Automatic Translational Correction of Multi-View Coronary Angiography Based on Auto-Annotation Data Generation

Yue Cao[1], Zhuo Zhang[1]*, Shuai Xiao[1], Jialin Li[1], Guipeng Lan[1], Jiabao Wen[1], Jiachen Yang[1]

[1]School of Electrical and Information Engineering, Tianjin University, Tianjin, China
z_zhuo@tju.edu.cn, caoyue2004@tju.edu.cn

Abstract

Multi-view automatic translational correction (ATC) in coronary angiography (CAG) is critical for intraoperative automatic diagnosis, in which deep learning playing a key role. However, heartbeat-induced soft matching errors and costly annotations make it difficult to build high-quality, large-scale datasets for calibration algorithm training. The training of clinical models is difficult to fulfill, as existing datasets differ significantly from real CAG in both style and structure. To address this challenge, we propose a novel high-quality data synthesis method for annotation-free ATC. We fully automated the construction of a labeled, high-fidelity dataset for training matching models. An evolutionary algorithm is introduced for global optimization of translation estimation, mitigating epipolar constraint violations caused by vascular deformation and enabling reliable correction across large viewpoint differences. Furthermore, a theoretical analysis is presented, demonstrating that error propagation between adjacent views is more accurate than direct estimation across distant views. Our experiments on clinical datasets demonstrate that our method not only significantly outperforms weakly supervised learning approaches, but also performs comparably to fully supervised methods. Moreover, it exhibits remarkable multicenter generalizability.

Code — https://github.com/CaoYue2004/AutomaticTranslationalCorrection

Introduction

Translational correction in coronary angiography (CAG) is vital for systems like Quantitative Flow Ratio (QFR) and 3D Quantitative Coronary Angiography (QCA) (Tu et al. 2011), as it compensates for unrecorded platform movement. Figure 1(a) shows the effect of platform movement on coronary angiography. It also plays a key role in real-time registration during Percutaneous Coronary Intervention (PCI) and serves as a foundation for multi-view synchronization and 3D reconstruction (Cheriet et al. 1999). As shown in Figure 1(b), 3D reconstruction requires precise registration, otherwise it will affect its accuracy. Despite growing research interest and the development of effective algorithms, high-quality training data remain scarce. Real CAG images are affected by cardiac motion and platform shifts, requiring costly expert annotations prone to error. Although voxel-based synthetic datasets (Zhu et al. 2025) are available, their structural regularity and modality gap limit their usefulness for direct model training. As a result, achieving stable and robust Automatic translation correction (ATC) remains an open challenge.

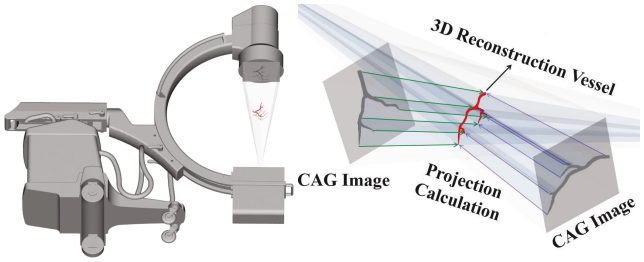

Figure 1: (a)Platform translation causes vessel displacement, which in turn leads to errors in CAG image. (b)The process of 3D reconstruction from CAG images.

Current ATC methods mainly use two multi-view matching strategies: centerline-based and keypoint-based approaches. The centerline-based method relies on vascular topology but is limited by the complexity of coronary anatomy and the difficulty of accurate centerline extraction (Haris et al. 1999; Xiao et al. 2013; Xiao et al. 2016). Keypoint-based methods have shown strong performance in medical image registration and have been clinically validated in static structures like the hip joint and retina(Bier et al. 2018; Grupp et al. 2019; Wang et al. 2024). However, their application to CAG is hindered by the need for large-scale datasets and the challenges posed by the dynamic nature of coronary vessels, which limits data quality and clinical feasibility.

Looking more closely, keypoint-based approaches for coronary angiography can be divided into manual and learning-based methods. Manual methods (Tu et al. 2011), such as hand-correction, are time-consuming and labor-intensive. Learning-based methods include supervised approaches (Wang et al. 2024; Fu et al. 2021), which require high-quality annotated data, and weakly supervised

*Corresponding author

approaches (Li et al. 2022; Wang et al. 2020; Zhou et al. 2021), which reduce manual effort but still suffer from limited matching accuracy due to reliance on epipolar constraints. Therefore, addressing dataset quality and matching errors is crucial for achieving ATC.

To overcome these challenges, we propose a fully annotation-free multi-view CAG ATC framework. We first address data scarcity by introducing a high-fidelity synthesis method using an Efficient Geometry-aware 3D Generative Model (EG3D) (Chan et al. 2021) model pre-trained on coronary data to generate realistic, view-consistent CAG images. Leveraging the generalization ability of SuperPoint (DeTone et al. 2018), we automatically track bifurcation points and reproject them to generate large-scale, high-confidence keypoint pairs, forming a robust training dataset. To handle epipolar constraint violations due to vascular deformation, we introduce a robust evolutionary algorithm with adaptive noise suppression. Compared to traditional least squares methods (Gu et al. 2006), it offers superior accuracy and resilience under large view differences and complex anatomy.

Experiments on Left Coronary Artery (LCA) and Right Coronary Artery (RCA) sequences show that our method, despite using no manual annotations, outperforms state-of-the-art weakly supervised approaches and matches the performance of fully supervised methods. Ablation studies confirm the importance of each module, and multi-center experiments demonstrate strong generalization across datasets.

In summary, the main contributions of this study are as follows:

- Proposed the first fully annotation-free CAG image translation correction framework, capable of end-to-end cross-view translation estimation, and for the first time applies generative models to downstream coronary angiography tasks.
- Introduced an innovative data synthesis method that addresses the challenges of CAG image dynamics and limited annotations, generating high-quality, photorealistic matching pairs.
- Presented a highly robust evolutionary matching algorithm that effectively mitigates the failure of epipolar constraints caused by vascular deformation, achieving high-precision image correction.
- Proposed a proof method demonstrating that small-angle error propagation outperforms direct large-angle computation, providing theoretical support for this work.

Related Work

Coronary Angiography Spatial Registration

Coronary angiography spatial registration is a fundamental step for multi-view synchronous detection and 3D reconstruction. It includes non-rigid and rigid registration. Non-rigid registration is mainly influenced by cardiac motion, and systems like QFR and 3D-QCA often ensure cardiac phase consistency by manually selecting end-diastolic frames (Tu et al. 2011). Deep learning-based methods (Xiao et al. 2016; N.Baka et al. 2014; Wu et al. 2022) have improved registration accuracy to some extent, but typically depend on complex preprocessing steps such as vessel centerline extraction, limiting automation. Rigid registration mainly suffers from platform movement errors, whereas rotations and translations of the X-ray source are recorded in DICOM files and can be corrected through coordinate transformations. Several studies (Wang et al. 2024; Liao et al. 2010; Fu et al. 2021) have effectively addressed translation errors by building optimization objective functions and iteratively refining point correspondences under epipolar constraints. However, these methods still rely on extracting and matching vessel contours or centerlines (Zhu et al. 2025). Traditional methods like SIFT (Lowe et al. 2004) and RANSAC (Fischler and Bolles 1981) rely heavily on low-level features (e.g., gradients, intensity), making them vulnerable in cases with sparse features or noise. Deep learning approaches offer more robust feature representations, especially for complex structures or low-quality images. Nevertheless, strongly supervised methods require large volumes of high-quality annotated data, which is often difficult to obtain in medical imaging due to data scarcity and high annotation costs. Weakly supervised strategies (Li et al. 2022; Wang et al. 2020; Zhou et al. 2021; Barroso et al. 2024) reduce the need for annotations but still suffer from limited matching accuracy.

Synthetic Datasets in Medical Imaging

Due to the scarcity of high-quality annotated data in medical imaging, generating realistic and diverse synthetic datasets has become a vital complementary approach for training deep learning models. Recent research has explored various synthesis methods. One category uses generative models like Generative Adversarial Network (GANs) (Yuan et al. 2024; Shen et al. 2023; Hou et al. 2023) and diffusion models (Pan et al. 2025; Guo et al. 2025; Ye et al. 2023; Moghadam et al. 2023; Packhauser et al. 2022) to learn image distributions and produce high-quality samples. Another category focuses on data augmentation through pixel-level transformations (Su et al. 2023) or label propagation (Cai et al. 2023). Many studies enhance synthetic data by combining limited annotations with data completion (Pan et al. 2025; Cai et al. 2023), perturbation-based augmentation (Su et al. 2023; Guo et al. 2025), or transfer learning (Yuan et al. 2024; Shen et al. 2023; Ye et al. 2023). Some (Hou et al. 2023; Packhauser et al. 2022) leverage weakly annotated image-report pairs to improve semantic consistency and clinical relevance without detailed manual labels. Others (Fu et al. 2024; Iyer et al. 2023) adopt purely physics-based modeling to avoid annotation altogether. Despite progress, achieving a balance between image fidelity, diversity, and automated accurate annotation remains a key challenge and focus in the field.

Optimization Algorithms

Traditional optimization methods like least squares (Gu et al. 2006) are computationally efficient and easy to implement but are sensitive to noise and outliers, and often struggle with high-dimensional, nonlinear, or complex problems.

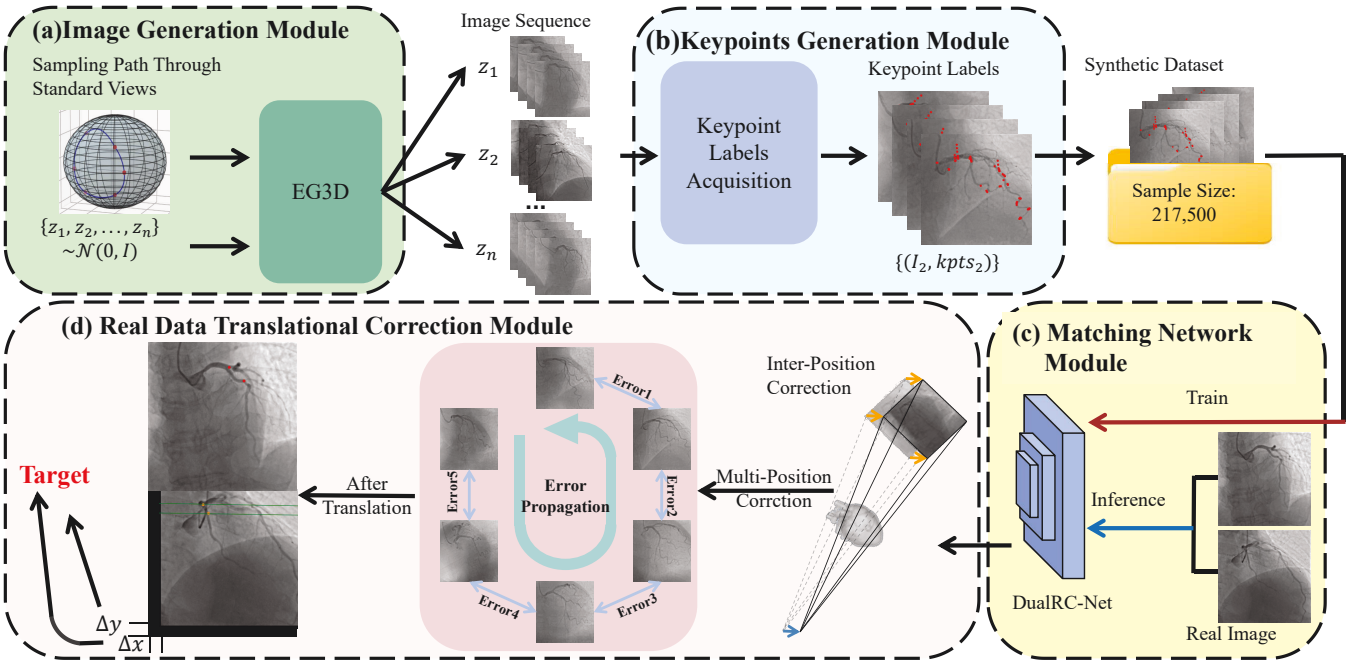

Figure 2: Overview of our proposed ATC pipeline.

In contrast, evolutionary algorithms—such as genetic algorithms (Katoch et al. 2020), particle swarm optimization (Shi et al. 1998), and differential evolution (Storn et al. 1997)—offer stronger global search capabilities and greater robustness, making them well-suited for complex optimization tasks. In coronary angiography analysis, vessel motion caused by the heartbeat disrupts geometric consistency between views, making robust spatial registration essential. Several studies have explored multi-view registration under epipolar constraints (Wang et al. 2024; Liao et al. 2010; Fu et al. 2021), with promising results. However, most rely on manually annotated feature points, which are sparse and limit robustness in real-world scenarios, ultimately affecting registration accuracy.

Method

To address the above issues, we propose an ATC framework, with the overall pipeline comprising four components as shown in Figure 2. (a) Image Generation Module: Design a sampling path across standard viewpoints and use EG3D models (Chan et al. 2021) (with different seeds) to generate image sequences. (b) Keypoint Generation Module: Extract keypoint labels from the sequences to build the dataset. (c) Matching Network Module: Train DualRC-Net (Li et al. 2020) on the synthetic dataset. (d) Real Data Translational Correction Module: Apply the matching network and evolutionary algorithm for adjacent view correction, followed by error propagation for full-view correction.

Preliminary

In the task of CAG image generation, we denote the real image as $I_{\text{real}} \in \mathbb{R}^{H \times W}$, and the generated image as $I_{\text{gen}} \in \mathbb{R}^{H \times W}$, with corresponding imaging parameters denoted as $[R|t]$, where $R \in \mathbb{R}^{3\times3}$ and $t \in \mathbb{R}^{3\times1}$.

During the image generation process, we first sample a latent vector $z \sim \mathcal{N}(0, I)$, which is then encoded into a latent code w by a mapping network (denoted as $Mapper$). Then, the generation network G synthesizes the image conditioned on both w and the imaging parameters $[R|t]$, generating the target image I_{gen}. This process can be formalized as:

$$I_{\text{gen}} = G(Mapper(z), [R|t]) \tag{1}$$

In CAG, typically 6 standard viewing angles are used for LCA, and 2 standard views for RCA (Green et al. 2016) to facilitate a comprehensive assessment of coronary status.

Labeled Sample Generation

To obtain multi-view images with consistent structure and semantic annotations, we construct a 2D interpolation curve (I.J. Schoenberg 1946) $\mathcal{C}(s) : [0, 1] \to \mathbb{R}^3$ around the EG3D model (Chan et al. 2021). The path passes through 6 standard projection views of LCA, denoted as $\{v_1^L, \ldots, v_6^L\}$, and 2 standard projection views of RCA, denoted as $\{v_1^R, \ldots, v_3^R\}$. We sample N points $\{s_i\}_{i=1}^N$ (where N is the number of samples) evenly along the curve, and render images $I_i = \text{Render}(\mathcal{M}, \mathcal{C}(s_i))$ at each sample point s_i (where $\mathcal{M}$ is the EG3D model (Chan et al. 2021)).

Next, we apply a keypoint detector $\mathcal{K}(\cdot)$ to each image sequence $\{I_i\}$ to obtain a set of keypoints $\mathcal{P}_i = \mathcal{K}(I_i)$. Among these, we retain only the keypoints detected in at least M views as the valid keypoint set, and obtain the associated matching keypoint set $\mathcal{P}^*$ across multiple views, which is used for subsequent 3D geometric annotation and training data construction.

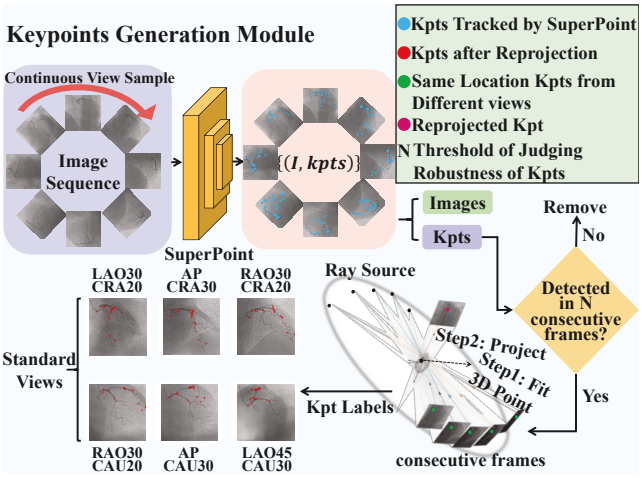

Figure 3: The framework of the proposed Keypoints Generation Module.

To compute the 3D position of each keypoint $p \in \mathcal{P}^*$, we perform multi-view triangulation using its stable detections across viewpoints. For the i-th image, the keypoint's pixel location $\mathbf{x}_i = (u_i, v_i)^\top$ is expressed in homogeneous coordinates as $\tilde{\mathbf{x}}_i = (u_i, v_i, 1)^\top$. The camera intrinsic matrix $\mathbf{K}_i \in \mathbb{R}^{3\times3}$, extrinsic matrix $[\mathbf{R}_i|\mathbf{t}_i]$, and X-ray source position $\mathbf{S}_i$ are derived from DICOM metadata. Using these, the 3D world coordinates of p are reconstructed:

$$\mathbf{P}_i = \mathbf{R}_i \cdot \mathbf{K}_i^{-1} \cdot \tilde{\mathbf{x}}_i + \mathbf{t}_i \tag{2}$$

From this, we construct a ray that originates from the X-ray source and passes through the keypoint p:

$$\ell_i(\alpha) = \mathbf{S}_i + \alpha \cdot (\mathbf{P}_i - \mathbf{S}_i), \quad \alpha > 0 \tag{3}$$

To compute the intersection in 3D space, we perform a least-squares optimization (Gu et al. 2006) over all rays $\{\ell_i\}$, estimating the point $\hat{\mathbf{X}}$ that best approximates the intersection of all rays:

$$\mathbf{v}_i = \frac{\mathbf{P}_i - \mathbf{S}_i}{\|\mathbf{P}_i - \mathbf{S}_i\|} \tag{4}$$

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \sum_i \left\|(\mathbf{I} - \mathbf{v}_i\mathbf{v}_i^\top)(\mathbf{X} - \mathbf{S}_i)\right\|^2 \tag{5}$$

where $\mathbf{v}_i$ is the unit direction vector of the ray. The result $\hat{\mathbf{X}}$ is the estimated 3D position of the keypoint in the world coordinate system. Finally, this 3D point is projected onto the reference canonical view using the perspective projection equation:

$$(u_s, v_s, 1)^\top = \frac{1}{\omega} \cdot \mathbf{K}_s \cdot \left(\mathbf{R}_s\hat{\mathbf{X}} + \mathbf{t}_s\right) \tag{6}$$

Among them, $\mathbf{K}_s$, $\mathbf{R}_s$, and $\mathbf{t}_s$ represent the intrinsic matrix, rotation matrix, and translation vector of the standard camera pose, respectively; ω is the normalization factor. Based on this, the keypoint coordinates in each view can be obtained.

Figure 3 illustrates the keypoint acquisition process.

Automatic Translational Correction Method

To achieve spatial consistency among multi-view images, we propose an automatic translational correction method consisting of two main modules. First, the matching model is trained on the synthetic dataset. Then, it is used to correct adjacent viewpoints, and the translational offsets for all viewpoints are computed via error propagation.

Training of the Matching Model. We train DualRC-Net (Li et al. 2020) for image matching. Its dual-resolution feature extractor captures both global context and local details, enhancing pixel-level matching across viewpoints. To further improve accuracy, a neighborhood consensus module evaluates local consistency in the feature space to refine point correspondences.

During training, images rendered from standard viewpoints are input, and 2D projections of 3D vascular keypoints are used as supervision signals to guide the learning of spatial correspondences.

Optimization of Translation Offsets. Based on epipolar geometry, corresponding points in two views must lie on their respective epipolar lines and satisfy the epipolar constraint (Merle et al. 1998). Using our trained keypoint matching model, we introduce an evolutionary algorithm with an adaptive fitness function to optimize translational offset estimation. For each adjacent view pair, one image is set as the reference. Matched keypoints in this view are used to compute the fundamental matrix and derive corresponding epipolar lines projected onto the target image. This yields keypoint–epipolar line pairs in the target view, enabling geometric consistency evaluation.

The algorithm initializes with zero translation and iteratively adjusts parameters to minimize a fitness function that combines matching error, epipolar constraint deviation, and translation magnitude. The function is formally defined as:

$$\begin{aligned}\arg\min_{\Delta x,\,\Delta y} F(l, p, \Delta x, \Delta y, T) = \frac{1}{n}\sum_{i=1}^{n} d_i - \alpha \cdot N \\ +\beta \cdot \frac{1}{n}\sum_{i=1}^{n}\sqrt{\Delta x^2 + \Delta y^2}\end{aligned} \tag{7}$$

Here, l represents the epipolar line equation, p denotes the coordinates of the corresponding keypoints, n is the total number of matched points, and d_i is the distance from the i-th point to its corresponding epipolar line. T refers to the threshold for the distance between a point and its epipolar line, and N is the number of points within this threshold. Δx and Δy are the translation parameters, while α and β are tunable hyperparameters that control the weights of the number of inliers and the magnitude of the translation, respectively.

Through the above optimization, the algorithm ensures the geometric consistency of epipolar lines and achieves high-precision and robust translational alignment across images.

Error Propagation. To improve matching accuracy between images with large viewpoint differences, we compute

translation vectors between adjacent frames and progressively propagate them to a reference image, deriving global translation for each view. This layer-by-layer propagation accumulates local registration information, enhancing alignment accuracy under large viewpoint variations.

This strategy significantly improves translation estimation accuracy between cross-view images. However, quantifying the improvement from error propagation and establishing its theoretical foundation remains an open question, which we address in the following section with a mathematical analysis comparing error propagation to direct cross-view error computation.

Theoretical Evidence

Based on the above error propagation strategy, we propose the following theoretical analysis to verify its effectiveness.

Assumption 1 *The observation error difference between adjacent view angles i and $i+1$ is denoted as $\varepsilon_{i,i+1}$, which follows a zero-mean Gaussian distribution $\varepsilon_{i,i+1} \sim \mathcal{N}(0, \Sigma_{i,i+1})$, where $\Sigma_{i,i+1}$ represents the covariance matrix of the observation error difference.*

Assumption 2 *The variance of the matching error increases rapidly with the visual angle difference θ_{ij}, i.e., the covariance matrix Σ_{ij} of the error difference is a Loewner convex function of the visual angle difference. We denote*

$$\Sigma_{ij} = f(\theta_{ij}), \quad f : [0, \pi] \to \mathbb{R}^{2\times 2} \tag{8}$$

where $f(\cdot)$ is a matrix-valued function of the visual angle difference θ_{ij}, and is Loewner convex in the positive semi-definite matrix sense.

Lemma (Jensen et al. 1906) *If the function f is Loewner convex, then for any two points $\theta_1, \theta_2 \in [0, \pi]$ and any $\lambda \in [0, 1]$, we have:*

$$f(\lambda\theta_1 + (1-\lambda)\theta_2) \preceq \lambda f(\theta_1) + (1-\lambda) f(\theta_2) \tag{9}$$

Corollary *Let X and Y be random vectors with covariance matrices Σ_X and Σ_Y, respectively. If $\Sigma_X \preceq \Sigma_Y$, then: $\mathrm{tr}(\Sigma_X) \leq \mathrm{tr}(\Sigma_Y)$, the expected squared norm $\mathbb{E}(\|X\|^2) \leq \mathbb{E}(\|Y\|^2)$.*

Proposition *Under Assumptions 1 and 2, the accumulated relative observation error calculation method results in a smaller average error variance than the direct cross-angle error calculation.*

Proof From Assumptions 2 and Lemma, we have:

$$\sum_{i=1}^{k-1} \Sigma_{i,i+1} = \sum_{i=0}^{k-1} f(\theta_i) \preceq f\left(\sum_{i=0}^{k-1} \theta_i\right) = \Sigma_{0,k} \tag{10}$$

where $\Sigma_{i,i+1}$ is the covariance matrix of the relative observation error between adjacent view angles i and $i+1$, and $\Sigma_{0,k}$ is the covariance matrix of the error between the 0-th and k-th view.

According to the corollary, if the covariance matrices follow the partial ordering relation, their traces also satisfy the same inequality. Thus, the accumulated method results in a smaller or equal trace compared to the direct method, indicating a smaller or equal expected error, i.e., lower average error variance.

Experiment

Dataset and Preprocessing

Training Set. The training set in this study consists of a fully synthetic dataset, including images generated by EG3D (Chan et al. 2021) and keypoint coordinates obtained through keypoint tracking and re-projection via SuperPoint (DeTone et al. 2018). Specifically, we selected seed points 1 to 500 from EG3D (Chan et al. 2021), with each seed point containing 6 (LCA) or 2 (RCA) standard views. For each view, 5 types of translation transformations (measured in pixels) were applied: (20, 10), (10, 20), (0, 0), (-20, -10), (-10, -20). Therefore, a total of 217,500 image pairs were generated for LCA, and 22,500 image pairs for RCA.

Test Set. The test set includes data from 458 real cases across five centers, distributed as follows: Tianjin Hangyi Cardiovascular Hospital (HY) (94 LCA cases, 300 RCA cases), Tieling Hospital of Traditional Chinese Medicine (TL) (3 LCA cases, 3 RCA cases), Shengjing Hospital of China Medical University (SJ) (4 LCA cases, 3 RCA cases), Dongguan People's Hospital (DG) (10 LCA cases, 4 RCA cases), and the 252nd Hospital of the PLA (252) (20 LCA cases, 17 RCA cases). For comparative and ablation experiments, data from HY were selected to validate the accuracy of this method; for multi-center experiments, data from the other centers were used to assess the generalization ability of the method.

Preprocessing. To ensure all test images are diastolic-phase angiographic frames, we performed the following data cleaning steps: (1) trained a ResNet (He et al. 2015) to classify angiograms as LCA or RCA; (2) trained another ResNet (He et al. 2015) to filter out non-angiographic frames; (3) used a pre-trained SuperPoint model (DeTone et al. 2018) for inter-frame keypoint tracking and selected the frame with the most dispersed keypoints (i.e., farthest from the centroid) as the diastolic frame.

Evaluation Metrics

The experiments used three evaluation metrics: 1) mean Translational Correction Error (mTCE, in millimeters), which was measured error reduction after ATC compared to manual correction as the ground truth; 2) mean Reprojection Error (Klein et al. 1998) (mRE, in millimeters), calculated as the distance between manually annotated points and their re-projected counterparts after deriving 3D keypoints and projecting them back onto the image; and 3) Success Rate (SR), which is the proportion of cases where mRE is within 8mm. All metrics are calculated based on the original image resolution (512×512 pixels) and actual size (158 mm × 158 mm).

Implementation Details

The EG3D model used in this study is a pre-trained model, trained on 220,000 coronary angiography images with camera parameters from HY. The training was conducted on an NVIDIA GeForce RTX 3080 Ti GPU. The generated samples exhibit high fidelity and multi-view consistency (FID = 8.21).

Method	LCA			RCA		
	mTCE↓	mRE↓	SR(%)↑	mTCE↓	mRE↓	SR(%)↑
Ground Truth	10.23	9.37	31.58	13.07	7.29	64.67
SIFT (Lowe et al. 2004)	16.55	9.63	34.04	11.79	5.88	83.33
PBO (Oinonen et al. 2010)	18.85	10.23	27.08	12.35	7.32	78.12
PoSFeat (Li et al. 2022)	$19.94_{\pm 1.59}$	$10.46_{\pm 0.24}$	$21.93_{\pm 5.32}$	$15.97_{\pm 2.64}$	$8.89_{\pm 1.26}$	$56.56_{\pm 3.19}$
CAPS (Wang et al. 2020)	$16.30_{\pm 0.31}$	$9.12_{\pm 0.23}$	$40.79_{\pm 1.32}$	$5.10_{\pm 0.31}$	$6.29_{\pm 0.13}$	$78.38_{\pm 2.16}$
Patch2Pix (Zhou et al. 2021)	$20.36_{\pm 3.28}$	$12.71_{\pm 1.21}$	$15.79_{\pm 7.91}$	$21.78_{\pm 2.34}$	$12.26_{\pm 1.00}$	$14.89_{\pm 4.94}$
Mickey (Barroso et al. 2024)	$16.99_{\pm 0.15}$	$9.66_{\pm 0.13}$	$35.96_{\pm 4.23}$	$13.17_{\pm 0.11}$	$7.32_{\pm 0.13}$	$61.84_{\pm 1.02}$
Supervise (Li et al. 2020)	$10.55_{\pm 0.62}$	$7.76_{\pm 0.23}$	$64.47_{\pm 3.12}$	$4.84_{\pm 0.73}$	$5.90_{\pm 0.45}$	$81.22_{\pm 1.00}$
Small Dataset (Li et al. 2020)	$13.53_{\pm 0.75}$	$7.63_{\pm 0.67}$	$61.84_{\pm 2.97}$	$5.34_{\pm 0.25}$	$4.53_{\pm 0.44}$	$83.78_{\pm 0.95}$
LSM (Gu et al. 2006)	$9.24_{\pm 0.50}$	$7.37_{\pm 0.05}$	$66.23_{\pm 3.31}$	$10.07_{\pm 0.94}$	$6.71_{\pm 0.16}$	$79.32_{\pm 1.21}$
Ours	$\mathbf{4.79_{\pm 0.16}}$	$\mathbf{6.19_{\pm 0.36}}$	$\mathbf{83.13_{\pm 2.11}}$	$\mathbf{3.64_{\pm 0.45}}$	$\mathbf{4.45_{\pm 0.12}}$	$\mathbf{96.56_{\pm 0.56}}$

Table 1: Results of comparison experiments.

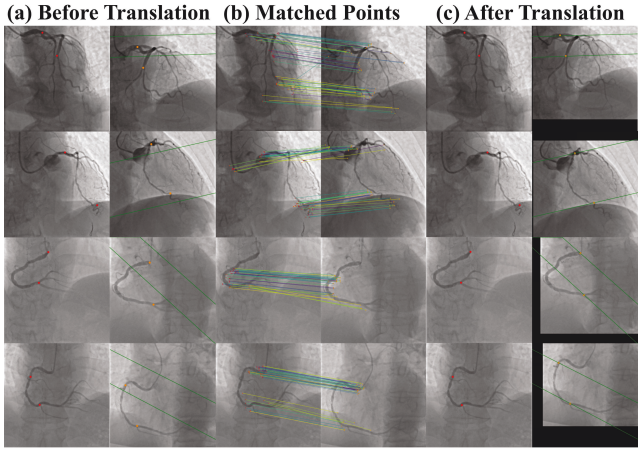

Figure 4: Qualitative results: (a) Multi-view correspondence analysis showing epipolar geometry constraints (red/orange points with green epipolar lines); (b) Color-coded correspondence matching between views; (c) Translational-corrected results evaluated through epipolar distance minimization (orange point-to-line alignment).

During the training of DualRC-Net, we used an NVIDIA L20 GPU with 48GB of memory, based on PyTorch framework and employing Adam optimizer. Separate models were trained for LCA and RCA. Initial learning rate was set to 0.001, and batch size was 32. The left coronary model was trained for 435,000 iterations, while the right coronary model was trained for 45,000 iterations. Learning rate was halved at the 87,000th and 9,000th iterations, respectively.

To reduce the randomness in experimental results, all model trainings were repeated 3 times, and their performance was evaluated separately to ensure result stability.

Comparative Experiments

To validate the effectiveness of our method in translation correction, we conducted comparative experiments. For keypoint matching, we evaluated traditional methods (SIFT (Lowe et al. 2004), PBO (Oinonen et al. 2010)) and deep learning approaches, categorized as weakly supervised (PosFeat (Li et al. 2022), CAPS (Wang et al. 2020), Patch2Pix (Zhou et al. 2021), Mickey (Barroso-Laguna et al. 2024)) and strongly supervised (DualRC-Net (Li et al. 2020)). Weakly supervised methods rely on epipolar constraints and are trained solely on synthetic EG3D (Chan et al. 2021) data, while the strongly supervised method uses manually annotated correspondences (details in the supplementary material). To ensure fairness, we retrained DualRC-Net (Li et al. 2020) on a randomly selected subset of 217,500 image pairs—matching our method's training size. For translation computation, we compared our evolutionary algorithm with the traditional least squares method (LSM) (Gu et al. 2006).

Table 1 summarizes the performance differences among various methods in keypoint matching and translation correction tasks. As shown, our method outperforms all other methods. Compared to the best-performing weakly supervised method, CAPS (Wang et al. 2020), our approach improves SR by 42.34% on LCA dataset and 18.18% on RCA dataset, significantly surpassing existing weakly supervised methods. When compared to strongly supervised models trained on datasets of the same scale (Supervise and Small Dataset), our method achieves comparable performance. This demonstrates that with the same amount of training data, our method can achieve results similar to supervised learning approaches without requiring manual annotations. More importantly, our method supports training with large-scale synthetic data. Under such conditions, our model outperforms the supervised model on all metrics. On LCA dataset, mTCE and mRE decrease by 56.30% and 16.76%, respectively, and SR increases by 18.66%. On RCA dataset, mTCE and mRE decrease by 9.18% and 19.89%, respectively, and SR increases by 15.34%.

Figure 4 demonstrates effectiveness of our method using 2 LCA and 2 RCA examples. In (b), the nearly parallel matching lines reflect the accuracy of the correspondences. In (c), after translation, most points lie almost exactly on the epipolar lines, indicating accuracy of the computed translation.

Ablation Study

To validate the effectiveness of each module and the rationale behind parameter choices, we conducted a series of ablation studies.

SPF	IWT	SAV	LCA			RCA		
			mTCE↓	mRE↓	SR(%)↑	mTCE↓	mRE↓	SR(%)↑
-	-	-	$18.33_{\pm 1.89}$	$9.15_{\pm 1.75}$	$38.16_{\pm 4.33}$	$6.30_{\pm 0.85}$	$5.80_{\pm 0.61}$	$82.63_{\pm 0.99}$
✓	-	-	$16.67_{\pm 2.31}$	$8.85_{\pm 1.38}$	$51.32_{\pm 3.52}$	-	-	-
✓	-	✓	$7.38_{\pm 1.13}$	$6.80_{\pm 1.02}$	$81.52_{\pm 2.31}$	$5.34_{\pm 0.38}$	$4.88_{\pm 0.25}$	$86.58_{\pm 1.53}$
✓	✓	-	$16.39_{\pm 1.34}$	$8.73_{\pm 1.26}$	$54.74_{\pm 2.79}$	-	-	-
✓	✓	✓	$\mathbf{4.79_{\pm 0.16}}$	$\mathbf{6.19_{\pm 0.36}}$	$\mathbf{83.13_{\pm 2.11}}$	$\mathbf{3.64_{\pm 0.45}}$	$\mathbf{4.45_{\pm 0.12}}$	$\mathbf{96.56_{\pm 0.56}}$

Table 2: Quantitative results of ablation analysis of different components.

method	Multi-Center (LCA)			HY (LCA)		
	GT	ours	improve	GT	ours	improve
mTCE	14.58	6.27±0.28	57.00%	10.23	4.79±0.16	53.18%
mRE	11.91	8.68±0.54	27.12%	9.37	6.19±0.36	33.94%
SR	27.03	70.27±3.89	43.24%	31.58	83.13±2.11	51.55%
method	Multi-Center (RCA)			HY (RCA)		
	GT	ours	improve	GT	ours	improve
mTCE	12.37	4.17±0.97	66.29%	13.37	3.64±0.45	72.15%
mRE	6.80	3.90±0.66	42.65%	7.29	4.45±0.12	38.96%
SR	92.59	100.00±0.00	7.41%	64.67	96.56±0.56	31.89%

Table 3: Performance comparison on LCA and RCA datasets under Multi-Center and HY settings.

Module Effectiveness. To validate our modules, we conducted ablation studies on three factors: 1) same-phase frame selection (SPF), 2) inclusion of translated images (IWT) in training, and 3) adjacent-view error propagation (SAV). As shown in Table 2, on the left coronary dataset, the progressive addition of the SPF, IWT, and SAV modules led to improvements in all metrics. Ultimately, compared to the baseline, our method reduced mTCE and mRE by 132.36% and 31.59%, respectively, while increasing SR by 44.97%. On the right coronary dataset, adding the SPF module increased SR by 3.93%, and further incorporating the SAV module boosted SR by 9.98%. These results demonstrate that each additional module we introduced effectively enhanced performance compared to the baseline.

Rationality of Parameter Selection. We performed ablation studies on three parameters: the number of consecutive frames for SuperPoint (DeTone et al. 2018) tracking, the number of matched point pairs, and the α and β values in the evolutionary algorithm's fitness function. As shown in the supplementary materials, optimal performance was achieved with 14 frames, 70 matched keypoints, and α = -0.5, β = 0.1.

Multi-Center Experiment

To demonstrate our generalizability, we conducted experiments on multi-center datasets. Although coronary structures are consistent, image style (e.g. transparency, contrast) varies across centers. As shown in Table 3, on LCA data the overall metrics after translation correction are slightly lower than on HY due to higher inherent errors, but improvement magnitude is similar. On RCA data, the pre-correction success rate was already 92.59%, leaving little room for improvement, with a modest 7.41% gain. Other metrics and trends are consistent with the HY. Figure 5 shows results of multi-center experiments. Regardless of image style, our method achieves reliable matching and accurate translation estimation, demonstrating strong generalization ability.

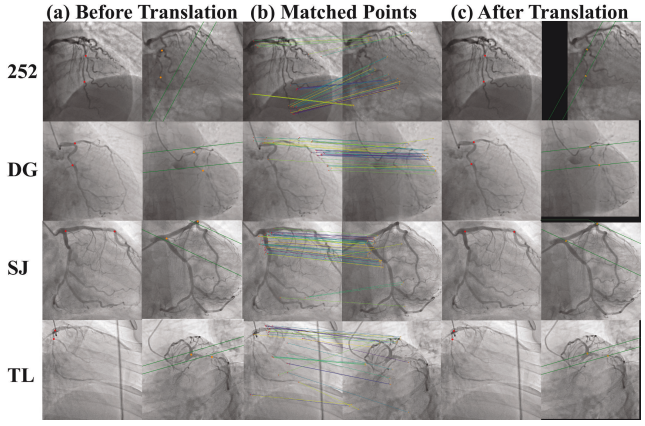

Figure 5: Multi-center experimental results.

Conclusion

We propose an ACT method for CAG images to address platform motion in coronary angiography. By auto-annotating data, we overcome the issue of limited manual annotations. Our evolutionary algorithm mitigates epipolar constraint failures due to vascular deformation, improving translation accuracy. Theoretical support for error propagation is also provided. Experiments show that our method outperforms weakly supervised approaches and rivals strongly supervised methods. Ablation studies validate the effectiveness of each module, and multi-center experiments demonstrate its strong generalization ability.

References

Baka, N.; Metz, C. T.; Schultz, C. J.; van Geuns, R.-J.; Niessen, W. J.; and van Walsum, T. 2014. Oriented Gaussian Mixture Models for Nonrigid 2D/3D Coronary Artery Registration. *IEEE Transactions on Medical Imaging*, 33(5): 1023–1034.

Barroso-Laguna, A.; Munukutla, S.; Prisacariu, V. A.; and Brachmann, E. 2024. Matching 2D Images in 3D: Metric Relative Pose from Metric Correspondences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4852–4863.

Bezerra, C. G.; Talou, G. D. M.; Bulant, C. A.; de Alencar Araripe Falcão, B.; Mariani, J.; Blanco, P. J.; Feijóo, R. A.; and Neto, P. A. L. 2015. Three-dimensional reconstruction of coronary arteries based on the integration of intravascular ultrasound and conventional angiography. *Revista Brasileira de Cardiologia Invasiva (English Edition)*, 23(2): 134–138.

Bier, B.; Unberath, M.; Zaech, J.-N.; Fotouhi, J.; Navab, N.; and Maier, A. 2018. *X-ray-transform Invariant Anatomical Landmark Detection for Pelvic Trauma Surgery: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV*, 55–63. ISBN 978-3-030-00936-6.

Cai, H.; Li, S.; Qi, L.; Yu, Q.; Shi, Y.; and Gao, Y. 2023. Orthogonal Annotation Benefits Barely-supervised Medical Image Segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3302–3311.

Chan, E. R.; Lin, C. Z.; Chan, M. A.; Nagano, K.; Pan, B.; De Mello, S.; Gallo, O.; Guibas, L. J.; Tremblay, J.; Khamis, S.; Karras, T.; and Wetzstein, G. 2022. Efficient Geometry-Aware 3D Generative Adversarial Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16123–16133.

Cheriet, F.; and Meunier, J. 1999. Self-calibration of a biplane X-ray imaging system for an optimal three dimensional reconstruction. *Computerized Medical Imaging and Graphics*, 23(3): 133–141.

DeTone, D.; Malisiewicz, T.; and Rabinovich, A. 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 337–33712.

Fischler, M. A.; and Bolles, R. C. 1987. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In Fischler, M. A.; and Firschein, O., eds., *Readings in Computer Vision*, 726–740. San Francisco (CA): Morgan Kaufmann. ISBN 978-0-08-051581-6.

Fu, X.; Li, Y.; Tang, F.; Li, J.; Zhao, M.; Teng, G.-J.; and Zhou, S. K. 2024. 3DGR-CAR: Coronary Artery Reconstruction from Ultra-sparse 2D X-Ray Views with a 3D Gaussians Representation. In Linguraru, M. G.; Dou, Q.; Feragen, A.; Giannarou, S.; Glocker, B.; Lekadir, K.; and Schnabel, J. A., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, 14–24. Cham: Springer Nature Switzerland. ISBN 978-3-031-72104-5.

Fu, Z.; Fu, Z.; Gong, Z.; Feng, X.; Gu, H.; Xie, R.; Zhang, J.; and Fei, J. 2021. Optimization For 3D Reconstruction Of Coronary Artery Tree By Two-stage Levenberg-Marquardt Algorithm. In *2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, 84–89.

Green, P.; Frobisher, P.; and Ramcharitar, S. 2016. Optimal angiographic views for invasive coronary angiography: A guide for trainees. *British Journal of Cardiology*, 23: 110–113.

Grupp, R. B.; Unberath, M.; Gao, C.; Hegeman, R. A.; Murphy, R. J.; Alexander, C. P.; Otake, Y.; McArthur, B. A.; Armand, M.; and Taylor, R. H. 2020. Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration. *International journal of computer assisted radiology and surgery*, 15(5): 759—769.

Guo, E.; Zhao, Z.; Wang, Z.; Chen, T.; Liu, Y.; and Zhou, L. 2025. DiN: Diffusion Model for Robust Medical VQA with Semantic Noisy Labels. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, 14337–14346.

Haris, K.; Efstratiadis, S.; Maglaveras, N.; Pappas, C.; Gourassas, J.; and Louridas, G. 1999. Model-based morphological segmentation and labeling of coronary angiograms. *IEEE Transactions on Medical Imaging*, 18(10): 1003–1015.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hou, Z.; Yan, R.; Wang, Q.; Lang, N.; and Zhou, X. 2023. Diversity-Preserving Chest Radiographs Generation from Reports in One Stage. In Greenspan, H.; Madabhushi, A.; Mousavi, P.; Salcudean, S.; Duncan, J.; Syeda-Mahmood, T.; and Taylor, R., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 482–492. Cham: Springer Nature Switzerland. ISBN 978-3-031-43904-9.

Iyer, K.; Nallamothu, B. K.; Figueroa, C. A.; and Nadakuditi, R. R. 2023. A multi-stage neural network approach for coronary 3D reconstruction from uncalibrated X-ray angiography images. *Scientific Reports*, 13(1): 17603.

Jensen, J. L. W. V. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1): 175–193.

Katoch, S.; Chauhan, S. S.; and Kumar, V. 2021. A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5): 8091–8126.

Li, K.; Wang, L.; Liu, L.; Ran, Q.; Xu, K.; and Guo, Y. 2022. Decoupling Makes Weakly Supervised Local Feature Better. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15817–15827.

Li, X.; Han, K.; Li, S.; and Prisacariu, V. 2020. Dual-Resolution Correspondence Networks. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 17346–17357. Curran Associates, Inc.

Liao, R.; Luc, D.; Sun, Y.; and Kirchberg, K. 2010. 3-D reconstruction of the coronary artery tree from multiple views of a rotational X-ray angiography. *The International Journal of Cardiovascular Imaging*, 26(7): 733–749.

Lowe, D. G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2): 91–110.

Lu, J.; Ma, S.; Liu, Y.; Ma, Y.; Mou, L.; Jiang, Y.; and Zhao, Y. 2025. Rethinking Data Augmentation for Single-Source Domain Generalization in OCT Image Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 1–14.

Merle, A.; Finet, G.; Lienard, J.; and Magnin, I. 1998. 3D reconstruction of the deformable coronary tree skeleton from two X-ray angiographic views. In *Computers in Cardiology 1998. Vol. 25 (Cat. No.98CH36292)*, 757–760.

Moghadam, P. A.; Van Dalen, S.; Martin, K. C.; Lennerz, J.; Yip, S.; Farahani, H.; and Bashashati, A. 2023. A Morphology Focused Diffusion Probabilistic Model for Synthesis of Histopathology Images. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1999–2008.

Oinonen, H.; Forsvik, H.; Ruusuvuori, P.; Yli-Harja, O.; Voipio, V.; and Huttunen, H. 2010. Identity verification based on vessel matching from fundus images. In *2010 IEEE International Conference on Image Processing*, 4089–4092.

Packhäuser, K.; Folle, L.; Thamm, F.; and Maier, A. 2023. Generation of Anonymous Chest Radiographs Using Latent Diffusion Models for Training Thoracic Abnormality Classification Systems. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 1–5.

Pan, S.; Liu, Y.; Zhao, L.; Chen, E. Z.; Chen, X.; Chen, T.; and Sun, S. 2025. Label-Efficient Data Augmentation with Video Diffusion Models for Guidewire Segmentation in Cardiac Fluoroscopy. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(6): 6308–6316.

Schoenberg, I. J. 1988. *Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions*, 3–57. Boston, MA: Birkhäuser Boston. ISBN 978-1-4899-0433-1.

Shi, Y.; and Eberhart, R. 1998. A modified particle swarm optimizer. In *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, 69–73.

Storn, R.; and Price, K. 1997. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4): 341–359.

Toutouzas, K.; Chatzizisis, Y. S.; Riga, M.; Giannopoulos, A.; Antoniadis, A. P.; Tu, S.; Fujino, Y.; Mitsouras, D.; Doulaverakis, C.; Tsampoulatidis, I.; Koutkias, V. G.; Bouki, K.; Li, Y.; Chouvarda, I.; Cheimariotis, G.; Maglaveras, N.; Kompatsiaris, I.; Nakamura, S.; Reiber, J. H.; Rybicki, F.; Karvounis, H.; Stefanadis, C.; Tousoulis, D.; and Giannoglou, G. D. 2015. Accurate and reproducible reconstruction of coronary arteries and endothelial shear stress calculation using 3D OCT: Comparative study to 3D IVUS and 3D QCA. *Atherosclerosis*, 240(2): 510–519.

Wang, Q.; Zhou, X.; Hariharan, B.; and Snavely, N. 2020. Learning Feature Descriptors Using Camera Pose Supervision. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 757–774. Cham: Springer International Publishing. ISBN 978-3-030-58452-8.

Wang, Y.; Wang, X.; Gu, Z.; Liu, W.; Ng, W. S.; Huang, W.; and Cheng, J. 2024. SuperJunction: Learning-Based Junction Detection for Retinal Image Registration. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1): 292–300.

Wang, Z.; Fu, Z.; and Jing, F. 2024. A Two-stage Angiographic Parameter Calibration Method for Coronary Artery 3D Reconstruction. In *2024 7th International Conference on Intelligent Robotics and Control Engineering (IRCE)*, 276–281.

Wu, W.; Zhang, J.; Peng, W.; Xie, H.; Zhang, S.; and Gu, L. 2022. CAR-Net: A Deep Learning-Based Deformation Model for 3D/2D Coronary Artery Registration. *IEEE Transactions on Medical Imaging*, 41(10): 2715–2727.

Xiangqian, G.; Hongwen, K.; and Hongxing, C. 2006. The least-square method in complex number domain. *Progress in Natural Science*, 16(3): 307–312.

Xiao, R.; Yang, J.; Fan, J.; Ai, D.; Wang, G.; and Wang, Y. 2016. Shape context and projection geometry constrained vasculature matching for 3D reconstruction of coronary artery. *Neurocomputing*, 195: 65–73. Learning for Medical Imaging.

Xiao, R.; Yang, J.; Goyal, M.; Liu, Y.; and Wang, Y. 2013. Automatic Vasculature Identification in Coronary Angiograms by Adaptive Geometrical Tracking. *Computational and Mathematical Methods in Medicine*, 2013.

Ye, J.; Ni, H.; Jin, P.; Huang, S. X.; and Xue, Y. 2023. Synthetic Augmentation with Large-Scale Unconditional Pre-training. In Greenspan, H.; Madabhushi, A.; Mousavi, P.; Salcudean, S.; Duncan, J.; Syeda-Mahmood, T.; and Taylor, R., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 754–764. Cham: Springer Nature Switzerland. ISBN 978-3-031-43895-0.

Yuan, Z.; Fang, Z.; Huang, Z.; Wu, F.; Yao, Y.-F.; and Li, Y. 2024. Adapting Pre-trained Generative Model to Medical Image for Data Augmentation. In Linguraru, M. G.; Dou, Q.; Feragen, A.; Giannarou, S.; Glocker, B.; Lekadir, K.; and Schnabel, J. A., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, 79–89. Cham: Springer Nature Switzerland. ISBN 978-3-031-72086-4.

Zhou, Q.; Sattler, T.; and Leal-Taixé, L. 2021. Patch2Pix: Epipolar-Guided Pixel-Level Correspondences. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4667–4676.

Zhu, Y.; Wang, Y.; Di, C.; Liu, H.; Liao, F.; and Ma, S. 2025. Sparse and transferable three-dimensional dynamic vascular reconstruction for instantaneous diagnosis. *Nature Machine Intelligence*, 7(5): 730–742.