial shear stress (ESS) Li et al., 2018; Kweon et al., 2018) and computational

* Corresponding authors at: School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China.
E-mail addresses: chumiao@sjtu.edu.cn (M. Chu), sxtu@sjtu.edu.cn (S. Tu).

https://doi.org/10.1016/j.media.2024.103432
Received 12 June 2024; Received in revised form 22 November 2024; Accepted 6 December 2024
Available online 15 December 2024

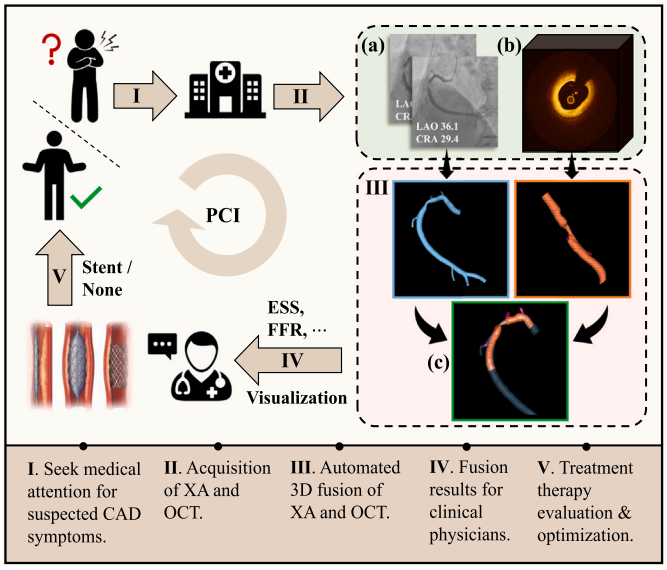

Fig. 1. Coronary artery disease diagnosis and treatment guided by the 3D fusion of XA and OCT. (a) XA images from two views and (b) OCT pullback, and (c) 3D fusion model of coronary tree.

physiology assessments (e.g., fractional flow reserve (FFR) Wang et al., 2018; Jiang et al., 2021).

The major challenges in developing an automated coronary cross-modal fusion model include: (1) Overcoming misalignment caused by the XA foreshortening effect and non-uniform OCT acquisition; (2) Effectively reconstructing major bifurcations along the main vessels; and (3) Eliminating the need for manual intervene in the intermediate stage. The current cross-modal alignment approaches can be briefly divided into "wire-based" and "wireless" categories. The "wire-based" methods track interventional devices like imaging probe on guidewires to mark alignment points, often using techniques like ECG-triggered fluoroscopy alongside intravascular imaging wire pullback to track the transducer under X-ray (Wang et al., 2013; Prasad et al., 2016; Wu et al., 2023) or a custom system designed to record landmarks (Houissa et al., 2019). The "wire-based" methods only work for data acquired under these specific tracing settings and may increase radiation dose and need extra manual manipulation or equipment in the workflow of PCI. In contrast, "wireless" methods are purely based on original images, with no need to change the standard workflow in the catheterization lab. The methods can be widely applied to previous data retrospectively. However, existing wireless methods often rely on simple equidistant mapping (Wahle et al., 2006; Tu et al., 2011; Wang et al., 2018; Toutouzas et al., 2015; Andrikos et al., 2017) without utilizing the inherent characteristics of the vessel contour, which can lead to misalignment issues (Kweon et al., 2018; Wu et al., 2020; Poon et al., 2023). Our previous work (Qin et al., 2021) implements a two-stage alignment for 2D XA and OCT. However, the study is limited by using traditional nonrigid point-matching method based on sole lumen diameter. The current study advances the previous work into 3D alignment with key advancements in alignment algorithm and achieve a fully automation workflow. Moreover, current fusion methods are restricted to specific bifurcations, such as the left main or RCA (Wu et al., 2020, 2023; Andrikos et al., 2017), and require additional imaging for each side branch, which limits their clinical applications. In addition, all reported frameworks rely on semi-automated reconstruction of SBs (Li et al., 2015; Kweon et al., 2018; Li et al., 2018) and the manual identification of alignment markers (Poon et al., 2023). Challenges caused by SB detection mistakes generated during automated analysis has not been addressed yet. Finally, there is a of direct quantitative metrics to evaluate the fusion accuracy, and thus only indirect clinical parameters such as ESS and FFR are used as surrogate measures of evaluation.

To overcome the above limitations, we propose a novel wireless, deep learning-based framework, named **Auto**mated **F**usion framework of **O**CT and **X**A (**AutoFOX**), which includes three fully automated procedures of Initial Reconstruction (IR), Co-Registration (CR), and Fusion Reconstruction (FR), as illustrated in Fig. 2. The IR procedure is responsible for automatically generating the essential coronary anatomic structures of XA and OCT. In the CR procedure, key issues of vessel alignment and lumen rotation registration are addressed. A novel 3D coronary vessel alignment network, named **Trans**former-based **C**oronary vessel **A**lignment **N**et (**TransCAN**), is proposed. TransCAN treats 3D vessels as sequential data, leverages dynamic time warping (DTW) (Müller, 2007) theory and Transformer (Vaswani et al., 2017) architecture to deeply integrate SB information in a multi-task manner, effectively overcoming misalignment issues. Then, the axial rotation errors between the two vessels are corrected. Finally, the FR procedure utilizes an innovative reconstruction algorithm for SB lumen and generate the final coronary tree fusion model. In addition, we defined various morphological metrics and use paired coronary computed tomography angiography (CTA) as the reference standard, these metrics will quantitatively assess the accuracy of the fusion model in a more direct way.

Overall, the contributions of our work are as follows:

- We introduce AutoFOX, a novel automated framework for coronary cross-modal 3D fusion of OCT and XA. The framework includes Initial Reconstruction (IR), Co-Registration (CR), and Fusion Reconstruction (FR).
- A multi-task deep learning model for 3D coronary vessel alignment, named TransCAN, is proposed. TransCAN leverages DTW theory and Transformer architecture to address misalignment challenges, while integrating side branch information through several novel modules.
- An innovative side branch lumen reconstruction algorithm is utilized to overcome challenges in fusion at bifurcations, which enhances the assessment of bifurcation lesions.
- We utilized paired CTA data as the reference standard and have defined various morphological metrics, providing a more accurate and direct method for assessing the quantitative precision of our 3D fusion model.

2. Related works

The proposed AutoFOX consist of a specially designed deep learning network for 3D vessel sequence alignment. It effectively overcomes misalignment issues by adapting DTW theory and leverages Transformer's strength in handling long-sequence interactions. Thus, we herein provide detailed reviews of the related works in DTW and Transformer models.

2.1. DTW-based sequence alignment

The 3D vessel alignment task can be regarded as a high-dimensional sequence alignment problem. The inherent constraints of the DTW algorithm perfectly meet the requirements of this task, including: (1) Monotonicity, maintains the time order of points, ensuring the path does not go backward in time; (2) Continuity, restricts the path transitions to adjacent points in time, preventing it from jumping in time; (3) Boundary conditions, ensures the warping path starts and ends at the beginning and end points of both sequences. DTW is based on dynamic programming and operates by constructing a cost map that compares each point of one sequence to each point of another sequence, the goal is to find a path through this map that minimizes the total cost, which represents the alignment between the two sequences. Therefore,

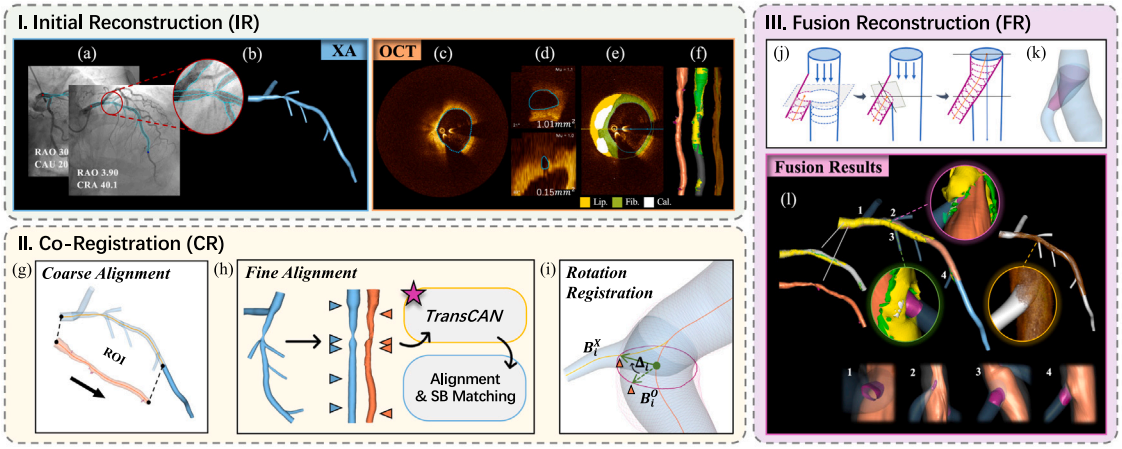

Fig. 2. Workflow diagram of AutoFOX. (I) Initial Reconstruction (IR) includes (a) segmentation of the coronary vessels from two angiographic views, resulting in (b) 3D-XA coronary tree. Meanwhile, IR provides (c) lumen segmentation (d) SB ostium detection, and (e) plaques identification, enabling the reconstruction of (f) 3D-OCT model with lipid, fibrous and calcification plaques reconstruction; (II) ; Co-Registration (CR) consists of (g) Coarse Alignment, (h) Fine Alignment via TransCAN, and (i) Rotational Registration; (III); Fusion Reconstruction (FR) includes (j) the reconstruction of the OCT SB ostium and (k) fusion into 3D-XA. (l) is the final fusion result with bifurcation structure, plaques and 3D-image.

DTW has the ability to handle sequences that are of different lengths and that may have non-linear distortions.

Algorithms like dDTW (Bork et al., 2013), weighted DTW (Jeong et al., 2011), shapeDTW (Zhao and Itti, 2018), etc., optimize the cost map of one-dimensional sequences. With the development of deep learning methods, learning-based cost maps will be applicable to alignment tasks for higher-dimensional, multimodal data, such as audio-to-video (Halperin et al., 2019), TCC (Dwibedi et al., 2019), LAV (Haresh et al., 2021), and VAVA (Liu et al., 2022) in video-to-video alignment, D^3TW (Chang et al., 2019) and DP-DTW (Chang et al., 2021) in video-action alignment, etc. Specifically, softDTW (Cuturi and Blondel, 2017) replaces DTW minimum with "soft minimum" to generate a "soft path" that can be used as a differentiable loss function and enable an end-to-end framework (Hadji et al., 2021; Xu et al., 2023). Increasing the softDTW smoothing parameter commonly decreases noise, yet it may sacrifice crucial data details, whilst lowering the parameter could yield more stringent alignment but poses a risk of overfitting (Cuturi and Blondel, 2017).

2.2. Transformer models with positional encoding

When 3D vessel contours are processed as high-dimensional sequences, the Transformer model enables more accurate and efficient feature integration and interaction through its attention mechanisms and positional encoding (PE). Compared to LSTM (Yu et al., 2019) series and other sequence processing models, Transformers tend to have faster training times and handle long-range dependencies more efficiently. The self-attention mechanism allows each element in the sequence to interact with every other element, a feature that enhances the model's ability to process sequences in parallel. Meanwhile, the cross-attention mechanism extends the Transformer's capabilities to tasks involving multiple sequences, especially enabling the model to align source and target sequences from different modalities. Therefore, the Transformer architecture can achieve state-of-the-art performance in a variety of domains (Dong et al., 2018; Gulati et al., 2020; Liu et al., 2021; Han et al., 2022).

However, the self-attention mechanism inherently discards the order of tokens in a sequence, as it treats inputs in a permutation-invariant manner. Some work (Dosovitskiy et al., 2020) incorporate absolute positional encoding to tokens, thereby infusing some order awareness into the model. Alternatively, relative positional encodings (Shaw et al., 2018; Dai et al., 2021) consider the relative order of tokens, maintaining translation-equivalence but leading to additional computational complexity. A new strategy, conditional positional encodings (Chu et al., 2021)dynamically generates positional encodings based on the local context of input tokens, offering a flexible approach to managing sequence order in Transformers.

3. Methods

To achieve automated cross-modal fusion, a series of necessary steps need to be implemented. First, coronary vessel models for each image modality must be reconstructed separately. Then, their correspondences are established by treating reconstructed vessels as sequential data. To establish optimal longitudinal correspondence between sequences, we adapt our previously proposed two-stage (coarse and fine) alignment strategy (Qin et al., 2021) to 3D alignment task, further enhancing it with deep learning models. Next, the relative axial rotational angle between vessels is determined. Finally, the reconstruction of the fusing model in 3D space is completed based on the alignment results, with optimized bifurcation structures. We proposed AutoFOX to implement the above cross-model fusion steps through three fully automated procedures, i.e., Initial Reconstruction (IR), Co-Registration (CR), and Fusion Reconstruction (FR). The AutoFOX workflow has already been implemented in a prototype software owned by Shanghai Jiao Tong University, with can be used to reproduce the data of this study.

3.1. Initial reconstruction

The IR procedure automatically generates essential coronary structures from both XA and OCT without any manual interaction. Two XA images at least 25 degrees apart are imported into AngioPlus Core software (version V3, Pulse Medical, Shanghai, China) to obtain main vessel and SB ostium segmentation. 3D-XA model is reconstructed based on the segmentation results (Çimen et al., 2016). AngioPlus is also used to generate the intraluminal structure of lumen, plaque and media layer on OCT, as illustrated in Fig. 2(I). The contours of 3D-XA and 3D-OCT are stored monotonically along the direction of blood flow slice by slice. Each slice has the same number of points, and the index values are continuous. The unit of vessel length is millimeters (mm), and the spacing between adjacent slices of the vessel is uniform in two images.

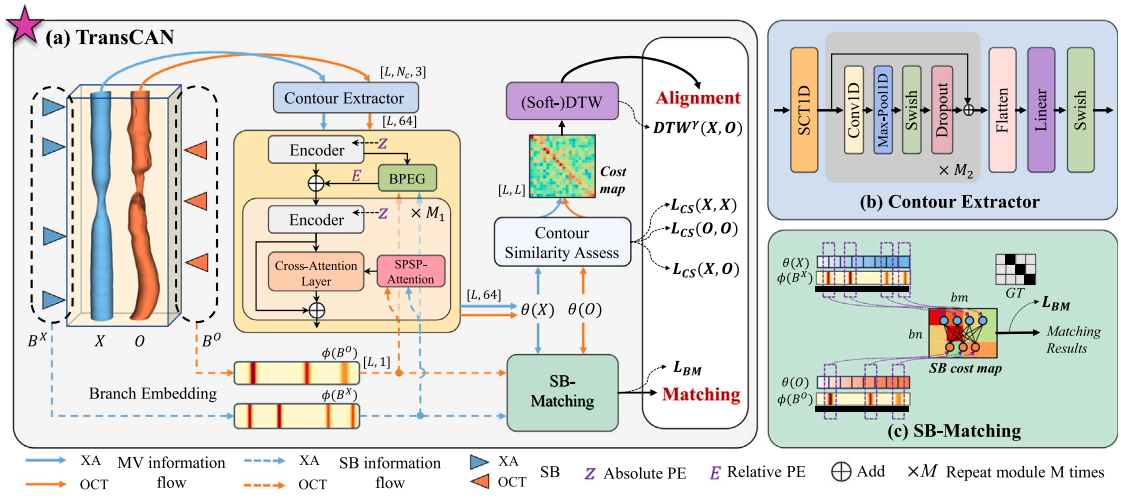

Fig. 3. The schematic of the design of (a) TransCAN and the details of (b) Contour Extractor module and (c) SB-Matching module in TransCAN.

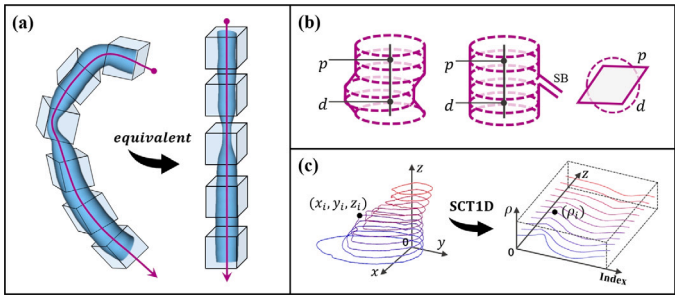

Fig. 4. Visual description of key challenges in fine-alignment. (a) The straightening of 3D-XA with cubes representing convolution block; (b) Example slices p and d have same area but differ in anatomical and topological characteristics. From left to right represent the two slices have spatial distortions, locating on opposite directions of a SB, and different contour shapes; (c) The SCT1D transforms 3D points into a 1D representation.

3.2. Co-registration: Coarse alignment

Given that OCT imaging typically captures only a segment of the entire artery, while XA usually provides a more complete view, the coarse alignment is primarily focused on accurately identifying the corresponding region of interest (ROI) on the XA image automatically. The MV lumen diameter and SB information of the two images are used to calculate the DTW distance within a sliding window. Due to the high time complexity of the naive DTW algorithm, we adopt the concept of Early Abandoning (Rakthanmanon et al., 2012) to enhance efficiency. The OCT sequence is slide along with the XA, and an XA segment with the same slice number as the OCT is generated at each step. The sliding score at each sliding window is calculated by combining the error accumulation of diameter and SBs information. Ultimately, the sliding window with the highest score is selected as the ROI.

3.3. Co-registration: TransCAN in fine alignment

The fine alignment aims to achieve slice-by-slice alignment of the two vascular sequences within the ROI. To this end, the fine alignment model TransCAN has been proposed, with the SB information deeply integrated. The optimization of position encoding and attention mechanisms, along with the incorporation of self-supervised learning and a multi-task framework, further ensures the alignment accuracy required for clinical applications. TransCAN's input data includes MV slices and SB instances from XA and OCT, represented as S^X, B^X, S^O and B^O, respectively. As shown in Fig. 3, the details of modules in TransCAN are as follows.

3.3.1. Vessel sequences preprocessing

Based on the characteristics of the vascular sequential data, we propose three preprocessing steps to standardize and simplify the computation.

First, the contours of the 3D-XA vessel within ROI are "straightened", as illustrated in Fig. 4(a). The spatial curvature information from 3D-XA is absent in OCT, and thus this information cannot contribute to the alignment between XA and OCT. Therefore, we sample features along the vessel centerline without considering its curvature, which reduces the data distribution differences between 3D-XA and 3D-OCT. Meanwhile, the consistency of feature format allows for a shared feature extraction module, which significantly lowers computational costs compared to using independent modules for each modality.

Second, considering the model's fixed-size input, we use linear interpolation to uniformly resample all vascular sequences to the same number of slices, L, as well as resampling the points in each vascular slice to the same number, N_c. Based on our statistical analysis of the average vessel length in the dataset, the value of L, N_c are chosen to be 224, 180 respectively. In this way, OCT slice interval spacing remains similar before and after resampling.

Finally, we observed that allowing the network to learn implicit features automatically from the stacked 3D contour point set, rather than relying on input explicit features like contour area, can more effectively overcome misalignment issues. Explicit features are directly measurable attributes that are often oversimplified and may not fully capture vessel complexity. In contrast, implicit features are complex, high-dimensional latent vectors extracted by the deep learning model, revealing hidden structural patterns beyond human visual interpretation. As shown in Fig. 4(a), for example, two contour slices, p and d, located at different positions within the vessel, should be distinguished based on their anatomical and topological differences, even if they have the same area value. Since the vessel has directional order in 3D space, using sequentially stacked 3D contours as input preserves the original anatomical and topological information. This enhances the network's understanding of vessel continuity and anatomical structure, and reduces the influence of local noise. The number of stacked slices in this study is set to 5, based on empirical experience and parameter tuning.

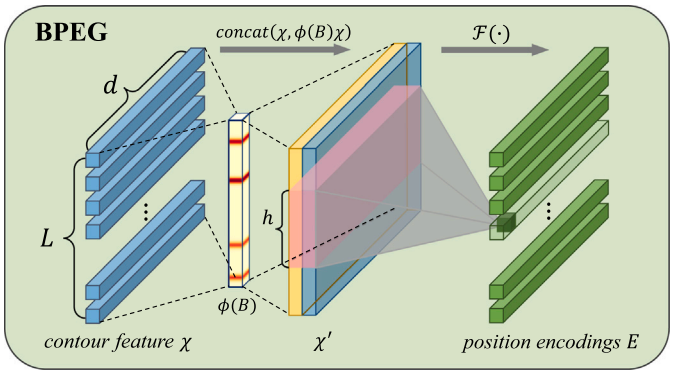

Fig. 5. Branch Position Encoder Generator (BPEG) module. L and d represent the length of sequence and contour feature χ, respectively. The $\mathcal{F}(\cdot)$ consists of a flattening operation followed by a 1D-Depthwise Separable Convolution with a kernel size of $h\times1$.

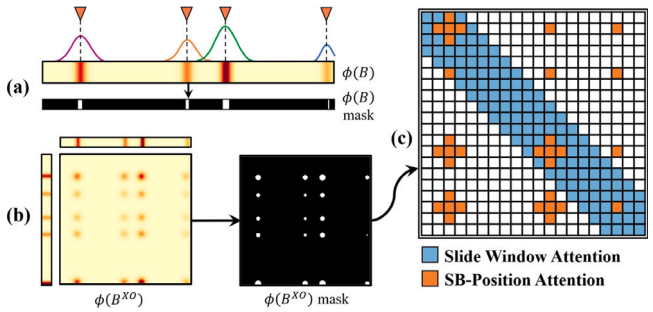

Fig. 6. Illustration of Branch Embedding and SPSP-Attention. (a) Branch Embedding $\phi(B)$ with its binarized mask; (b) SB-position attention using $\phi(B^{XO})$ with its binarized mask; (c) SPSP-attention mask combines Slide Window attention and SB-Position attention.

3.3.2. Contour extractor and branch embedding

The Contour Extractor module is proposed to extract high-dimensional feature representations from the stacked 3D vascular contour slices for downstream tasks, as shown in Fig. 3(a). The MV slices from XA and OCT $S^X = \{P_1^x, P_2^x, \dots, P_L^x\}$ and $S^O = \{P_1^o, P_2^o, \dots, P_L^o\}$ are first fed into the Contour Extractor module slice by slice, where $P_k = \left[(x_i, y_i, z_i)\right]_{i=1:N_c}^{\text{Slice } k}$ represents the 3D contour consisting of a total of N_c points for the kth slice in the vessel sequence. Inspired by spherical coordinate transform in 2D space (SCT2D) (Yang et al., 2020), we developed a 1D method, terms as SCT1D, which transforms P_k into 1D point set $\left[(\rho_i)\right]_{i=1:N_C}^{\text{Slice } k}$ under polar coordinates, where ρ is the distance from the pole (x_c, y_c), as shown in Fig. 4(c). For OCT, (x_c, y_c) is the center of the imaging catheter, while it is the centroid of each slice contour for XA. Benefiting from this, the Contour Extractor can be designed as a 1D-CNN module, which reduces computational cost and memory use.

Since vessel slices are perpendicular to the Z-axis, we use slice indices to replace z-values, which are incorporated as absolute position encoding. Similarly, with equal points N_c per slice, point indices can replace angle values in 2D polar coordinates. The index starting point for all slices aligns with the vessel's lateral line, allowing the model to learn invariance despite small offsets. After SCT1D and Contour Extractor, the MV slices data are transformed from $S \in \mathbb{R}^{L\times C\times N_c\times 3}$ to $S' \in \mathbb{R}^{L\times d}$, where d is the feature vector length and set to 64.

Since the SB information has not been fully utilized at this stage, we propose the Branch Embedding module to effectively integrate SB characteristics into the MV's feature sequence and to better represent the interaction range between different SBs. This module encodes the discrete SB information into a continuous representation using 1D Gaussian kernels, which is ultimately represented as an embedding vector $\phi(B) \in \mathbb{R}^{L\times1}$ of the same length as the MV sequence. We consider the longitudinal position and ostium area to be the most reliable information for SB identification, therefore the discrete SB information can be represented as $B^O = \left\{\left(BI_1^\circ, BS_1^\circ\right), \left(BI_2^\circ, BS_2^\circ\right), \dots, \left(BI_{bn}^o, BS_{bn}^o\right)\right\}$ for OCT and $B^X = \left\{\left(BI_1^x, BS_1^x\right), \left(BI_2^x, BS_2^x\right), \dots, \left(BI_{bm}^x, BS_{bm}^x\right)\right\}$ for XA, where $BI_k \in [0, L]$ is the the slice index along the MV corresponding to the position of the kth SB, and BS_k represents the area sizes of the kth SB ostium. Here, bn and bm denote the number of SBs detected in OCT and XA, respectively. Taking OCT as an example, the B^O generates $\phi(B^O)$ through the process defined by Eqs. (1) and (2) as shown in Fig. 6(a).

$$g(i)_k = BS_k^o \times \exp\left(-\frac{(i - BI_k^o)^2}{2\times\sigma^2}\right) \tag{1}$$

$$\phi(B^O)[i] = max\{g(i)_1, g(i)_2, \dots, g(i)_{bn}\} \tag{2}$$

where $g(i)_k$ represents the Gaussian distribution generated by the kth SB in the embedding vector, and σ represents the radius of the Gaussian kernel, which we set to 1, BI_k^o and BS_k^o represent the mean and amplitude of the distribution, respectively. This design positions the Gaussian distribution curve at the center of the SB ostium and reflects its area size. Since the final value of the ith element in $\phi(B)$ is influenced by the contributions of all SBs, we apply the max operation ensuring that the value depends on the SB contributing the most, preventing abnormal values that could arise from direct summation.

3.3.3. Position encoding and branch position encoding generator

Position encoding is the key to overcome low robustness in aligning problems with significant deformations. The normalized z-values from original sequence are used as the absolute position encoding $Z = \{\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_L\}$, and are applied at the very beginning of each Transformer Encoder layer to ensure the **permutation-variance**. Furthermore, the relative position encoding emphasizes approximate **translation-equivariance** in vascular alignment, adjusting for "translation" errors from coarse alignment or foreshortening effects.

$$E = \mathcal{F}[concat(\chi, \phi(B)\chi)] \tag{3}$$

Inspired by PEG (Chu et al., 2021), we propose a Branch Position Encoding Generator (BPEG). The output of the first Transformer Encoder is denoted as $\chi \in \mathbb{R}^{L\times d}$, and the purpose of BPEG is to combine χ and $\phi(B)$ to dynamically generate the relative positional encoding E. BPEG mainly consists of two steps, as shown in Eq. (3) and Fig. 5. First, the fusion of SB and MV information is achieved by performing element-wise multiplication between $\phi(B)$ and χ, followed by concatenation to form the new feature $\chi' \in \mathbb{R}^{L\times d\times2}$. Next, a function $\mathcal{F}(\cdot)$ is used to learn the contextual information from χ' and map it to $E \in \mathbb{R}^{L\times d}$. The function $\mathcal{F}(\cdot)$ consists of a flattening operation followed by a 1D-Depthwise Separable Convolution (Chollet, 2017) with a kernel size of $h\times1$. In our experiments, we found that setting h to 9 provides the best balance between local receptive field size and computational cost.

BPEG is placed after the first Transformer Encoder, and the benefit of this is that after the first Encoder, BPEG can leverage the global contextual information captured by the model to dynamically generate E (Chu et al., 2021). Consequently, E will combine with the main branch features through an Add operation, transmitting more effective features to subsequent Encoders and cross-attention calculations.

3.3.4. Attention mechanism

TransCAN utilizes the **self-attention** mechanism to learn the dynamic correlations and complex patterns within vascular sequences effectively and uses **cross-attention** mechanism to facilitate information interaction between XA and OCT. However, due to the unique characteristics of vascular sequences, dense computation in cross attention is unnecessary. On the contrary, focusing the computation on

more relevant areas can significantly reduce computational complexity and minimize the risk of overfitting (Huang et al., 2019). Therefore, we designed and incorporated a **sparse attention** mechanism named **S**lide window **P**lus **S**B-**P**osition (SPSP)-Attention.

One major characteristic of the vascular sequence is its strict monotonicity. It is unnecessary to compute the correlations between slices that are too far apart, especially after the coarse alignment. Therefore, using sliding window attention (Beltagy et al., 2020) can effectively improve computational efficiency. Additionally, we consider SBs as key regions for information interaction. Based on this, we designed SB-position attention of $\phi(B^{XO}) \in \mathbb{R}^{L\times L}$ via outer product of $\phi(B^X)$ and $\phi(B^O)$, which is binarized into attentional mask, as shown in Fig. 6(b). This allows SB regions to contribute to attention score calculation, with larger SB involving more feature slices. Finally, we combined the two components into the SPSP-Attention mask $SP \in \mathbb{R}^{L\times L}$ (Fig. 6(c)). This approach ensures the calculation of correlations between important features while reducing the computational complexity from $O(L^2)$ to approximately $O(Lw)$, where w is the window size and is set to 48.

$$\mathrm{ATTN}_{\mathrm{SP}}(Q_X, K_O, V_O) = \mathrm{softmax}\left(Q_X K_O^T \cdot \mathrm{SP}\right) V_O \tag{4}$$

For the feature sequences from XA, the calculation process of cross-attention based on SPSP-Attention is shown in Eq. (4). Here, Q_X is the query function from the linear transformation of XA features, and K_O and V_O are the key and value functions from the linear transformation of OCT features, respectively.

3.3.5. SB-matching sub-task module

To eliminate the axial rotational error between vessels, matched SBs are commonly used to calculate the relative angle. However, false and missed detections in IR may result in discrepancies in the number of detected SBs across different modalities, making straightforward one-to-one matching unreliable. Given that the alignment task and the SB matching task can complement each other, an auxiliary sub-task, i.e., SB-Matching, is designed to strengthen the feature correlations at the correctly matched SBs.

First, we obtain the outputs of XA and OCT from the main TransCAN network, denoted as $\theta(X) \in \mathbb{R}^{L\times d}$ and $\theta(O) \in \mathbb{R}^{L\times d}$, respectively. Next, as illustrated in Fig. 3(c), feature slices corresponding to the SBs are selected for SB-Matching from these outputs. The $\phi(B)$ binarized mask can be used to determine the position and number of these SB feature slices, as shown in Fig. 6(a). Each SB is ultimately represented by the average of selected multi-slice features. Subsequently, we formulated the SB matching as a binary classification task and constructed the SB cost map $A^B \in \mathbb{R}^{bm\times bn}$ based on cosine similarity (Eq. (5)). Cost map A^B measures the spatial similarity between SB features derived from the two modalities, which is normalized to a range of [0, 1]. The binary ground truth (GT) $G \in \mathbb{R}^{bm\times bn}$ of the SB matching relationships can be constructed from alignment annotation. Thus, we can calculate the difference between A^B and G using cross-entropy $H_B(\cdot)$, as shown in Eq. (6).

$$A^B_{ij} = \frac{1}{2}\left[\frac{\bar{\Theta}^X_i \cdot (\bar{\Theta}^O_j)^T}{\|\bar{\Theta}^X_i\|\|\bar{\Theta}^O_j\|} + 1\right] \tag{5}$$

$$\begin{aligned} H_B\left(A^B, G\right) = \frac{1}{bm\times bn}\sum_{i=1}^{bm}\sum_{j=1}^{bn}\Big[&A^B_{ij}\cdot\log(\delta(G_{ij})) \\ &+ (1-A^B_{ij})\cdot\log(1-\delta(G_{ij}))\Big] \end{aligned} \tag{6}$$

where $\bar{\Theta}^X_i \in \mathbb{R}^{1\times d}$ and $\bar{\Theta}^O_j \in \mathbb{R}^{1\times d}$ represent the average feature at the ith SB position of $\theta(X)$ and jth SB position of $\theta(O)$, respectively. δ represents the sigmoid function. The scoring range of A^B is between 0 and 1.

3.3.6. Alignment with DTWs

DTW algorithm is used to generate the final slice-by-slice alignment results within the vessel ROIs. The output $\theta(X)$ and $\theta(O)$ from TransCAN can be dynamically used to generate the cost map $A^M \in \mathbb{R}^{L\times L}$ for DTW in the form of cosine similarity, as defined in Eq. (7). The values in A^M are normalized to [0, 1] to reflect the similarity between different slices of XA and OCT, the closer their features are, the lower the score is.

$$A^M = \frac{1}{2}\left[1 - \frac{\theta(X)\cdot\theta(O)^T}{\|\theta(X)\|\|\theta(O)\|}\right] \tag{7}$$

$$\min{}^{\gamma}\{a_1, a_2, \ldots, a_n\} = -\gamma\log\sum_{i=1}^{n}\exp\left(-\frac{a_i}{\gamma}\right) \tag{8}$$

$$\mu(i,j) = A^M_{ij} + \min{}^{\gamma}\{\mu(i,j-1), \mu(i-1,j), \mu(i-1,j-1)\} \tag{9}$$

$$\mathrm{DTW}^{\gamma}(X,O) = \mu(L,L) \tag{10}$$

The soft-DTW replaces the discrete *min* operator in DTW by the smoothed min^{γ} in Eq. (8), where γ is the smoothing parameter. Soft-DTW generates the optimal dynamic path from (0,0) to (L, L) within A^M according to the rules in Eq. (9) and ultimately calculating the total cost $\mathrm{DTW}^{\gamma}(X,O)$ in Eq. (10). This end-to-end mode is referred to as softTransCAN, while the one integrating naive DTW is named nTransCAN for distinction.

3.3.7. Loss functions

To ensure that the dynamic path generated by the cost map A^M is more accurate through supervised learning, we designed a contour similarity assessment (CSA) unit to compute the loss functions for TransCAN.

First, in order to ensure that the information in the cost map A^M accurately reflects the similarity between $\theta(X)$ and $\theta(O)$, we defined the cross-align loss function $L_{CS}(X,O)$ in a double margin form (Lin et al., 2015), as shown in Eq. (11). In this form, the correctly corresponding vessel slices are treated as positive sample pairs and will be brought closer in the feature space, while those that do not correspond are treated as negative sample pairs and will be pushed apart. The annotated fine alignment can be represented as the GT cross-align path shown in Fig. 7(b). However, the positive sample pairs constructed based on this path are too sparse. Therefore, we apply Gaussian kernels centered on the path to generate the GT cross-align matrix $M \in \mathbb{R}^{L\times L}$, as illustrated as Fig. 7(c).

Second, considering the prior information of the vessel itself, the correlation between slices should be inversely proportional to their distance. Therefore, the self-align matrix $M^D \in \mathbb{R}^{L\times L}$ shown in Fig. 7(a) conforms to a Gaussian distribution attenuation along the diagonal of the matrix. We adopted self-supervised learning (Haresh et al., 2021) to calculate the self-align losses $L_{CS}(X,X)$ and $L_{CS}(O,O)$ for $\theta(X)$ and $\theta(O)$, respectively. They are similar to the definition in Eq. (11), except that they use M^D to construct sample pairs, guiding the generation of a cost map computed from the self-outer product. Ultimately, we combine them into the contour similarity assessment loss function L_{CSA}, which is defined as shown in Eq. (12).

$$\begin{aligned} L_{cs}(X,O) = \frac{1}{L^2}\sum_{i=1}^{L}\sum_{j=1}^{L}\Big[&M'_{ij}\cdot\max(0, m_{pos} - A^M) \\ &+ (1-M'_{ij})\cdot\max(0, A^M - m_{neg})\Big] \end{aligned} \tag{11}$$

$$L_{CSA} = L_{CS}(X,X) + L_{CS}(O,O) + L_{CS}(X,O) \tag{12}$$

where the σ here is set to 0.5, m_{pos} and m_{neg} are set as the margin values for positive and negative samples, guiding how the model differentiates between correctly and incorrectly identified samples. To ensure the training stability, m_{pos} and m_{neg} are set to 0.98 and 0.02, respectively, which proves to be more effective compared to using 1 and 0 (Lin et al., 2015). The Gaussian kernel used to construct M and M^D has an amplitude of 1.0 and a standard deviation of 0.5.

Subsequently, the SB-Matching loss L_{BM} is constructed using $H_B(\cdot)$ and defined in Eq. (13). It is used to guide the SB-Matching sub-task

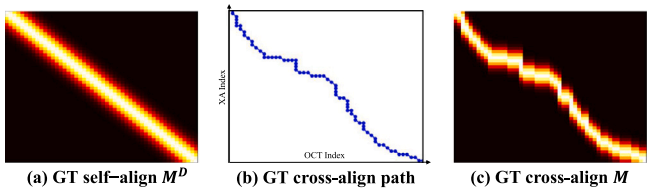

Fig. 7. Alignment matrices. (a) The GT self-align matrix M^D; (b) The GT cross-align path generated from annotation; (c) The GT cross-align matrix M.

and indirectly assist the main task. Meanwhile, the soft-DTW loss is used in the softTransCAN mode to guide the generation of the optimal dynamic path. Finally, the total loss L_{Total} consists of L_{CSA}, L_{BM} and soft-DTW loss (Eq. (14)).

$$L_{BM} = -H_B(A^B, G) \tag{13}$$

$$L_{Total} = \lambda_1 L_{CSA} + \lambda_2 L_{BM} + \lambda_3 \mathrm{DTW}^{\gamma}(X, O) \tag{14}$$

where λ_1, λ_2 and λ_3 are chosen as 1, 0.2, and 0.01 respectively, to ensure that the different loss values are in a similar order of magnitude. For the nTransCAN mode, λ_3 is set to 0.

3.4. Co-Registration: Lumen rotation registration

After the coarse and fine alignment stages are completed, we only obtain the precise correspondence between the two types of vascular slices, but their relative rotation angles in 3D space remain unknown. Therefore, we designed a lumen rotation registration stage, mainly utilizing the SB ostium azimuth angle and the similarity between contours.

With the known alignment and the spatial normal vectors of the 3D-XA centerline, we can use an affine transformation matrix to transform the OCT slice into the 3D-XA space. Simultaneously, with the correct matching relationship of the SBs, we calculate the difference in azimuth angle relative to the contour center for the ith SB pair, B_i^X and B_i^O, as angle Δ_i, and the azimuth angle difference for the $(i+1)$-th SB pair as angle Δ_{i+1}. Thus, the slices between two pairs of SBs vary according to the linear rule of $(\Delta_i - \Delta_{i+1})/N$, where N is the slice number between two pairs. The best angle for fine tuning of each slice is selected from 10° self-rotation interval with maximum IoU.

3.5. Fusion reconstruction

Based on the results of CR, the OCT main vessel lumen is accurately mapped onto the 3D-XA coronary tree model. Subsequently, FR refines the reconstruction of side branch lumens, significantly improving fusion accuracy at bifurcations. This enhancement not only improves visualization but also enables more accurate quantitative assessment of bifurcation lesions. The original segmentation results of the OCT branch lumen are perpendicular to the Z-axis, making them unsuitable for direct alignment with XA branch. By estimating the centerline's trajectory and calculating the normal vectors, we create new slices perpendicular to the centerline. This method, however, introduces an unknown fan-shaped area, as shown in Fig. 8(a), due to contours extending into the MV lumen with an "unknown boundary" during reconstruction.

We adopted a lumen reconstruction based on contour interpolation. Using the last known branch contour slice Is as the start, and a slice Ie of the MV contour above the branch as the end, a Bezier curve is fit from the SB centerline to the Z-axis. Then, as shown in the process of Fig. 8(b), a complete interpolated lumen is formed with linear growth of contour point at a step length of $(\rho_s - \rho_e)/N$, where N is the number of interpolations, and ρ_s and ρ_e correspond to the radius of the start and end points, respectively. Equidistant mapping is used for the alignment

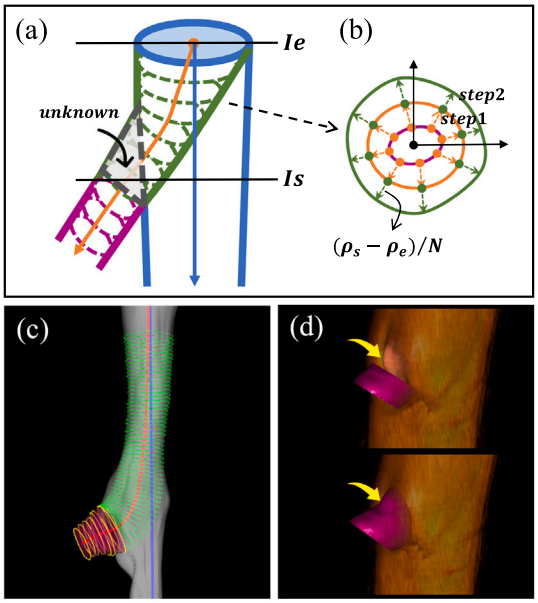

Fig. 8. Process of side branch lumen reconstruction. (a) Interpolation of OCT SB lumen contours; (b) Gradual growth of contours during the interpolation ; (c) Interpolation effect of the new OCT SB lumen; (d) Repair and completion at the SB ostium after reconstruction.

at these locations. Finally, as illustrated in Figs. 8(c) and 8(d), we can repair the unknown fan-shaped area and complete the fusion at the ostium of bifurcation. The refined reconstruction of the artery enables accurate hemodynamic assessments such as endothelial shear stress and fractional flow reserve.

4. Experiments and results

4.1. Datasets

The performance of AutoFOX is evaluated both internally for alignment accuracy of TransCAN and externally for morphological accuracy of the fusion model, using data from real world clinical practice with no overlap between them.

- To develop and evaluate TransCAN for vessel alignment, paired XA and OCT images of 278 patients from core lab (CardHemo, Med-X Research Institute, Shanghai Jiao Tong University) were used. Experienced analysts at the core lab performed the data annotation using the AngioPlus Core software (version V3, Pulse Medical, Shanghai, China), generating 55,104 alignment pairs. The dataset was split into training, validation, and test sets in a ratio of 7:1:2 at vessel level. We followed the standard practice by using training data to learn model parameters, tuning hyperparameters and selecting the model based on validation data. Test data is reserved solely for evaluating model once it is finalized. Model performance and ablation studies were reported on the test data.
- For the external validation of AutoFOX fusion model's morphological accuracy, we used an independent dataset of 67 patients with coronary CTA, XA, and OCT images. The data were provided by two sites: OLV Clinic, Aalst, Belgium (site1, 50 patients), and Fujian Medical University Union Hospital, Fuzhou, China (site2, 16 patients). The ethic committees of these two hospitals approved the retrospective analysis of these datasets. Patients provided written informed consent. The CTA model is used as the reference standard for morphology assessment, which is automatically generated by CtaPlus Core software (version V2, Pulse Medical, Shanghai, China).

The time intervals between different image modalities acquisition were within 3 months in majority (86.57%) of the study population, and the rest were less than 6 months. All images were acquired prior to any coronary intervention. Additionally, the analyzed CTA and XA images were synchronized to either end-diastole or end-systole to ensure temporal synchrony.

4.2. Experimental settings and evaluation metrics

4.2.1. Experimental settings

TransCAN is implemented by PyTorch 1.12 and executed on NVIDIA A100 Core GPU 80G, under Ubuntu 20.04 environment. Adam is used as the optimizer, with an initial learning rate of 0.001 for the main alignment task and 0.02 for the SB-Matching sub-task in synchronized training. The learning rate decayed at a rate of 0.95 every 10 epochs. The 3D reconstruction in AutoFOX utilized the ITK 5.3.0 and VTK 9.2.6 libraries based on C++.

4.2.2. Alignment metrics

To evaluate the alignment accuracy, we designed alignment metrics from three aspects: coarse alignment, fine alignment, and side branch matching. The metrics include average distance error value and precision value.

For each slice of 3D-OCT and its corresponding slice in 3D-XA, we calculated the absolute distance difference between the prediction of TransCAN and the GT. The coarse alignment error εCA, is the average distance of the ROI endpoints. For fine alignment, based on our previous work (Qin et al., 2021), some key anatomical positions of particular interest in clinical analysis have been identified. In addition to SBs, these include stenotic lesions, severe foreshortening positions, aneurysm, etc., and their average error εFA^K was calculated. Furthermore, the evaluation was extended to the overall sequence within the ROI, and the average error εFA was computed.

To evaluate alignment performance, we set different thresholds τ settings according to clinical requirement: 1.0 mm, 1.5 mm, and 2.0 mm. An alignment is considered successful when the error between corresponding pairs is less than the set threshold τ. Precision is calculated by determining the proportion of correctly aligned pairs (those with error $\leq \tau$) relative to the total number of evaluated pairs, for coarse alignment, noted as P_{CA}, and for refine alignment at the overall sequence level and at key clinical anatomical positions, noted as P_{FA} and P_{FA^K} respectively. Additionally, to evaluate the accuracy of SB-Matching module, we calculated the proportion of correctly matched OCT SBs out of the total number of OCT SBs, denoted as P_{BM}.

4.2.3. Alignment comparison methods

Due to the lack of diversity in reported 3D vessel alignment methods, we adapted several methods from similar tasks for comparison. The alignment comparison methods we selected can be mainly divided into three groups. The first group includes traditional algorithms based on vessel diameter curves, such as the widely used equidistant mapping (EM) and the point cloud registration algorithm TPS-RPM (Chui and Rangarajan, 2003), which estimates the non-rigid deformation between two diameter curves to achieve alignment. Additionally, the DTW series is included, comprising naive DTW (Bork et al., 2013), weighted derivative DTW (wdDTW) (Jeong et al., 2011), and shapeDTW (Zhao and Itti, 2018).

In the second group, we adapted several state-of-the-art methods from video-to-video tasks for comparison, including TCC (Dwibedi et al., 2019), LAV (Haresh et al., 2021), and VAVA (Liu et al., 2022). Since the video encoders in these methods are not directly applicable to our vessel data dimensions, we utilized the Contour Extractor module from this study, combined with two deep learning backbones suitable for handling sequence information—LSTM (Yu et al., 2019) and Transformer (Vaswani et al., 2017)—to replace the video embedding extraction process in the original methods. And the third group consists of the two model modes we proposed, nTransCAN and softTransCAN.

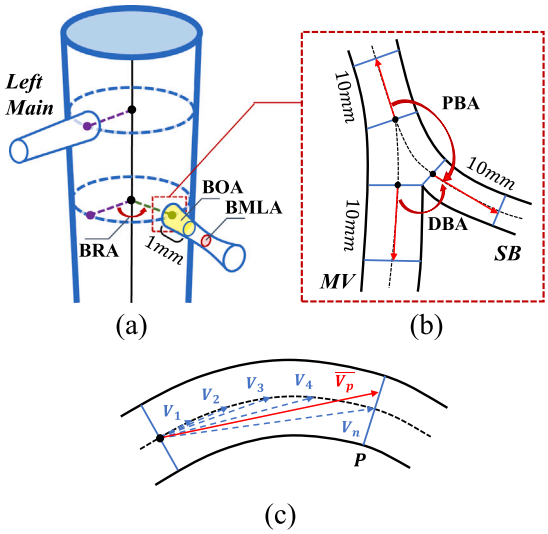

Fig. 9. Diagram of morphological metrics. (a) Definitions of BRA, BOA and BMLA; (b) Definitions of PBA and DBA; (c) Definition of directional vector $\bar{V}_p$ pointing towards P.

4.2.4. Morphological metrics

In order to quantitatively measure the morphological accuracy of the fusion model and the CTA reference model, we used SBs as the primary markers and designed morphological metrics, which are illustrated in Fig. 9 and defined as follows:

- Proximal bifurcation angle (PBA) and the distal bifurcation angle (DBA): PBA and DBA are used to represent the angles between the SB and the proximal or distal of the MV at the bifurcation, respectively (Wang et al., 2023). As shown in Fig. 9(b) and (c), to measure the degree of bifurcation angle, we selected all points on the centerline within a 10 mm range to participate in the calculation of a certain directional vector. These points form sub-vectors V_1, V_2, ⋯, V_n, and the average directional vector $\bar{V}_p$ pointing towards P is calculated using Eq. (15).

$$\bar{V}_p = \frac{1}{n}(|V_1| + |V_2| + |V_3| + \cdots + |V_n|) \tag{15}$$

  PBA is crucial to understand how blood is diverted from the MV into the SBs, which is a key to assess hemodynamic changes, while DBA indicates the blood flow distribution between branch vessels and the potential distribution of stress on the vessel walls.
- Branch rotation angle (BRA): BRA calculate the orientation between given SB relative and the first branch, like left main vessel. The directional vector for calculating points from the center of the slice where the SB is located to its ostium position. BRA can reflect the accuracy of rotation registration algorithm.
- Branch ostia area (BOA) and Branch minimal lumen area (BMLA): BOA measures the average area within a 1.0 mm segment at the SB ostium, while BMLA identifies the minimal area within a 3 mm segment. Both BOA and BMLA reflects SB lumen reconstruction accuracy and aids in evaluating branch stenosis severity.

The errors of the above metrics, i.e., εDBA, εPBA, εBRA, εBOA and $\varepsilon BMLA$, between the AutoFOX fusion model and the paired CTA reference model are calculated to assess morphological accuracy. Given the 3D tomographic property of CTA, it is used as the reference for assessing the side branch structure, especially for evaluating bifurcation angles. To account for potential concerns with CTA resolution in luminal measurements, both absolute values and correlation analysis have been conducted to evaluate the consistency of lumen structure in 3D space.

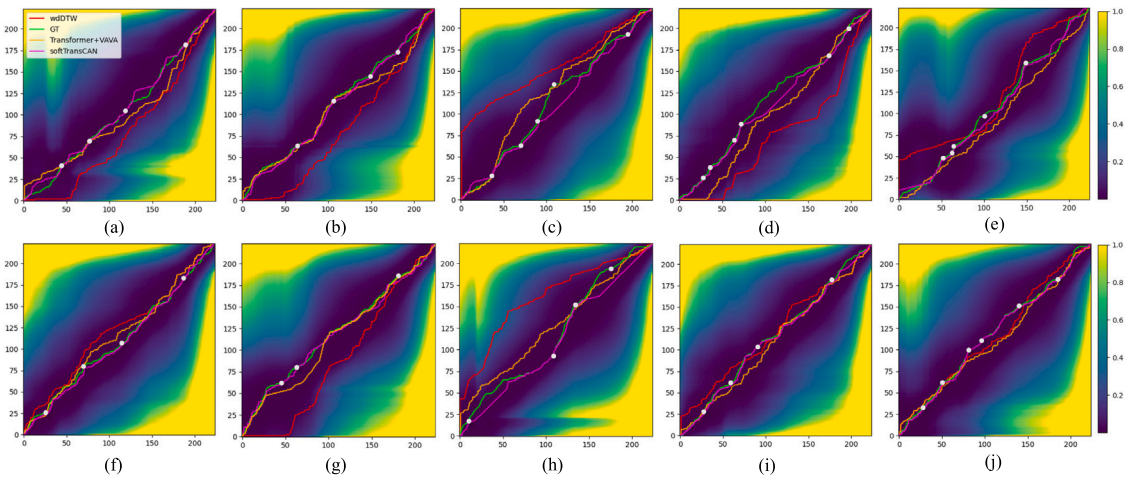

Fig. 10. Alignment results display of different methods. (a) to (j) are the alignment results of ten samples from the validation set. The background in the images shows the cost predicted by softTransCAN. The dynamic paths are from wdDTW, Transformer+VAVA, softTransCAN and GT. The paths are generated according to their respective cost maps. The white dots represent key anatomical positions.

Table 1
Alignment results of TransCAN and comparison methods.

Methods	εFA (mm)	εFA^K (mm)	P_{FA}%			P_{FA^K}%		
			τ = 1.0 mm	τ = 1.5 mm	τ = 2.0 mm	τ = 1.0 mm	τ = 1.5 mm	τ = 2.0 mm
EM	2.18 ± 1.64	1.98 ± 1.59	41.85 ± 5.37	53.30 ± 5.58	61.34 ± 6.37	40.76 ± 5.67	55.71 ± 5.99	62.46 ± 6.56
TPS-RPM	2.29 ± 1.43	1.79 ± 1.24	39.25 ± 5.09	51.52 ± 5.23	60.79 ± 6.10	42.23 ± 4.93	58.65 ± 5.36	79.47 ± 5.18
DTW	3.03 ± 1.45	2.43 ± 1.37	22.69 ± 4.08	32.11 ± 3.92	43.72 ± 4.01	31.67 ± 3.79	45.75 ± 3.81	59.24 ± 3.96
shapeDTW	2.08 ± 1.25	1.68 ± 1.19	40.63 ± 3.74	58.17 ± 3.88	63.89 ± 4.32	44.28 ± 3.19	59.82 ± 3.30	80.94 ± 4.12
wdDTW	1.73 ± 1.13	1.45 ± 1.06	45.01 ± 3.56	62.29 ± 3.94	78.96 ± 3.89	52.49 ± 2.99	73.90 ± 3.29	84.16 ± 2.74
LSTM + TCC	1.54 ± 1.03	1.29 ± 0.96	51.70 ± 2.14	71.56 ± 2.96	82.86 ± 2.55	58.04 ± 2.52	76.83 ± 2.68	86.22 ± 2.05
LSTM + LAV	1.46 ± 1.07	1.25 ± 1.01	51.67 ± 2.21	72.13 ± 2.64	83.23 ± 2.31	58.65 ± 2.37	77.71 ± 2.40	86.80 ± 1.93
LSTM + VAVA	1.42 ± 0.92	1.23 ± 0.84	52.20 ± 2.16	73.65 ± 2.59	83.58 ± 1.96	58.94 ± 2.39	78.30 ± 2.33	87.10 ± 2.00
TF[a] + TCC	1.34 ± 1.06	1.15 ± 0.89	54.75 ± 2.24	76.79 ± 2.48	84.16 ± 2.24	60.12 ± 2.45	80.64 ± 2.49	87.39 ± 2.11
TF[a] + LAV	1.30 ± 0.99	1.12 ± 0.95	55.43 ± 2.32	77.29 ± 2.03	84.34 ± 1.82	60.70 ± 2.35	81.23 ± 2.12	87.39 ± 1.94
TF[a] + VAVA	1.28 ± 0.90	1.10 ± 0.77	55.98 ± 2.10	77.37 ± 1.97	84.39 ± 1.71	61.00 ± 2.08	81.23 ± 1.83	87.39 ± 1.71
softTransCAN	1.09 ± 0.92	**0.82 ± 0.69**	63.05 ± 2.28	81.70 ± 1.94	87.09 ± 1.85	**71.26 ± 2.03**	**83.87 ± 1.75**	**87.98 ± 1.70**
nTransCAN	**0.99 ± 0.81**	0.85 ± 0.75	**66.99 ± 2.33**	**82.24 ± 1.86**	**87.16 ± 1.62**	70.67 ± 2.14	83.58 ± 1.92	**87.98 ± 1.84**

[a] TF = Transformer.

In this study, the CTA model was automatically generated by the CtaPlus Core software (version V2, Pulse Medical, Shanghai, China). In detail, the software enables 3D instance segmentation and reconstruction of both main vessels and side branches for CTCA, based on which ordered vessel centerlines are computed via 3D skeletonization. The data format of centerlines derived by CtaPlus software are in line with that derived by AutoFOX, making sure the morphology metrics of angle measurement are consistent. For lumen measurements on CTCA, multiplanar reformation views are first generated based on centerlines. CtaPlus uses an additional model to refine the contours, which are then transformed back into 3D space for measurement. Additionally, the 3D vessel contour of side branch consists of ordered slices perpendicular to the centerline, which is consistent with AutoFOX. Therefore, the measurements of BOA and BMLA are in the same way as AutoFOX. Moreover, the CTA model and the AutoFOX fusion model are aligned to the same phase of the cardiac cycle, which allows for accurate measurement comparison.

4.3. Results analysis

4.3.1. Alignment performance of TransCAN

Table 1 summarizes the alignment results on test set of nTransCAN and softTransCAN in comparison. Fig. 10 gives some examples for alignment results and Fig. 11 shows the impact of different alignment methods on the fusion model for focal lesion, diffused lesion and aneurysm.

In the test set, the εCA was 1.34 ± 1.12 mm with an precision of 80.49% at τ=1.5 mm, and we observed the errors did not significantly affect the fine alignment process. Both nTransCAN and softTransCAN demonstrated high fine alignment accuracy, with εFA of 0.99±0.81 mm and 1.09±0.88 mm for the overall sequence, and 0.85±0.74 mm and 0.82± 0.69 mm at 341 key positions pairs. At the threshold level τ=1.0 mm, nTransCAN and softTransCAN outperformed the best of other methods by 11.01% and 7.07% at P_{FA} and by 9.67% and 10.26% at P_{FA^K}, respectively. nTransCAN achieved the highest accuracy at the overall sequence level, while the end-to-end learning allowed softTransCAN to optimize key position alignment at a cost of an over smooth dynamic paths at non-key locations. In addition, a comparative analysis was conducted to evaluate the performance improvement by adding fine alignment to coarse alignment, where the EM method represents the scenario of using coarse alignment alone. Results indicate that fine alignment improves the alignment accuracy at key positions from 40.76% to 71.26%. Among a total of 288 SB pairs in the test set, the P_{BM} achieves an accuracy of 90.28%.

4.3.2. Learning curve analysis of TransCAN

Although softTransCAN performs slightly below nTransCAN at the overall sequence level, it has the potential for improved performance on larger datasets. As the learning curve illustrated in Fig. 12, we tracked the changes in average alignment error and precision at τ=1.0 mm for TransCAN and softTransCAN across different dataset scales: 40%, 50%,

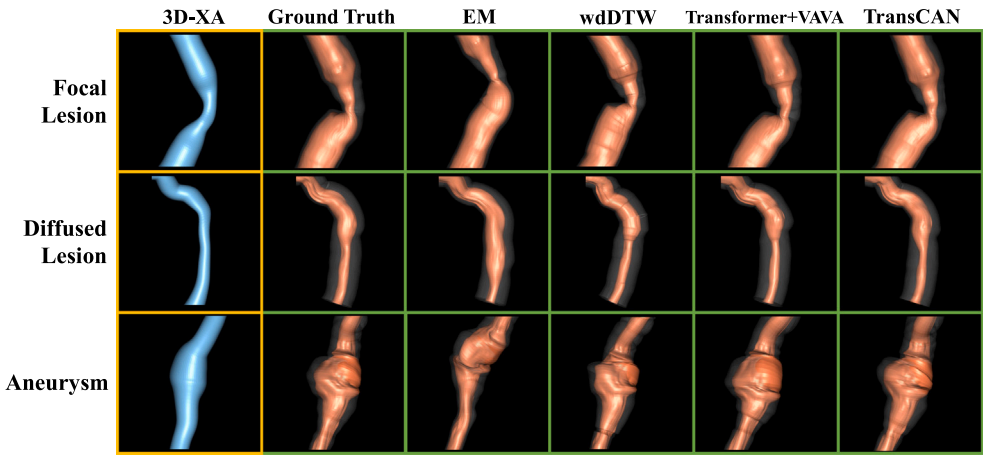

Fig. 11. The impact of different alignment methods on the fusion model for focal lesion, diffused lesion and aneurysm.

Table 2
Performance of ablation study on TransCAN.

Cross-attention	SPSP-attention	BPEG	Z	SB-Matching	εFA (mm)	εFA^K (mm)	P_{FA}%	P_{FA^K}%	P_{BM}%
			✓		1.22 ± 0.92	1.08 ± 0.95	58.12 ± 2.39	61.29 ± 2.29	–
		✓	✓	✓	1.19 ± 1.03	0.99 ± 0.95	59.20 ± 2.24	62.28 ± 2.36	82.87 ± 3.75
✓		✓	✓	✓	1.08 ± 0.88	0.89 ± 0.73	63.02 ± 2.10	69.21 ± 2.28	87.50 ± 4.34
✓	✓		✓		1.14 ± 0.90	1.01 ± 0.79	61.88 ± 2.22	65.69 ± 2.56	–
✓	✓			✓	1.17 ± 0.86	0.95 ± 0.89	59.82 ± 2.34	67.45 ± 2.27	79.17 ± 3.12
✓	✓		✓	✓	1.13 ± 0.91	0.94 ± 0.91	61.96 ± 2.41	67.45 ± 2.43	80.09 ± 2.95
✓	✓	✓	✓		1.06 ± 0.80	0.90 ± 0.83	63.74 ± 2.27	68.91 ± 2.07	–
✓	✓	✓	✓	✓	**0.99 ± 0.81**	**0.85 ± 0.75**	**66.99 ± 2.33**	**70.67 ± 2.14**	**90.28 ± 3.96**

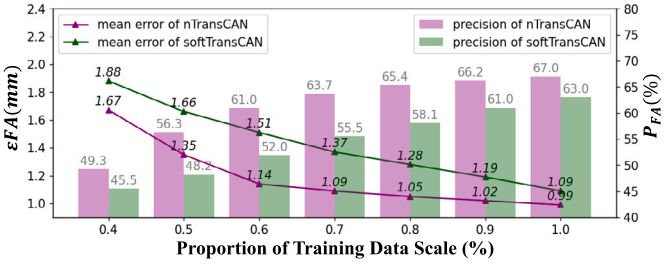

Fig. 12. The learning curve of nTransCAN and softTransCAN with changes in training data scale.

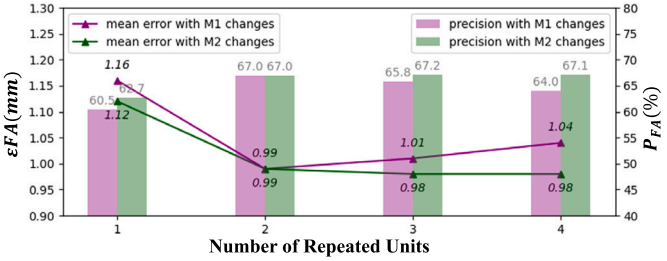

Fig. 13. The accuracy of nTransCAN with changes of the unit repetition number M_1 and M_2.

60%, 70%, 80%, 90%, and the entire training set. The performance of nTransCAN achieved a saturation after using more than 40% data sets, whereas softTransCAN is still in a clear ascending phase. Therefore, we believe that nTransCAN has nearly reached complete fitting at the current dataset scale and anticipate that the accuracy of softTransCAN may surpass that of nTransCAN at a certain turning point as the data scale increases.

Table 3
Performance of ablation study on CSA loss.

Methods	εFA (mm)	P_{FA}%
nTransCAN w/o self-align	1.05 ± 0.90	63.30 ± 2.16
nTransCAN	**0.99 ± 0.81**	**66.99 ± 2.33**
softTransCAN w/o self-align	1.14 ± 0.92	61.84 ± 2.25
softTransCAN w/o cross-align	1.23 ± 1.02	58.06 ± 2.39
softTransCAN w/o CSA	1.31 ± 0.99	55.29 ± 2.41
softTransCAN	**1.09 ± 0.92**	**63.05 ± 2.28**

Table 4
Performance of smooth parameter in SoftTransCAN.

γ	εFA (mm)	P_{FA}%
0.1	1.14 ± 0.90	61.88 ± 2.27
0.2	**1.09 ± 0.92**	**63.05 ± 2.28**
0.3	1.23 ± 1.05	58.04 ± 2.33
0.5	1.87 ± 1.14	44.14 ± 3.62
≥ 0.8	2.18 ± 2.25	41.85 ± 5.37

4.3.3. Ablation study on TransCAN

We conducted ablation study on key modules in TransCAN and the effect of contour similarity assess loss. Table 2 shows the results from ablation study of TransCAN, here we choose nTransCAN mode as reference. All threshold level is set to τ=1.0 mm. At the overall sequence level, the application of positional encoding, including BPEG and Z, reduced the error by 0.14 mm and 5.03% in total. Additionally, the incorporation of branch information using the SB-Matching and BPEG modules collectively reduced the error by 0.15 mm and 5.11%. The SPSP-attention mainly mitigates the model's overfitting issue and results in a precision improvement of 0.09 mm and 3.97%. At key positions level, the contributions of positional encoding and branch information are nearly identical, they improved the P_{FA^K} by 3.22% in precision, and SPSP-attention also brought an 1.46% improvement. TransCAN's side branch matching accuracy reached 90.28%. Notably, BPEG had a significant impact on the SB-Matching results,

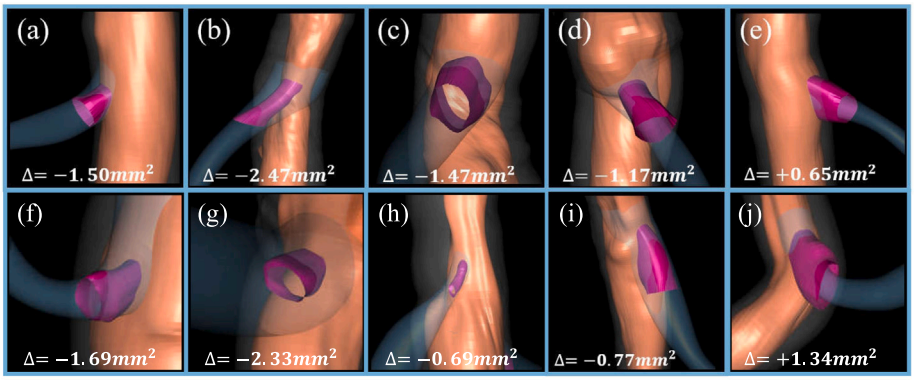

Fig. 14. Visualization of side branch ostia generated by AutoFOX (purple) overlapped with side branch of 3D-XA (transparent blue), and Δ is the BOA difference between AutoFOX and 3D-XA.

Table 5
Morphological results of AutoFOX.

Center	Patient Number	SB Number	εPBA (°)	εDBA (°)	εBRA (°)	εBOA (mm^2)	$\varepsilon BMLA$ (mm^2)
Site 1	50	158	2.83 ± 1.80	8.35 ± 4.19	5.78 ± 3.82	0.29 ± 0.01	0.15 ± 0.01
Site 2	17	58	2.26 ± 1.53	7.93 ± 3.97	6.34 ± 3.54	0.35 ± 0.01	0.14 ± 0.01
All	67	216	2.68 ± 1.78	8.23 ± 4.12	5.93 ± 3.75	0.31 ± 0.01	0.15 ± 0.01

Table 6
Time cost of AutoFOX.

Procedures	Time (s)	
	Duration	GPU speedup
IR	37.8	↑ 24.1
IR with plaques	54.1 (+16.3)	↑ 42.8 (+18.7)
CR	3.7	↑ 2.9
FR	2.9	–
FR with plaques	13.4 (+10.5)	–
Total	44.4	↑ 27.0
Total with plaques	71.2 (+26.8)	↑ 45.7 (+18.7)

contributing an improvement of 10.19%, while the cross-attention and SPSP-attention also brought an enhancement of 4.63% and 2.78%, respectively.

We also designed an ablation experiment to determine the optimal number of repeatable units M_1 and M_2 in TransCAN, as shown in Figs. 3(a) and 3(b). We evaluated the accuracy variations when they were set to 1, 2, 3, and 4, respectively. The results indicate that increasing M_1 and M_2 from 1 to 2 yields a significant improvement in model accuracy (Fig. 13). However, when they exceed 2, accuracy does not improve significantly, Therefore, both M_1 and M_2 are set to 2.

Furthermore, Table 3 demonstrates the effectiveness of the CSA loss. The self-supervised mechanism brought an accuracy improvement of 3.69% and 1.21% for the nTansCAN and softTransCAN, respectively. After removing the cross-align loss, softTransCAN used the unsupervised CSA loss can achieve an average error of 1.23 mm. The total CSA loss implementation significantly enhanced the precision of softTransCAN by 7.76%. Additionally, Table 4 presents the comparative experimental results for different γ in soft-DTW loss. A smaller γ value results in higher sensitivity to the alignment path but can lead to unstable gradient computations. As γ increases, the dynamic path becomes smoother, and the inclination towards the diagonal becomes more pronounced. When γ exceeds 0.8, it entirely degenerates into EM alignment. We observed the lowest mean error for softTransCAN when γ was set to 0.2.

4.3.4. Morphological results of AutoFOX

Table 5 primarily shows the statistical results on the model's morphological metrics in two external validation sets from multi-centers and we did not observe significant distribution differences across them. A total of 216 SBs were matched in both the AutoFOX fusion model and the CTA reference model. Fig. 14 illustrates the details in the reconstructed SB ostia obtained by AutoFOX and Fig. 15 displays some examples of the fusion results. The average errors for the angular metrics are small with εPBA, εDBA and εBRA only 2.68 ± 1.78°, 8.23 ± 4.12° and 5.93 ± 3.75°, εBOA and $\varepsilon BMLA$ 0.31 ± 0.01 mm^2 and 0.15 ± 0.01 mm^2, respectively. This also indicates that the fusion model has high feasibility for tasks such as coronary vessel hemodynamic assessment. In addition, we conducted a sensitivity analysis by including only patients with images acquired within 3 months interval and the results remained consistent. Specifically, εPBA, εDBA and εBRA were 2.61±1.71°, 8.27±4.09° and 5.88±3.70°, while εBOA and $\varepsilon BMLA$ were 0.30 ± 0.01 mm^2 and 0.15 ± 0.01 mm^2.

The experimental results demonstrate that the AutoFOX fusion model shows high morphological consistency with the CTA reference model. Compared to 3D-XA, the fusion model significantly enhances the correlation and agreement in BOA ($r = 0.71$ vs $r = 0.52$) and BMLA ($r = 0.84$ vs $r = 0.60$) with their reference standard value, as shown in Fig. 16.

4.3.5. The time cost of AutoFOX

Table 6 displays the time consumption of 67 samples in external validation sets using AutoFOX. Focusing solely on lumen fusion, IR is the most time-consuming procedure because it includes segmentation models for multiple images and various objects, averaging 37.8 s (ranging from 22.6 s to 49.1 s). CR and FR are faster, taking an average of 3.7 s (2.7 s to 6.1 s) and 2.9 s (2.2 s to 4.3 s) respectively. The total time duration with lumen only is 44.4 s, fusion involving plaques affects IR and FR procedures with 16.3 s and 10.5 s, respectively, and adds 26.8 s to the total time. GPU speedup applies to the segmentation model in IR and TransCAN in CR, respectively saving 27.0 s and 18.7 s for lumen and plaques processing times.

Although the current running speed of the framework does not meet the requirements for intraoperative real-time use, it ensures high efficiency in preoperative planning and postoperative evaluation without significantly adding to the overall diagnostic time cost. Our testing environment was equipped with an Intel® Core i7-11700K @ 3.60 GHz, NVIDIA RTX A4000, and 32 GB RAM.

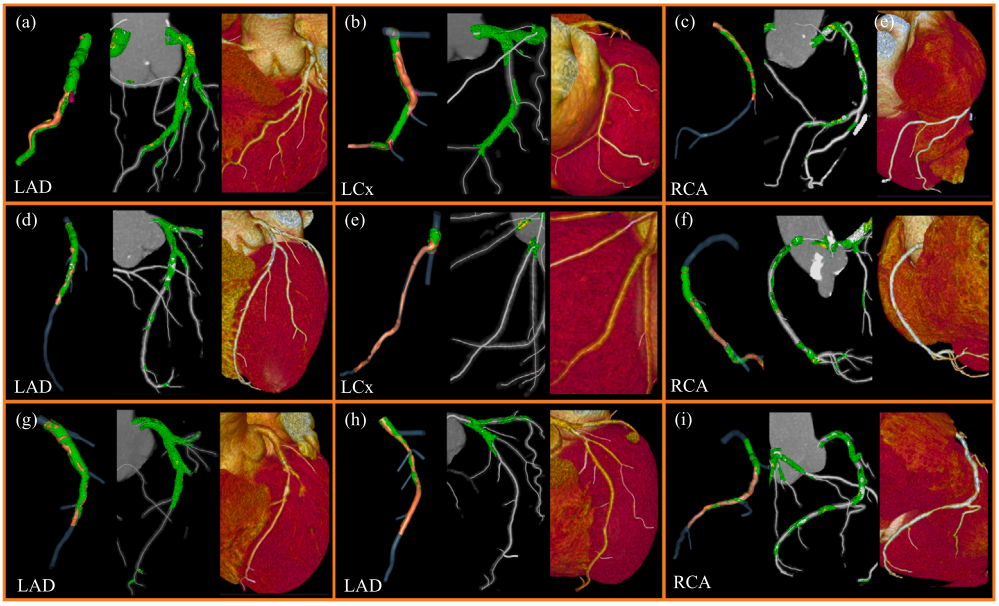

Fig. 15. The reconstructed fusion model by the AutoFOX. (a) to (i) are paired comparison between the fusion model (left), the reference CTA vessel tree model (middle) and the whole heart CTA image (right). There is high consistency in structure and intraluminal information distribution between the AutoFOX model and the CTA model in 3D space. Orange indicates OCT lumen and green indicates fibrous plaques.

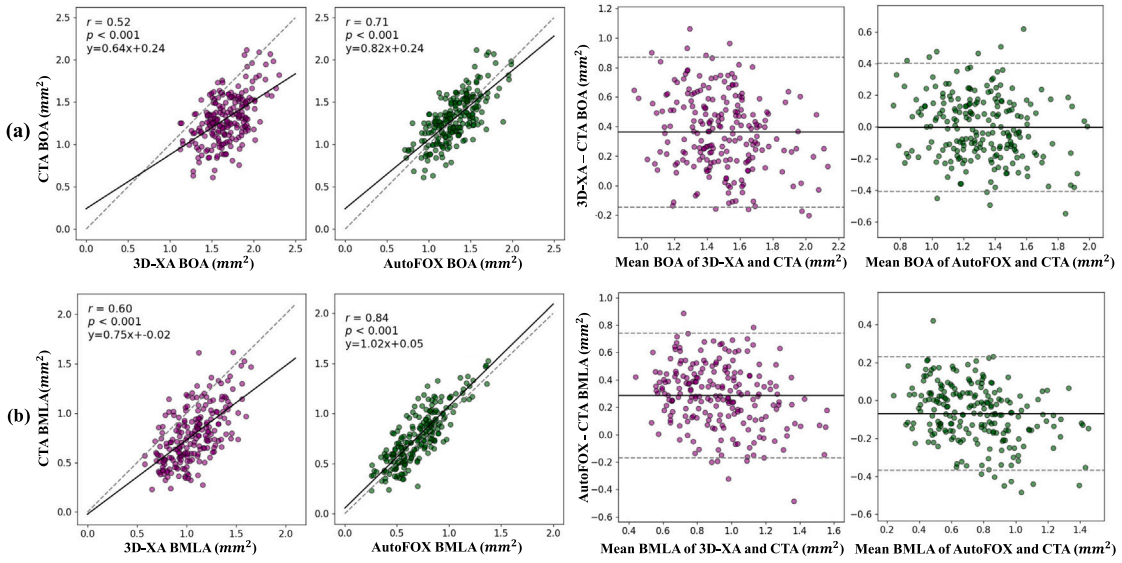

Fig. 16. Correlation and agreement between 3D-XA and CTA-model in (a) BOA and (b) BMLA.

5. Conclusion and discussion

This paper presents AutoFOX, a fully automated cross-modal 3D fusion framework for coronary X-ray Angiography and OCT through three procedures: Initial Reconstruction, Co-Registration, and Fusion Reconstruction. AutoFOX overcomes the limitations of existing vascular 3D alignment and fusion methods with a dedicated designed multi-task model TransCAN. AutoFOX treats vascular contours as sequential data, straightens 3D-XA to remove redundant curvature, and reduces model parameters through 3D to 1D transformation. Furthermore, the reconstruction algorithm is refined for SB lumen, which enhances the assessment of bifurcation lesions.

TransCAN shows the highest alignment accuracy compared with other methods. The novelty of the TransCAN lies in the deep integration of SB information: The SB-Matching sub-task enhances the matching of SB features; The BPEG module provides SB-weighted relative position encoding; and the SPSP-attention reduces the computational complexity of cross-attention while effectively maintaining information interaction at the SBs. Two TransCAN modes, nTransCAN and soft-TransCAN are proposed and compared. The former achieves optimal accuracy at the overall sequence level, while the latter performs better at clinically key positions and has greater scalability in larger datasets. The proposed CSA loss effectively enhances model robustness. Although some modules bring only minor numerical improvements, they are still of great clinical significance considering their proportion in the small lesion length and SB ostium length. Ultimately, we evaluate the fusion model with an independent multi-center dataset through 5 morphological metrics using the paired CTA as the reference standard. High morphological consistency is observed between the CTA and the fusion model generated by AutoFOX, especially for clinically significant BOA and BMLA. In addition to smaller difference in absolute luminal measurements, the improved correlation with CTA further demonstrates

that the integration of OCT enhances the fusion model's consistency with the actual lumen structure.

One major advantage of our work is the full automation of the entire framework without the need of manual intervene. This is contributed by the special designed module within AutoFOX, making it robust to the potential noise of the upstream output. Finally, the advanced 3D fusion model offers significant application value in guiding percutaneous coronary intervention for CAD patients, particularly in complex bifurcation lesions. By enhancing lesion visualization, this technology enables cardiologists to make more informed decisions during procedures, optimizing treatment strategies and potentially reducing complications. Given the accumulated evidence supporting imaging-based computational physiology assessments, the fusion of OCT and XA can enhance the evaluation accuracy of key parameters such as fractional flow reserve (FFR) and endothelial shear stress (ESS) by utilizing the precise geometry modeling provided by Auto-FOX. Furthermore, the fused 3D coronary artery tree also paves the way for investigating hemodynamic mechanisms in CAD development, potentially uncovering novel insights into disease progression and informing more targeted therapeutic interventions. As for the analysis speed by AutoFOX, although the current speed falls short of meeting intraoperative real-time requirements, it maintains high efficiency in preoperative planning and postoperative evaluation without significantly increasing the overall diagnostic time cost. Future study on quantitative comparison of plaque distribution is worthy of investigation based on further improvement of CTA model plaque detection performance.

Notably, our framework is not limited to any specific image content and modalities, we believe it could serve as a universal framework for fusion tasks across various 3D vascular-like structures, such as bronchial tube, gastrointestinal tract, renal artery, cerebral vessel, etc., thereby is spotential to expand into broader clinical application scenarios. However, it should be noted that several hyperparameters in this study are specifically tailored to the characteristics of coronary data, such as the number of sampling points and network layers, which may not be universally applicable when transferring to other tasks.

CRediT authorship contribution statement

Chunming Li: Writing – original draft, Methodology, Formal analysis. **Yuchuan Qiao:** Writing – review & editing. **Wei Yu:** Software, Methodology. **Yingguang Li:** Software, Methodology. **Yankai Chen:** Software, Methodology. **Zehao Fan:** Methodology. **Runguo Wei:** Methodology. **Botao Yang:** Methodology. **Zhiqing Wang:** Formal analysis. **Xuesong Lu:** Writing – review & editing. **Lianglong Chen:** Data curation. **Carlos Collet:** Data curation. **Miao Chu:** Writing – review & editing, Writing – original draft, Supervision. **Shengxian Tu:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: S Tu is the co-founder of Pulse Medical, reports research grants and consultancy from Pulse Medical. All other authors report no competing interests.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China (Grant No. 82020108015 and 82327808 to ST and 82302285 to MC); and in part by the Startup Fund for Young Faculty at Shanghai Jiao Tong University, China (23X010501994).

Data availability

The data that has been used is confidential.

References

Andrikos, I.O., Sakellarios, A.I., Siogkas, P.K., Rigas, G., Exarchos, T.P., Athanasiou, L.S., Karanasos, A., Toutouzas, K., Tousoulis, D., Michalis, L.K., et al., 2017. A novel hybrid approach for reconstruction of coronary bifurcations using angiography and OCT. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC, IEEE, pp. 588–591.

Beltagy, I., Peters, M.E., Cohan, A., 2020. Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.

Bezerra, H.G., Costa, M.A., Guagliumi, G., Rollins, A.M., Simon, D.I., 2009. Intracoronary optical coherence tomography: a comprehensive review: clinical and research applications. JACC Cardiovasc. Interv. 2 (11), 1035–1046.

Bork, C., Ng, K., Liu, Y., Yee, A., Pohlscheidt, M., 2013. Chromatographic peak alignment using derivative dynamic time warping. Biotechnol. Prog. 29 (2), 394–402.

Chang, C.Y., Huang, D.A., Sui, Y., Fei-Fei, L., Niebles, J.C., 2019. D3tw: Discriminative differentiable dynamic time warping for weakly supervised action alignment and segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3546–3555.

Chang, X., Tung, F., Mori, G., 2021. Learning discriminative prototypes with dynamic time warping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8395–8404.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1251–1258.

Chu, X., Tian, Z., Zhang, B., Wang, X., Wei, X., Xia, H., Shen, C., 2021. Conditional positional encodings for vision transformers. arXiv preprint arXiv:2102.10882.

Chui, H., Rangarajan, A., 2003. A new point matching algorithm for non-rigid registration. Comput. Vis. Image Underst. 89 (2–3), 114–141.

Çimen, S., Gooya, A., Grass, M., Frangi, A.F., 2016. Reconstruction of coronary arteries from X-ray angiography: A review. Med. Image Anal. 32, 46–68.

Cuturi, M., Blondel, M., 2017. Soft-dtw: a differentiable loss function for time-series. In: International Conference on Machine Learning. PMLR, pp. 894–903.

Dai, Z., Liu, H., Le, Q.V., Tan, M., 2021. Coatnet: Marrying convolution and attention for all data sizes. In: Advances in Neural Information Processing Systems, vol. 34, pp. 3965–3977.

Dong, L., Xu, S., Xu, B., 2018. Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 5884–5888.

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

Dwibedi, D., Aytar, Y., Tompson, J., Sermanet, P., Zisserman, A., 2019. Temporal cycle-consistency learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1801–1810.

Gulati, A., Qin, J., Chiu, C.C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., et al., 2020. Conformer: Convolution-augmented transformer for speech recognition. arXiv preprint arXiv:2005.08100.

Hadji, I., Derpanis, K.G., Jepson, A.D., 2021. Representation learning via global temporal alignment and cycle-consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11068–11077.

Halperin, T., Ephrat, A., Peleg, S., 2019. Dynamic temporal alignment of speech to lips. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 3980–3984.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. IEEE Trans. Pattern Anal. Mach. Intell. 45 (1), 87–110.

Haresh, S., Kumar, S., Coskun, H., Syed, S.N., Konin, A., Zia, Z., Tran, Q.H., 2021. Learning by aligning videos in time. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5548–5558.

Houissa, K., Ryan, N., Escaned, J., Cruden, N.L., Uren, N., Slots, T., Kayaert, P., Carlier, S.G., 2019. Validation of a novel system for co-registration of coronary angiographic and intravascular ultrasound imaging. Cardiovasc. Revascularization Med. 20 (9), 775–781.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W., 2019. Ccnet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 603–612.

Jeong, Y.S., Jeong, M.K., Omitaomu, O.A., 2011. Weighted dynamic time warping for time series classification. Pattern Recognit. 44 (9), 2231–2240.

Jiang, J., Feng, L., Li, C., Xia, Y., He, J., Leng, X., Dong, L., Hu, X., Wang, J., Xiang, J., 2021. Fractional flow reserve for coronary stenosis assessment derived from fusion of intravascular ultrasound and X-ray angiography. Quant. Imaging Med. Surg. 11 (11), 4543.

Kweon, J., Kang, S.J., Kim, Y.H., Lee, J.G., Han, S., Ha, H., Yang, D.H., Kang, J.W., Lim, T.H., Kwon, O., et al., 2018. Impact of coronary lumen reconstruction on the estimation of endothelial shear stress: In vivo comparison of three-dimensional quantitative coronary angiography and three-dimensional fusion combining optical coherent tomography. Eur. Heart J.-Cardiovasc. Imaging 19 (10), 1134–1141.

Li, Y., Gutiérrez-Chico, J.L., Holm, N.R., Yang, W., Hebsgaard, L., Christiansen, E.H., Mæng, M., Lassen, J.F., Yan, F., Reiber, J.H., et al., 2015. Impact of side branch modeling on computation of endothelial shear stress in coronary artery disease: coronary tree reconstruction by fusion of 3D angiography and OCT. J. Am. Coll. Cardiol. 66 (2), 125–135.

Li, Y., Li, Z., Holck, E.N., Xu, B., Karanasos, A., Fei, Z., Chang, Y., Chu, M., Dijkstra, J., Christiansen, E.H., et al., 2018. Local flow patterns after implantation of bioresorbable vascular scaffold in coronary bifurcations–novel findings by computational fluid dynamics. Circ. J. 82 (6), 1575–1583.

Lin, J., Morere, O., Chandrasekhar, V., Veillard, A., Goh, H., 2015. Deephash: Getting regularization, depth and fine-tuning right. arXiv preprint arXiv:1501.04711.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.

Liu, W., Tekin, B., Coskun, H., Vineet, V., Fua, P., Pollefeys, M., 2022. Learning to align sequential actions in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2181–2191.

Martin, S.S., Aday, A.W., Almarzooq, Z.I., Anderson, C.A., Arora, P., Avery, C.L., Baker-Smith, C.M., Barone Gibbs, B., Beaton, A.Z., Boehme, A.K., et al., 2024. 2024 heart disease and stroke statistics: A report of US and global data from the American heart association. Circulation.

Müller, M., 2007. Dynamic time warping. Inf. Retr. Music Motion 69–84.

Poon, E.K., Wu, X., Dijkstra, J., O'Leary, N., Torii, R., Reiber, J.H., Bourantas, C.V., Barlis, P., Onuma, Y., Serruys, P.W., 2023. Angiography and optical coherence tomography derived shear stress: are they equivalent in my opinion? Int. J. Cardiovasc. Imaging 39 (10), 1953–1961.

Prasad, M., Cassar, A., Fetterly, K.A., Bell, M., Theessen, H., Ecabert, O., Bresnahan, J.F., Lerman, A., 2016. Co-registration of angiography and intravascular ultrasound images through image-based device tracking. Catheter. Cardiovasc. Interv. 88 (7), 1077–1082.

Qin, H., Li, C., Li, Y., Huang, J., Yang, F., Kubo, T., Akasaka, T., Xiao, C., Gutiérrez-Chico, J.L., Tu, S., 2021. Automatic coregistration between coronary angiography and intravascular optical coherence tomography: feasibility and accuracy. JACC Asia 1 (2), 274–278.

Räber, L., Mintz, G.S., Koskinas, K.C., Johnson, T.W., Holm, N.R., Onuma, Y., Radu, M.D., Joner, M., Yu, B., Jia, H., et al., 2018. Clinical use of intracoronary imaging. Part 1: guidance and optimization of coronary interventions. An expert consensus document of the European Association of Percutaneous Cardiovascular Interventions. Eur. Heart J. 39 (35), 3281–3300.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E., 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 262–270.

Shaw, P., Uszkoreit, J., Vaswani, A., 2018. Self-attention with relative position representations. arXiv preprint arXiv:1803.02155.

Toutouzas, K., Chatzizisis, Y.S., Riga, M., Giannopoulos, A., Antoniadis, A.P., Tu, S., Fujino, Y., Mitsouras, D., Doulaverakis, C., Tsampoulatidis, I., et al., 2015. Accurate and reproducible reconstruction of coronary arteries and endothelial shear stress calculation using 3D OCT: comparative study to 3D IVUS and 3D QCA. Atherosclerosis 240 (2), 510–519.

Tu, S., Holm, N.R., Koning, G., Huang, Z., Reiber, J.H., 2011. Fusion of 3d qca and ivus/oct. Int. J. Cardiovasc. Imaging 27, 197–207.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), In: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc.

Wahle, A., Lopez, J.J., Olszewski, M.E., Vigmostad, S.C., Chandran, K.B., Rossen, J.D., Sonka, M., 2006. Plaque development, vessel curvature, and wall shear stress in coronary arteries assessed by X-ray angiography and intravascular ultrasound. Med. Image Anal. 10 (4), 615–631.

Wang, P., Ecabert, O., Chen, T., Wels, M., Rieber, J., Ostermeier, M., Comaniciu, D., 2013. Image-based co-registration of angiography and intravascular ultrasound images. IEEE Trans. Med. Imaging 32 (12), 2238–2249.

Wang, X., Peng, C., Liu, X., Pan, Z., 2018. Functional assessment of stenotic coronary artery in 3D geometric reconstruction from fusion of intravascular ultrasound and X-ray angiography. IEEE Access 6, 53330–53341.

Wang, Z., Yang, J., Li, C., Huang, J., Fezzi, S., Chen, E., Cai, W., Stankovic, G., Wijns, W., Chen, L., et al., 2023. Dynamic assessment of the left main-left circumflex bending angle: Implications for ostial left circumflex artery in-stent restenosis after successful two-stent PCI. Int. J. Cardiol. 378, 11–19.

Wu, W., Oguz, U.M., Banga, A., Zhao, S., Thota, A.K., Gadamidi, V.K., Vasa, C.H., Harmouch, K.M., Naser, A., Tieliwaerdi, X., et al., 2023. 3D reconstruction of coronary artery bifurcations from intravascular ultrasound and angiography. Sci. Rep. 13 (1), 13031.

Wu, W., Samant, S., de Zwart, G., Zhao, S., Khan, B., Ahmad, M., Bologna, M., Watanabe, Y., Murasato, Y., Burzotta, F., et al., 2020. 3D reconstruction of coronary artery bifurcations from coronary angiography and optical coherence tomography: feasibility, validation, and reproducibility. Sci. Rep. 10 (1), 18049.

Xu, M., Garg, S., Milford, M., Gould, S., 2023. Deep declarative dynamic time warping for end-to-end learning of alignment paths. arXiv preprint arXiv:2303.10778.

Yang, H., Zhen, X., Chi, Y., Zhang, L., Hua, X.S., 2020. Cpr-gcn: Conditional partial-residual graph convolutional network in automated anatomical labeling of coronary arteries. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3803–3811.

Yu, Y., Si, X., Hu, C., Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. 31 (7), 1235–1270.

Zhao, J., Itti, L., 2018. shapedtw: Shape dynamic time warping. Pattern Recognit. 74, 171–184.