IEEE Journal of Translational Engineering in Health and Medicine

Received 16 March 2023; revised 5 October 2023; accepted 23 October 2023.
Date of publication 1 November 2023; date of current version 24 November 2023.

Digital Object Identifier 10.1109/JTEHM.2023.3329031

TTN: Topological Transformer Network for Automated Coronary Artery Branch Labeling in Cardiac CT Angiography

YUYANG ZHANG[1], GONGNING LUO[1], (Member, IEEE), WEI WANG[1,2], (Member, IEEE), SHAODONG CAO[3], SUYU DONG[4], (Member, IEEE), DAREN YU[5], XIAOYUN WANG[5], AND KUANQUAN WANG[1], (Senior Member, IEEE)

[1]Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518000, China
[3]Department of Radiology, The Fourth Hospital of Harbin Medical University, Harbin 150001, China
[4]College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China
[5]Department of Cardiology, The Fourth Hospital of Harbin Medical University, Harbin 150001, China

CORRESPONDING AUTHOR: K. WANG (wangkq@hit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62272135, Grant 62372135, Grant 62001144, and Grant 62001141; and in part by the Science and Technology Innovation Committee of Shenzhen Municipality under Grant JCYJ20210324131800002 and Grant RCBS20210609103820029.

ABSTRACT Objective: Existing methods for automated coronary artery branch labeling in cardiac CT angiography face two limitations: 1) inability to model overall correlation of branches, since differences between branches cannot be captured directly. 2) a serious class imbalance between main and side branches. Methods and procedures: Inspired by the application of Transformer in sequence data, we propose a topological Transformer network (TTN), which solves the vessel branch labeling from a novel perspective of sequence labeling learning. TTN detects differences between branches by establishing their overall correlation. A topological encoding that represents the positions of vessel segments in the artery tree, is proposed to assist the model in classifying branches. Also, a segment-depth loss is introduced to solve the class imbalance between main and side branches. Results: On a dataset with 325 CCTA, our method obtains the best overall result on all branches, the best result on side branches, and a competitive result on main branches. Conclusion: TTN solves two limitations in existing methods perfectly, thus achieving the best result in coronary artery branch labeling task. It is the first Transformer based vessel branch labeling method and is notably different from previous methods. Clinical impact: This Pre-Clinical Research can be integrated into a computer-aided diagnosis system to generate cardiovascular disease diagnosis report, assisting clinicians in locating the atherosclerotic plaques.

INDEX TERMS Cardiac CT angiography, vessel branch labeling, transformer.

I. INTRODUCTION

Cardiovascular diseases are the leading cause of human mortality [1]. Benefiting from the non-invasion and high sensitivity, cardiac CT angiography (CCTA) has been widely used to diagnose cardiovascular diseases for suspicious patients [2]. To locate atherosclerotic plaques, clinicians need to annotate anatomical names of coronary artery branches manually. Automated coronary artery branch labeling (CABL) in CCTA can assist clinicians in locating atherosclerotic plaques, reducing a lot of manual labors. Nonetheless, among subjects the numbers, lengths and sizes of coronary artery branches, and the directions that branches span, are completely different as shown in Fig.1, which is the main challenge for automated CABL methods.

Two significant bottlenecks exists in current methods: 1) Inability to model **the overall correlation of branches**. As illustrated in Fig.2, utilizing the tree-structure or graph-structure, TreeLab-Net [3] and CPR-GCN [4] only construct the local correlation between adjacent branches. Nonadjacent branches are indirectly connected through intermediate branches, precluding direct capture of differences between nonadjacent branches. 2) A severe **class imbalance between**

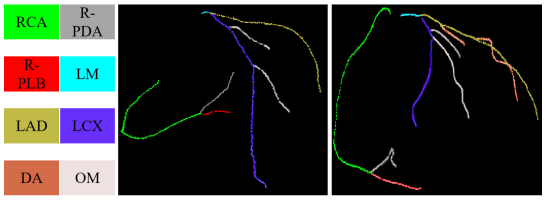

FIGURE 1. The directions, the numbers and the shapes of coronary artery branches are completely different between these two subjects. Branches shown are right coronary artery (RCA), right posterior descending artery (R-PDA), right posterior lateral branch (R-PLB), left main artery (LM), left anterior descending artery (LAD), left circumflex artery (LCX), diagonal artery (DA), obtuse marginal artery (OM).

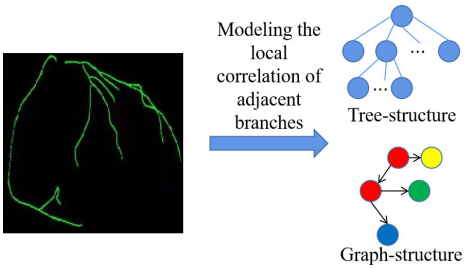

FIGURE 2. Existing methods only construct the local correlation between adjacent branches, nonadjacent branches are indirectly connected through intermediate branches. The differences between nonadjacent branches cannot be captured directly.

the main and side branches. The number of side branches, e.g., OM and DA, is much less than the number of main branches, e.g., LM, LAD and LCX. This issue causes the loss function to be overly sensitive to errors in the main branches.

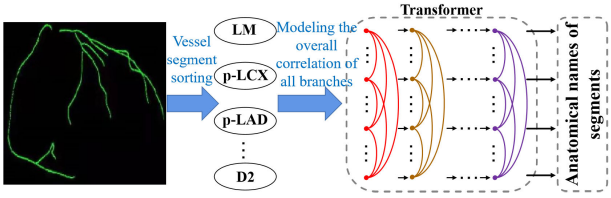

FIGURE 3. The coronary artery tree can be treated as a sequence data, e.g., LM, proximal-LCX (p-LCX), proximal-LAD (p-LAD), …, D2, in which each segment is a element. By constructing the overall correlation of all branches, Transformer can capture the differences between all branches directly, strengthening the ability to classify branches.

The coronary artery can be treated as a **sequence data** due to its anatomical topology as shown in Fig.3. According to [5], coronary artery branches of clinical relevance are left main (LM), left descending artery (LAD), left circumex artery (LCX), ramus-intermedius (RI), obtuse margin (OM), diagonal artery (DA), left posterior descending artery (L-PDA), right coronary artery (RCA), right posterior lateral branches (R-PLB) and right posterior descending artery (R-PDA). LAD, LCX, and RI originate from LM. R-PLB and R-PDA originate from RCA. DA's originate from LAD, OM's and L-PDA originate from LCX. Commonly, RCA, LM, LAD, and LCX are considered as the main branches. RI, OM's, DA's, L-PDA, R-PLB and R-PDA are considered as the side branches. Hence the Transformer [6] based model is a perfect fit for the CABL task as illustrated in Fig.3. Transformer can construct the overall correlation of all branches and thus captures the differences between branches directly.

To address the limitations in existing methods, we propose a topological Transformer network (TTN), which solves the CABL task from a novel perspective of sequence labeling learning. By constructing the overall correlation of branches, the TTN can discern the differences between them directly, thereby distinguishing the branches accurately. A segment-depth loss is calculated between the reference and predicted labels to solve the class imbalance between main and side branches. The contributions and innovations are as follows:

- The TTN considers the vessel branch labeling task as an end-to-end parallel sequence prediction problem. To our best knowledge, it is the first Transformer based vessel branch labeling method and notably differs from previous methods. On a CCTA dataset with 325 subjects, the TTN obtained the best overall result on all branches, and the best result on the side branches compared with other methods.
- The feature of each vessel segment in this study consists of the geometric feature extracted from the centerline tree and the image feature extracted from the CCTA image. The geometric feature used in existing methods lacks extra image information of vessel segments that can also be exploited to classify branches.
- The topological encoding can uniquely represent the position of a vessel segment in the artery tree. It assists the model in classifying branches by incorporating the positional differences of vessel segments into the model.
- The segment-depth loss effectively solves the class imbalance between the main and side branches by incorporating the anatomical topology of the artery tree into the training process.

II. RELATED WORK

Many anatomical labeling methods have been proposed for abdominal arteries [7], [8], [9], [10], coronary arteries [3], [4], [11], [12], [13], [14], airways [15] and brain arteries [16], [17], [18]. Existing methods can be divided into three categories.

Knowledge based methods first build a standard model according to the anatomical topology and prior knowledge obtained from a few samples. Then the test samples are aligned with the standard model to get the label of each branch. However, missing branches will affect the anatomical topology of samples and degrade model's performance. Cao et al. [13] built a model for the two sub-trees respectively. Then the anatomical name of each branch was obtained by matching the test samples to the models. Using a statistical model, Yang et al. [12] first matched the main branches, and the side branches were identified afterward. Gulsun et al. [14] calculated the geodesic paths between test samples and a reference anatomy model, matching test samples to the reference anatomy model.

IEEE Journal of Translational Engineering in Health and Medicine

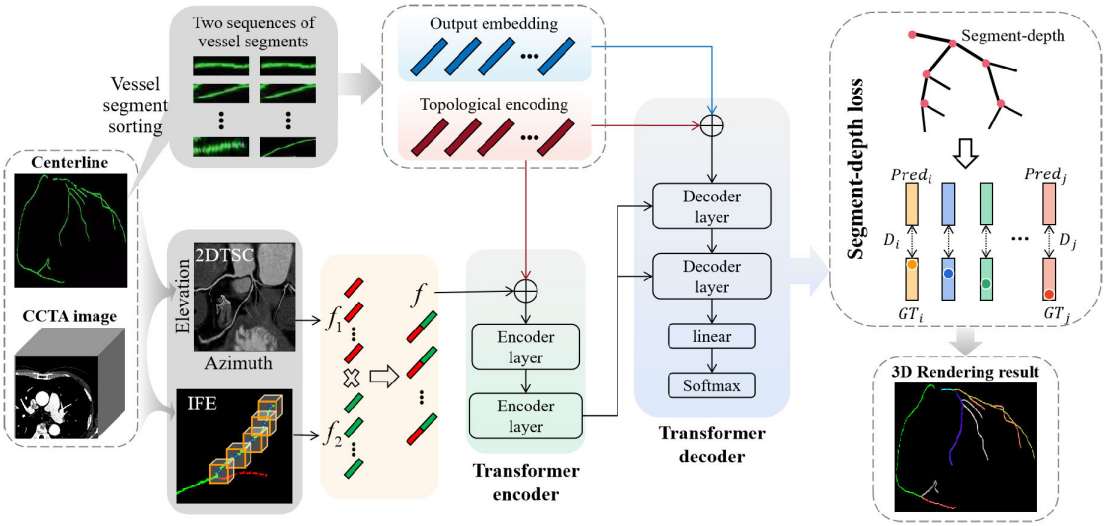

FIGURE 4. The overall framework of TTN. For each segment, 2DTSC (2D transform spherical coordinates) extracts the geometric feature f_1 from centerline tree, IFE (image feature extractor) extracts the image feature f_2 in CCTA image, then f_1 and f_2 are concatenated as the combined feature f. The topological encoding of each segment is added to f and output embedding respectively, as the inputs of Transformer encoder and decoder. Also, the encoded features of vessel segments (output of encoder) are passed into each layer of the decoder. Finally, model performs the classification on each segment. The segment-depth loss is computed between the reference and the predicted labels.

Alternatively, some works are based on **traditional machine learning**. To correct the topological errors of results, such methods always need post-processing. Matsuzaki et al. [7] firstly identified the anatomical names of several branches through a rule-based pre-processing. Rest branches were assigned using a multi-class AdaBoost classifier [19]. Bogunovic et al. [16] was firstly trained on some pre-labeled samples to learn the features of bifurcation and anatomical topology of the tree structure. Then the anatomical name of each branch was assigned using maximum a posteriori probability (MAP) estimate [20]. Kitasaka et al. [10] adopted the conditional random field [21] for the representation of abdominal vessel tree and adopted an adaptive gradient algorithm [22] for structure learning. The anatomical name assignment of vessel branches was performed using MAP.

Recently, Wu et al. [3] introduced a **deep learning based** method TreeLab-Net, consisting of multi-layer perception (MLP) [23] and bidirectional tree-structural Long Short-Term Memory (Bi-TreeLSTM) [24]. TreeLab-Net only employed the geometric features from centerlines and put the image features in images aside. Due to the tree-topology, missing branches will disrupt the indices of nodes in the model, causing incorrect labeling. Also with the tree-structure, TreeLab-Net cannot model the overall correlation of branches, messages are only passed between adjacent branches. The farther branches are to the artery tree root, the less information they raise. In addition, the class imbalance issue between the main and side branches causes an unsatisfactory performance on the side branches.

Yang et al. [4] presented another **deep learning based** method CPR-GCN. It combines the graph convolutional network (GCN) [25] and the bidirectional Long Short-Term Memory (BiLSTM) [26]. As we all know, the graph model is transductive [27]. Hence CPR-GCN cannot adaptively change its structure to match the test data with different topology from the training data. In addition, as a graph model [28] CPR-GCN also suffers from the shallow structure [29], i.e., deep network will degrade the model's performance. So CPR-GCN is prone to underfitting. The class imbalance between the main and side branches also exists in CPR-GCN, leading to an unsatisfactory performance on the side branches.

III. METHODOLOGY

The architecture of TTN is shown in Fig.4. For each segment, 2DTSC (2D transform spherical coordinates) extracts the geometric feature from the centerline tree, IFE (image feature extractor) extracts the image feature in the image. The two features are concatenated as the combined feature. The topological encoding is added to the combined feature and output embedding respectively, as inputs of the Transformer encoder and decoder. The topological encoding represents the positions of vessel segments in the artery tree, assisting the model in classifying branches. The segment-depth loss can greatly solves the class imbalance between the main and side branches, it incorporates the anatomical topology of the artery tree into the training process.

A. COMBINED FEATURE OF VESSEL SEGMENTS

The geometric feature used in existing methods lacks extra image information such as the sizes, lengths and surrounding textures of vessel segments that also contains great information which can be exploited to identify different branches. The

vessel segment's feature in this work is a concatenation of both the geometric and image features.

1) GEOMETRIC FEATURE

To reduce the variance of centerline points' coordinates across different subjects, we transform the Cartesian coordinates of centerline points in each CCTA into a spherical coordinates [3]. For a CCTA, origin of the spherical coordinate system (SCS) is defined as the bounding box center of its centerline points, the direction from origin to aorta valve is defined as the north pole of SCS. In detail, a rotation matrix transforming z-axis of Cartesian coordinate system (CCS) to the north pole of SCS, is computed. It is also used to transform the x-axis of CCS to the prime meridian of SCS. Centerline points' Cartesian coordinates are transformed into the spherical coordinates (*radius*, *azimuth*, *elevation*) afterward. We abandon *radius* and define (*azimuth*, *elevation*) as centerline points' 2D transform spherical coordinates (2DTSC), as shown in Fig.4.

2DTSC normalizes the variance of centerline points across different subjects from three aspects: 1) For each CCTA, the global transformation of coordinates is normalized by treating the origin of SCS as the bounding box center of centerline points; 2) For each subject, the north pole of SCS points toward aorta valve, positioning the aorta above centerline points in 2DTSC; 3) Distance variance from the origin of SCS to centerline points is eliminated by ignoring the radius dimension in spherical coordinates.

For each vessel segment, three features were selected and concatenated as the geometric feature: 1) Parent-child angles in both CCS and 2DTSC spaces, i.e., the angle between the tangential direction at parent segment's end and the tangential direction at child segment's start; 2) The 2DTSC coordinates of start-point, center-point and endpoint; 3) The direction pointing from start-point to endpoint, and the tangential direction at start-point in 2DTSC.

2) IMAGE FEATURE

The artery branches have a sequential dependency in 3D CCTA, we propose an image feature extractor (IFE) consisting of 3D CNN [30] and a self-attention layer [31], extracting the tubular image features of vessel segments in CCTA images. An illustration of extracting the image feature of R-PDA is shown in Fig.5.

For each CCTA, control points are sampled on all vessel segments $(P_1, P_2, \ldots, P_n)$ equidistantly by a sampling distance s. An image patch I_k^i is a 3D cube centered at control point i on segment P_k in CCTA. The size of I_k^i is $r \times r \times r$. The image feature in patch I_k^i is extracted by 3D CNN. For all subjects, image features in the cubes set I_k of vessel segment P_k are padded to a fixed length to ensure the mini-batch manner in training process. The fixed-length image features set is as the input of self-attention layer and the output is treated as the image feature of segment P_k. The weights of

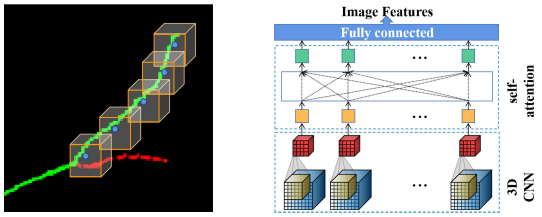

FIGURE 5. Illustration of the image feature extractor that extracts the image features of vessel segments in 3D CCTA, which consists of a 3D CNN and a self-attention layer. For each segment, a series of 3D patches extracted along the centerline in CCTA are the input of 3D CNNs, then the outputs are summarized by the self-attention layer as the image feature of this segment.

3D CNN and self-attention are the same across all subjects and all vessel segments.

It is worth noting that the image feature is of less importance than the geometric feature [4]. Hence the image feature is downweighted by a scalar $\mu \in (0, 1)$ to reduce its weight in the combined feature.

B. TOPOLOGICAL ENCODING

The coronary artery exhibits a tree-structural anatomy, each vessel segment takes up a specific position as shown in Fig.1. We argue that such position information can be exploited to identify different branches. In addition, original Transformer is permutation-invariant [32], i.e., there is no spatial difference among different elements in a sequence. We propose a topological encoding that represents the positions of vessel segments in the artery tree, incorporating the positional differences of vessel segments into the model and assisting the model in classify branches. The topological encoding passes the coronary artery's topology to both Transformer encoder and decoder.

Firstly, we use Breadth First Search (BFS) [33] and Depth First Search (DFS) [34] to sort vessel segments respectively, obtaining two segment sequences. An example is shown in Fig.6, segments shown are LM, proximal-LCX (p-LCX), proximal-LAD (p-LAD), middle-LCX (m-LCX), OM1, middle-LAD (m-LAD), D1, distal-LCX (d-LCX), OM2, distal-LAD (d-LAD), D2. Using BFS, the sequence of the segments is LM, proximal-LCX, proximal-LAD, ..., R-PLB, R-PDA; while using DFS, the sequence is LM, proximal-LCX, middle-LCX, ..., R-PLB, R-PDA.

For each segment in the sequence obtained by BFS, we construct a $d/2$ channel encoding using *sin* and *cos* functions with different frequencies:

$$PE_{(pos,i)} = \begin{cases} sin(pos \cdot \omega_k), & i = 2k \\ cos(pos \cdot \omega_k), & i = 2k+1 \end{cases} \quad (1)$$

where $\omega_k = 1/10000^{2k/\frac{d}{2}}$; pos is the position of the segment in the sequence. The same operation is applied to the sequence obtained by DFS, constructing another $d/2$ channel encoding. We concatenate these two $d/2$ channel encodings, obtaining a d channel topological encoding of each segment.

IEEE Journal of Translational Engineering in Health and Medicine

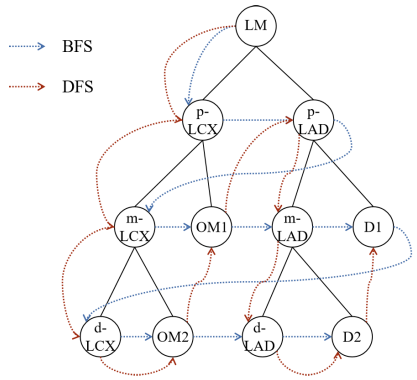

FIGURE 6. Two sequences of vessel segments obtained by applying BFS algorithm and DFS algorithm to the same coronary artery respectively.

Using both BFS and DFS makes our topological encoding determine the positions of segments in the artery tree exactly.

C. SEGMENT-DEPTH LOSS

In the coronary artery, the number of side branches, e.g., OM and DA, is much less than the number of main branches, e.g., LM, LAD and LCX. Existing methods do not focus on this issue, resulting in an unsatisfactory performance on the side branches. To solve this issue, we propose a segment-depth loss which incorporates the anatomical topology of coronary artery into the training process.

The entire coronary artery exhibits a tree-structural anatomy, in which the side branches are farther to the tree root than main branches. As shown in Fig.7, compared with main branches, there are more bifurcation points between the side branches and the tree root.

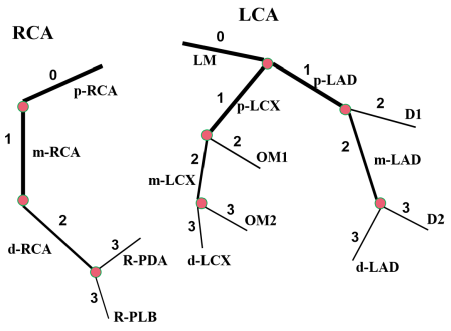

FIGURE 7. Illustration of the segment-depth of vessel segments. Red points indicate the bifurcation points. The segment-depth of a vessel segment is defined as the number of bifurcation points from artery root to its endpoint.

As illustrated in Fig.7, the segment-depth of a vessel segment is defined as the number of bifurcation points from the artery root to the endpoint of this segment. For instance, the segment-depth of d-LAD and OM1 is 4 and 3 respectively. We represent the segment-depth as D and add it to original

Algorithm 1 Optimization Process of TTN

Require: 3D CCTA I, segment centerlines $(P_1, P_2, \ldots, P_n)$, segment-depth $(D_1, D_2, \ldots, D_n)$, output embedding $(OE_1, OE_2, \ldots, OE_n)$

Ensure: $GT_1, GT_2, \ldots, GT_n$

$S \leftarrow$ Build_sequence$(P_1, P_2, \ldots P_n)$;

$f_1^1, f_1^2, \ldots, f_1^n \leftarrow$ 2DTSC$(P_1, P_2, \ldots P_n)$;

$I_1, I_2, \ldots, I_n \leftarrow I \otimes (P_1, P_2, \ldots, P_n)$;

$f_2^1, f_2^2, \ldots f_2^n \leftarrow$ IFE$(I_1, I_2, \ldots I_n)$;

$f \leftarrow$ Concat(f_1, f_2)

$TE \leftarrow$ BFS&DFS$(P_1, P_2, \ldots, P_n)$

$Pred_1, Pred_2, \ldots, Pred_n \leftarrow$ Transformer$(f \oplus TE, OE \oplus TE)$;

loss $\leftarrow$Segment-depth loss$(Pred, GT)$;

cross entropy loss [35]. The segment-depth loss is defined as:

$$L = \frac{1}{N}\sum_i D_i L_i = -\frac{1}{N}\sum_i D_i \sum_{c=1}^{M} y_{ic} log(p_{ic}) \qquad (2)$$

where N is the number of all segments, M is the number of all classes, i represents the i-th segment, D_i is the segment-depth of segment i. y_{ic} equals to 1, if the true class of segment i is c, otherwise equals to 0. p_{ic} is the predicted probability indicating that segment i belongs to class c.

Our segment-depth loss pays more attention to the side branches due to the anatomical topology of artery tree. As shown in Fig.7, the segment-depth of side branches is higher than the segment-depth of main branches. In addition, side branches a smaller number, such as OM2 and D2, have higher depth D than side branches with a larger number, such as OM1 and D1. Therefore, among side branches, segment-depth loss also pays more attention to the fewer ones.

D. OPTIMIZATION AND APPLICATION

The training process is shown in Alg.1. A training sample is composed of the CCTA image I, vessel segment centerlines $P_1, P_2, \ldots, P_n$, segment-depth $D_1, D_2, \ldots, D_n$, output embedding $OE_1, OE_2, \ldots, OE_n$ and ground truth labels $GT_1, GT_2, \ldots, GT_n$. For each training sample, all vessel segments are sorted into a sequence data S using BFS algorithm in advance. Along each segment, a series of 3D patches I_n is extracted as the input of IFE. The geometric features f_1 of vessel segments extracted from centerline tree by 2DTSC are concatenated with the image features f_2 output by IFE, as the combined features f of vessel segments. The topological encoding TE of vessel segments are constructed using both BFS and DFS, and added to f as the input of Transformer encoder. Meanwhile, TE are added to the output embedding OE of vessel segments as the input of Transformer decoder. The model then performs classification on all segments $P_1, P_2, \ldots, P_n$. The segment-depth loss is computed between the reference and predicted labels, i.e., $GT_1, GT_2, \ldots, GT_n$ and $Pred_1, Pred_2, \ldots, Pred_n$.

In the application process, all vessel segments of a test sample are sorted into a sequence data using BFS algorithm in advance. A series of 3D patches extracted along each segment in CCTA are the input of IFE. The geometric features of

vessel segments extracted from centerline tree via 2DTSC are concatenated with the image features output by IFE, as the combined features of vessel segments. The topological encoding of vessel segments are constructed using both BFS and DFS, and added to the combined features as the input of Transformer encoder. A beginning token is passed into the decoder, then the decoder outputs the label of the first segment. The output embedding and the topological encoding of the first segment are added and passed into the decoder, then the decoder outputs the label of the second segment. This process is repeated until the label of the last vessel segment is output.

IV. EXPERIMENT CONFIGURATIONS

A. DATA AND EVALUATION METRICS

Until now, a public CCTA dataset with anatomical name annotations of the coronary artery branches is not available. In this work, a CCTA dataset with 325 subjects was collected from the Department of Radiology of The Fourth Hospital of Harbin Medical University, the protocol number of ethical approval is 20220156. In this dataset, the largest, smallest and average number of branches is 11, 6 and 8.89. After separated at bifurcation points, the largest, smallest and average number of vessel segments is 16, 9 and 14.41.

To guarantee the reliability of centerline trees, centerlines were manually annotated by three experienced radiologists rather than using an automated or semi-automated system. Each radiologist independently annotated 1/3 dataset. They first annotated the root of the centerline tree as the center of the coronary ostium (i.e. the point where the coronary artery originates from the aorta), and the endpoint of each artery branch as the most distal point where the artery was still distinguishable from the background. Then they annotated the centerlines in the axial, coronal or sagittal slices. The centerline point was annotated at the center of gravity of the artery lumen in each slice. After the manual annotation each radiologist independently detected all 325 cases for any possible disagreements or errors, e.g., centerlines deviating from the center of gravity of the artery lumen, or confusions caused by the coronary artery bridging, stenosis or image artifact. The detection was performed utilizing the stretched multi-planar vessel reconstruction images constructed with the annotated centerlines [36]. They took discussions on controversial cases and reached an agreement on how to refine these cases to obtain the final centerlines. Then they annotated the anatomical names and segment-depth of vessel segments alone, and took discussion on the inconsistent ones to obtain the final results.

We divided all 325 CCTA into five subsets randomly and equally, the five-fold cross-validation was used to train and test all models. The evaluation was performed on all branches, based on the reference labels and the predicted labels. For each class, the recall rate is defined as $Recall = \frac{t_p}{t_p+f_p}$, the precision is $Precision = \frac{t_p}{t_p+f_n}$, the F1 score is $F1 = 2\frac{Precision \times Recall}{Precision+Recall}$. Also, the average of these metrics on all classes were used to evaluate the overall performance, $\text{AvgRecall} = \frac{1}{n}\sum_{i=1}^{n}\text{Recall}$, $\text{AvgPrecision} = \frac{1}{n}\sum_{i=1}^{n}\text{Precision}$, $\text{AvgF1} = \frac{1}{n}\sum_{i=1}^{n}\text{F1}$.

B. NETWORK ARCHITECTURE AND HYPER-PARAMETERS

The detailed architecture of TTN is shown in Table1. It contains two trainable components, a 3D CNN followed by a self-attention layer and a lightweight Transformer followed by a FC layer, which are jointly trained in end-to-end.

TABLE 1. The detailed architecture of TTN model.

Block	Details
3D CNN	kernel size=3, in channel=1, out channel=16 maxpooling size=2 kernel size=3, in channel=16, out channel=32 maxpooling size=2 kernel size=3, in channel=32, out channel=64 maxpooling size=2
Self-attention	layer=1, head=1, hidden size=64
Encoder	layer=2, head=2, hidden size=128
Decoder	layer=2, head=2, hidden size=128
Fully connected	out channel=64 out channel=# of classes

Original images had a slice spacing of 0.5mm and an in-plane resolution of $(0.44 \pm 0.04) \times (0.44 \pm 0.04)\text{mm}^2$. To guarantee the isotropy of 3D patches in image feature extraction, all images were resampled with a voxel size of 0.5mm^3. The radius of coronary artery branches is generally between 0 and 3 mm [37] (0 and 6 voxels), the size of 3D patches in CCTA passed into IFE was chosen as $12 \times 12 \times 12$ voxels to cover the angels, sizes, and surrounding textures of branches. The control points on centerlines in all subjects were sampled at the same distance $s = 10$ voxels to make the 3D patches overlap each other, keeping the sequential image feature along vessel segments. Coefficient downweighting the image feature was set as $\mu = 0.4$ experimentally. The model was implemented based on PyTorch. Adam optimizer [38] was used to update the model's parameters, initial learning rate was 10^{-3} with 10^{-5} decay. Batch size was 8 and the model was trained up to 1200 epochs on an NVIDIA Geforce 3090 GPU.

V. RESULTS

We further reproduced three state-of-the-art CABL methods, AICT [13], TreeLab-Net [3] and CPR-GCN [4]. For fair comparison, the hyper-parameters of all methods were tuned to give the best results with trial and error. As shown in Table2, our TTN obtains the best results of AvgRecall(0.894), AvgPrecision(0.869) and AvgF1(0.880). In addition, TTN achieves the best results of all metrics on all side branches. For instance, it obtains 0.922 *Recall*, 0.895 *Precision* and 0.909 $F1$ on DA branch, which is superior to other methods with a large margin. It also achieves a competitive result on the main branches, e.g., 0.912 *Recall*, 0.985 *Precision* and 0.947 $F1$ on RCA.

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 2. Results of TTN and other methods. All of these methods were evaluated on 325 subjects using a five-fold cross-validation. Recall, precision and F1 score on each branch are listed. The averages of these metrics on all branches are also used to evaluate the overall performance.

Method	Metrics	RCA	R-PDA	LM	LAD	DA	LCX	OM	L-PDA	RI	R-PLB	Avg(std)
AICT [13]	Recall	0.935	0.786	0.96	0.936	0.754	0.911	0.743	0.794	0.771	0.835	0.842±0.085
	Precision	0.957	0.743	1.00	0.905	0.834	0.846	0.821	0.725	0.771	0.802	0.840±0.090
	F1	0.946	0.764	0.980	0.920	0.792	0.878	0.780	0.758	0.771	0.818	0.841±0.083
TreeLab-Net [3]	Recall	0.941	0.771	0.944	0.925	0.790	0.936	0.766	0.794	0.75	0.773	0.839±0.085
	Precision	0.949	0.761	0.915	0.928	0.805	0.925	0.815	0.725	0.735	0.758	0.831±0.089
	F1	0.945	0.766	0.929	0.927	0.798	0.931	0.790	0.758	0.742	0.765	0.835±0.086
CPR-GCN [4]	Recall	0.929	0.829	0.912	0.923	0.754	0.949	0.737	0.778	0.792	0.722	0.832±0.088
	Precision	0.937	0.773	0.927	0.940	0.783	0.907	0.769	0.710	0.717	0.737	0.820±0.096
	F1	0.933	0.800	0.919	0.931	0.768	0.928	0.752	0.742	0.752	0.729	0.826±0.090
TTN	Recall	0.912	0.900	0.880	0.853	0.922	0.890	0.934	0.873	0.833	0.938	**0.894±0.035**
	Precision	0.985	0.778	0.965	0.912	0.895	0.833	0.897	0.797	0.816	0.813	**0.869±0.072**
	F1	0.947	0.834	0.921	0.882	0.909	0.861	0.915	0.833	0.825	0.871	**0.880±0.042**

TABLE 3. Ablation study results of TTN. All of the image feature, the topological encoding and the segment-depth (SD) loss are essential parts of TTN. All of the ablation experiments were evaluated on 325 subjects using a five-fold cross-validation. Recall, precision and F1 score on each branch are listed. The averages of these metrics on all classes are also used to evaluate the overall performance.

Ablation study	Metrics	RCA	R-PDA	LM	LAD	DA	LCX	OM	L-PDA	RI	R-PLB	Avg(std)
Without image feature	Recall	0.895	0.814	0.848	0.859	0.892	0.860	0.910	0.825	0.750	0.876	0.853±0.047
	Precision	0.966	0.704	0.876	0.904	0.866	0.832	0.899	0.754	0.720	0.759	0.828±0.089
	F1	0.930	0.755	0.862	0.881	0.879	0.846	0.905	0.788	0.735	0.813	0.839±0.065
Without topological encoding	Recall	0.856	0.614	0.792	0.813	0.862	0.839	0.892	0.794	0.729	0.804	0.800±0.079
	Precision	0.924	0.581	0.805	0.894	0.832	0.811	0.861	0.725	0.603	0.655	0.769±0.122
	F1	0.889	0.597	0.798	0.852	0.847	0.825	0.876	0.758	0.660	0.722	0.782±0.097
Without SD loss	Recall	0.941	0.829	0.928	0.885	0.838	0.924	0.838	0.762	0.708	0.835	0.849±0.074
	Precision	0.965	0.763	0.906	0.925	0.824	0.855	0.909	0.727	0.694	0.81	0.838±0.090
	F1	0.953	0.795	0.917	0.905	0.831	0.888	0.872	0.744	0.701	0.822	0.843±0.080
TTN	Recall	0.912	0.900	0.880	0.853	0.922	0.890	0.934	0.873	0.833	0.938	**0.894±0.035**
	Precision	0.985	0.778	0.965	0.912	0.895	0.833	0.897	0.797	0.816	0.813	**0.869±0.072**
	F1	0.947	0.834	0.921	0.882	0.909	0.861	0.915	0.833	0.825	0.871	**0.880±0.042**

The superiority of TTN derives from two aspects: 1) The overall correlation of branches allows the differences between all branches to be directly captured, side branches can be more easily classified. 2) The segment-depth loss greatly alleviates the class imbalance issue, strengthening the ability of TTN to classify side branches. As a conventional method, AICT [13] cannot leverage the advantage of large-scale data. Also, it only considers geometric feature, leaving the image feature and the topology of artery tree aside. TreeLab-Net [3] only utilizes the geometric features from centerlines, without the image features in CCTA. Due to the tree structure, missing branches will change the topology of model, damaging performances. CPR-GCN [4] presents a fixed graph structure, it cannot be adaptable to the test data with different anatomy from the training data, limiting the application in clinic. Also, CPR-GCN summarizes the image features using BiLSTM which is more time-consuming than self-attention.

We compare the normalized confusion matrices [39] to analysis the misclassification of each class in detail (Fig.8 and Fig.9). In other methods, there are high misclassifications on all side branches, R-PDA, DA, OM, L-PDA, RI and R-PLB. While TTN achieves the best result, only making 6% R-PDA, 3% D, 2% OM, 6% L-PDA, 10%RI, 4% R-PLB misclassified as R-PLB, OM, L-PDA, LCX, LCX, R-PDA respectively.

To validate the performance of TTN on diseased cases, we analyze the results on cases with different degrees of stenosis. In the dataset there are 57 cases with no stenosis, 89 cases with mild stenosis (<25%), 84 cases with mild stenosis (25%-50%), 64 cases with moderate stenosis (50%-70%) and 31 cases with severe stenosis (>70%). Table4 lists the results on cases with different degrees of stenosis. We found that the presence of stenosis had a less pronounced effect. In cases with severe stenosis the TTN obtains a result of AvgRecall(0.892), AvgPrecision(0.871) and AvgF1(0.880); while in cases with no stenosis the TTN obtains a result of AvgRecall(0.899), AvgPrecision(0.870) and AvgF1(0.883).

Fig.10 qualitatively demonstrates the performance of TTN. In case 1, 2 and 3, TTN generates correct labels for all branches, while other methods make various errors, especially on the side branches. In case 2, AICT [13] makes OM1 and middle-LAD misclassified as L-PDA and DA; TreeLab-Net [3] makes proximal-LCX and OM2 misclassified as RI and LCX; CPR-GCN [4] makes LM, OM1, DA misclassified as RI, LAD, LM respectively.

VI. ABLATION STUDY

A. IMAGE FEATURE

The image feature contains great information that can be exploited to classify branches, since the sizes, lengths and surrounding textures of different branches vary greatly. As shown in Table3, removing the image feature, metrics will drop on all branches. For instance, the Recall, Precision and F1 are 0.895, 0.966, 0.930 on RCA, compared with 0.912, 0.985, 0.947 in our TTN. 10% R-PDA and 7% R-PLB are misclassified as R-PLB and R-PDA respectively (Fig.8), while TTN only makes 6% R-PDA and 4% R-PLB misclassified as R-PLB and R-PDA respectively (Fig.9). Although

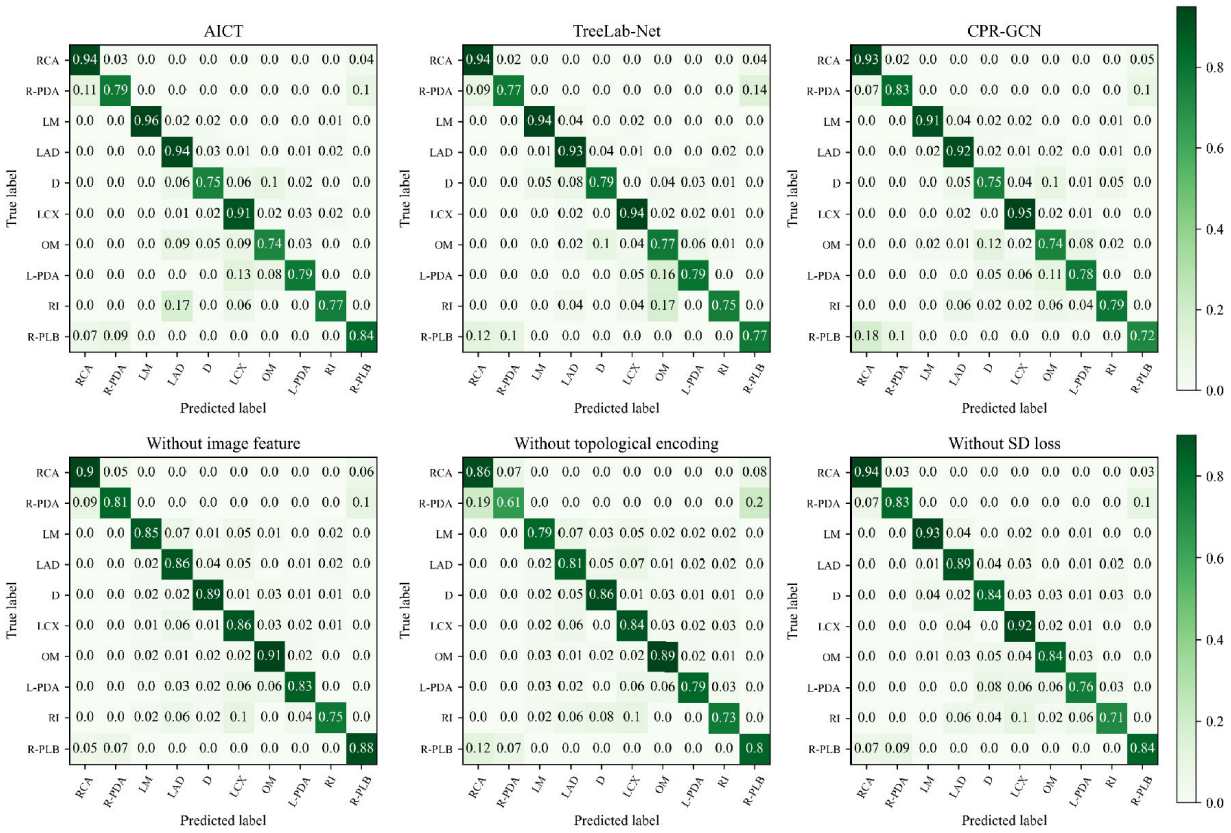

FIGURE 8. Normalized confusion matrices of branch labeling results of other methods and ablation study. The misclassification of each class can be observed directly. Other methods produce a large number of misclassifications on the side branches, and the ablation models also make considerable misclassifications on all branches. The number of each predicted class is normalized by the overall number of true class.

the image feature contains extra image information, it is of less importance than the geometric feature. In case 4 and 5 (Fig.10), model without the image feature predicts the labels for all segments correctly. In case 6, it only makes L-PDA misclassified as DA.

B. TOPOLOGICAL ENCODING

The topological encoding is an important characteristic of vessel segments, it represents the positions of vessel segments in the artery tree. To utilize this topological characteristic, topological encoding is added to the combined feature and output embedding of each segment. As shown in Table3, without the topological encoding, metrics on all branches drop dramatically. 20% R-PDA and 12% R-PLB are misclassified as R-PLB and R-PDA (Fig.8), while TTN only makes 6% R-PDA and 4% R-PLB misclassified as R-PLB and R-PDA respectively (Fig.9). In some subjects, the brother-segments (originating from the same parent segment) are hard to identify accurately, since they have the similar features (both the geometric feature and the image feature). The topological encoding provides a positional specificity for each segment to solve the brother-segment confusion. As shown in Fig.10, in all samples, removing the topological encoding makes proximal-LAD, RI, proximal-LAD

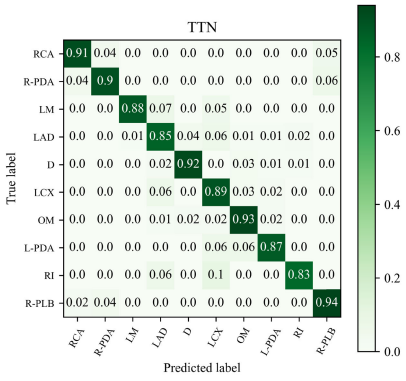

FIGURE 9. Normalized confusion matrix of branch labeling result of TTN, which produces fewer misclassifications on the side branches than other methods and also maintains fewer overall misclassifications on all branches than ablation models.

misclassified as RI, LAD, RI, since RI and proximal-LAD are brother-segments.

C. SEGMENT-DEPTH LOSS

Benefiting from the segment-depth loss, our TTN outperforms other methods on the side branches with a large margin. As shown in Table3, removing the segment-depth

IEEE Journal of Translational Engineering in Health and Medicine

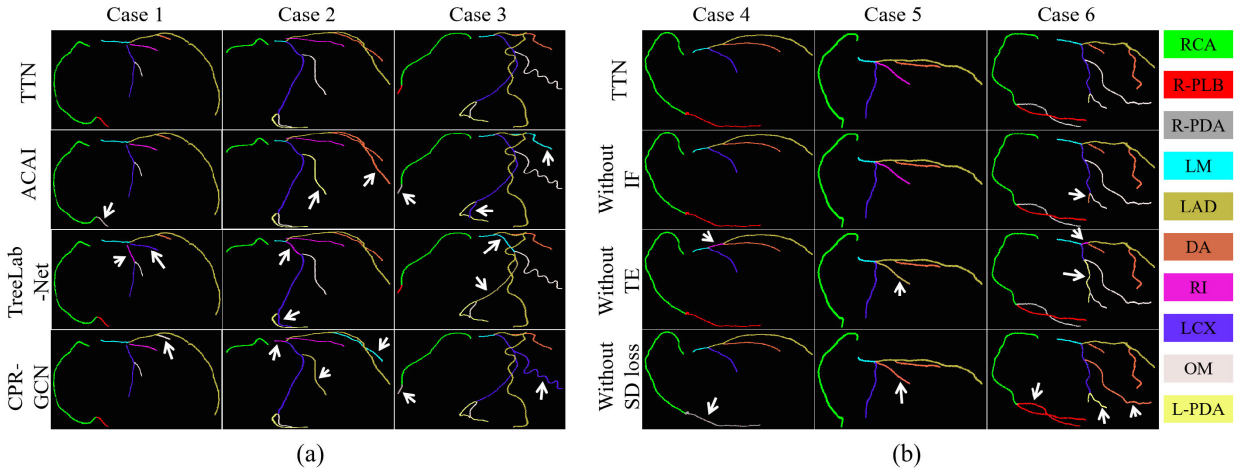

FIGURE 10. White arrows indicate incorrect labeling. (a) 3D rendered examples labeled with different methods. TTN generates correct labels for all branches, while other methods make various errors, especially on the side branches. (b) 3D rendered examples labeled in the ablation study. TTN generates correct labels for all branches, while other ablation models make various errors.

TABLE 4. Results of TTN on cases with different degrees of stenosis. There are 57 cases with no stenosis, 89 cases with mild stenosis (<25%), 84 cases with mild stenosis (25%-50%), 64 cases with moderate stenosis (50%-70%) and 31 cases with severe stenosis (>70%). Recall, precision and F1 score on each branch are listed. The averages of these metrics on all branches are also used to evaluate the overall performance.

Degree of stenosis	Metrics	RCA	R-PDA	LM	LAD	DA	LCX	OM	L-PDA	RI	R-PLB	Avg(std)
None	Recall	0.910	0.911	0.913	0.851	0.927	0.896	0.940	0.867	0.837	0.934	**0.899±0.035**
	Precision	0.983	0.766	0.956	0.902	0.906	0.842	0.911	0.805	0.810	0.821	0.870±0.071
	F1	0.945	0.832	0.934	0.876	0.916	0.868	0.925	0.835	0.823	0.874	**0.883±0.045**
Mild(<25%)	Recall	0.909	0.898	0.881	0.858	0.919	0.901	0.931	0.870	0.831	0.940	0.894±0.034
	Precision	0.988	0.784	0.962	0.910	0.887	0.835	0.904	0.791	0.820	0.806	0.869±0.072
	F1	0.947	0.837	0.920	0.883	0.903	0.867	0.917	0.829	0.825	0.868	0.880±0.042
Mild(25%-50%)	Recall	0.913	0.904	0.882	0.849	0.925	0.883	0.933	0.873	0.834	0.935	0.893±0.035
	Precision	0.982	0.772	0.965	0.922	0.895	0.837	0.892	0.793	0.817	0.813	0.869±0.073
	F1	0.946	0.833	0.922	0.884	0.910	0.859	0.912	0.831	0.825	0.870	0.879±0.042
Moderate(50%-70%)	Recall	0.917	0.881	0.879	0.862	0.920	0.886	0.930	0.878	0.835	0.933	0.892±0.032
	Precision	0.989	0.782	0.968	0.905	0.891	0.827	0.894	0.813	0.808	0.810	0.869±0.072
	F1	0.952	0.829	0.921	0.883	0.905	0.855	0.912	0.844	0.821	0.867	0.879±0.043
Severe(>70%)	Recall	0.908	0.894	0.877	0.857	0.918	0.885	0.935	0.877	0.826	0.941	0.892±0.035
	Precision	0.981	0.780	0.968	0.916	0.901	0.844	0.882	0.800	0.814	0.819	**0.871±0.070**
	F1	0.943	0.833	0.920	0.886	0.909	0.864	0.908	0.837	0.820	0.876	0.880±0.041

loss, the performance drops dramatically on all side branches, e.g., 11.1% Recall, 7% Precision and 8.9% F1 drop on L-PDA compared with TTN. The normalized confusion matrices in Fig.8 and Fig.9 further demonstrate the effectiveness of segment-depth loss in handling the class imbalance. Removing the segment-depth loss makes 10% R-PDA, 4% D, 5% OM, 8% L-PDA, 10% RI, 9% R-PLB misclassified as R-PLB, LM, D, D, LCX, R-PDA respectively. While TTN only makes 6% R-PDA, 3% D, 2% OM, 6% L-PDA, 10% RI, 4% R-PLB misclassified as R-PLB, OM, LCX, OM, LCX, R-PDA. In Fig.10, without the segment-depth loss, only the side branches are misclassified. In case 4 and 5, R-PLB and RI are misclassified as R-PDA and DA. In case 6, R-PDA, OM1 and OM2 are misclassified as R-PLB, DA and L-PDA.

D. COEFFICIENT FOR DOWNWEIGHTING IMAGE FEATURE

To identify the optimal value of μ that downweights the image feature, experiments set is conducted using various $\mu = \{0.2, 0.3, 0.4, 0.5\}$. We select the $F1$ metric on each branch as the evaluation measure, since it can reflect both *Recall* and *Precision*. Fig.11 illustrates that the performance of TTN can be improved by increasing the value of μ from 0.2 to 0.4. However, with $\mu = 0.5$, the performance has degraded. For $\mu = \{0.2, 0.3\}$, the weight of image feature is too small to provide sufficient image information of vessel segments. For $\mu = 0.5$, the weight of image feature is too

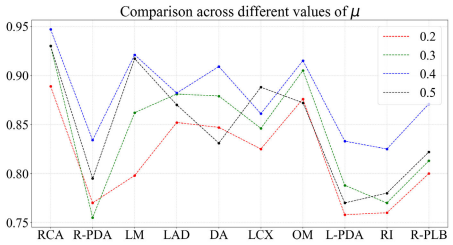

FIGURE 11. Ablation study of the value of μ. *F1* on each branch, which can reflect both *Recall* and *Precision*, is selected as the evaluation measure. 0.2, 0.3, 0.4 and 0.5 are the values of μ.

large, reducing the weight of geometric feature that is of more importance than the image feature. Hence $\mu = 0.4$ is selected as the scaling factor of the image feature in this work.

VII. DISCUSSION AND CONCLUSION

In this work besides the geometric feature, the representation feature of vessel segments also contains extra image feature extracted in the 3D CCTA image. The sizes, lengths and surrounding textures of vessel segments, which are usually ignored in existing methods, can be exploited to classify branches. Ablation study suggests that even though the geometric feature is important, the image feature weighs more than just semantic information. Since the image feature is of less importance than geometric feature [4], we downweight the image feature by a scalar $\mu \in (0, 1)$.

The coronary artery exhibits a tree-structural anatomy, each vessel segment takes up a specific position as shown in Fig.1. Such anatomical topology is a powerful tool used to classify branches. The topological encoding constructed by both BFS and DFS algorithms uniquely indicates a specific vessel segment. It is worth noting that the generation of topological encoding is not limited to our method, other forms can also be explored, as long as uniquely indicating a specific segment.

Both experiment results and ablation study demonstrate that the segment-depth loss alleviates the class imbalance issue significantly. It increases the weight of side branches in the loss function. As far as we know, existing methods have not focused on the class imbalance between main and side branches. The segment-depth loss greatly improves the performance of TTN on the side branches. However, it is at the expense of a slight decrease in the performance on main branches.

In this study we trained the model separately for the two subtrees. TTN is able to label the two subtrees using only one model by combining them into a single sequence. Training one model is more time-saving than training the two subtrees separately. Also in clinical diagnosis, using a single model to label the entire coronary artery is more flexible and efficient. However, Transformer-based models are more likely to "make mistakes" on the long sequences than on the short sequences. With training the two subtrees separately, each model only processes a short sequence. Combing the two subtrees into a long sequence brings a great challenge to the model to achieve a high-accuracy labeling result. Hence, to achieve a high-accuracy labeling results in this work, we trained the TTN model separately for the right and left subtrees.

The quality of branch labeling highly depends on the quality of centerline extraction. Missing, shortened and wrongly extracted branches degrade the labeling result significantly. In this study, to guarantee the reliability of centerline trees, centerlines of all 325 subjects are manually annotated by three experts. In future work, we will investigate whether the method could find out the missing, shortened or erroneously extracted branches automatically.

Our TTN solves the CABL task from a novel perspective of sequence labeling learning. It has handled two limitations in existing methods perfectly, i.e., lack of the overall correlation of branches and the serious class imbalance between main and side branches. To our best knowledge, TTN is the first Transformer based vessel branch labeling method and is notably different from previous methods. The model is trained in end-to-end. Given a CCTA dataset with 325 subjects, benefiting from the constructed overall correlation of branches, along with the image feature, the topological encoding and the segment-depth loss, our TTN obtains the best overall result on all branches, as well as the best result on side branches compared with other methods.

REFERENCES

[1] F. D. Fuchs and P. K. Whelton, "High blood pressure and cardiovascular disease are linked," *Hypertension*, vol. 75, no. 2, pp. 285–292, 2020.
[2] J. Leipsic et al., "SCCT guidelines for the interpretation and reporting of coronary CT angiography: A report of the society of cardiovascular computed tomography guidelines committee," *J. Cardiovascular Comput. Tomogr.*, vol. 8, no. 5, pp. 342–358, Sep. 2014.
[3] D. Wu et al., "Automated anatomical labeling of coronary arteries via bidirectional tree LSTMs," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 2, pp. 271–280, Feb. 2019.
[4] H. Yang, X. Zhen, Y. Chi, L. Zhang, and X.-S. Hua, "CPR-GCN: Conditional partial-residual graph convolutional network in automated anatomical labeling of coronary arteries," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3802–3810.
[5] G. L. Raff et al., "SCCT guidelines for the interpretation and reporting of coronary computed tomographic angiography," *J. Cardiovascular Comput. Tomogr.*, vol. 3, no. 2, pp. 122–136, Mar. 2009.
[6] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[7] T. Matsuzaki, M. Oda, T. Kitasaka, Y. Hayashi, K. Misawa, and K. Mori, "Automated anatomical labeling of abdominal arteries and hepatic portal system extracted from abdominal CT volumes," *Med. Image Anal.*, vol. 20, no. 1, pp. 152–161, Feb. 2015.
[8] W. Zhang, J. Liu, J. Yao, and R. M. Summers, "Automatic anatomical labeling of abdominal arteries for small bowel evaluation on 3D CT scans," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 210–213.
[9] B. H. Hoang et al., "A study on automated anatomical labeling to arteries concerning with colon from 3D abdominal CT images," *Proc. SPIE*, vol. 7962, Mar. 2011, Art. no. 79623R.
[10] T. Kitasaka et al., "Automatic anatomical labeling of arteries and veins using conditional random fields," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 6, pp. 1041–1048, Jun. 2017.
[11] A. Akinyemi, S. Murphy, I. Poole, and C. Roberts, "Automatic labelling of coronary arteries," in *Proc. 17th Eur. Signal Process. Conf.*, Aug. 2009, pp. 1562–1566.
[12] G. Yang et al., "Automatic coronary artery tree labeling in coronary computed tomographic angiography datasets," in *Proc. Comput. Cardiol.*, Sep. 2011, pp. 109–112.
[13] Q. Cao et al., "Automatic identification of coronary tree anatomy in coronary computed tomography angiography," *Int. J. Cardiovascular Imag.*, vol. 33, no. 11, pp. 1809–1819, Nov. 2017.
[14] M. A. Gülsün et al., "CTA coronary labeling through efficient geodesics between trees using anatomy priors," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14–18, 2014, Proceedings, Part II 17*. Springer, 2014, pp. 521–528.
[15] S. Gu, Z. Wang, J. M. Siegfried, D. Wilson, W. L. Bigbee, and J. Pu, "Automated lobe-based airway labeling," *Int. J. Biomed. Imag.*, vol. 2012, pp. 1–9, Jan. 2012.
[16] H. Bogunovic, J. M. Pozo, R. Cárdenes, L. S. Román, and A. F. Frangi, "Anatomical labeling of the circle of Willis using maximum a posteriori probability estimation," *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1587–1599, Sep. 2013.
[17] M. Bilgel, S. Roy, A. Carass, P. A. Nyquist, and J. L. Prince, "Automated anatomical labeling of the cerebral arteries using belief propagation," *Proc. SPIE*, vol. 8669, Mar. 2013, Art. no. 866918.

IEEE Journal of Translational Engineering in Health and Medicine

[18] D. Robben et al., "Simultaneous segmentation and anatomical labeling of the cerebral vasculature," *Med. Image Anal.*, vol. 32, pp. 201–215, Aug. 2016.

[19] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. Berlin, Germany: Springer, 2013, pp. 37–52.

[20] E. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 3, pp. 185–192, Sep. 1987.

[21] Y. Wang, K.-F. Loe, and J.-K. Wu, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289, Feb. 2006.

[22] F. Dondelinger, D. Husmeier, S. Rogers, and M. Filippone, "Ode parameter inference using adaptive gradient matching with Gaussian processes," in *Proc. Artif. Intell. Statist.*, 2013, pp. 216–228.

[23] E. A. Zanaty, "Support vector machines (SVMs) versus multilayer perception (MLP) in data classification," *Egyptian Informat. J.*, vol. 13, no. 3, pp. 177–183, Nov. 2012.

[24] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," 2015, *arXiv:1503.00075*.

[25] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[26] P. Zhou et al., "Attention-based bidirectional long short-term memory networks for relation classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, 2016, pp. 207–212.

[27] G. Ciano, A. Rossi, M. Bianchini, and F. Scarselli, "On inductive–transductive learning with graph neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 758–769, Feb. 2022.

[28] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model is a machine learning technique that can be used to learn the representations of graphs," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[29] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[30] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[31] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 7354–7363.

[32] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16239–16248.

[33] L. Luo, M. Wong, and W.-M. Hwu, "An effective GPU implementation of breadth-first search," in *Proc. 47th Design Autom. Conf.*, Jun. 2010, pp. 52–55.

[34] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM J. Comput.*, vol. 1, no. 2, pp. 146–160, Jun. 1972.

[35] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[36] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, "Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier," *Med. Image Anal.*, vol. 51, pp. 46–60, Jan. 2019.

[37] E. A. Hulten, S. Carbonaro, S. P. Petrillo, J. D. Mitchell, and T. C. Villines, "Prognostic value of cardiac computed tomography angiography: A systematic review and meta-analysis," *J. Amer. College Cardiol.*, vol. 57, no. 10, pp. 1237–1247, 2011.

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

[39] P. J. Hardin and J. M. Shumway, "Statistical significance and normalized confusion matrices," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 735–739, 1997.