Response letter for the manuscript titled ‘**Multi-graph Graph Matching for Coronary Artery Semantic Labeling in Invasive Coronary Angiograms**’

Dear Editor and Reviewers,

We are grateful for the comments and suggestions from the editors and the reviewers, which are crucial for improving our work. We have revised the manuscript to address the reviewers’ comments fully. Our point-by-point reply to the review comments is summarized below. In this document, the original reviewers’ comments are in **black**; our responses are in **blue**; the quotations in the revised manuscript are in **red**.

Reviewers' comments:

EiC: While you are revising your paper, here is a list of points worth checking, which we find author's overlook. I will check that these are adhered to before your paper is approved for publication, assuming the revision satisfies the Associate Editor and Reviewers.

Reply: Thanks. We have double-checked and revised the manuscript according to the list below from the editor-in-chief.

a) Make sure your title is succinct and grammatical. It should ideally not exceed 10-15 words.

Reply: Thanks for your suggestion. Our current title, "Multi-graph Graph Matching for Coronary Artery Semantic Labeling in Invasive Coronary Angiograms," has 12 words that precisely convey the scope of our work.

b) Make sure your conclusions reflect on the strengths and weaknesses of your work, how others in the field can benefit from it and thoroughly discus future work. The conclusions should be different in content from the abstract and be rather longer too.

Reply: Thanks for the suggestion. Our current form of the conclusions follows the suggestions. Specifically, we included a summary of our findings, discussed limitations, and outlined future work.

c) Take a careful look at your bibliography and how you cite papers listed in it. Make sure it is current and cites recent work. Please cite a variety of different sources of literature. Please do not make excessive citation to arXiv papers, or papers from a single conference series. Do not cite large groups of papers without individually commenting on them. So we discourage " In prior work [1,2,3,4,5,6] …". Your bibliography should only exceptionally exceed about 40 items.

Reply: Thanks for the suggestions. We have checked the reference section, which meets the publication requirements. Besides, we have added several references accordingly and corrected the format of the citations.

d) You may have originally written your paper with a different audience in mind. Please make sure the revised version is relevant to the readership of Pattern Recognition. To this end, please make sure you cite RECENT work from the field of pattern recognition that will be relevant to our readership.

Reply: Thanks for the suggestion. Our study is related to deep learning on graphs and medical image processing for coronary arteries semantic labeling using invasive coronary angiograms, aiming at the potential readers in pattern recognition, especially for the readers who are interested in graph neural network, graph matching, image processing and pattern recognition.

e) Do not exceed the page limits or violate the format, i.e. double spaced SINGLE column with a maximum of 35 pages for a regular paper and 40 pages for a review.

Reply: Our paper has a total page of 35, containing a title page, manuscript, acknowledgement and reference, which meets the publication requirements.

**Reviewer's Responses to Questions**

1. Are the objectives and the rationale of the study clearly stated?

Please provide suggestions to the author(s) on how to improve the clarity of the objectives and rationale of the study. Please number each suggestion so that author(s) can more easily respond.

Reviewer #1: Yes.

The authors highlight the formidable challenges associated with segmenting coronary arteries in invasive coronary angiography (ICA) images. These challenges include the low contrast between the foreground vessels and the background, which hinders accurate foreground extraction, leading to incomplete vascular delineation.

Concurrently, the authors note that pixel-to-pixel based approaches struggle with accurately categorizing coronary arteries. This difficulty arises from the morphological and pixel intensity feature similarities between different arterial segments, which can lead to ambiguities and misclassifications when relying solely on pixel-level information.

The authors posit that existing graph matching-based methodologies, which are confined to comparing two vessel trees derived from ICA for coronary artery semantic labeling, fail to encompass the full intricacy of arterial anatomy. Such a binary comparison does not sufficiently account for the nuanced complexities inherent in the vascular structure, suggesting a need for a more sophisticated approach that can better represent the rich anatomical details of the coronary arteries.

Reply: Thank you for your thoughtful and detailed feedback. We greatly appreciate your recognition of the challenges associated with coronary artery semantic segmentation and labeling in ICA images. Your summary of the study's objectives and rationale highlights the importance of addressing these challenges, and we are grateful for your insightful analysis of the existing methodologies and their limitations. Thank you again for your time and valuable feedback.

Reviewer #2: The authors introduce a multi-graph matching method for coronary artery semantic labeling, utilizing intra- and inter-graph feature interactions to enhance representation extraction. However, as the authors acknowledge, pixel-to-pixel methods struggle with identifying small branches. The proposed graph-based method seems to use segmentation as a pre-processing step to identify these small branches, which cannot be considered a contribution of the proposed graph matching algorithm.

Reply: Our primary objective is to achieve accurate coronary artery semantic labeling, and we acknowledge that the successful extraction of individual arterial branches is a significant aspect of this task. We would like to clarify that our primary objective is to perform semantic labeling for small branches, rather than focusing on their extraction. The goal is to assign accurate semantic labels to these branches, even in the presence of challenges such as vessel overlap, missing segments, and variations in image acquisition views, ensuring that the anatomical relationships and connections are accurately captured.

Most existing studies focus on segmentation of only major arteries, as follows.

* Jun et al. proposed a T-Net [1], a nested encoder-decoder, to extract three major arteries, including LCX, LAD, and the right coronary artery (RCA) for the right ventricle. However, side branches such as OM1, OM2, D1, and D2 were not included.
* Zhang et al. proposed a Progressive Perception Learning framework [2] to extract the three main coronary arteries, including LCX, LAD, and RCA using ICAs.
* Xian et al. employed fully convolutional networks to extract LCX, LAD, and RCA using ICA videos [3].

Existing studies have primarily focused on the extraction of major arteries, often ignoring the side branches. Additionally, their models can only extract one type of artery at a time. In contrast, our model can extract multiple arteries simultaneously, providing a comprehensive understanding of the coronary artery vasculature and offering more information for clinical applications. We added these related papers to the related work section in the revised manuscript.

In addition, classifying coronary arterial branches accurately remains challenging due to the morphological similarity among different types of arteries, the variability in anatomical structures and varying arterial anatomy under different projection view angles.

*1) Morphological similarity between different arterial branches*

Pixel-intensity-based models face challenges in distinguishing between individual arterial segments and generating semantic segmentation due to the morphological similarity among branches in the coronary vascular tree and the overlap of arteries in 2D (projection) ICAs, as shown in Figure RL1.

A collage of images of a vein

Description automatically generated

Figure RL1. Challenging examples for coronary artery semantic labeling.

Figure RL1 highlights these challenges, where yellow rectangles indicate arterial overlaps in 2D. In these three examples, the LAD branches from the first two ICAs exhibit similar morphological and pixelwise features to the D branch of the third ICA. This demonstrates the difficulties in coronary semantic segmentation caused by morphological similarity.

*2) Varying anatomy across different ICA projection view angles*

ICA is a key diagnostic procedure used to visualize the coronary arteries and detect any blockages or abnormalities. In clinical practice, several different angiographic views are commonly utilized to obtain comprehensive images of the coronary anatomy during this procedure. The varying anatomy across different ICA projection view angles ensures a thorough assessment of the coronary vasculature from multiple perspectives; however, identifying arterial branches is challenging. The key views used in ICA include:

* Left Anterior Oblique (LAO): The x-ray beam is directed from the patient's left side. The patient's body is rotated to the right, and the heart is imaged from the left anterior aspect. This view helps in visualizing the LMA artery, LAD and LCX.
* Right Anterior Oblique (RAO): The x-ray beam is directed from the patient's right side, with the patient's body rotated to the left. This view provides detailed images of the right coronary artery (RCA) and parts of the left coronary system.
* Anteroposterior (AP): The x-ray beam is directed from the front of the patient to the back (anteroposterior direction). This straightforward projection is useful for assessing the coronary arteries and can be particularly helpful for evaluating stents and grafts.
* Caudal (CAU): The x-ray beam is angled towards the patient's feet. This view is particularly useful for visualizing the left main coronary artery bifurcation and the proximal segments of the LAD and LCX.
* Cranial (CRA): The x-ray beam is angled towards the patient's head. This view is useful for visualizing the distal segments of the coronary arteries, including the mid and distal LAD.

We add the description of ICA image scanning technique to the revised manuscript in the introduction section.

(Section 1. Introduction)

In clinical practice, rotating the X-ray beam horizontally generates three primary view angles: Left Anterior Oblique (LAO), Right Anterior Oblique (RAO), and Anteroposterior (AP). When the X-ray beam is rotated vertically, it produces two additional projection angles: Caudal (CAU) and Cranial (CRA). By combining these horizontal and vertical projections, six commonly used view angles are created: LAO\_CAU, RAO\_CAU, AP\_CAU, LAO\_CRA, RAO\_CRA, and AP\_CRA. Each of these views provides distinct and complementary insights into coronary anatomy, which are crucial for accurate diagnosis and treatment planning. The selection of view angles is based on the coronary segments of interest and the specific clinical context. However, as illustrated in Figure 1, the location, orientation, and positioning of arterial branches can vary considerably between individuals. This variability underscores the challenges in accurately extracting coronary arteries, given the significant anatomical differences across patients.

A collage of images of a person's body

Description automatically generated

**Figure 1**. Representative examples of ICAs under different view angles with corresponding 3D artery visualization using an example 3D Coronary computed tomography angiography (CCTA) images with annotated arteries.

Figure 1 displays example ICA images captured from different viewing angles, highlighted in the central scatter plot. Each dot in the scatter plot represents a distinct ICA image within our dataset, with color variations indicating different view angles. The x-axis (Primary Angle) represents the horizontal rotation of the x-ray beam, associated with RAO, AP, and LAO projection angles, while the y-axis (Secondary Angle) represents the vertical rotation, corresponding to CRA and CAU projection angles. The surrounding images reveal that even when the same view angle is used, there is considerable variability in the spatial arrangement and topology of the coronary arteries. These observations underscore the anatomical differences across individuals and the inherent difficulty of accurately interpreting ICA images. As a result, even when using manually or semi-automatically generated ICA graphs, precise semantic labeling of coronary arteries remains a significant challenge. While distinguishing coronary arteries is relatively straightforward using 3D CCTA images due to their clear spatial representation, identifying individual branches using 2D ICA images remains challenging because of vessel overlap and view-dependent anatomical variations.

(Section 2. Related Work)

Modern deep learning based approaches for extracting individual coronary arteries are classified into two categories: semantic segmentation and semantic labeling. Semantic segmentation is a pixel-level labeling task where the goal is to classify each pixel in an image into a specific class/category. For example, in the context of coronary artery segmentation, semantic segmentation would involve labeling each pixel in an angiogram as belonging to one of the coronary arteries (LAD, LCX, and right coronary artery) or background. Pixel-to-pixel based approaches achieved impressive results in segmenting main arteries. Jun et al. proposed a T-Net [1], a nested encoder-decoder, to extract three major arteries, including LCX, LAD, and the right coronary artery (RCA). However, side branches such as OM1, OM2, D1, and D2 were not included. Zhang et al. proposed a Progressive Perception Learning framework [2] to extract the three main coronary arteries, including LCX, LAD, and RCA using ICAs. Similarly, Xian et al. employed fully convolutional networks to extract LCX, LAD, and RCA using ICA videos [3]. Existing studies have primarily focused on the extraction of major arteries, often ignoring the side branches. Additionally, their models can only extract one type of artery at a time. In contrast, our model can extract multiple arteries simultaneously, providing a comprehensive understanding of the coronary artery vasculature and offering more information for clinical applications.

2. If applicable, is the application/theory/method/study reported in sufficient detail to allow for its replicability and/or reproducibility?

Please provide suggestions to the author(s) on how to improve the replicability/reproducibility of their study. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1: Mark as appropriate with an X:

Yes [X] No [] N/A []

Provide further comments here:

The authors proposed a Multi-graph Graph Matching (MGM) based approach for semantic labeling of coronary arteries in invasive coronary angiograms (ICAs). The method section of the article describes in detail how to achieve semantic labeling of coronary arteries by constructing arterial graphs, performing feature embedding, graph matching, and stenosis detection. In addition, the model training and testing process and how to evaluate the performance of the model are also introduced.

The manuscript describes the method in some detail and has a certain reproducibility reference, but it is encouraged to open source the code and data.

Reply: Thanks for your positive comments and completed summarization. To further help readers duplicate the results, we released our implementation alone with deidentified example ICAs for better understanding. Our code, instruction, illustration and running environment are available at GitHub (<https://github.com/MIILab-MTU/MGM_ICA>). We added this into the introduction section.

(Section 1. Introduction)

The code for the proposed method is available at <https://github.com/MIILab-MTU/MGM_ICA>.

Reviewer #2: Mark as appropriate with an X:

Yes [X] No [] N/A []

Provide further comments here:

The algorithm is straightforward to follow.

Reply: Thanks for your positive comments. As the datasets belonged to Nanjing First Medical University and the Medical University of South Carolina, the access to these datasets should be restricted. However, in the revised manuscript, we uploaded the implementation and provide an example along with Jupyter notebook to duplicate the results and show the results of the example coronary artery semantic labeling using the pre-processed ICAs.

Our open-sourced repository is available at GitHub (https://github.com/MIILab-MTU/MGM\_ICA).

(Section 1. Introduction)

The code for the proposed method is available at <https://github.com/MIILab-MTU/MGM_ICA>.

3. If applicable, are statistical analyses, controls, sampling mechanism, and statistical reporting (e.g., P-values, CIs, effect sizes) appropriate and well described?

Please clearly indicate if the manuscript requires additional peer review by a statistician. Kindly provide suggestions to the author(s) on how to improve the statistical analyses, controls, sampling mechanism, or statistical reporting. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1: Mark as appropriate with an X:

Yes [] No [X] N/A []

Provide further comments here:

Reviewer #2: Mark as appropriate with an X:

Yes [] No [X] N/A []

Provide further comments here:

The p-value should be provided to indicate the statistical significance.

Reply: Thanks for pointing out the issue related to the statistical significance in the performance evaluation. In the revised manuscript, we employed the Student's *t*-Test to evaluate the statistical significance of the proposed method compared to existing peer methods. Specifically, for each peer method and the proposed MGM, we performed 5-fold cross-validation to assess model performance. We have included the detailed statistical test results in the revised manuscript in Section 4.4.

(Section 4.4. Comparison with Existing Methods)

Furthermore, we conducted a Student's *t*-Test to assess whether the performance differences presented in Table 5 are statistically significant. The null hypothesis assumes that there is no significant difference in performance between the two models, while the alternative hypothesis posits that there is a statistically significant difference. We evaluated the statistical significance of performance differences between the proposed MGM and each baseline model for coronary artery semantic labeling using ICAs, considering metrics such as ACC, REC, PREC, and F1. These results are illustrated in Figure S1 in the supplementary materials.

Based on the observations from Figure S1, significant differences were observed across all evaluation metrics (ACC, REC, PREC, and F1) for coronary artery semantic labeling when comparing the proposed MGM model to the baseline models. The statistical analysis confirmed that these differences are not random, with the *t*-Test showing a statistically significant difference between the proposed model and each of the baseline methods.

(Supplementary materials. Section 1)

A group of colorful bars

Description automatically generated with medium confidence

**Figure S1.** The ACCs, RECs, PRECs and F1s achieved by the proposed MGM and baseline models were compared. The p-values of the student t-test are shown in the horizontal lines.

4. Could the manuscript benefit from additional tables or figures, or from improving or removing (some of the) existing ones?

Please provide specific suggestions for improvements, removals, or additions of figures or tables. Please number each suggestion so that author(s) can more easily respond.

Reviewer #1: The authors provide a thorough and well-reasoned argument, but I would like to see the performance of MGM on a public dataset, which would improve the reproducibility of the paper.

orCaScore dataset: https://orcascore.grand-challenge.org/Data/

Reply: Thank you for your thoughtful feedback. We appreciate your suggestions and would like to address the points you raised:

1. Dataset Concerns: We acknowledge that the dataset used in this paper is 2D ICA, while the provided dataset, orCaScore, is for coronary artery semantic labeling using 3D Coronary Computed Tomography Angiography (CCTA). CCTA offers clearer, comprehensive images and benefits from automated algorithms and additional radiomics features, making semantic labeling more accurate and less subjective. While ICA remains the gold standard for assessing vascular anatomical information, it faces challenges due to morphological similarities, varying projection angles, and subjective interpretation [4]. Our proposed MGM is specifically aimed to address the unique challenges posed by ICA, such as managing projection angle variability and morphological similarity through specialized algorithms. Applying this method to CCTA, where these challenges are not as prevalent, would not only be redundant but also overlook the optimized algorithms already available for CCTA.
2. Future Research Direction: We are excited to confirm that applying the proposed method to arterial semantic labeling using CCTA images is indeed a key direction for our future research. This will allow us to extend the applicability of graph matching to more comprehensive clinical tasks and applications and improve its robustness.

Thank you again for your valuable suggestion.

Reviewer #2: My concerns primarily relate to the application in clinical practice. Given that semantic segmentation is a highly challenging task, the authors should demonstrate that the proposed labeling process remains stable even when errors occur in the segmentation results. In fact, the pre-processing step of segmentation significantly reduces the challenges of semantic labeling. With accurate segmentation results, the semantic labeling process becomes much easier to address.

Reply: Thanks for your question. The manual or semi-automated segmentation of coronary arteries is the first step in artery identification. Coronary artery segmentation poses challenges at three distinct image levels: surrounding, local and semantic [2]. At the surrounding level, the low contrast between foreground vessels and the background is problematic. This is primarily caused by X-ray radiation dosage and contrast agent limitations, hindering accurate foreground extraction and leading to incomplete vascular segment delineation. Additionally, difficulties in excluding non-vessel areas arise due to their similar appearance to vascular segments, such as catheter outlines, spines, and ribs, complicating stenosis severity assessment. At the local level, local ambiguity near coronary vessel boundaries poses a challenge. High-frequency detail loss during the 3D to 2D projection in ICA imaging results in smooth grayscale boundaries instead of distinct edges, adversely affecting vessel boundary delineation. This ambiguity is particularly critical near stenosis areas, impacting the accuracy of stenosis severity assessment.

Furthermore, due to patient-specific anatomy, the variations of projection angles, and contrast dye degradation, recognizing the type of individual coronary arteries remains challenging. According to training guidelines provided by the American College of Cardiology Core Cardiology Training Symposium, it is recommended that trainees spend 4 months and achieve a minimum cumulative number of 100 cases to understand coronary anatomy sufficiently to perform independent diagnostic cardiac catheterization [5]. Additionally, the European Society of Cardiology recommends that a general cardiologist assist with or perform 300 coronary angiographies and interpret 1000 investigations [6]. Thus, manual annotation of coronary arteries is time-consuming and challenging. Therefore, automated identification of individual arteries is essential for improving efficiency and accuracy in cardiac diagnostics, especially for junior cardiologists and trainees.

Admittedly, building an end-to-end model for coronary artery semantic segmentation is more straightforward because it directly obtains pixel-to-pixel semantic mapping. However, as described in the introduction and discussion sections, the topology and anatomy vary significantly between patients, while the pixel intensities remain similar for different types of arteries. This makes it challenging for CNN-based end-to-end models to differentiate individual arteries.

(Section 2. Related Work)

Pixel-to-pixel based approaches achieved impressive results in segmenting main arteries. Jun et al. proposed a T-Net [1], a nested encoder-decoder, to extract three major arteries, including LCX, LAD, and the right coronary artery (RCA). However, side branches such as OM1, OM2, D1, and D2 were not included. Zhang et al. proposed a Progressive Perception Learning framework [2] to extract the three main coronary arteries, including LCX, LAD, and RCA using ICAs. Similarly, Xian et al. employed fully convolutional networks to extract LCX, LAD, and RCA using ICA videos [3]. Existing studies have primarily focused on the extraction of major arteries, often ignoring the side branches. Additionally, their models can only extract one type of artery at a time. In contrast, our model can extract multiple arteries simultaneously, providing a comprehensive understanding of the coronary artery vasculature and offering more information for clinical applications.

6. Have the authors clearly emphasized the strengths of their study/theory/methods/argument?

Please provide suggestions to the author(s) on how to better emphasize the strengths of their study. Please number each suggestion so that the author(s) can more easily respond.

Reviewer #1: The MGM method was compared with four other deep learning-based coronary artery semantic labeling methods, including association graph-based graph matching network (AGMN), edge-attention graph matching network (EAGMN), neural graph matching (NGM), and bidirectional tree LSTM (BiTL).

The author fully discusses the superiority of MGM over previous graph matching methods. But I suggest providing some comparisons with graph convolutional network solutions. like:

1. Topology-preserving automatic labeling of coronary arteries via anatomy-aware connection classifier. https://doi.org/10.48550/arXiv.2307.11959
2. TaG-Net: Topology-Aware Graph Network for Centerline-Based Vessel Labeling.

https://ieeexplore.ieee.org/document/10032183

Reply: Thanks for recommending these two papers related to vessel labeling according to centerline and topology-aware/preserving technology. We carefully reviewed these two papers and found that both of them are related to 3D artery/vessel semantic labeling using CCTA, where the anatomy is based on 3D coordinates, which is rotation free and irrelevant to the projection view of image capture procedure.

In contrast, the proposed work in our paper is related to 2D artery semantic labeling, in which the vessel overlapped, and ICA capture projection angles may significantly influence the anatomy visualization of the coronary arteries. As illustrated in Question 1, the arterial anatomy varies significantly due to the image capture projection angles and this makes it challenging to perform semantic labeling for coronary artery using ICA; however, arterial semantic labeling using CCTA is not influenced by the overlap and view difference since the arteries are presented in 3D without overlap and view difference.

For the first paper titled ‘topology-preserving automatic labeling of coronary arteries via anatomy-aware connection classifier’, the author mentioned that the different categories of coronary arteries have anatomically predetermined connections; for instance, LAD and LCX originate from LMA, while the diagonal branches are raised from LAD branch. The ResUNet based encoder and transformer were employed to extract the features for different arterial segments and a topology-aware connection classifier was proposed to identify the correct artery-to-artery connection. As mentioned before, this paper is related to coronary artery semantic labeling using 3D CCTA images which didn’t consider the view projection difference. Though they showed impressive performance with an ACC greater than 0.9, they are not suitable for coronary artery semantic labeling using 2D ICAs. Understanding the relationships between arteries and their spatial orientation requires a three-dimensional perspective; however, coronary artery labeling using 2D images in ICA is challenging. In addition, although the authors released their Github repository (https://github.com/zutsusemi/MICCAI2023-TopoLab-Labels/), no code implementation was provided.

For the second recommended paper, titled ‘TaG-Net: Topology-Aware Graph Network for Centerline-Based Vessel Labeling’, the author proposed a novel topology-aware graph network (TaG-Net) for vessel labeling and applied to head and neck vessels segmentation and semantic labeling using 3D computed tomography angiography (CTA) images. Though the proposed TaG-Net achieved impressive performance compared to the peer methods, it is still unsuitable for coronary artery semantic labeling task using 2D ICAs due to the internet differences between 3D and 2D image acquisitions and anatomy variations under different projection view angles.

In our current implementation, we have considered the topology constrain as a post processing procedure to optimize the graph matching results. In the revised manuscript, we added the detailed information for considering the topology prior knowledge as the post processing procedure, as shown in section 3.4.

An example of adding the topology prior knowledge as the post processing procedure is shown in Figure RL2. According to the coronary arterial anatomy, the validated connections for the LMA branch include LAD1 and LCX1, representing the first segments of the LAD and LCX. For other branches, the validated connected arterial segments are depicted in Figure RL2 (b). Suppose denotes the index of D2 in Figure S1 (a). The graph matching results indicate that the predicted label of D2 is D1. And the predicted labels of LAD2 and LAD3 are correctly identified as LAD2 and LAD3, respectively, as shown in Figure RL2 (c). The results reveal that LAD2 and LAD3 are accurately predicted; however, the prediction of D2 is incorrect. Consequently, D2 is an unconfident prediction and the graph matching result for D2 in this case will not be considered during the majority voting.

A screen shot of a computer

Description automatically generated

**Figure RL2**. Example of coronary artery anatomy and the look-up table for structural loss quantification.

Furthermore, these two papers also inspired us for further building the coronary artery semantic labeling using ICAs while considering the topology constraint as a prior knowledge during the model training. We cited these two papers and mentioned that we will further consider these methods in our future work.

(Section 3.4. Testing)

The coronary artery system adheres to a strict anatomical topology dictated by cardiovascular anatomy. For instance, the initial segments of the LCX and the LAD are connected to the LMA. Additionally, the side D branches are connected to the LAD, while the side OM branches are connected to the LCX. To ensure that predicted arterial labels align with this anatomical structure, a weighted penalty is applied to the graph-matching results if the predictions deviate from the expected anatomical connections.

To achieve this, we developed a look-up table that encodes the known physical connectivity between various arterial branches. If the predicted arterial labels match the connectivity patterns specified in the look-up table, the prediction is considered highly confident, which is further used for majority voting in line 10 of the Algorithm 2. Otherwise, the graph matching results will be discarded.

(Section 4.8. Clinical Application and Future Work)

Cardiologists often base their assessment of blockages on the presence and severity of lesions within these main branches [7]. For example, to differentiate between various types of complex lesions, such as bifurcation, calcified, chronic total occlusions, and unprotected left main coronary artery lesions, it is required to extract LMA from ICA first and provide the extracted arterial segment for cardiologists for screening. Thus, semantic segmentation of the coronary arterial tree to extract individual coronary arterial branches is important. Additionally, improving topology-aware methods [8,9] for 2D ICAs by integrating view-invariant features or leveraging multi-modal imaging strategies could enhance robustness against projection differences and spatial variability. Furthermore, it will be important to validate our approach on large-scale and multi-centered datasets to ensure its robustness and generalizability.

Reviewer #2: The authors clearly highlight their contributions to the proposed multi-graph matching method, particularly in comparison to pixel-to-pixel based methods.

Reply: Thank you for your positive feedback and for recognizing our contributions, especially the comparison of our proposed multi-graph matching method to pixel-to-pixel-based approaches. We are pleased that our efforts to clearly articulate the advantages and improvements offered by our method have been effective.

7. Have the authors clearly stated the limitations of their study/theory/methods/argument?

Please list the limitations that the author(s) need to add or emphasize. Please number each limitation so that author(s) can more easily respond.

Reviewer #1: I think the author could further elaborate on the following:

In Multi-graph Graph Matching (MGM) algorithms, cycle consistency is crucial to ensure accurate matching. However, as the number of involved graphs increases, maintaining this consistency becomes more difficult, which can lead to performance degradation.

Reply: Thanks for your comments. The performance will degrade when increasing the number of graphs because of the following two reasons:

1) As the number of graphs involved in the matching task increases, maintaining cycle consistency becomes increasingly challenging. The growing complexity is inherent to the combinatorial nature of MGM, where each additional graph introduces new pairwise and higher-order relationships that must align. This added complexity results in slight performance degradation as ensuring complete consistency across all cycles becomes computationally infeasible [10].

2) The number of testing cases for voting is limited. As the number of compared graphs increases, the number of validated groups of graphs decreases. For example, if , we need to identify 5 ICAs in the template set that have the same number of coronary arterial segments under the same projection view angles. Due to the limited number of template ICAs, the number of ICAs in the testing set failing to find validated ICAs in the template set increases, as shown in Table RL1.

Table RL1. The number of testing ICAs that cannot find validated ICAs in the template set under different values of .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| m | 3 | 4 | 5 | 6 | 7 |
| # of ICAs cannot find m-1 validated ICAs in the template set | 7 | 13 | 29 | 29 | 44 |

In our experiments, if we could not find validated ICAs in the template set, we duplicated them until we had template ICAs for testing. Our coronary artery semantic labeling algorithm relies on majority voting, which is shown in line 10 in Algorithm 2. If we could not find more than 1 MGM pair, the majority voting degrades into "a singular decision," which limits the performance.

We added the corresponding discussion in Section 4.3 in the revised manuscript.

(Section 4.3 MGM for Coronary Artery Semantic Labeling)

The performance of our MGM algorithm is impacted by two primary factors. First, as the number of graphs involved in the matching task increases, maintaining cycle consistency becomes increasingly difficult. This is due to the combinatorial complexity of MGM, where each additional graph introduces new pairwise and higher-order relationships that must align. Ensuring complete consistency across all cycles becomes computationally challenging, leading to slight performance degradation as the number of graphs grows. This computational issue is well-documented in the literature [10], as maintaining these relationships across numerous graphs becomes infeasible.

Second, the number of testing cases available for majority voting is limited. When more graphs are compared (e.g., ), fewer valid groups of graphs can be identified due to the limited number of ICAs available in the template set that satisfy the required validation criteria (same number of arterial segments and view angles). In cases where fewer than valid ICAs are found, they are duplicated to maintain testing consistency. However, if only a single MGM pair can be validated, majority voting degrades into a singular decision, which reduces the robustness and performance of the coronary artery semantic labeling algorithm. These limitations highlight how both the complexity of MGM and the scarcity of matching templates affect overall performance.

The article mentioned that the model performed poorly on ICA images of certain specific viewpoints (such as RAO CRA), which may be due to the limited number of training samples for these viewpoints. This suggests that the model may be sensitive to the viewpoint distribution in the training data.

Reply: Thanks for pointing out the issue. The model's performance on certain viewpoints, like the RAO CRA, which is affected by the limited representation of these specific angles in the training dataset. This highlights an important consideration: the model's ability to generalize to new viewpoints is highly dependent on the diversity and distribution of the data it was trained on. If certain viewpoints are underrepresented, the model may struggle to learn the distinguishing features for those angles, leading to poorer performance. Expanding the dataset to include a wider range of viewpoints, or using data augmentation techniques to simulate these angles, could potentially improve the model's robustness and overall performance across all viewpoints. We added this to the limitations and future work section.

(Section 4.8. Clinical Application and Future Work)

Cardiologists often base their assessment of blockages on the presence and severity of lesions within these main branches [7]. For example, to differentiate between various types of complex lesions, such as bifurcation, calcified, chronic total occlusions, and unprotected left main coronary artery lesions, it is required to extract LMA from ICA first and provide the extracted arterial segment for cardiologists for screening. Thus, semantic segmentation of the coronary arterial tree to extract individual coronary arterial branches is important. Additionally, improving topology-aware methods [8,9] for 2D ICAs by integrating view-invariant features or leveraging multi-modal imaging strategies could enhance robustness against projection differences and spatial variability. Furthermore, it will be important to validate our approach on large-scale and multi-centered datasets to ensure its robustness and generalizability.

Although cross-center testing showed good generalization of the model, these tests were still limited to two specific medical centers. The generalization ability of the model to a wider geographical, ethnic, and population distribution remains an open question.

Reply: While cross-center testing demonstrated good generalization of the model, it was limited to only two specific medical centers. At this time, we are constrained by the availability of data from a broader range of geographical, ethnic, and population distributions. Expanding the dataset to encompass more diverse medical centers and populations is a key focus for our future work. We recognize that achieving true generalization across varied patient demographics is essential, and we are actively exploring ways to address this in upcoming phases of the project. However, we didn’t find any public dataset with annotated LMA, LAD, LCX, D and OM branches for coronary artery semantic labeling using ICAs.

Reviewer #2: The authors are encouraged to discuss the stability of their method in the face of errors arising from pre-processing segmentation results.

Reply: Thanks for your question. In the revise manuscript, we added the robustness test to evaluate the performance for different methods with different levels of the corrupted datasets.

(Section 4.6 Robustness test)

The proposed MGM was trained and evaluated solely using the 'optimal' individual graphs. However, we cannot ensure that the binary segmentation model will consistently produce accurate arterial contours for all ICAs, especially considering the potential degradation of contrast dye. To assess the robustness of the proposed model, we conducted two experiments aimed at evaluating its performance under various conditions.

**(i) Arterial segments of varying lengths**. We evaluated the coronary artery semantic labeling performance on arteries of different lengths. The results of this evaluation are presented in Table 9. The centerline length is expressed in terms of the number of pixels.

**Table 9**. The accuracy of coronary artery semantic labeling for arterial centerlines of varying lengths, with the centerline lengths given in pixels and the corresponding range provided.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Centerline length | LMA | LAD | LCX | D | OM |
| (0,50] | 0.9934±0.0081, 381 | 0.9895±0.0086, 325 | 0.9645±0.0260, 332 | 0.9604±0.0488, 66 | 0.8759±0.0715, 88 |
| (50,100] | 1.0000±0.0000, 204 | 0.9759±0.0033, 473 | 0.9259±0.0150, 486 | 0.9695±0.0223, 294 | 0.8892±0.0571, 247 |
| (100,150] | 1.0000±0.0000, 21 | 0.9591±0.0281, 353 | 0.9332±0.0489, 363 | 0.9684±0.0111, 294 | 0.9337±0.0510, 271 |
| (150,200] | 1.0000±0.0000, 1 | 0.9166±0.0287, 244 | 0.9501±0.0185, 203 | 0.9178±0.0410, 215 | 0.9165±0.0351, 263 |
| (200,250] | -, 0 | 0.9147±0.0611, 151 | 0.8954±0.1029, 144 | 0.9286±0.0221, 167 | 0.9118±0.0686, 135 |
| (250,300] | -, 0 | 0.9456±0.0274, 94 | 0.9494±0.0277, 96 | 0.8478±0.0856, 77 | 0.8923±0.0646, 60 |
| (300,350] | -, 0 | 0.9857±0.0286, 76 | 0.9267±0.0904, 45 | 1.0000±0.0000, 33 | 0.8788±0.1497, 29 |
| (350,500] | -, 0 | 0.9846±0.0308, 54 | 0.9470±0.0663, 51 | 1.0000±0.0000, 19 | 0.8267±0.2585, 24 |

The results presented in **Table 9** show that the model performs well for coronary artery semantic labeling, especially for shorter arterial centerlines (≤50 pixels). In particular, the LMA segment achieves nearly perfect accuracy across all length ranges, indicating that the model handles larger arteries well. However, accuracy for smaller arteries like OM and D is lower, particularly for shorter segments, reflecting the challenges posed by smaller artery sizes and their more complex projections in 2D images.

As centerline length increases, accuracy tends to decrease, especially for segments like LAD and LCX. For example, LAD accuracy drops significantly after the (100,150] range. This decline could be attributed to increased complexity in segmenting longer arterial segments and the potential degradation of contrast dye in longer centerlines. Smaller arteries like OM also show increased variability in performance as centerline length increases, with the lowest accuracy observed in the longest centerline ranges (350,500] for OM.

These results suggest that the model is more effective for shorter, larger arterial segments, but improvements are needed for handling longer and smaller arteries. Future work should focus on improving segmentation accuracy for these more challenging cases, especially by incorporating more diverse data and advanced techniques to handle the complexities introduced by longer centerlines and smaller arteries.

**(ii) Graph matching using incomplete trees**. We conducted robustness tests by randomly removing parts of the arterial segments and evaluating the model's performance on the resulting incomplete vascular graphs. Specifically, we generated corrupted ICA-based vascular graphs by randomly deleting portions of arterial segments from the ICAs in the test set, while leaving the template set intact. To ensure graph connectivity, each deleted segment had to include at least one endpoint. The impact on model performance was assessed by randomly removing 5%, 10%, 15%, 20%, 25%, and 30% of the arterial segments, and the results were compared to those of baseline methods. The accuracy (ACC), precision (PREC), recall (REC), and F1 score are shown in Figure 5.

A screenshot of a graph

Description automatically generated

**Figure 5**. Achieved ACC, PREC, REC, and F1 of the proposed MGM, AGMN, EAGMN, NGM and IPCA using different corrupted ICAs. The horizontal axis indicates the probability of deleting an artery segment randomly.

The results in Figure 5 demonstrate the robustness of the proposed model, MGM, compared to baseline methods (AGMN, EAGMN, NGM, and IPCA) under varying levels of arterial segment deletion. Across all metrics—ACC, PREC, REC, and F1 score—MGM consistently outperforms other models, maintaining high performance even as the deletion rate increases. Notably, MGM exhibits minimal degradation in performance, highlighting its resilience to incomplete vascular graphs. In contrast, the baseline methods experience varying degrees of performance decline, with IPCA showing the steepest drop across all metrics.

At higher deletion rates (25% and 30%), MGM's advantage becomes more pronounced, reflecting its ability to handle severe data loss effectively. This is evident in its stable F1 score, which balances precision and recall better than competing models. AGMN, EAGMN, and NGM show moderate robustness but cannot match MGM’s consistent performance, while IPCA struggles significantly under higher corruption levels. These findings underscore MGM’s effectiveness and reliability in maintaining performance when faced with incomplete vascular graphs, making it a highly suitable model for practical scenarios involving noisy or missing data.

However, if the arterial segment contains two bifurcation points and is removed, it results in the separation of the individual arterial graph into two separate graphs, thereby breaking the continuity of the vascular tree. In such cases, human intervention becomes necessary.

8. Does the manuscript structure, flow or writing need improving (e.g., the addition of subheadings, shortening of text, reorganization of sections, or moving details from one section to another)?

Please provide suggestions to the author(s) on how to improve the manuscript structure and flow. Please number each suggestion so that author(s) can more easily respond.

Reviewer #1: Yes, I think the author's writing and expression are fluent enough.

Reviewer #2: The manuscript is easy to follow, and the clarification is clear.

Reply: Thank you for your feedback for both Reviewer #1 and Reviewer #2. We appreciate your positive comments regarding the fluency of the writing and the clarity of the manuscript. Based on your suggestions, we believe the structure and flow of the manuscript are well-executed, and we will continue to ensure clarity and coherence in the final version.

9. Could the manuscript benefit from language editing?

Reviewer #1: Yes

Reviewer #2: No

Reply: Thanks for your suggestion. We have double-checked the grammar issues and improved the English writing with the help of a native speaker in the revised version. Here are the detailed revisions.

Original text: Furthermore, due to patient specific anatomy, the variations of projection angles, and contrast dye degradation, recognizing the type of individual coronary arteries remains challenging.

Revised text: Furthermore, due to patient specific anatomy, variations of projection angles, and contrast dye degradation, recognizing the type of individual coronary arteries remains challenging.

Original text: Pixel-to-pixel based approaches face difficulties in identifying categories of coronary arteries due to the morphological and pixel-intensity based feature similarities between different arterial segments.

Revised text: Pixel-to-pixel-based approaches face difficulties in identifying categories of coronary arteries due to the morphological and pixel-intensity based feature similarities between different arterial segments.

Original text: 121 features, including the topology, pixel characteristics, and positional attributes, as detailed in [10], are extracted from each node.

Revised text: One hundred and twenty-one features, including the topology, pixel characteristics, and positional attributes, as detailed in [10], are extracted from each node.

Original text: Since the Sinkhorn operators has already shown effective and efficient performance in permutation prediction [19], we employ a Sinkhorn operator [20] to convert the node similarity matrix into a doubly-stochastic matrix as node affinity, as shown in Eq. 2

Revised text: Since the Sinkhorn operators have already shown effective and efficient performance in permutation prediction [19], we employ a Sinkhorn operator [20] to convert the node similarity matrix into a doubly-stochastic matrix as node affinity, as shown in Eq. 2.

Original text: However, using the combination function to find all possible graph matching pairs would generate a great number of matched pairs, which results in slow prediction procedure during the testing, hinder the clinical applicability of the MGM in practice.

Revised text: However, using the combination function to find all possible graph matching pairs would generate a great number of matched pairs, which results in a slow prediction procedure during the testing, hinder the clinical applicability of the MGM in practice.

Original text: The reason to increase the number of ICAs in the template set is due to the limited number of ICAs in the uni-site test.

Revised text: The reason for increasing the number of ICAs in the template set is due to the limited number of ICAs in the uni-site test.

Original text: We choose 600 graph matching set and derive the feature importance as shown in Figure 4.

Revised text: We chose 600 graph matching set and derive the feature importance as shown in Figure 4.

**AE**: This paper has merits in proposing a novel graph matching-based method for coronary artery segmentation in 3D, as recognized by the reviewers in their comments provided to the detailed questions. However, the paper needs substantial review especially at experimental level, with particular attention to evaluating the impact of the pre-processing stage and more systematic (and reproducible) generalization tests, deepening the understanding of error cases related to viewpoint changes. I invite the authors to consider a revision of this submission.

Reply: Thank you for your insightful feedback. We appreciate the recognition of the novelty of our graph matching-based method for coronary artery segmentation. We would like to clarify that our task specifically focuses on semantic labeling of coronary arteries using 2D ICA images, rather than 3D CCTA images. Semantic labeling in 2D ICA images is inherently more challenging than in 3D CCTA due to several factors, including the loss of 3D spatial context, reduced contrast between vessels and surrounding structures, and the complexity of interpreting artery types from a 2D projection.

In addition, 2D ICA images often suffer from projection artifacts, like boundary smoothing and ambiguity near stenosis areas, as described in the surrounding and local challenges. These issues are compounded by the variation in patient anatomy and projection angles, making semantic labeling more complex compared to 3D CCTA, where more detailed spatial information is available.

We understand that the impact of preprocessing and systematic generalization tests are critical for the robustness of the model, especially in handling viewpoint changes, and we performed the robustness tests in the revised manuscript. Expanding our testing to more diverse datasets, including different medical centers, will be an important future direction, as will improving our handling of error cases and viewpoint variations.

Thank you again for your valuable suggestions. I believe our revised manuscript has well reflected these improvements.

**Reference**

[1] T.J. Jun, J. Kweon, Y.-H. Kim, D. Kim, T-Net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography, Neural Netw. 128 (2020) 216–233. https://doi.org/10.1016/j.neunet.2020.05.002.

[2] H. Zhang, Z. Gao, D. Zhang, W.K. Hau, H. Zhang, Progressive Perception Learning for Main Coronary Segmentation in X-ray Angiography, IEEE Trans. Med. Imaging (2022) 1–1. https://doi.org/10.1109/TMI.2022.3219126.

[3] Z. Xian, X. Wang, S. Yan, D. Yang, J. Chen, C. Peng, Main Coronary Vessel Segmentation Using Deep Learning in Smart Medical, Math. Probl. Eng. 2020 (2020) 1–9. https://doi.org/10.1155/2020/8858344.

[4] C. Zhao, M. Esposito, Z. Xu, W. Zhou, HAGMN-UQ: Hyper Association Graph Matching Network with Uncertainty Quantification for Coronary Artery Semantic Labeling, Med. Image Anal. (2024) 103374. https://doi.org/10.1016/j.media.2024.103374.

[5] A.K. Jacobs, J.D. Babb, J.W. Hirshfeld, D.R. Holmes, Society for Cardiovascular Angiography and Interventions, Task force 3: training in diagnostic and interventional cardiac catheterization endorsed by the Society for Cardiovascular Angiography and Interventions, J. Am. Coll. Cardiol. 51 (2008) 355–361. https://doi.org/10.1016/j.jacc.2007.11.011.

[6] U.J. Jensen, J. Jensen, G. Olivecrona, G. Ahlberg, B. Lagerquist, P. Tornvall, The role of a simulator-based course in coronary angiography on performance in real life cath lab, BMC Med. Educ. 14 (2014) 49. https://doi.org/10.1186/1472-6920-14-49.

[7] S.F. Iftikhar, M.A. Bishop, P. Hu, Complex Coronary Artery Lesions, in: StatPearls, StatPearls Publishing, Treasure Island (FL), 2024. http://www.ncbi.nlm.nih.gov/books/NBK539899/ (accessed March 27, 2024).

[8] L. Yao, F. Shi, S. Wang, X. Zhang, Z. Xue, X. Cao, Y. Zhan, L. Chen, Y. Chen, B. Song, Q. Wang, D. Shen, TaG-Net: Topology-Aware Graph Network for Centerline-Based Vessel Labeling, IEEE Trans. Med. Imaging 42 (2023) 3155–3166. https://doi.org/10.1109/TMI.2023.3240825.

[9] Z. Zhang, Z. Zhao, D. Wang, S. Zhao, Y. Liu, J. Liu, L. Wang, Topology-Preserving Automatic Labeling of Coronary Arteries via Anatomy-aware Connection Classifier, (2023). https://doi.org/10.48550/arXiv.2307.11959.

[10] J. Yan, Y. Li, W. Liu, H. Zha, X. Yang, S.M. Chu, Graduated Consistency-Regularized Optimization for Multi-graph Matching, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Comput. Vis. – ECCV 2014, Springer International Publishing, Cham, 2014: pp. 407–422. https://doi.org/10.1007/978-3-319-10590-1\_27.