Graph-PAVNet: A Graph-Based Learning Framework for Pulmonary Artery and Vein Separation Using Multimodal Feature Sampling

Qingya Li[12], Ye Yuan[12], Lu Liu[12], Nan Bao[41], Lisheng Xu[5], Wenjun Tan[123(✉)]

[1] Key Research Laboratory of Intelligent Computing of Medical Images, Ministry of Education, Northeastern University, China
[2] School of Computer Science and Engineering, Northeastern University, China
[3] National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China
tanwenjun@cse.neu.edu.cn
[4] College of Medicine and Biological Information Engineering, Northeastern University, China
[5] College of Information Science and Engineering, Northeastern University, China

Abstract. Pulmonary artery-vein separation is critical for clinical diagnosis and treatment planning. However, existing pixel- or voxel-based methods often produce fragmented predictions, significantly reducing clinical confidence. To address above problems, we propose Graph-PAVNet, a graph structure learning framework designed for PA/PV separation. First, our Light Vessel Structured Modelling (LVSM) module constructs a topology-aware vascular graph by leveraging the inherent structural and semantic relationships within the vascular network. LVSM shifts from traditional voxel-level predictions to topology-based branch-level inference, effectively resolving prediction discontinuity. However, it is challenging for a single graph to do the separation task. Due to this issue, we propose the Modal Feature Sampling (MFS) module. MFS enriches node features by constructing a hybrid Real-Virtual (RV) feature matrix that integrates multi-source information. It also employs a dynamic feature weighting mechanism to achieve cross-modal complementarity, overcoming the challenges posed by modal discrepancies. For hierarchical inference, the Hierarchical Graph Attention Network (HGAT) stratifies nodes by vascular generation order (main to peripheral branches) and employs hierarchical masking to enforce structured inter-layer propagation. At last, we introduce a novel metric: Branch Misprediction Coefficient (BMC) to better evaluate the clinical relevance and branch inconsistency. Experimental results show that our method outperforms existing approaches in both quantitative accuracy and clinical interpretability, offering a new paradigm for pulmonary artery-vein separation.

Keywords: Pulmonary Artery-Vein Separation, Graph Structures, Multimodality Feature Sampling.

---

https://github.com/Lqy1018/Graph-PAVNet
Q. Li and Y. Yuan—Both authors contributed equally to this work.

1 Introduction

Pulmonary arteries (PA) and veins (PV) serve critical functions in cardiopulmonary circulation by facilitating oxygen exchange and maintaining hemodynamic equilibrium. The accurate identification of pulmonary artery-vein holds paramount clinical significance for diagnostic evaluation and therapeutic management. However, due to the complex and variable morphology of the pulmonary vasculature, precise separation of PA/PV presents persistent challenges in medical imaging analysis. With the rapid advancement of deep learning methods, some new techniques have emerged as predominant solutions for PA/PV differentiation challenges, including convolutional neural networks (CNNs [1,2,3]), recurrent neural networks (RNNs [4,5]), and more advanced variants such as U-Net [6], V-Net [7], etc. These models, by learning from extensive medical imaging data, are capable of automatically identifying the image features of PAs and PVs and thereby assisting clinicians in diagnostic tasks.

Despite the successes of deep learning models in PA/PV separation, pixel-level prediction methods face two critical limitations. First, these models often produce fragmented outputs, as seen in **Fig. 1**, particularly in low-contrast regions or fine vascular branches, leading to structural discontinuities and artifacts. Second, these errors propagate into clinical applications: disrupted vascular topology compromises quantitative metrics (e.g., vessel diameter, bifurcation analysis), rendering the results unreliable for diagnostic or interventional planning.

To address these challenges, researchers have begun to explore the integration of graph-based methods into medical image analysis. Unlike the grid-like structure processed by CNNs [1,2,3], graphs represent vessels as interconnected nodes and edges, preserving continuity and hierarchical dependencies. Some studies have combined Graph Convolutional Networks (GCNs) with CNNs to improve segmentation outcomes [8]. For post-processing, graphs can also be used to connect disjointed segments [9]. Hybrid frameworks further demonstrate that incorporating geometric priors with deep learning can enhance the robustness of tubular structure segmentation.

Based on the above analysis, this paper introduces the first end-to-end trainable framework for pulmonary artery/vein (PA/PV) separation. Our approach innovatively integrates vascular topology graphs with hierarchical graph neural networks. The key contributions are as follows: **(1) Light Vessel Structured Modelling (LVSM)**: We encode the vascular skeleton as a sparse graph, where nodes represent bifurcation points or endpoints, and edges represent vessel segments. This significantly reduces computational complexity while preserving anatomical topology. It also supports interactive corrections by clinicians using graph editing tools. **(2) Modal Feature Sampling (MFS)**: We focus on the authenticity of points and edges in the graph structure. Instead of relying solely on concrete image features and prior features, we transform the feature sampling to abstract connections between real and virtual nodes and edges. This enhances the representation of vascular structures. **(3) Hierarchical Graph Attention Network (HGAT)**: We utilize a hierarchical gated propagation mechanism. Nodes are explicitly stratified based on the order of vessel generation (main branches to peripheral branches). A hierarchical mask controls inter-layer information exchange, ensuring that the network captures both local and global dependencies effectively.

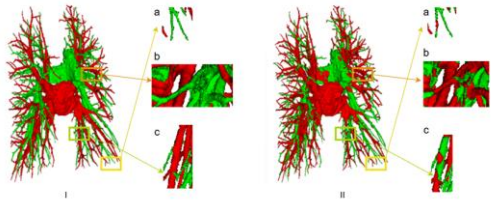

Fig. 1. The challenges of prediction discontinuity. I. II. represent label and prediction separately, while a-c shows different parts of vascular trees.

2 Method

The overall workflow is illustrated in **Fig. 2**, which sequentially comprises the following key components: LVSM module constructs sparse vascular graph from vessel skeleton; MFS module extracts and integrates heterogeneous features from multiple sources, generating fused node embeddings; HGAT establishes a multi-level graph attention architecture based on seed nodes for node attribute refinement. Next, we will explain each part in details.

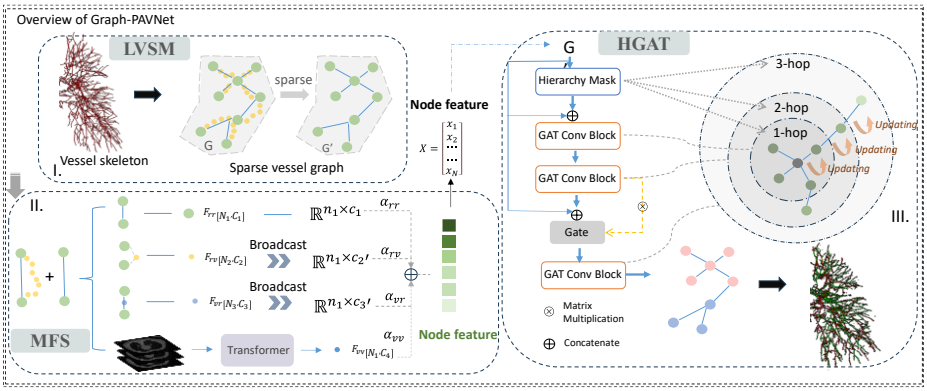

Fig. 2. The overview of Graph-PAVNet.

2.1 Light Vessel Structured Modelling (LVSM)

Inspired by Charbonnier et al. [10], after obtaining the vascular volume, we simplify it into a set of vessel skeleton points $S = \{s_k \in \mathbb{R}^3 | k = 1, \ldots, N\}$. The key nodes set $V = \{v_i\}$ is composed of bifurcation points $B = \{b_p\}$ and endpoints $L = \{l_q\}$, i.e., $V = B \cup L$. The edge set $E = \{e_{ij}\}$ is generated according to the following rules:

For each bifurcation point $b_p \in B$, we search for adjacent key nodes $v_j \in V$ along its connected paths. An edge e_{ij} is established if $\exists$ continuous path $P_{b_p \to v_j} \subseteq S$ with $\min_{S_k \in P_{b_p \to v_j}} d(S_k, \partial\Omega) > \tau$, where τ denotes the vessel radius threshold and $\partial\Omega$ represents vascular boundaries. To further simplify the graph structure and aggregate nodes, we construct a new graph $G_{simplified} = (V, E_{simplified})$ after constructing

$G = (V, E)$. The edge set $E_{simplified}$ only retains the direct connections between bifurcation points and endpoints, defined as follows:

For each edge $e_{ij} \in$ E, its corresponding simplified edge $e_{simplified} \in E_{simplified}$ is defined as $E_{simplified} = \{e_{ij}^{simplified} = (v_i, v_j) | e_{ij} \in \text{E}\}$, where $v_i, v_j \in$ V are the key nodes at the end of the edge.

Through the above steps and formulations, the vascular skeleton is modeled as a simplified graph structure $G_{simplified}$, which only keeps bifurcation points and endpoints. This significantly reduces data volume and computational complexity while preserving the essential topological relationships of the vascular network.

2.2 Modal Feature Sampling (MFS)

In the task of PA/PV separation, relying on one single feature often fails to capture the complexity and diversity of vascular structures. Additionally, the simplification of graph structures in the previous step can lead to the loss of local morphological information. Thus, it is necessary to supplement the model with additional features to ensure comprehensive information for accurate modeling. Inspired by discussions on prior knowledge in various studies(like centerline [11]), we propose a multimodal feature sampling strategy to mitigate the limitations of insufficient features and reduce the coupling challenges between topological structure modeling and image feature representation.

Feature Domain. This module constructs a comprehensive feature space for the simplified graph $G_{simplified} = (V, E_{simplified})$ by defining a quadruplet feature space. Specifically, we categorize feature information into four modal types based on the authenticity of nodes and edges: Features from real nodes and real edges (F_{rr}), Features from real nodes and virtual edges (F_{rv}), Features from virtual nodes and real edges (F_{vr}), and Features from virtual nodes and virtual edges (F_{vv}).

F_{rr} captures the anatomical features of true vascular nodes, reflecting geometric and topological properties. The mathematical formulation is:

$$F_{rr_{[N_1,C_1]}} = \left[\frac{1}{|N(v_i)|}\sum_{v_k \in N(v_i)} cos\theta_{ijk}, \mathbb{I}_{lobe}(v_i), \int_{P_{ij}} k(s)ds, \frac{\partial \rho}{\partial s}|_{s=0}\right] \quad (1)$$

where θ_{ijk} is the angle between adjacent edges, $\mathbb{I}_{lobe}$ is the lobe label indicator function, $k(s)$ is the path curvature, and ρ is the vessel radius. The notation $[N_1, C_1]$ represents the dimensionality of the feature matrix. Additionally, F_{rr} includes 3D coordinates, branch directions, tracheal labels, etc.

F_{rv} quantifies the differences between true vessels and their simplified representations, describing local deformations and morphological changes. The differential geometric differences between original and simplified edges are calculated as:

$$F_{rv_{[N_2,C_2]}} = [\frac{\| P_{ij}^{real} - P_{ij}^{virtual} \|}{L_{ij}}, \frac{\int k_{real}(s)ds}{\int k_{virtual}(s)ds}, \angle(n_{real}, n_{virtual})] \quad (2)$$

where L_{ij} is the edge length, $k(s)$ is the path curvature, and n is the normal vector.

F_{vr}extracts geometric features from the simplified graph, providing global topological information. The definition is:

$$F_{vr_{[N_3,C_3]}} = [\| v_i - v_j \|, \frac{\rho_i + \rho_j}{2}, \angle(t_{ij}, z)] \quad (3)$$

where ρ_i is the node radius, t_{ij} is the edge direction vector, and z is a global reference attribute. Additional features include Manhattan distance and the number of path nodes in the original graph.

F_{vv} mitigates local texture loss via a 3D orthogonal plane Transformer. For each node $v_i = (x, y, z)$, we sample 27-point neighborhoods across axial/coronal/sagittal planes:

$$P_{axis} = \bigcup_{i,j,k\epsilon\{-1,0,1\}}(x + i\Delta x, y + j\Delta y, z + k\Delta z) \tag{4}$$

A Transformer encoder with multi-head self-attention processes the 9-patch grids per plane:

$$F_{vv_{[N_1,C_4]}} = \bigoplus_{p=1}^{9} \frac{1}{n}\sum_{i=1}^{n} T_i^{(L)} \in \mathbb{R}^{9d} \tag{5}$$

where $\bigoplus$ denotes channel-wise concatenation and $T_i^{(L)}$ represents final-layer token embeddings

Feature Fusion. To integrate modality-specific information from virtual nodes and edges into real nodes and align features with diverse notations, we present a topology-aware propagation mechanism. At first, for alignment of features from different parts, the features of virtual nodes and edges are broadcast into their connected real nodes through the graph structure.

$$F_m \in \mathbb{R}^{n_m \times c_m} \rightarrow \mathbb{R}^{n_1 \times c_k} \quad \forall m \in \{rr, rv, vr, vv\} \tag{6}$$

For final fused feature matrix $\mathbb{X}$, we employ a gated hierarchical attention mechanism, where α_* represents modality-specific attention weights computed through a learnable gating network.

$$\mathbb{X} = \alpha_{rr} \cdot F_{rr} + \alpha_{rv} \cdot F_{rv} + \alpha_{vr} \cdot F_{vr} + \alpha_{vv} \cdot F_{vv} \tag{7}$$

2.3 Hierarchical Graph Attention Network (HGAT)

The hierarchical confusion in vascular segmentation, where conventional GATs [12] set nodes as the same place, motivates our design of HGAT—a model which emphasizes more on node relationships. To establish anatomically meaningful hierarchies, we initialize arteriovenous seed nodes at the pulmonary hilum and compute minimum hop counts via Breadth-First Search (BFS), generating hierarchical mask. It helps with: (1) as topological priors to amplify attention weights for proximal nodes during feature aggregation, and (2) as training curricula that phase-in node subsets from 0-hop to 2-hop neighborhoods, emulating semi-supervised propagation dynamics. As seen as Fig2. III., building upon this hierarchy, the network architecture cascades three functionally specialized GAT modules: initial feature fusion ($h_i = [x_i \parallel m_i]$) explicitly binds vascular semantics with topological positions(seed mask); subsequent local (0-1 hop) and global (2+ hop) propagation layers progressively resolve spatial dependencies, where hierarchical gating ($g_i = \sigma(W[h_i \parallel c_i]$, c_i is the hop) dynamically modulates cross-hop information flow. By these steps (hierarchical encoding, progressive propagation, and dynamic gating), HGAT effectively addresses the hierarchy confusion in GATs.

2.4 Evaluation Coefficient

To evaluate the topological accuracy and anatomical plausibility of PAV separation, we introduce a novel quantitative metric:

Branch Misprediction Coefficient (BMC).

$$BMC = \frac{\sum_{v \epsilon s} \sum_{pred_n \epsilon N(v)} \mathbb{I}(C_v \neq C_n)}{2} \quad (8)$$

Where v is a point in the set s, $N(v)$ is the neighboring nodes of v, $\mathbb{I}(C_v \neq C_n)$ is an indicator function that returns 1 if the attributes of two points are different. This coefficient reflects the severity of the prediction fracture.

3 Experiments and Results

3.1 Experiment Setup

Datasets and Evaluation Metrics. The dataset used in the experiment is from the ISICDM2020 competition dataset [13] and consists of 24 sets of CT scans with slice counts ranging from 192 to 576, slice thicknesses ranging from 0.75 to 1.5 mm, a voltage of 120 kVp, and a tube current ranging from 88 to 215 mA. These CT scans are randomly split into 4 subgroups (each subgroup has 6 sets: 4 sets for Training, 1 set for validation, 1 set for Testing). The accuracy metrics include: S-Dice, BMC, Recall, PPV(Positive Predictive Value), and node prediction accuracy.

Implementation Details. Our experiments were implemented on the basis of PyTorch-Geometric and MATLAB. Specifically, the graph structure establishment and initial attribute calculation are realized in MATLAB, which are saved as txt files and then transferred to the Python project for subsequent feature fusion and node classification. The project was trained and tested on an NVIDIA Tesla T4 machine. In order to be fair, our comparison experiment also divided the training and testing into four subgroups, and the training and testing were carried out with the same data. Prior to the comparison experiment, we preprocessed the original CT images to [-1000,900] and cropped them to [128, 128, 128] size CT blocks using a sliding window. The optimizer of the model used the Adam optimizer, the initial learning rate was set to 0.005, and the loss function was set to a custom loss function weighted by the seed node loss function and the total loss function. In addition, we set a simple clustering function as postprocess for node prediction and the following results are the test results after clustering.

Table 1. The results of ablation study for Feature Sampling.

Feature(Acc[%]↑)	Case-01	Case-02	Case-03	Case-04	Case-05	Case-06
F_{rr}	0.7194	0.7423	0.5995	0.6607	0.6337	0.6461
$F_{rr} + F_{rv}$	0.7684	0.7560	0.6269	0.6723	0.7542	0.6579
$F_{rr} + F_{rv} + F_{vr}$	0.7840	0.8303	0.6488	0.5759	0.8956	0.6346
$F_{rr} + F_{rv} + F_{vr} + F_{vv}$	**0.8140**	**0.8581**	**0.7716**	**0.6819**	**0.8989**	**0.7993**

Ablation Study. To validate the effectiveness of the proposed multimodal feature sampling mechanism, we conducted ablation experiments on four modalities, using node

accuracy as evaluation coefficient. The results are summarized in **Table 1**. From the experimental results, we can observe that for each case, when compared with F_{rr} only, the feature matrix incorporating both virtual points and virtual edges all shows improvement. The average accuracy increase of improvement is about 10%. This aligns well with our theoretical hypothesis: More comprehensive, More effective. But also, we found that there was a significant individualized difference in the data results, with a 21.7% difference between the final results of case-01 and case-04. This also shows that the robustness of existing methods still needs to be improved.

Comparison with State-of-the-Art Methods. In the context of PAV separation task, we also compared our approach with several state-of-the-art models, including 3D V-Net[7], UNETR[14], nnUNet[15], TransUNet[16], U-KAN[17], and Mamba-UNet[18]. The experimental results are compiled in **Table 2**, with evaluation metrics of Dice, PPV, Recall, and before-mentioned BFC for disruption, all in skeletonized form. It can be found that our graph-based framework demonstrates superior performance on certain vascular structures (e.g., veins, with improvements of 1.6% in both Dice and PPV). Additionally, Recall has increased by 2.4% and 3.4%, respectively. This evidence substantiates that our approach not only matches but also surpasses existing methods in certain aspects. Notably, under the BFC coefficient assessment, our method significantly reduces the number of disruption artifacts to fewer than 20 instances. This substantial improvement enhances the clinical applicability of our method and underscores the immense potential of graph-based learning in tasks with well-defined topological relationships.

Table 2. The comparison with the state-of-the-art models.

Method	S-Dice[%]↑		S-PPV[%]↑		S-Recall[%]↑		BFC↓
	PA	PV	PA	PV	PA	PV	
3D VNet	0.7164	0.5929	0.7051	0.7711	0.7280	0.6816	266
UNETR	0.6692	0.6014	0.5833	0.8055	0.7847	0.6854	556
nnUNet	**0.9062**	0.8859	0.9698	0.9478	0.8504	0.8316	74
UNet++	0.6315	0.6407	0.6088	0.6634	0.6560	0.6300	918
TransUNet	0.7545	0.6785	0.8302	0.8528	0.6915	0.5643	650
U-KAN	0.8673	0.8411	0.9270	0.9011	0.8149	0.7891	386
Mamba-UNet	0.8992	0.8717	**0.9866**	0.9644	0.8261	0.7952	122
Ours	0.8875	**0.9025**	0.9625	**0.9805**	**0.8752**	**0.8657**	**12**

Furthermore, we projected the skeletonized results back into the three-dimensional volume to visually compare the outcomes with existing methods. To simulate potential manual corrections in clinical practice, here we employed an interactive approach to adjust the node prediction results before recreation. **Fig. 3** illustrates the prediction outcomes of the current best-performing model for the pulmonary artery and vein separation task, alongside a comparison of the classification of a small segment of peripheral vessels. It is evident that the predictions suffer from issues of prediction discontinuities. When compared to the ground truth labels, these discontinuities are not only present in the peripheral vessels but also in the main vascular trunks. Our method demonstrates

superior continuity in predictions, with fewer discontinuities in the branching structures. Additionally, the graph-based approach significantly reduces the number of inference units, enabling the integration of human-assisted corrections. This not only enhances the precision of the predictions but also provides reliable support for clinical applications.

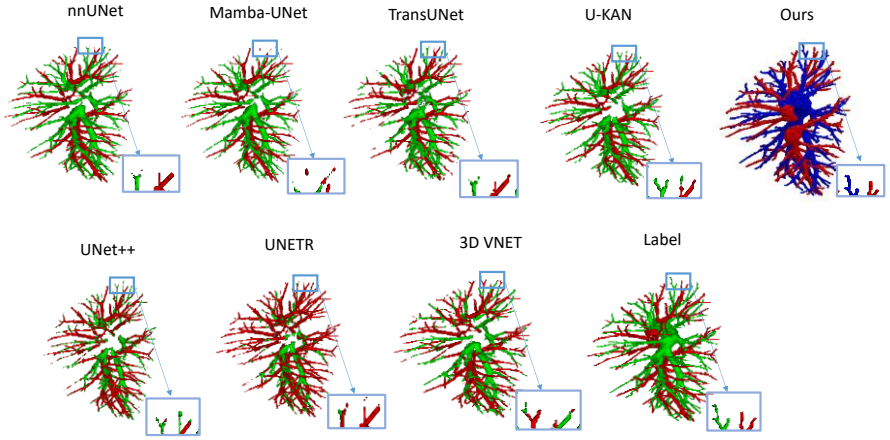

Fig. 3. The 3D visualization of PAV separation.

4 Conclusion

In this paper, due to the unavoidable fragmented prediction results of existing deep learning models, we attempt to transform the inference unit into graph structure to improve this situation. Based on this idea, we propose the first graph-centric deep learning framework in the context of pulmonary artery-vein separation tasks. This framework is sequentially divided into several essential parts: graph structure construction (LVSM), feature sampling and supplementation (MSF), and node prediction (HGAT). Building on existing experiments, we find that graph structures can serve as a new paradigm for PAV separation tasks. Compared to images, graph structures offer improvements in data volume, clinical interpretability, post-processing, and fracture as well. However, we also observe that when compared to classical image algorithm models, the difference between graphs during multi-graph training of our models can result in poorer generalization capabilities. Feature design and compression also require manual intervention, which demonstrates the potential and hidden challenges of graph structures. Our next research direction is trying to enhance the robustness of multi-graph models and design more reasonable feature attributes to reduce manual intervention.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62471122) and Fundamental Research Funds for the Central Universities(N25GFZ017, N25BJD005).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Nardelli P, Jimenez-Carretero D, Bermejo-Pelaez D, et al. Pulmonary artery–vein classification in CT images using deep learning[J]. IEEE transactions on medical imaging, 2018, 37(11): 2428-2440.
2. Pu J, Leader J K, Sechrist J, et al. Automated identification of pulmonary arteries and veins depicted in non-contrast chest CT scans[J]. Medical image analysis, 2022, 77: 102367.
3. Guo, J. *et al.* (2022). AANet: Artery-Aware Network for Pulmonary Embolism Detection in CTPA Images. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, vol 13431. Springer, Cham. https://doi.org/10.1007/978-3-031-16431-6_45
4. Qi H, Wu M, Ke S, et al. Deep Residual W-Unit Learning with Semantic Embedding for Automatic Pulmonary CT Artery-Vein Separation[C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 3275-3279.
5. Qi H, Wu M, Wang A, et al. Automatic CT pulmonary artery-vein segmentation using 3D RSU and semantic embedding[C]//Medical Imaging 2024: Clinical and Biomedical Imaging. SPIE, 2024, 12930: 664-669.
6. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer international publishing, 2015: 234-241.
7. Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). Ieee, 2016: 565-571.
8. Zhai Z, Staring M, Zhou X, et al. Linking convolutional neural networks with graph convolutional networks: application in pulmonary artery-vein separation[C]//Graph Learning in Medical Imaging: First International Workshop, GLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings 1. Springer International Publishing, 2019: 36-43.
9. Zhou Q, Tan W, Li Q, et al. A new segment method for pulmonary artery and vein[J]. Health Information Science and Systems, 2023, 11(1): 47.
10. Charbonnier J P, Brink M, Ciompi F, et al. Automatic pulmonary artery-vein separation and classification in computed tomography using tree partitioning and peripheral vessel matching[J]. IEEE transactions on medical imaging, 2015, 35(3): 882-892.
11. Pan L, Li Z, Shen Z, et al. Learning multi-view and centerline topology connectivity information for pulmonary artery–vein separation[J]. Computers in Biology and Medicine, 2023, 155: 106669.
12. Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.
13. Tan W, Li X, Zhou Q. Lung imaging anatomical structure segmentation data set and its application[J]. Chin. J. Image Graphics, 2021, 26(09): 2111-2120.

14. Hatamizadeh A, Tang Y, Nath V, et al. Unetr: Transformers for 3d medical image segmentation[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 574-584.
15. Isensee F, Petersen J, Klein A, et al. nnUNet: Self-adapting framework for UNet-based medical image segmentation[J]. arXiv preprint arXiv:1809.10486, 2018.
16. Chen J, Mei J, Li X, et al. 3D TransUNet: Advancing medical image segmentation through vision transformers[J]. arXiv preprint arXiv:2310.07781, 2023.
17. Li C, Liu X, Li W, et al. U-kan makes strong backbone for medical image segmentation and generation[J]. arXiv preprint arXiv:2406.02918, 2024.
18. Wang Z, Zheng J Q, Zhang Y, et al. Mamba-unet: Unet-like pure visual mamba for medical image segmentation[J]. arXiv preprint arXiv:2402.05079, 2024.