Anatomy-Grounded Synthetic Coronary Angiography for Geometry-Informed Multi-View Matching

In Kyu Lee[1,2⋆†(✉)], Sumin Seo[1,3,4⋆†], and Jaesik Min[1]

[1] Medipixel, Inc., Seoul, Republic of Korea
[2] University of California San Diego, La Jolla, USA
[3] Computational Health Center & Helmholtz AI, Helmholtz Munich, Neuherberg, Germany
[4] Munich Center for Machine Learning (MCML), Munich, Germany
kl002@ucsd.edu; {sumin.seo,jaesik.min}@medipixel.io

Abstract. Accurate correspondence matching across multiple angiographic views is the prerequisite for 3D coronary reconstruction and interventional guidance. However, the development of robust deep learning models for this task has been stifled by a fundamental data bottleneck. Obtaining ground truth for matching tasks in angiography pairs is prohibitively expensive and hard to scale. To overcome this barrier, we introduce a physically-grounded data generation framework that synthesizes high-fidelity Digital Reconstructed Radiographs (DRRs) from 3D Coronary CT Angiography (CCTA) volumes. Our framework generates dense, highly accurate 3D-to-2D correspondence labels by simulating realistic C-arm acquisition geometry on patient anatomy at zero human cost. Leveraging this dense supervision, we propose a Geometry-Informed Matching Module (GIMM) that integrates global feature and anatomical structure into correspondence learning. Unlike real angiography where assessment relies on subjective human annotation, our dataset provides 2D correspondence labels with paired images, allowing human-free evaluation. We comprehensively evaluate our method on the proposed CT-derived DRR dataset and demonstrate improvements over other matching baseline models. As the first public benchmark dataset for coronary correspondence matching, we will release our dataset and code in http://github.com/medipixel/GIMM.

Keywords: Local feature matching · Synthetic data generation · CT · Coronary angiography.

1 Introduction

Multi-view coronary angiography is routinely used for diagnosis and interventional planning in coronary artery disease. Accurate cross-view correspondence

⋆ Equal contribution. † Work performed while affiliated with Medipixel, Inc.

is a prerequisite for 3D reconstruction, quantification, and geometry-aware guidance during percutaneous coronary interventions (PCI) [2,9]. While classical geometric reconstruction pipelines rely on handcrafted features and epipolar constraints, recent advances in deep learning matching models have demonstrated strong performance in natural image domains [1,11].

However, transferring these advances to medical imaging remains difficult. Unlike natural images, X-ray projections suffer from overlapping anatomical structures and an inherent lack of 3D ground truth. Traditional feature correspondence estimation, which is fundamental to visual localization and pose estimation [7,8,11,12,16], relies heavily on appearance similarity rather than complex projective geometry. For example, SuperGlue [11] introduced a graph neural network-based matching framework that refines sparse keypoint correspondences through attention-based context aggregation. Subsequently, LoFTR [14] eliminated the need for explicit keypoint detection by adopting a detector-free, coarse-to-fine Transformer architecture for dense feature matching.

While effective in natural domain, they struggle in coronary angiography due to repetitive tubular structures, severe viewpoint variations, and projection-induced ambiguities, making dense correspondence estimation particularly challenging due to an inherent lack of 3D ground truth.

Establishing geometry-grounded, pixel-level correspondences requires expert annotation, which is expensive and yields only sparse labeling. Moreover, due to the projective nature of X-ray imaging, a 2D point corresponds to a 3D ray, making ground-truth verification inherently ambiguous. To address this limitation, we propose a physically grounded data generation framework that synthesizes Digital Reconstructed Radiographs (DRRs) [13,17] directly from patient-specific 3D Coronary CT Angiography (CCTA). By simulating clinically realistic C-arm acquisition geometries, our framework produces multi-view angiographic projections that preserve complex coronary anatomy. Crucially, this approach generates data from actual patient anatomy and includes precise 3D-to-2D correspondence labels, entirely eliminating the need for manual annotation.

Leveraging this dense anatomical supervision, we introduce a Geometry-Informed Matching Module (GIMM), which integrates geometric priors to resolve angiographic ambiguities. Our module improves matching in a coarse-to-fine manner: coarsely by incorporating global view information via C-arm angles to orient the network macroscopically, and finely by utilizing epipolar information to constrain and guide localized pixel-level matches. This dual integration of global structure and anatomical consistency significantly improves robustness across challenging view configurations.

Unlike real angiography, where evaluation relies on subjective human annotation [6], our dataset provides complete 3D ground truth, enabling objective and automated assessment of correspondence quality. We comprehensively evaluate our approach against other deep learning based matching models using both 2D and 3D distance errors. This work establishes the first public benchmark dataset for coronary correspondence matching.

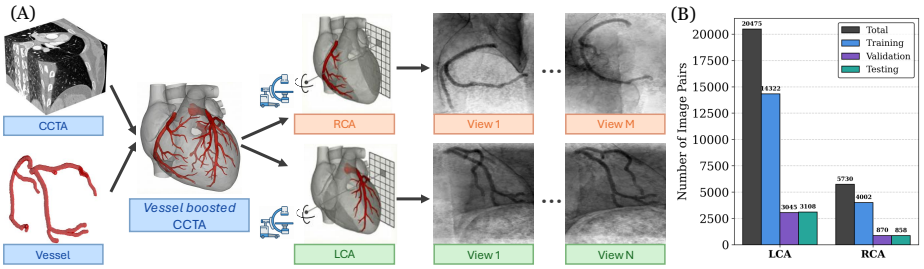

Fig. 1: (A) Overview of the CT DRR pipeline. (B) CT DRR data distribution.

In summary, this work makes three core contributions: (1) A physically grounded CT-derived DRR dataset with dense 3D-2D correspondence labels; (2) A geometry-informed matching module that leverages C-arm angle information and epipolar constraints for robust coarse-to-fine correspondence learning; (3) A comprehensive evaluation protocol enabling human-free correspondence assessment.

2 Method

2.1 CT-Derived DRR Dataset

To enable scalable and anatomically consistent supervision for multi-view correspondence learning, we construct a CT DRR dataset from CCTA with vessel segmentation masks, as illustrated in Fig. 1 (A). Given a CCTA volume $V(x)$, DRRs are generated using a forward projection model that simulates C-arm X-ray acquisition. For each detector pixel, intensity is computed by integrating attenuation coefficients along each X-ray ray according to Beer-Lambert law [5]: $I(u) = I_0 \exp(-\int_{R(u)} \mu(x)dx)$, where $R(u)$ denotes the ray corresponding to detector coordinate u, and $\mu(x)$ is the attenuation coefficient at location x.

To approximate contrast enhanced clinical angiography, we modulate the attenuation coefficients using anatomical masks. The coronary tree is separated into isolated right coronary artery (RCA) and left coronary artery (LCA) volumes. By amplifying the attenuation values within these masks and identifying bone structures via intensity thresholding, we realistically simulate contrast agent injection and radiographic anatomy. To reflect standard diagnostic protocols, we simulate clinical C-arm projection angles (LAO/RAO and cranial/caudal). We generate 4 routine views for RCA and 7 for LCA per patient scan. Multi-view pairs are then constructed from all unordered combinations within each isolated system, yielding 6 RCA pairs and 21 LCA pairs per scan.

2.2 Geometry-Informed Matching Module

To leverage the dense anatomical supervision of our dataset, we propose a Geometry-Informed Matching Module (GIMM) integrated into a coarse-to-fine

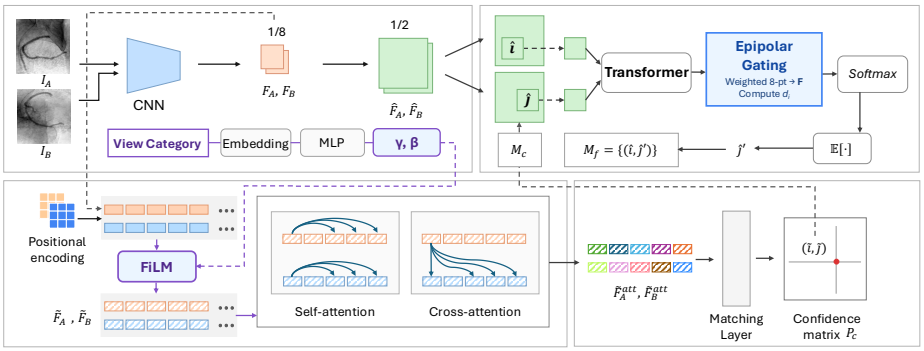

Fig. 2: Overall architecture of Geometry-Informed Matching Module (GIMM).

correspondence matching framework, as depicted in Fig. 2. Throughout the paper, F_k, $\hat{F}_k$, and $\tilde{F}_k$ denote the coarse-level, fine-level, and class-modulated features of image k, respectively. We denote the predicted correspondence pair by $(\hat{i}, \hat{j}')$, and the coarse and fine correspondence sets by $\mathcal{M}_c$ and $\mathcal{M}_f$. The coarse stage utilizes a shared ResNet backbone [4] and Transformer [15] to establish globally consistent coarse matches. The fine stage then processes local feature patches around each coarse match and estimates the final sub-pixel correspondence as the expected location over the local matching probability distribution, i.e., $\hat{j}' = \mathbb{E}[j']$, yielding the final fine-level matches $\mathcal{M}_f = (\hat{i}, \hat{j}')$.

Our GIMM explicitly enhances this baseline architecture. At the coarse level, a class conditioning mechanism injects acquisition-view priors into the feature representations to guide view-aware learning. At the fine level, an epipolar gating mechanism restricts matching range to geometrically plausible regions. Together, these modifications promote anatomically and geometrically consistent multi-view matching.

Coarse-to-Fine Matching for Paired Correspondences. Following [14], we adopt a coarse-to-fine matching strategy. In standard matching, the coarse correspondence matrix is built by warping grid points with depth and pose and selecting mutual nearest neighbors. For coronary angiography, reliable depth and pose are unavailable, so we replace the warp-based supervision with point-pair correspondence labels. Specifically, we use the point-pair labels to directly supervise both the coarse-level mapping and the fine-level 2D pixel correspondences. We applied the same modification to all baselines to ensure a fair comparison.

Class-Conditioning Module. Coronary angiographic projections acquired under different C-arm configurations exhibit distinct global anatomical layouts. For example, LAO and RAO views emphasize different vessel segments and overlap patterns. We hypothesize that explicitly incorporating view class information can provide a useful global prior during coarse correspondence establishment.

Let I_k denote the input image from view k, and let $F_k \in \mathbb{R}^{N\times C}$ denote the corresponding coarse-level feature tokens extracted by the backbone, where N is the number of spatial tokens and C is the feature dimension. Each image is associated with a discrete view class label $c_k \in \{1, ..., K\}$, corresponding to clinically defined projection categories.

We encode the view class label using a learnable embedding vector, $e_k \in \mathbb{R}^d$ with d embedding dimension. The embedding is then mapped to feature modulation parameters using two learnable functions, where $\gamma_k, \beta_k \in \mathbb{R}^C$. We adopt a Feature-wise Linear Modulation (FiLM) [10] to inject the global prior into coarse features:

$$\tilde{F}_k = \gamma_k \odot F_k + \beta_k \tag{1}$$

where $\odot$ denotes channel-wise multiplication. This modulation adaptively rescales and shifts feature channels based on the projection class, allowing the model to condition its matching behavior on view-specific anatomical configurations.

The modulated features $\tilde{F}_k$ are subsequently passed to the coarse Transformer module, producing Transformer-processed features $\tilde{F}_k^{\text{att}}$. By integrating acquisition-view information directly into feature representations, the proposed class-conditioning module encourages globally coherent matching across anatomically distinct projections.

Epipolar Geometry Module. During the fine matching phase, we apply epipolar gating to the similarity matrix. Given tentative correspondences $\{x_i, y_i, x_i', y_i'\}_{i=1}^N$ with confidence weights w_i, we estimate a fundamental matrix $\mathbf{F} \in \mathbb{R}^{3\times 3}$ by a weighted 8-point algorithm [3,18]. Each correspondence defines a row $a_i = [x_i'x_i, x_i'y_i, x_i', y_i'x_i, y_i'y_i, y_i', x_i, y_i, 1]$, and stacking all rows yields $\mathbf{A} \in \mathbb{R}^{N\times 9}$. We solve

$$\min_f \|\mathbf{W}\mathbf{A}f\|^2, \quad \mathbf{W} = \text{diag}(\sqrt{w_i}), \tag{2}$$

where $f \in \mathbb{R}^9$ is the column-wise vectorization of the normalized fundamental matrix $\tilde{\mathbf{F}}$. For numerical stability, we normalize points with weighted centroids μ_0, μ_1, and weighted mean distances d_0, d_1. The normalization transforms are

$$\mathbf{T}_0 = \begin{bmatrix} s_0 & 0 & -s_0\mu_0^x \\ 0 & s_0 & -s_0\mu_0^y \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{T}_1 = \begin{bmatrix} s_1 & 0 & -s_1\mu_1^x \\ 0 & s_1 & -s_1\mu_1^y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where $s_0 = \sqrt{2}/d_0, s_1 = \sqrt{2}/d_1$. We compute the weighted solution on normalized points to obtain $\tilde{\mathbf{F}}$, then enforce $\text{rank}(\tilde{\mathbf{F}}) = 2$ by singular value decomposition, and denormalize $\mathbf{F} = \mathbf{T}_1^\top \tilde{\mathbf{F}} \mathbf{T}_0$. The estimated $\mathbf{F}$ is used to compute the symmetric epipolar distance d_i for each predicted correspondence. A binary inlier mask is applied based on a distance threshold δ. The matching confidence is then gated, so that geometrically inconsistent correspondences are explicitly removed rather than softly attenuated. We also consider two relaxed variants in our ablation studies: *soft gating* $\exp(-d_i/\tau^2)$ and *logit gating* $\sigma(-d_i/\tau^2)$.

Table 1: Main result on synthetic CT-derived DRR and real coronary angiography (top-$K = 20$).

Method	Synthetic (2D)			Synthetic (3D)	Real	
	Dist. (px)	Prec@3px	Prec@5px	Dist. (mm)	F1@5px	F1@10px
SuperGlue	7.249	0.425	0.691	4.251	0.151	0.273
ASpanFormer	5.678	0.515	**0.878**	1.478	0.149	0.272
LoFTR	5.394	0.505	0.772	1.395	0.170	0.300
QuadTree	5.102	0.536	0.795	1.238	0.168	0.308
Ours	**4.738**	**0.547**	0.813	**1.171**	**0.175**	**0.332**

Table 2: Ablation study on the proposed components (top-$K = 20$).

Method	Class Cond.	Epi. Gating	2D Dist. (px)
Quadtree			5.102
+ Class Conditioning	✓		5.038
+ Epipolar Gating (Soft)	✓	✓(Soft)	5.001
+ Epipolar Gating (Logit)	✓	✓(Logit)	4.887
Ours	✓	✓(Hard)	**4.738**

3 Experiments

3.1 CT-DRR Dataset

Our synthetic generation pipeline utilizes CCTA scans and vessel annotations from the publicly available ImageCAS dataset [19]. For each view pair, 3D-to-2D projection correspondences are automatically computed using the known 3D vessel geometry and projection matrices, enabling precise and anatomically consistent supervision without manual annotation. The resulting dataset comprises 26,205 multi-view image pairs. To prevent cross-patient information leakage, the data is strictly divided at the patient level (70% train, 15% validation, 15% test), as detailed in Fig. 1 (B).

3.2 Implementation Details

The proposed model was trained for 20 epochs using a batch size of 12. Network optimization was performed using the AdamW optimizer with a learning rate of $8e-3$. To supervise the training process, we employed a focal loss for the coarse-level matching and an L_2 loss for the fine-level localization. For the FiLM layers, we utilized a feature embedding dimension d of 64. Within the epipolar gating module, the epipolar temperature parameter τ was set to 2.0, and matches were filtered using an epipolar distance threshold δ of 2.0 pixels.

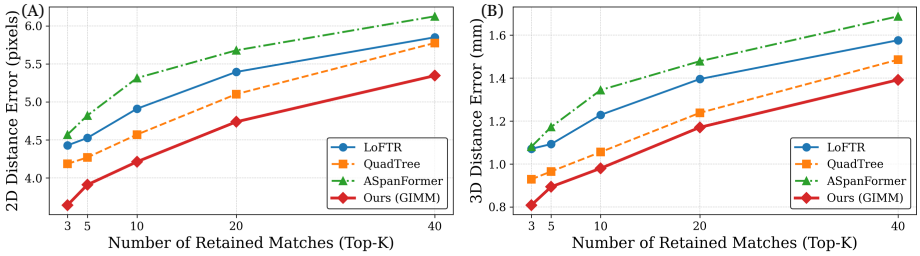

Fig. 3: Comparison of our proposed GIMM against baseline methods at various top-K: (A) mean 2D error and (B) mean 3D distance.

3.3 Results

Distance Metrics. We compute 2D and 3D correspondence errors using ground truth centerlines to strictly penalize anatomical mismatches. For 2D projection error, predictions within the source vessel mask are mapped to the nearest source centerline to calculate the Euclidean pixel distance to the corresponding target centerline. For 3D distance error, the valid matches in both vessel masks are mapped to their nearest 2D centerlines, and the Euclidean distance between their corresponding 3D spatial coordinates is computed.

Quantitative Results. Table 1 summarizes the performance of our method against baseline matching models on both the CT-derived DRR dataset and sparse real angiography validation. For the DRR evaluation, the proposed method achieves the lowest 2D average distance error of 4.738 pixels, outperforming the strongest baseline, QuadTree, and achieves the highest localization precision at a 3-pixel threshold. In 3D, our model achieves the lowest average distance error of 1.171 mm, demonstrating improved anatomical consistency by integrating C-arm view conditioning and epipolar geometry.

To further assess in real coronary angiography, we performed sparse quantitative validation on 150 real angiography image pairs using manually annotated landmarks. Since dense pixel-level correspondence annotation is infeasible in real angiography, this landmark-based evaluation provides a practical but limited estimate of real-data performance. GIMM achieves the highest F1 score at both thresholds, with F1@5px of 0.175 and F1@10px of 0.332. These results suggest that training on DRRs improves transfer to real angiography, although larger-scale real-data validation remains necessary before claims of clinical deployment.

Qualitative Results. Fig. 4 shows qualitative improvements in correspondence matching over SuperGlue and QuadTree. Beyond higher match density, our method preferentially selects anatomically structured points such as vessel center line-aligned regions and bifurcations, resulting in geometrically coherent matching in both LCA and RCA images. Despite the domain gap between syn-

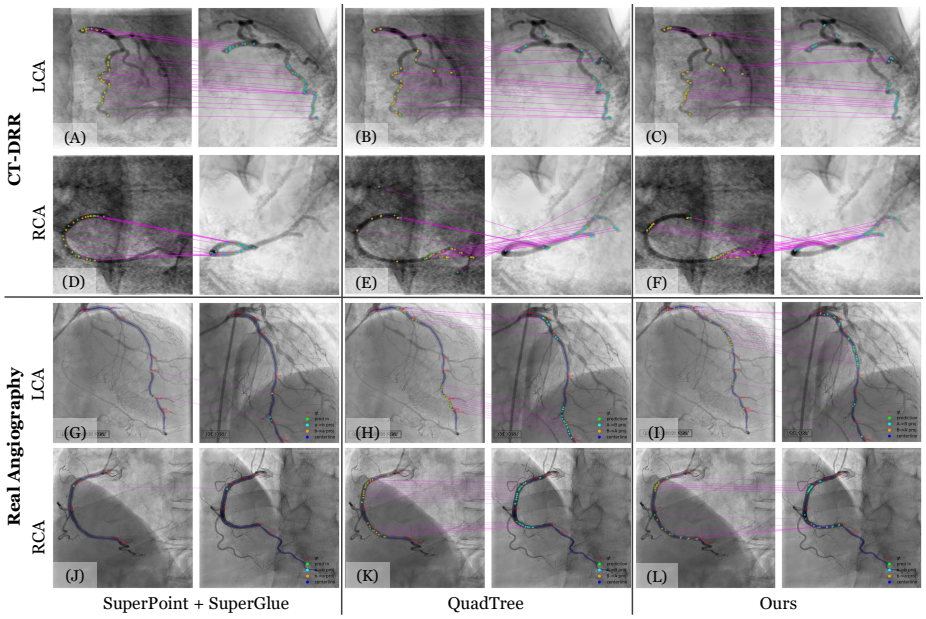

Fig. 4: Qualitative results. Predicted matches are shown in green, with magenta lines indicating paired correspondences. Orange and cyan markers represent right-to-left and left-to-right projections, respectively.

thetic and real angiography, our method maintains coherent, distributed structural alignment and demonstrates robustness to appearance variation.

3.4 Ablation Studies

Robustness Across Confidence Thresholds (Top-K Analysis). Standard recall metrics are inherently ill-posed for continuous tubular structures due to discrete feature grids. To evaluate the trade-off between match quantity and geometric quality, we compare top-K predictions, as shown in Fig. 3. As K increases, models must output lower-confidence matches. GIMM consistently outperforms all baselines across every threshold. This confirms GIMM reliably maintains high geometric accuracy across the entire confidence spectrum, rather than artificially inflating precision by outputting only highly certain predictions.

Effectiveness of Proposed Components. To validate the individual contributions of GIMM, we conduct a step-wise ablation (Table 2). Introducing Class Conditioning yields a reduction in 2D distance error by 0.064. Integrating C-arm projection geometry via Epipolar Gating further improves performance. While applying a soft epipolar mask lowers the error to 5.001px, applying the constraint directly at the logit level sharply improves precision to 4.887px. Finally, integrat-

ing both semantic conditioning and hard epipolar gating yields the best overall 2D error of 4.738px, demonstrating the synergy of these proposed modules.

4 Conclusion

We present an anatomy-grounded CT-derived DRR benchmark for dense correspondence matching in coronary angiography. We further propose GIMM, a view-aware and geometry-informed matching framework that leverages clinical projection priors and epipolar constraints for anatomically consistent multi-view correspondence. Experiments on the synthetic DRR benchmark demonstrate that our method consistently outperforms existing matching baselines in both 2D and 3D metrics. Sparse quantitative validation on real angiography using manually annotated landmarks further suggests improved transfer to clinical images, although dense real-data validation remains infeasible due to the prohibitive cost of pixel-level correspondence annotation. Future work will focus on larger-scale real angiography validation, improved DRR realism, and downstream evaluation in 3D coronary reconstruction.

Acknowledgments. The retrospective collection of clinical angiography data from patients at Soonchunhyang University Cheonan Hospital (Cheonan, Republic of Korea; before 2024) was approved by the Institutional Review Board.

Disclosure of Interests. In Kyu Lee and Sumin Seo performed this work while employed by Medipixel, Inc. and are currently affiliated with UC San Diego and Helmholtz Munich/MCML, respectively. Jaesik Min is a current employee of Medipixel, Inc.

References

1. Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: Roma: Robust dense feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19790–19800 (2024)
2. Girasis, C., Schuurbiers, J., Muramatsu, T., Aben, J.P., Onuma, Y., Soekhradj, S., Morel, M.A., Van Geuns, R.J., Wentzel, J.J., Serruys, P.W.: Advanced three-dimensional quantitative coronary angiographic assessment of bifurcation lesions: methodology and phantom validation. EuroIntervention **8**(12), 1451–1460 (2013)
3. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Transactions on pattern analysis and machine intelligence **19**(6), 580–593 (1997)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
5. Kak, A.C., Slaney, M.: Principles of computerized tomographic imaging. SIAM (2001)
6. Kim, H.W., Noh, S.C., Kim, S.H., Chu, H.W., Jung, C.H., Kang, S.H.: Effective descriptor extraction strategies for correspondence matching in coronary angiography images. Scientific Reports **14**(1), 18630 (2024)
7. Leroy, V., Cabon, Y., Revaud, J.: Grounding image matching in 3d with mast3r. In: European conference on computer vision. pp. 71–91. Springer (2024)

8. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision. vol. 2, pp. 1150–1157. IEEE (1999)
9. Nishi, T., Kitahara, H., Fujimoto, Y., Nakayama, T., Sugimoto, K., Takahara, M., Kobayashi, Y.: Comparison of 3-dimensional and 2-dimensional quantitative coronary angiography and intravascular ultrasound for functional assessment of coronary lesions. Journal of Cardiology **69**(1), 280–286 (2017)
10. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
11. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4938–4947 (2020)
12. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016)
13. Siddon, R.L.: Fast calculation of the exact radiological path for a three-dimensional ct array. Medical physics **12**(2), 252–255 (1985)
14. Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: Loftr: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8922–8931 (2021)
15. Tang, S., Zhang, J., Zhu, S., Tan, P.: Quadtree Attention for Vision Transformers. In: International Conference on Learning Representations
16. Weinzaepfel, P., Csurka, G., Cabon, Y., Humenberger, M.: Visual localization by learning objects-of-interest dense match regression. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5634–5643 (2019)
17. Wu, Y., Jena, R., Gulsun, M., Singh, V., Sharma, P., Gee, J.C.: Towards Establishing Dense Correspondence on Multiview Coronary Angiography: From Point-to-Point to Curve-to-Curve Query Matching. arXiv preprint arXiv:2312.11593 (2023)
18. Yi, K.M., Trulls, E., Ono, Y., Lepetit, V., Salzmann, M., Fua, P.: Learning to find good correspondences. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2666–2674 (2018)
19. Zeng, A., Wu, C., Lin, G., Xie, W., Hong, J., Huang, M., Zhuang, J., Bi, S., Pan, D., Ullah, N., et al.: Imagecas: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. Computerized Medical Imaging and Graphics **109**, 102287 (2023)