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 Do not cite large groups of papers without individually commenting on them.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 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**

Note: In order to effectively convey your recommendations for improvement to the author(s), and help editors make well-informed and efficient decisions, we ask you to answer the following specific questions about the manuscript and provide additional suggestions where appropriate.

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.

Reviewer #3: Yes, after the revision, the objectives are clear to me.

A: Thank you to both reviewers for your positive feedback.

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:

Reviewer #3: Mark as appropriate with an X:

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

Provide further comments here:

The manuscript describes the method in some detail and provides the source code in the reviewer report. However, the authors seem to provide only training code without testing code. If possible, please provide the testing code too.

A: Thanks for your inquires. In our GitHub repository, we have updated the workflow of the proposed model, adding both of the training and test code. Please see our implementation for model training from lines X to X in <https://github.com/MIILab-MTU/MGM_ICA/blob/main/train_artery_mgm.py> and implementation for model testing from lines X to X in the same python script.

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 [X] No [] N/A []

Provide further comments here:

Reviewer #3: Mark as appropriate with an X:

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

Provide further comments here:

The authors provide Student-t test as suggested. However, it lacks details about the calculation procedure of the p-value. I suggest the authors clarify this.

A: Thank you for pointing out this issue. We investigated the difference between the Student's t-test and the paired t-test, and have revised the corresponding section accordingly. Please see our response to your second question below.

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: See my Comments to Author.

Reviewer #3: The authors provide a thorough and well-reasoned argument, validated by extensive data sets, including publicly available data. I have no further comments.

A: Thank you to both reviewers for your positive feedback.

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: See my Comments to Author.

Reviewer #3: The authors clearly highlight their contributions to the proposed method, especially in comparison to pixel-based methods.

A: Thanks for your suggestions. In the revised manuscript, we further illustrate the advantage of the segment-classification based method compared to the pixel-based method for coronary artery semantic labeling.

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: See my Comments to Author.

Reviewer #3: Yes, they clearly stated the limitations of their study.

A: Thank you to both reviewers for your positive feedback.

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 #3: The paper is well written and clear as it is.

A: Thank you to both reviewers for your positive feedback.

9. Could the manuscript benefit from language editing?

Reviewer #1: No

Reviewer #3: No

A: Thank you to both reviewers for your positive feedback.

AE: The authors have considerably reviewed and improved the manuscript, which is noted by the reviewers. However, as highlighted, the paper still lacks a thorough comparison with more recent ICA segmentation methods, which would make a proper actualization and contextualization of this work within the literature.

A: Thank you for your thoughtful questions. In the previous version of the manuscript, we compared the proposed MGM method with existing approaches, two of which were our prior publications: AGMN [1] and EGAMN [2]. As in response to the second question raised by Reviewer #1, we conducted an updated literature review and observed that studies focusing on coronary artery semantic labeling remain limited. This is primarily due to the inherent challenges of differentiating 3D arterial branches from 2D ICA images..

In clinical practice, Coronary CT Angiography (CCTA) is generally recommended for patients with stable chest pain and intermediate risk of coronary artery disease (CAD), as it is non-invasive and provides detailed anatomical visualization [3]. In contrast, Invasive Coronary Angiography (ICA)—the imaging modality used in our study—is employed when CCTA results are inconclusive or when there is high clinical suspicion of severe CAD, as ICA enables both direct visualization and potential intervention [4].

Most existing publications focus on coronary artery semantic labeling using CCTA. However, as highlighted in our introduction: “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 due to vessel overlap and view-dependent anatomical variations.”

Thus, we narrowed the scope of our literature review to ICA-based coronary artery semantic labeling studies published within the past three years. Using the keyword ‘coronary artery semantic segmentation’, we identified the following most relevant works:

1. Zhang et al. [5] presented proposed the Progressive Perception Learning framework to segment the three main coronary arteries (LAD, LCX, RCA) in ICA. While their approach effectively extracted multiscale features, it excluded side branches (e.g., OM, LMA, D), which are essential for revascularization planning [6].
2. Park et al. [7] proposed an ensemble of encoder-decoder CNN models for major vessel segmentation in ICA. The method targeted LAD, LCX, and RCA, but did not address side branches or the LMA.

We further expanded our search criteria to include relevant works from 2020 onward, identifying the following additional studies:

1. Zhao et al. (2022) [8] (our pilot study) applied radiomics and SVM to classify LMA, LCX, LAD, OM, and D branches using ICA. However, it lacked consideration of graph connectivity and view angle information, resulting in a limited accuracy of 0.7033. We excluded this work from direct comparison due to methodological differences (traditional ML vs. deep learning) and a significant performance gap.
2. Jun et al. (2020) introduced T-Net for main vessel segmentation in ICA; this work is cited in our Related Work section [9].
3. Xian et al. (2020) [10] employed a U-Net-based model for main branch segmentation (LCX, LAD, RCA) in ICA.
4. Zhang et al. (2021) [11] presented a dual-branch multi-scale attention network (DMAN), achieving strong performance in main branch segmentation, including LCX, LAD, and RCA, with an average Dice coefficient of 0.9160.

Based on this comprehensive review, we found no existing studies addressing semantic labeling of both main and side branches in ICA images. The proposed MGM model, with its released code on GitHub, contributes a novel direction in this field. In the revised manuscript, we have also included a comparison between MGM and our additional prior work, HAGMN-UQ [12]. Results demonstrate that leveraging multiple graphs simultaneously yields significant improvements in graph matching performance.

**Reviewer #1**: Thank you for your comprehensive response and the authors' diligent efforts in addressing the majority of my initial concerns. While the revisions have significantly improved the manuscript, two critical issues remain that require further clarification:

1. Limited Dataset Validation

The current experimental validation appears restricted to proprietary in-house datasets. To strengthen the methodological credibility and ensure broader applicability, I strongly recommend demonstrating the MGM framework's performance on established, widely recognized public datasets. Comparative analysis using standardized benchmarks would substantially enhance the reproducibility and comparative value of the research findings.

A: Thank you very much for your thoughtful feedback. We fully agree that evaluation on publicly available datasets is crucial for ensuring methodological transparency and reproducibility. In response, we conducted an extensive search for publicly available ICA datasets related to coronary artery analysis. Our findings are summarized below:

1. ARCADE dataset (https://zenodo.org/records/8386059 ). The Automatic Region-based Coronary Artery Disease diagnostics using X-ray angiography imagEs (ARCADE) dataset contains 1,200 annotated coronary arterial trees, including stenotic regions. However, semantic segmentation results or vessel-level labels are not provided.
2. CADICA dataset [13]. This dataset consists of patient video recordings and associated disease-related metadata. Unfortunately, it does not include binary arterial masks or semantic labels necessary for vessel-level analysis.

To the best of our knowledge, no publicly available datasets currently offer the specific annotations required for evaluating coronary artery semantic labeling, particularly in the context of our proposed MGM framework. As a result, we relied on proprietary in-house datasets for this study. Additionally, our recent survey paper on machine learning applications in ICA [14] did not identify any public datasets suitable for semantic labeling of coronary arteries.

We recognize the importance of public benchmarks and are actively seeking opportunities to contribute to or collaborate on the development of such datasets in the future. We hope that our work can serve as a step in that direction. To support reproducibility, we have illustrated the complete data processing pipeline in Section 3.1 of the manuscript and have provided a sample processed ICA image in our GitHub repository. We believe this will assist readers in replicating the preprocessing steps and facilitate future validation efforts.

It should be noted that the dataset used in this study is integrated into a software product. As part of our commercial agreement with a catheter manufacturer, our ICA dataset and its annotations are subject to contractual confidentiality and cannot be shared in the public domain.

2. Insufficient Benchmarking Context

The comparative analysis primarily references methodologies from [10], [11] (developed by the authors' team) alongside older approaches [28], [34]. To properly position MGM within the current research landscape, it is essential to include comparisons with contemporary state-of-the-art methods published within the last 2-3 years. This expansion would more effectively demonstrate the method's competitive advantages and novel contributions to the field.

A: Thanks for pointing out this issue.

For our own proposed methods, beside AGMN in [10] and EAGMN [11], we further expanded our research and developed the Hypergraph Associated Graph Matching Network with Uncertainty Quantification (HAGMN-UQ) [12] for coronary artery semantic labeling. HAGMN-UQ leverages hypergraphs, which not only extends representation capabilities beyond pairwise relationships, but also improves the robustness and accuracy of the graph matching by enabling the modeling of higher-order associations. HAGMN-UQ achieved an accuracy of 0.9211, a precision of 0.9255, a recall of 0.9257 and a macro average F1 score of 0.9256 using the same datasets. The presented study, MGM in this paper, further improved the performance, as the MGM achieved accuracy, precision, recall and macro F1 of 0.9471, 9506, 0.9506, and 0.9506, respectively. We added HAGMN-UQ into the revised manuscript to further illustrate the superior performance of the presented MGM method.

In clinical practice, CCTA (Coronary CT Angiography) is recommended for patients with stable chest pain and intermediate risk of coronary artery disease (CAD) because it's non-invasive and provides detailed anatomical visualization of coronary arteries [3]. On the contrary, Invasive Coronary Angiography (ICA), which is the imaging modality used in this study, is employed when CCTA results are inconclusive or when there's high suspicion of severe CAD, as it allows direct visualization and potential immediate intervention [4].

Most of the publications are related to the coronary artery semantic labeling using CCTA. However, as mentioned in the introduction, 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. Due to the imaging acquisition scanner and the characteristic of the ICA, the coronary artery semantic labeling is far more challenging using ICA compared to using CCTA in

1. 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.
2. Additionally, difficulties in excluding non-vessel areas arise due to their similar appearance to vascular segments, such as catheter outlines, spines, and ribs, complicating vessel segmentation and stenosis severity assessment.
3. 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 steps, adversely affecting vessel boundary delineation. This ambiguity is particularly critical in the limited number of pixels near stenosis areas, impacting the accuracy of vessel identification and stenosis severity assessment.
4. Furthermore, due to patient specific anatomy, variations of projection angles, and contrast dye degradation, recognizing the type of individual coronary arteries remains challenging.

Thus, we narrowed the scope of our literature review to ICA-based coronary artery semantic labeling studies published within the past three years. Using the keyword ‘coronary artery semantic segmentation’, we identified the following most relevant works:

1. Zhang et al. [5] presented proposed the Progressive Perception Learning framework to segment the three main coronary arteries (LAD, LCX, RCA) in ICA. While their approach effectively extracted multiscale features, it excluded side branches (e.g., OM, LMA, D), which are essential for revascularization planning [6].
2. Park et al. [7] proposed developed an ensemble of encoder-decoder CNN models for major vessel segmentation in ICA. The method targeted LAD, LCX, and RCA, but did not address side branches or the LMA.

We further expanded our search criteria to include relevant works from 2020 onward, identifying the following additional studies:

1. Zhao et al. (2022) [8] (our pilot study) applied radiomics and SVM to classify LMA, LCX, LAD, OM, and D branches using ICA. However, it lacked consideration of graph connectivity and view angle information, resulting in a limited accuracy of 0.7033. We excluded this work from direct comparison due to methodological differences (traditional ML vs. deep learning) and a significant performance gap.
2. Jun et al. (2020) introduced T-Net for main vessel segmentation in ICA; this work is cited in our Related Work section [9].
3. Xian et al. (2020) [10] employed a U-Net-based model for main branch segmentation (LCX, LAD, RCA) in ICA..
4. Zhang et al. (2021) [11] presented a dual-branch multi-scale attention network (DMAN), achieving strong performance in main branch segmentation, with an average Dice coefficient of 0.9160.

Based on this comprehensive review, we found no existing studies addressing semantic labeling of both main and side branches in ICA images. The proposed MGM model, with its released code on GitHub, contributes a novel direction in this field. In the revised manuscript, we have also included a comparison between MGM and our additional prior work, HAGMN-UQ [12]. Results demonstrate that leveraging multiple graphs simultaneously yields significant improvements in graph matching performance.

(Section 4.4 Comparison with Competing Methods)

Our MGM approach is compared with four deep learning-based methods for coronary artery semantic labeling:

* Association Graph-based Graph Matching Network (AGMN) [1], utilizing the association graph for two-graph matching;
* Edge Attention Graph Matching Network (EAGMN) [2], an extension of AGMN with added edge attention;
* Hyper Association Graph Matching Network with Uncertainty Quantification (HAGMN-UQ) [12], which converts the ICA-derived graph into a hyper association graph for coronary artery matching and employs uncertainty quantification to measure the trustworthiness of the graph matching results.
* Neural Graph Matching (NGM) [15], using the association graph-induced affinity matrix for two-graph matching;
* Bidirectional Tree-LSTM (BiTL) [16], employing tree-structured bidirectional LSTM for hierarchical feature extraction along with the vascular tree.

**Table 5**. Achieved performance for coronary artery semantic labeling using MGM. The bold texts indicate the best performance achieved in the corresponding metric and types of arterial segments.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Metric | LMA | LAD | LCX | D | OM | Macro Avg |
| BiTL | ACC | 0.6000±0.4899 | 0.9385±0.0082 | 0.6770±0.2134 | 0.7395±0.3706 | 0.5324±0.3152 | 0.7291±0.0728 |
| AGMN | 0.9907±0.0031 | 0.8730±0.0440 | 0.8646±0.0274 | 0.8320±0.0391 | 0.8080±0.0411 | 0.8639±0.0182 |
| EAGMN | 0.9942±0.0064 | 0.8931±0.0295 | 0.8843±0.0393 | 0.8518±0.0373 | 0.7968±0.0431 | 0.8767±0.0188 |
| NGM | 0.9885±0.0122 | 0.9223±0.0341 | 0.8967±0.0324 | 0.9007±0.0466 | 0.8408±0.0480 | 0.9039±0.0354 |
| HAGMN-UQ | 0.9767±0.0123 | 0.9382±0.0064 | 0.9041±0.0213 | 0.9239±0.0115 | 0.8364±0.0283 | 0.9211±0.0117 |
| MGM | **0.9962±0.0047** | **0.9618±0.0094** | **0.9396±0.0244** | **0.9459±0.0126** | **0.9096±0.0357** | **0.9471±0.0175** |
| BiTL | PREC | 0.6000±0.4899 | 0.8911±0.0661 | 0.6244±0.0906 | 0.6047±0.3027 | 0.5739±0.2876 | 0.6588±0.1504 |
| AGMN | 0.9915±0.0045 | 0.8779±0.0212 | 0.8689±0.0183 | 0.8219±0.0225 | 0.8040±0.0257 | 0.8728±0.0161 |
| EAGMN | 0.9923±0.0094 | 0.8973±0.0221 | 0.8759±0.0179 | 0.8447±0.0300 | 0.8105±0.0276 | 0.8841±0.0166 |
| NGM | 0.9893±0.0120 | 0.9207±0.0332 | 0.8962±0.0327 | 0.8912±0.0488 | 0.8530±0.0445 | 0.9101±0.0337 |
| HAGMN-UQ | 0.9867±0.0099 | 0.9400±0.0151 | 0.9126±0.0147 | 0.9134±0.0194 | 0.8746±0.0174 | 0.9255±0.0112 |
| MGM | **0.9962±0.0047** | **0.9612±0.0094** | **0.9402±0.0244** | **0.9449±0.0127** | **0.9105±0.0356** | **0.9506±0.0163** |
| BiTL | REC | 0.6000±0.4899 | 0.8574±0.1210 | 0.6780±0.1679 | 0.5115±0.2560 | 0.6991±0.3759 | 0.6692±0.1795 |
| AGMN | 0.9923±0.0097 | 0.8847±0.0188 | 0.8739±0.0224 | 0.8158±0.0501 | 0.8015±0.0301 | 0.8736±0.0163 |
| EAGMN | 0.9904±0.0125 | 0.9017±0.0192 | 0.8690±0.0208 | 0.8386±0.0332 | 0.8284±0.0520 | 0.8856±0.0176 |
| NGM | 0.9901±0.0119 | 0.9192±0.0329 | 0.8957±0.0330 | 0.8820±0.0514 | 0.8657±0.0418 | 0.9105±0.0337 |
| HAGMN-UQ | 0.9886±0.0112 | 0.9436±0.0129 | 0.9083±0.0149 | 0.9168±0.0181 | 0.8709±0.0153 | 0.9257±0.0111 |
| MGM | **0.9962±0.0047** | **0.9605±0.0096** | **0.9409±0.0243** | **0.9440±0.0129** | **0.9115±0.0354** | **0.9506±0.0163** |
| BiTL | F1 | 0.6000±0.4899 | 0.9385±0.0082 | 0.6770±0.2134 | 0.7395±0.3706 | 0.5324±0.3152 | 0.6975±0.1160 |
| AGMN | 0.9907±0.0031 | 0.8730±0.0440 | 0.8646±0.0274 | 0.8320±0.0391 | 0.8080±0.0411 | 0.8737±0.0150 |
| EAGMN | 0.9942±0.0064 | 0.8931±0.0295 | 0.8843±0.0393 | 0.8518±0.0373 | 0.7968±0.0431 | 0.8840±0.0170 |
| NGM | 0.9885±0.0122 | 0.9223±0.0341 | 0.8967±0.0324 | 0.9007±0.0466 | 0.8408±0.0480 | 0.9098±0.0337 |
| HAGMN-UQ | 0.9877±0.0104 | 0.9418±0.0139 | 0.9105±0.0147 | 0.9151±0.0187 | 0.8728±0.0163 | 0.9256±0.0111 |
| MGM | **0.9962±0.0047** | **0.9618±0.0094** | **0.9396±0.0244** | **0.9459±0.0126** | **0.9096±0.0357** | **0.9506±0.0163** |

We further replaced Figure 5 for the robustness test, and it now reads:

(Section 4.6 Robustness test)

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 5**. Achieved ACC, PREC, REC, and F1 of the proposed MGM, AGMN, EAGMN, NGM, IPCA and HAGMN-UQ 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, IPCA and HAGMN-UQ) 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.

**Reviewer #3**: After the revision, the authors have addressed all the reviewers' comments in a correct way. I have no major comments after reading the contents. However, there are some minor comments to improve the manuscript.

1. I suggest the authors to do cosmetic revisions about the Pseudo-code, Figure and Tables to align the width of the paper (e.g., Table 4).

A: Thank you for your suggestions. We have adjusted the size of the tables and figures in the revised manuscript to strictly follow the template provided by Pattern Recognition. Specifically, we rearranged Tables 4, 5, and 9 to fit within the left and right margins. We also resized all figures and algorithm descriptions in the manuscript to ensure they fall within the same margins.

2. In this study, you considered Student t-test. However, since all the methods are performed on the same test set, why not considering paired t-test? Could you please clarity the reasons?

A: Thanks for pointing out the issue. We further investigate the difference between student t-test and paired t-test. Below is the comparison between the student t-test and paired t-test.

Student's t-test (Independent t-test)

* Used when: You're comparing the means of two independent groups.
* Example: Comparing the test scores of two different classes of students.
* Assumption: The two groups are not related and the variances are assumed equal (or adjusted if unequal).

Paired t-test (Dependent t-test)

* Used when: You're comparing the means of two related (paired) groups.
* Example: Measuring blood pressure before and after treatment on the same patients.
* Assumption: The differences between the paired observations are normally distributed.

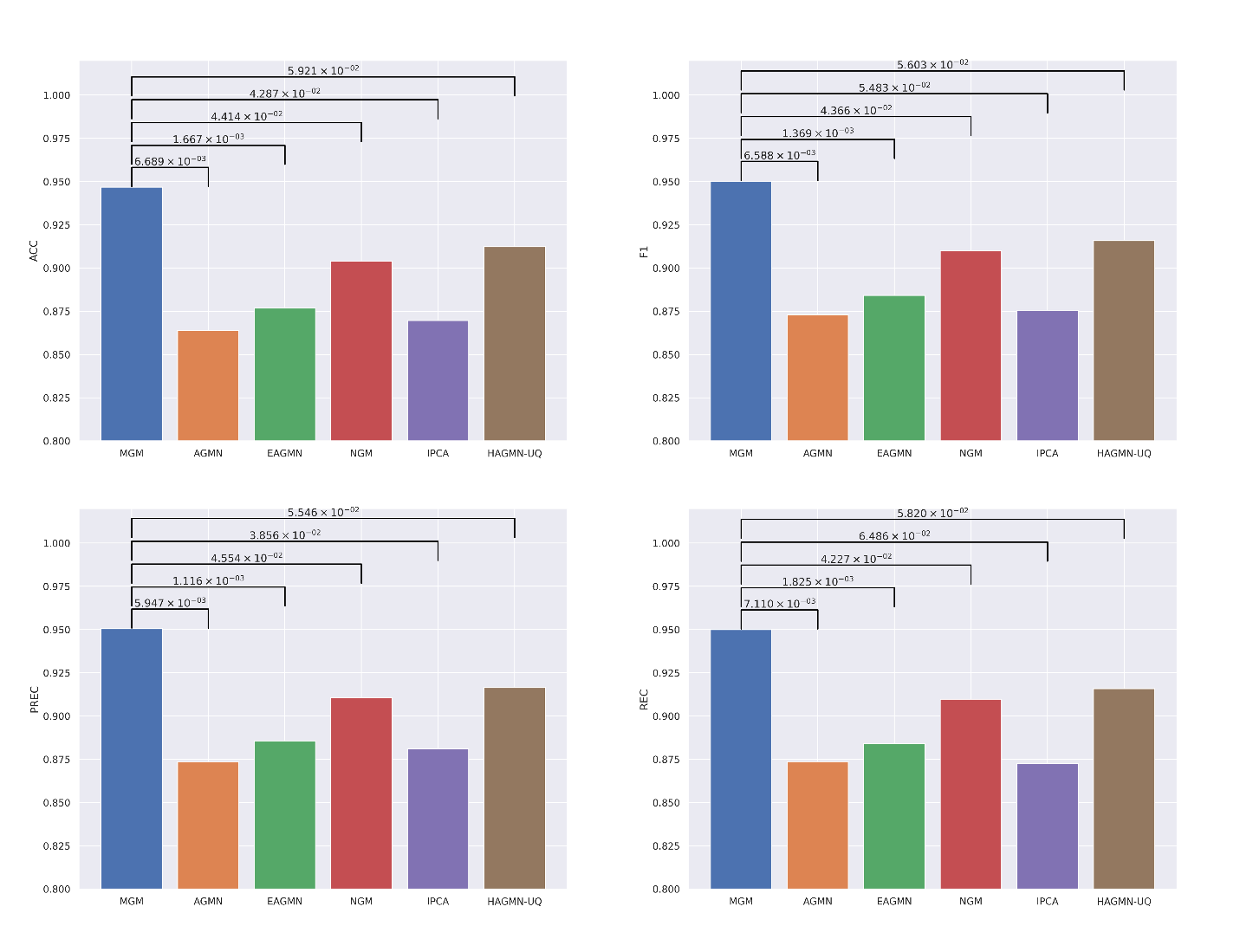
Since each data point contains two related measurements—one from each method—and both methods were evaluated on the same test splits or folds, a paired t-test is the more appropriate choice. This approach accounts for data variability and focuses on the actual performance differences between the methods. Accordingly, in the revised manuscript, we have adopted the paired t-test to assess the statistical significance of the performance improvements achieved by the proposed MGM model over existing baseline methods. For implementation, we utilized the ttest\_rel function from the scipy.stats library.

The comparison and analysis now read:

(Section 4.4 Comparison with Competing Methods)

Furthermore, we conducted a paired t-test to assess whether the performance differences presented in Table 5 are statistically significant. The null hypothesis assumes no significant difference in performance between the two models, while the alternative hypothesis posits 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.

And the figures to show the statistical significance using paired t-test in Figure S1 now reads:



**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.

3. Please clarify the specific configurations of the test metrics you are considering. For example, F1 score, is it macro F1 or weighted F1? The lack of test code made it hard for me to guess.

A: Thanks for your question. In our implementation, we evaluated the proposed MGM using macro F1. We updated our GitHub repository (https://github.com/MIILab-MTU/MGM\_ICA). For the main program shown in train\_artery\_mgm.py, we added a command line parameter ‘procedure’ to control the running flow. If it was set as ‘test\_one\_sample’, then it will show the demo of the MGM using 3 graphs per graph matching and output the matching results; if it was set as ‘test’, it goes to the entire testing procedure.

For the model evaluation, we employed the scikit-learn library to evaluate the artery semantic labeling performance. The evaluation code is shown in utils.py from lines 134 to 190.

Overall, I suggest the authors to address those comments, and I suggest a minor revision for this manuscript.

A: Thank you for your consideration. We hope that the revised manuscript has addressed all of your concerns.

**References**

[1] C. Zhao, Z. Xu, J. Jiang, M. Esposito, D. Pienta, G.-U. Hung, W. Zhou, AGMN: Association graph-based graph matching network for coronary artery semantic labeling on invasive coronary angiograms, Pattern Recognit. 143 (2023) 109789. https://doi.org/10.1016/j.patcog.2023.109789.

[2] C. Zhao, Z. Xu, G.-U. Hung, W. Zhou, EAGMN: Coronary artery semantic labeling using edge attention graph matching network, Comput. Biol. Med. 166 (2023) 107469. https://doi.org/10.1016/j.compbiomed.2023.107469.

[3] K. Yoshida, Y. Tanabe, T. Hosokawa, T. Morikawa, N. Fukuyama, Y. Kobayashi, T. Kouchi, N. Kawaguchi, M. Matsuda, T. Kido, T. Kido, Coronary computed tomography angiography for clinical practice, Jpn. J. Radiol. 42 (2024) 555–580. https://doi.org/10.1007/s11604-024-01543-1.

[4] Z. Li, Y. Zhang, G. Liu, H. Shao, W. Li, X. Tang, A robust coronary artery identification and centerline extraction method in angiographies, Biomed. Signal Process. Control 16 (2015) 1–8.

[5] 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.

[6] H.Y. Kim, J.-H. Doh, H.-S. Lim, C.-W. Nam, E.-S. Shin, B.-K. Koo, J.M. Lee, T.K. Park, J.H. Yang, Y.B. Song, J.-Y. Hahn, S.H. Choi, H.-C. Gwon, S.-H. Lee, S.M. Kim, Y. Choe, J.-H. Choi, Identification of Coronary Artery Side Branch Supplying Myocardial Mass That May Benefit From Revascularization, JACC Cardiovasc. Interv. 10 (2017) 571–581. https://doi.org/10.1016/j.jcin.2016.11.033.

[7] J. Park, J. Kweon, Y.I. Kim, I. Back, J. Chae, J. Roh, D. Kang, P.H. Lee, J. Ahn, S. Kang, D. Park, S. Lee, C.W. Lee, S. Park, S. Park, Y. Kim, Selective ensemble methods for deep learning segmentation of major vessels in invasive coronary angiography, Med. Phys. 50 (2023) 7822–7839. https://doi.org/10.1002/mp.16554.

[8] C. Zhao, R. Bober, H. Tang, J. Tang, M. Dong, C. Zhang, Z. He, M. Esposito, Z. Xu, W. Zhou, Semantic Segmentation to Extract Coronary Arteries in Invasive Coronary Angiograms, J. Adv. Appl. Comput. Math. 9 (2022) 76–85. https://doi.org/10.15377/2409-5761.2022.09.6.

[9] 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.

[10] 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.

[11] H. Zhang, D. Zhang, Z. Gao, H. Zhang, Joint Segmentation and Quantification of Main Coronary Vessels Using Dual-Branch Multi-scale Attention Network, in: M. De Bruijne, P.C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Med. Image Comput. Comput. Assist. Interv. – MICCAI 2021, Springer International Publishing, Cham, 2021: pp. 369–378. https://doi.org/10.1007/978-3-030-87193-2\_35.

[12] 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. 99 (2025) 103374. https://doi.org/10.1016/j.media.2024.103374.

[13] A. Jiménez‐Partinen, M.A. Molina‐Cabello, K. Thurnhofer‐Hemsi, E.J. Palomo, J. Rodríguez‐Capitán, A.I. Molina‐Ramos, M. Jiménez‐Navarro, CADICA : A new dataset for coronary artery disease detection by using invasive coronary angiography, Expert Syst. 41 (2024) e13708. https://doi.org/10.1111/exsy.13708.

[14] P. Baral, C. Zhao, M. Esposito, W. Zhou, A Review on Machine Learning for Arterial Extraction and Quantitative Assessment on Invasive Coronary Angiograms, Curr. Cardiovasc. Imaging Rep. (2024). https://doi.org/10.1007/s12410-024-09596-6.

[15] R. Wang, J. Yan, X. Yang, Neural Graph Matching Network: Learning Lawler’s Quadratic Assignment Problem With Extension to Hypergraph and Multiple-Graph Matching, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 5261–5279. https://doi.org/10.1109/TPAMI.2021.3078053.

[16] Q. Cao, A. Broersen, M.A. de Graaf, P.H. Kitslaar, G. Yang, A.J. Scholte, B.P.F. Lelieveldt, J.H.C. Reiber, J. Dijkstra, Automatic identification of coronary tree anatomy in coronary computed tomography angiography, Int. J. Cardiovasc. Imaging 33 (2017) 1809–1819. https://doi.org/10.1007/s10554-017-1169-0.