

DeepCA: Deep Learning-based 3D Coronary Artery Tree Reconstruction from Two 2D Non-simultaneous X-ray Angiography Projections

Yiying Wang¹ Abhirup Banerjee^{1,2} Robin P. Choudhury² Vicente Grau¹

¹Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

²Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK

Abstract

Cardiovascular diseases (CVDs) are the most common cause of death worldwide. Invasive x-ray coronary angiography (ICA) is one of the most important imaging modalities for the diagnosis of CVDs. ICA typically acquires only two 2D projections, which makes the 3D geometry of coronary vessels difficult to interpret, thus requiring 3D coronary artery tree reconstruction from two projections. State-of-the-art approaches require significant manual interactions and cannot correct the non-rigid cardiac and respiratory motions between non-simultaneous projections. In this study, we propose a novel deep learning pipeline named DeepCA. We leverage the Wasserstein conditional generative adversarial network with gradient penalty, latent convolutional transformer layers, and a dynamic snake convolutional critic to implicitly compensate for the non-rigid motion and provide 3D coronary artery tree reconstruction. Through simulating projections from coronary computed tomography angiography (CCTA), we achieve the generalisation of 3D coronary tree reconstruction on real non-simultaneous ICA projections. We incorporate an application-specific evaluation metric to validate our proposed model on both a CCTA dataset and a real ICA dataset, together with Chamfer ℓ_2 distance. The results demonstrate promising performance of our DeepCA model in vessel topology preservation, recovery of missing features, and generalisation ability to real ICA data. To the best of our knowledge, this is the first study that leverages deep learning to achieve 3D coronary tree reconstruction from two real non-simultaneous x-ray angiographic projections. The implementation of this work is available at: <https://github.com/WangStephen/DeepCA>.

1. Introduction

Cardiovascular diseases (CVDs) represent a group of disorders of the heart and blood vessels. They are the most common cause of death worldwide, representing 32% of all

global deaths [34] and posing a serious burden to healthcare systems. Invasive x-ray coronary angiography (ICA) is one of the most important imaging modalities for the diagnosis of CVDs and is most commonly utilised during real-time cardiac interventions [22]. However, ICA acquires only 2D projections of the coronary tree, which makes it difficult for cardiologists in clinical practice to understand the global 3D coronary vascular structure. This is complicated by the vessel overlap, foreshortening, complex vascular structure, and the artifacts caused by cardiac and respiratory motions and possible patient and device movements. These may all negatively affect the physicians' ability to locate the artery stenosis areas and navigate during clinical interventions [11]. To alleviate this, cardiologists tend to use extra doses of contrast and x-rays to get more views for assessing the lesion. However, potential chemotoxic adverse reactions of radiographic contrast and x-ray radiation risk restrict the number of projections acquired to typically 2-5 projections. Therefore, the development of 3D coronary tree reconstruction based on limited (two) 2D coronary angiography projections has large clinical significance [8].

3D coronary tree reconstruction from ICA images poses significant challenges: the complex vascular shape and limited projections provide limited information on 3D vessel structures. Most importantly, due to the non-simultaneous image acquisition, significant cardiac and respiratory non-rigid motions cause vessels to misalign between projections, which aggravates the difficulties of 3D reconstruction. Biplane x-ray angiography systems can simultaneously capture two projections and hence, unaffected by such non-rigid motions; however, they are expensive for clinical usage. Many traditional methods have been proposed for 3D coronary tree reconstruction from 2D non-simultaneous x-ray projections [8]. However, they usually require substantial manual annotation and often cannot correct non-rigid cardiac motion. In terms of 3D coronary tree reconstruction using deep learning, previous methods have typically used synthetic data, coronary computed tomography angiography (CCTA), or ICA data from bi-planar scans, none of which suffer from non-rigid motion between projections

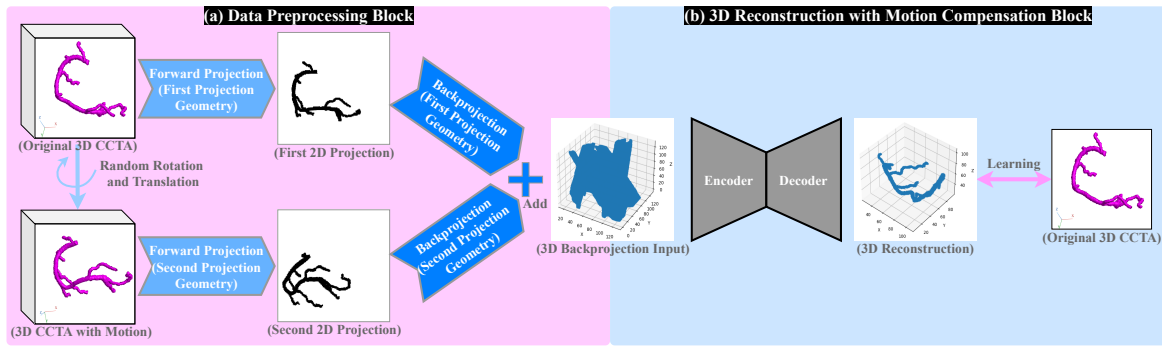


Figure 1. The overall workflow of our proposed DeepCA pipeline consists of a data preprocessing block and a 3D reconstruction with motion compensation block. (a) The data preprocessing block generates two simulated ICA projections from 3D CCTA data, including simulated motion between projections, and then produces the 3D model input via performing backprojection on the two simulated projections. (b) The 3D reconstruction with motion compensation block receives the 3D backprojection input to train a deep neural network for 3D coronary tree reconstruction learned from the CCTA data, implicitly compensating for any motion.

[5, 17, 19, 24, 28, 30, 32]. This limitation makes previous methods ill-suited for real non-simultaneous ICA acquisitions. Despite the improvement in deep neural networks, 3D coronary tree reconstruction from limited non-simultaneous angiographic projections has remained an open problem.

In this paper, we propose a novel deep learning pipeline named *DeepCA*, leveraging the Wasserstein conditional generative adversarial network with gradient penalty, latent convolutional transformer layers, and a dynamic snake convolutional critic to implicitly compensate for the non-rigid motion to achieve 3D coronary artery tree reconstruction from two real non-simultaneous ICA projections. To resemble real non-simultaneous ICA projections, we simulate 2D projections in different planes from CCTA data containing real coronary tree geometries, with a rigid transformation applied to the CCTA data before forward projection on the second projection plane. We then use these simulated projections to learn from the CCTA ground truth to enable generalisation to real non-simultaneous ICA projections. In this way, we overcome the problems of both the limited number of real paired ICA data with projection geometry information and the unavailable 3D ground truth for real ICA data. We focus on the right coronary artery (RCA) in this study, because RCA undergoes more compressive strain and is affected more by motion artifacts than other coronary vessels. We provide an application-specific evaluation method to tackle the deformation in 3D reconstructions, the unavailability of 3D ground truth for real ICA scans, and the motion between projection planes, together with Chamfer ℓ_2 distance. We validate our proposed model on a CCTA dataset and a real ICA dataset (the out-of-distribution domain), in comparison to four other models. The evaluation results demonstrate the promising performance of our proposed model in vessel topology preservation, recovery of missing features, and generalisation ability in 3D coronary

tree reconstruction from real non-simultaneous ICA projections. The main contributions of this work are:

1. **3D coronary tree reconstruction using deep learning:** To the best of our knowledge, this is the first study that leverages deep learning to achieve 3D coronary tree reconstruction from two real non-simultaneous angiography projections.
2. **Generalisation:** Through simulating projections from CCTA data, we achieve generalisation on 3D coronary tree reconstruction from two real non-simultaneous ICA projections.
3. **Extensive evaluation:** We use a specific metric designated for this problem in the absence of motion-free 3D ground truth, which provides a baseline for future improvement in this area.

2. Related Works

2.1. 3D Coronary Tree Reconstruction from Limited Views

Much effort has been devoted to 3D coronary tree reconstruction from ICA projections by conventional mathematical methods [8]. Several methods have been attempted for 3D reconstruction from limited views, including iterative reconstruction [23], non-uniform rational basis splines [10, 29], and point-cloud method [2]. However, most of these methods are dependent on traditional stereo-vision algorithms, which usually require significant manual interactions such as label annotations, and they often cannot correct non-rigid cardiac motion [1, 3].

Few studies have attempted deep learning-based 3D coronary tree reconstruction from several projections. In [17, 28], a supervised learning model was trained using 9 projections simulated from 3D synthetic coronary tree

data. Iyer *et al.* [19] proposed a vessel generator to generate 3D synthetic coronary tree data, in order to train supervised multi-stage networks based on multi-layer perceptrons to predict both vessel centrelines and radii simultaneously, from 3 projections. Wang *et al.* [30] used CCTA data to simulate projections without motion and train a weakly-supervised adversarial learning model for 3D reconstruction from two projections. Maas *et al.* [24] proposed a neural radiance fields (NeRF)-based model to tackle the problem without involving 3D ground truth in training from 4 projections. Wang *et al.* [32] proposed an implicit neural representation-based method to reconstruct 3D coronary tree from two simulated CCTA projections. All of these aforementioned studies only used synthetic data or CCTA data for both training and testing without taking motion between projections into account, and were never applied to real non-simultaneous ICA projections. Bransby *et al.* [5] used real ICA data to reconstruct a single coronary tree branch; but their acquisitions were based on bi-planar scans and so, no motion occurred between two projections.

To the best of our knowledge, our proposed method is the first to leverage deep learning to solve the problem of motion artifacts that exist between real non-simultaneous ICA projections acquired from single-plane x-ray angiography systems.

2.2. 3D Reconstruction from Limited Views in Medical Images

Many deep learning algorithms have demonstrated promising results on the limited-angle computed tomography (CT) reconstruction problem recently, giving it potential for clinical diagnosis [31]. In [33], a differentiable filtered backprojection-based neural network with an image domain U-Net was presented to achieve satisfactory reconstruction results from limited projections, while a fast self-supervised solution was proposed in [38] with promising results in terms of both speed and quality. However, the minimum number of projections required for the aforementioned methods were 145 and 50, respectively. Based on conditional generative adversarial networks, X2CT-GAN [35] was proposed to reconstruct CT from bi-planar x-rays. Further to this method, the CCX-RayNet was proposed with an addition of transformers [27], achieving more accurate results than X2CT-GAN. An end-to-end encoder and decoder convolutional neural network was adopted in [20] for 3D knee bones reconstruction from bi-planar projections. Cafaro *et al.* [6] proposed an unsupervised generative model with prior knowledge of anatomic structures to reconstruct 3D tomographic images of head and neck from bi-planar x-rays. However, these methods tackle the 3D reconstruction problem from simultaneous projections that do not suffer from non-rigid motion between projections. Park *et al.* [25] proposed a framework for 3D teeth reconstruction from

one panoramic radiograph using neural implicit functions, where the teeth are first segmented in the radiograph followed by 3D teeth reconstruction. However, their training involved the 3D ground truth, which is usually unavailable for our problem of 3D coronary tree reconstruction from two non-simultaneous ICA projections.

3. Proposed Pipeline

Our proposed DeepCA method consists of two blocks: a data preprocessing block and a 3D reconstruction with motion compensation block, as illustrated in Fig. 1. In the data preprocessing block, we generate two simulated ICA projections based on 3D CCTA data, with simulated motion on the second projection plane, and then apply backprojection on them to produce the input to the model at the next block. In the 3D reconstruction with motion compensation block, we map the 3D backprojection input to the CCTA data for 3D coronary tree reconstruction via training a deep neural network, implicitly compensating for any motion.

3.1. Data Preprocessing Block

Breathing and cardiac motions introduce deformations to the coronary tree between projections. We use deep learning to implicitly compensate for these motion artifacts. In the generation of simulated projections, we introduce rigid transformations to the CCTA data before performing forward projection on the second projection plane to simulate motion, as illustrated in Fig. 1. We use the projection geometry of real coronary angiography to simulate the two cone-beam forward projections. Details of the projection geometry parameters are provided in the supplementary material.

Since contrast injections used in real ICA projections change image intensity values between projections, we first segment vessels from the images to binarise them, where points on vessels are assigned as 1 and background as 0. In order to generalise to real ICA projections, we binarise the simulated projections with a threshold of 0 as well, i.e. any points with values greater than 0 are set to 1 and 0 otherwise. Using the known projection geometry, we perform backprojection on both binary simulated projections separately. We binarise the two 3D backprojections with a threshold of 0 and add them together to generate a single 3D input to our model.

3.2. 3D Reconstruction with Motion Compensation Block

We train a model to map the 3D backprojection result to its corresponding CCTA ground truth. Our DeepCA model architecture is based on the Wasserstein conditional generative adversarial network (WCGAN) with gradient penalty, latent convolutional transformer layers, and a dynamic snake convolutional critic, as illustrated in Fig. 2. Via mapping the input with non-aligned projections to 3D

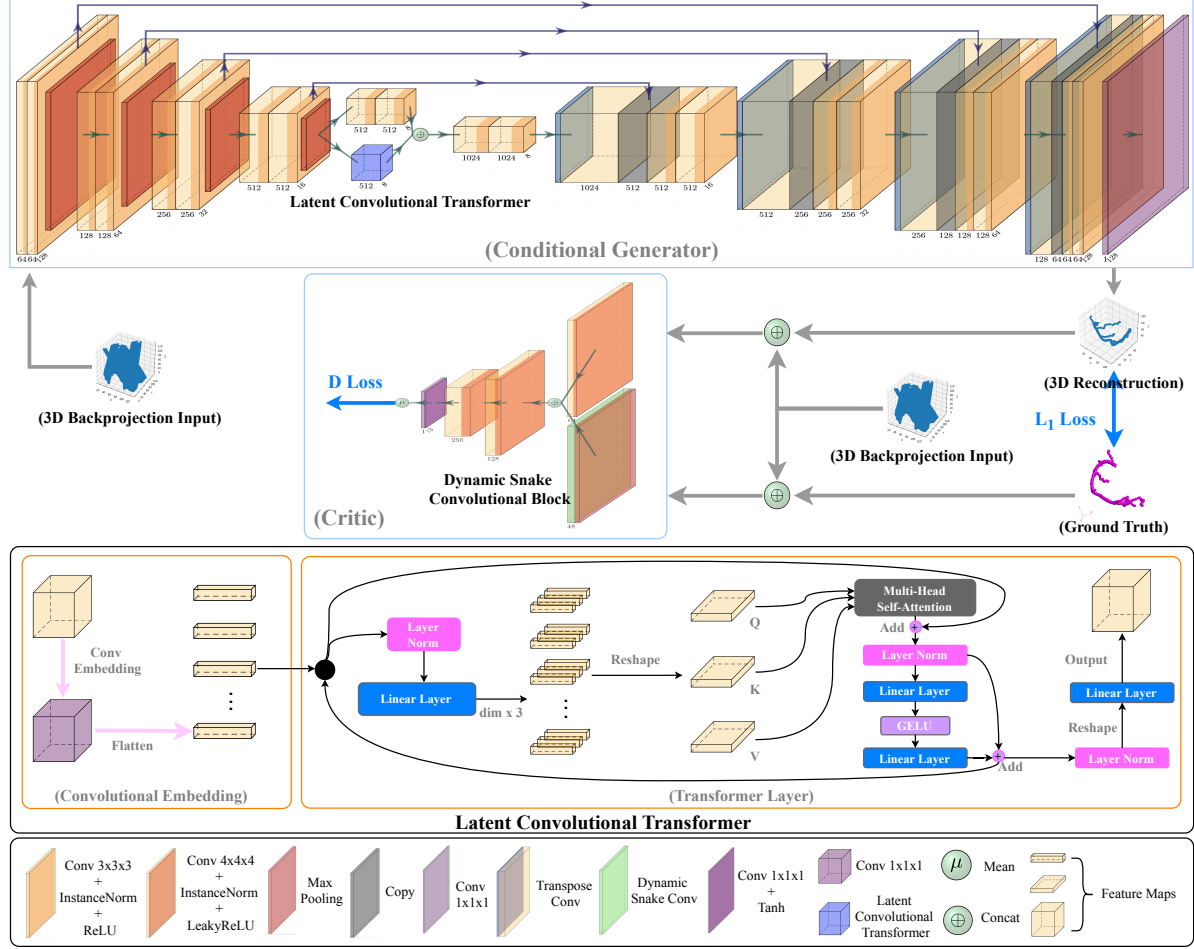


Figure 2. The proposed DeepCA model architecture includes a conditional generator and a critic. The conditional generator is based on 3D U-Net with additional proposed convolutional transformer layers in the latent space. The generator produces corresponding reconstructed results according to the input condition. The latent convolutional transformers are built on convolutional embeddings following 8 transformer layers. The predicted results and the corresponding ground truth are concatenated with the input separately, which are then sent to the critic. The proposed critic uses both dynamic snake convolution and traditional convolution at the first layer to extract both global tubular and local features, and then applies several downsamplings to generate the critic loss.

coronary tree data, most motion artifacts are corrected by our model. With the critic used, any residual uncorrected deformations are adjusted, while ensuring the connectedness of the coronary tree structures in the reconstructions and increasing the model’s elastic generalisation capacity. So when generalising to real ICA projections, the non-rigid motion is compensated implicitly.

WCGAN with gradient penalty (WCGAN-GP). The conditional structure of the GAN [18] enables us to generate a desired output from a specific input, and the Wasserstein adversarial objective with an additional gradient penalty constraint [12] improves training stability. The model consists of an encoder-decoder generator G and a critic D . The 3D backprojection result x is the input to the gener-

ator G , which has the 3D U-Net [7] as backbone, producing the predicted reconstruction \hat{y} as output. To ensure strict learning from the 3D backprojection input x to the corresponding ground truth y , we keep the input x without any added noise, in contrast to previous style transfer applications. The predicted reconstruction \hat{y} and the corresponding ground truth y are then concatenated (\oplus) with the conditional input x , respectively.

$$\hat{y} = G(x), \quad \hat{y}_x = \hat{y} \oplus x, \quad y_x = y \oplus x. \quad (1)$$

Next, \hat{y}_x and y_x are used in the critic to approximate the Wasserstein distance (or, Earth-Mover distance) $W((\mathbb{P}_r)_{y_x \sim \mathbb{P}_r}, (\mathbb{P}_g)_{\hat{y}_x \sim \mathbb{P}_g})$ and gradient penalty constraint $GP(\mathbb{P}_{\hat{y}_x})$ for each data batch, where \mathbb{P}_r is the conditional ground truth data distribution, \mathbb{P}_g is the conditional model

generation distribution, and $\mathbb{P}_{\tilde{\mathbf{y}}_x}$ is the distribution sampling uniformly along straight lines between pairs of points sampled from the distributions \mathbb{P}_r and \mathbb{P}_g .

$$W((\mathbb{P}_r)_{\mathbf{y}_x \sim \mathbb{P}_r}, (\mathbb{P}_g)_{\tilde{\mathbf{y}}_x \sim \mathbb{P}_g}) = \mathbb{E}_{\mathbf{y}_x \sim \mathbb{P}_r} [D(\mathbf{y}_x)] - \mathbb{E}_{\tilde{\mathbf{y}}_x \sim \mathbb{P}_g} [D(\tilde{\mathbf{y}}_x)]. \quad (2)$$

$$GP(\mathbb{P}_{\tilde{\mathbf{y}}_x}) = \mathbb{E}_{\tilde{\mathbf{y}}_x \sim \mathbb{P}_{\tilde{\mathbf{y}}_x}} [(\|\nabla_{\tilde{\mathbf{y}}_x} D(\tilde{\mathbf{y}}_x)\|_2 - 1)^2],$$

where $\tilde{\mathbf{y}}_x = \epsilon \mathbf{y}_x + (1 - \epsilon) \hat{\mathbf{y}}_x$, $\epsilon \in U[0, 1]$. (3)

During training, the generator G tries to minimise $W(\mathbb{P}_r, \mathbb{P}_g)$ between distributions \mathbb{P}_r and \mathbb{P}_g , while the critic D tries to maximise this distance along with minimising the constraint $GP(\mathbb{P}_{\tilde{\mathbf{y}}_x})$. The objective function of the WCGAN-GP is presented in Eq. (4), where λ_1 is the penalty coefficient. We use $\lambda_1 = 10$.

$$\mathcal{L}_{\text{WCGAN-GP}}(G, D) = \arg \min_G \max_D W((\mathbb{P}_r)_{\mathbf{y}_x \sim \mathbb{P}_r}, (\mathbb{P}_g)_{\tilde{\mathbf{y}}_x \sim \mathbb{P}_g}) + \lambda_1 \min_D GP(\mathbb{P}_{\tilde{\mathbf{y}}_x}). \quad (4)$$

We additionally impose an ℓ_1 loss to enforce the reconstruction to align with the ground truth. Our final objective function \mathcal{L} is presented in Eq. (6).

$$\mathcal{L}_{\ell_1}(G) = \mathbb{E}_{\hat{\mathbf{y}}, \mathbf{y}} (\|\mathbf{y} - \hat{\mathbf{y}}\|_1), \quad (5)$$

$$\mathcal{L} = \mathcal{L}_{\text{WCGAN-GP}}(G, D) + \lambda \mathcal{L}_{\ell_1}(G). \quad (6)$$

The hyperparameter λ is set to 100 after fine-tuning. The number of critic iterations per generator iteration is set to 2.

Latent convolutional transformer layers (CTLs). Inspired by 2D compact transformers [13], we implement our 3D latent CTLs that use convolutional embeddings following transformer layers in the latent space. Since our data contain a large empty background representing non-vessel regions, the usual patch embeddings are not suitable, while in the latent space, the transformer can help us extract the relations between feature maps and enforce their importance via attention modules. We do not require positional embeddings for feature maps as they are order invariant. The latent CTLs consider the max pooling results $f_{\text{input}} \in \mathbb{R}^{N \times C \times H \times W \times D}$ from the last layer as input, where N is the batch size, C denotes the number of channels, and H, W, D stand for height, width, and depth, respectively. We next perform convolutional embeddings on the latent feature maps $f_{\text{embeddings}} \in \mathbb{R}^{N \times (H \times W \times D) \times C} =$

$\text{reshape}(\text{Conv}_{1 \times 1 \times 1}(f_{\text{input}}))$. The feature map embeddings then go through the transformer layers as follows:

$$(Q, K, V) \in \mathbb{R}^{3 \times N \times (H \times W \times D) \times C} = \text{reshape}(\text{LinearLayer}(\text{LayerNorm}(f_{\text{embeddings}}))), \quad (7)$$

$$\mathbf{z}' \in \mathbb{R}^{N \times (H \times W \times D) \times C} = \text{LayerNorm}(\text{MHSA}(Q, K, V) + f_{\text{embeddings}}), \quad (8)$$

$$\mathbf{z} \in \mathbb{R}^{N \times (H \times W \times D) \times C} = \text{LinearLayer}(\text{GELU}(\text{LinearLayer}(\mathbf{z}')) + \mathbf{z}'), \quad (9)$$

where MHSA denotes a multi-head self-attention module, GELU the activation layer of the Gaussian error linear unit, \mathbf{z}' the intermediate result after MHSA, \mathbf{z} the output of one transformer layer, and Q, K, V stand for query, keys, and values vectors, respectively. We use 8 attention heads and 8 transformer layers, where the output \mathbf{z} for one layer will replace the $f_{\text{embeddings}}$ for the next layer.

After using the transformers to encode the relations between the latent feature maps, we use a linear layer to do final mappings on these embedded feature maps and reshape the results as the same size as input.

$$f_{\text{output}} \in \mathbb{R}^{N \times C \times H \times W \times D} = \text{reshape}(\text{LinearLayer}(\text{reshape}(\text{LayerNorm}(\mathbf{z}))))). \quad (10)$$

Dynamic snake convolutional (DSConv) critic. The coronary tree is formed of quasi-tubular topological structures, and the traditional convolutional kernel is not optimal for recognising thin local structures and variable global morphologies. DSConv kernels [26] are designed specifically for such structures. Differing from the traditional 3D kernel with size $3 \times 3 \times 3$, DSConv flattens the whole kernel along different axes with random offsets to generate a dynamic snake-shaped kernel as illustrated in Fig. 3. For 3D data, DSConv generates three dynamic feature maps for X-, Y-, and Z-axes respectively and then concatenates these three feature maps together as the output. Due to its nature of dynamic offsets, DSConv can extract tubular features more efficiently. We use the DSConv in the first layer of the critic to extract global tubular features and concurrently perform traditional convolution to extract essential local features as well. We then concatenate these global and local features together, and apply downsamplings to calculate the critic loss. This design of the critic can effectively distinguish vessel tubular structures.

4. Experimental Settings

4.1. Datasets

We use a public CCTA dataset [36] containing 3D binary segmented coronary trees for our study, and split the coro-

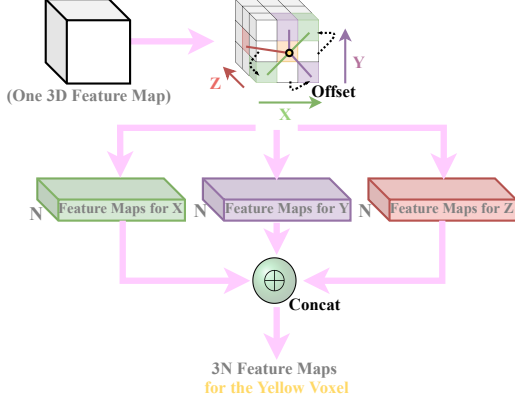


Figure 3. The dynamic snake convolution (DSConv). For each voxel (the yellow voxel in the figure as an example) in the feature map, the DSConv flattens the whole kernel along different axes with random offsets to extract different dynamic feature maps for X-, Y-, and Z-axes separately and then concatenates these feature maps together as the convolution output.

nary trees into RCA and the left coronary artery. We use 879 segmented RCA data in total, dividing them into 75% training, 15% validation, and 10% test datasets. We perform the cone-beam forward projection using the TIGRE toolbox [4]. The volume size is $128 \times 128 \times 128$, and the detector size is 512×512 . Details of other projection geometry parameters are provided in the supplementary material.

We also collect a clinical ICA dataset of 8 patients for evaluation, who were admitted at the Oxford John Radcliffe Hospital with suspected coronary stenosis and provided informed consent. For two of the patients, three ICA projections were captured, while for the rest, there were two ICA projections. Our model takes 3D backprojection from binary projections as input, so these clinical ICA data are pre-segmented and then backprojected before evaluation.

4.2. Baseline Models and Implementation Details

As there is no equivalent previous work on this problem using deep learning, we implement four models as baselines for comparative analyses. We replace the 3D U-Net in WCGAN-GP with Unet++ [39] (termed as Un2+), with Unet+++ [16] (termed as Un3+), and with DSConv Net [26] (termed as DSCN). We also implement the 3D convolutional vision transformer GAN [37] (termed as CVTG). We use Adam optimiser for both generator and critic [21], with an initial learning rate of 10^{-4} . The training was performed with a batch size of 3 on NVIDIA Quadro RTX 8000.

4.3. Metrics

We adopt the overlap using a sweeping distance threshold ($Or(d)$) for evaluation metric, where d is the distance threshold in mm unit, as proposed in ‘A Coronary Artery

Reconstruction Challenge’ [9]. $Or(d) \in [0, 1]$ with 0 representing no overlap and 1 the perfect match; the metric is equivalent to the Dice score when $d = 0$. The different d values allow us to measure reconstructions under different degrees of deformation. Let us assume the set of all vessel points is \mathcal{P}_{target} in the target data and \mathcal{P}_{pred} in the prediction. Given a threshold d , every point $\mathbf{p} \in \mathcal{P}_{target}$ is marked as belonging to the set $TPR(d)$ of true positives of the reference if there is at least one point $\mathbf{u} \in \mathcal{P}_{pred}$ satisfying $distance(\mathbf{p}, \mathbf{u}) \leq d$ and to the set $FN(d)$ of false negatives otherwise. Points $\mathbf{u} \in \mathcal{P}_{pred}$ are labelled as belonging to the set $TPM(d)$ of true positives of the tested method if there is at least one $\mathbf{p} \in \mathcal{P}_{target}$ satisfying $distance(\mathbf{u}, \mathbf{p}) \leq d$ and to the set $FP(d)$ of false positives otherwise. The overlap $Or(d)$ for a certain distance threshold d can then be calculated as:

$$Or(d) = \frac{|TPM(d)| + |TPR(d)|}{|TPM(d)| + |TPR(d)| + |FN(d)| + |FP(d)|}. \quad (11)$$

In addition, we use the Chamfer ℓ_2 distance ($CD(\ell_2)$) for measuring the corresponding voxel-wise or pixel-wise prediction errors (mm) in either 3D or 2D data according to their voxel or pixel spacing.

We evaluate the models on both the CCTA test dataset and the out-of-distribution real clinical ICA dataset. For the CCTA test dataset, we directly validate the results in 3D space after rigidly registering the ground truth to the predicted reconstruction using $Or(d)$ with $d = \{1, 2\} mm$ and $CD(\ell_2)$. Since the training ground truth is the original CCTA data used to generate the first projection, there is no motion between the original ICA data and the reprojections of the predicted reconstructions on the first projection plane, while it exists on other projection planes. For this reason, we measure the Dice score between the ICA data and reprojections on the first projection plane (same as $Or(0)$). For the second and any additional projection planes, we first rigidly register the ICA data to the reprojections. We then compute the $Or(d)$ with $d = \{1, 2\} mm$ and $CD(\ell_2)$ between them. All the 3D reconstruction results are binarised with a threshold of 0.5 before evaluation, reprojection, and visualisation. All the 2D reprojections are binarised with a threshold of 0 before evaluation and visualisation.

5. Results and Discussion

5.1. Analysis on 3D CCTA Test Dataset

As demonstrated in Table 1, our proposed DeepCA model achieves the best performance in all metrics on the CCTA test dataset compared to the 4 baseline models. This indicates our model can better capture the vessel topological structures in the source domain of the CCTA data. We also visualise the corresponding voxel-wise prediction errors in terms of $CD(\ell_2)$ between the ground truth and 3D

Table 1. Quantitative performance of our proposed DeepCA model and 4 baseline models in terms of $Ot(d)$ (%) and $CD(\ell_2)$ (mm). The Dice score (%) is equivalent to $Ot(0)$ (%). All values represent mean (\pm standard deviation), and the best results are annotated in **bold**.

Model	3D CCTA Test Dataset			2D Real Clinical ICA Dataset (Unseen Domain)						
	$Ot(1)$	$Ot(2)$	$CD(\ell_2)$	1 st	2 nd Projection			Additional Projection		
				Dice	$Ot(1)$	$Ot(2)$	$CD(\ell_2)$	$Ot(1)$	$Ot(2)$	$CD(\ell_2)$
Un2+ [39]	51.99 (± 12.17)	64.75 (± 13.28)	4.64 (± 2.01)	65.42 (± 6.68)	22.97 (± 10.82)	31.05 (± 12.76)	12.79 (± 4.61)	25.25 (± 18.40)	39.37 (± 20.75)	8.49 (± 3.26)
Un3+ [16]	55.10 (± 10.72)	68.69 (± 10.96)	4.49 (± 1.67)	62.23 (± 9.49)	22.27 (± 8.30)	30.19 (± 10.51)	12.20 (± 3.77)	32.92 (± 23.99)	42.37 (± 26.30)	7.70 (± 1.59)
DSCN [26]	61.74 (± 14.28)	72.21 (± 13.31)	3.49 (± 1.68)	83.59 (± 4.01)	30.66 (± 11.30)	42.74 (± 11.68)	7.81 (± 2.19)	53.26 (± 6.74)	67.90 (± 4.77)	3.92 (± 0.48)
CVTG [37]	61.53 (± 11.49)	73.71 (± 10.75)	3.51 (± 1.36)	76.84 (± 5.73)	32.98 (± 12.29)	44.85 (± 15.85)	8.23 (± 3.73)	49.50 (± 0.82)	64.63 (± 3.05)	4.22 (± 0.51)
DeepCA	64.21 (± 10.78)	76.25 (± 9.72)	3.22 (± 1.20)	83.31 (± 4.32)	45.70 (± 6.79)	58.39 (± 8.42)	4.51 (± 1.29)	58.58 (± 4.01)	72.88 (± 0.90)	2.81 (± 0.06)

predictions, as illustrated in Fig. 4. We can see that our proposed model can reconstruct all the branches, though the reconstructed sinoatrial nodal arteries (marked by the black boxes) have larger offsets compared to the ground truth due to deformations. Additional qualitative results for all the models on the CCTA dataset are provided in the supplementary material.

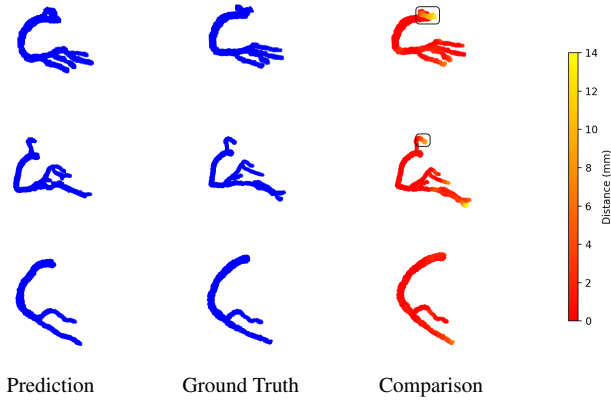


Figure 4. Three 3D reconstruction results on the CCTA test dataset by our DeepCA model. From top to bottom: three CCTA samples. From left to right: predicted reconstruction, ground truth, and the corresponding voxel-wise prediction errors in terms of $CD(\ell_2)$.

5.2. Analysis on 2D Clinical ICA Dataset

From the quantitative results presented in Table 1, we can observe that our proposed DeepCA model attains the best performance on real ICA data in all metrics on both the second and additional projection planes compared to the 4 baseline models. For the first projection plane, our model achieves the second best performance, with only 0.33% behind the DSCN in terms of Dice score. However, in the

second and additional projection planes, which are mostly affected by motion, our method performs the best. In particular, the results for the additional projection plane are the most significant, since the model is trained only based on two projections. Our proposed model presents large improvements of 38.57% and 9.99% in terms of $Ot(1)$, 42.25% and 28.32% in terms of $CD(\ell_2)$, for the second and additional projection planes respectively, compared to the best baseline models. This indicates our model has a better generalisation ability to the unseen domain.

An interesting qualitative example is presented in Fig. 5, where we observe a missing vascular section at the middle of the main RCA branch in the original ICA data, as marked by the pink box. This section is successfully recovered during 3D reconstruction, demonstrating our model’s generative ability to recover missing vascular structures that may be missing due to acquisition or panning/zooming errors.

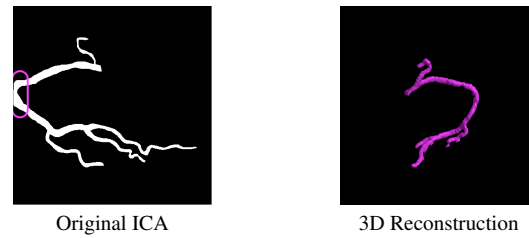


Figure 5. An example of 3D reconstruction for the RCA branch of a patient. Left: the original ICA data. Right: 3D reconstruction by our proposed DeepCA model.

Figure 6 shows three qualitative results by our proposed DeepCA model on real ICA data. The qualitative results for all the models on all the ICA data are provided in the supplementary material. We find that our model can reconstruct almost all the branches, especially the posterior descending arteries. Overall, the results illustrate good vessel

connectivity, though there exist some broken acute marginal branches as marked by the yellow boxes in Fig. 6. The reason behind this may be that those areas are affected heavily by motion, causing incomplete reconstruction. We also note that in the second projection of P_3 in Fig. 6, the reconstructed posterior descending arteries are shorter as marked by the blue box. This may be the result of non-simultaneous ICA acquisitions where the contrast agent arrives at different distances between projections causing vessel differences in the different angiographic scans.

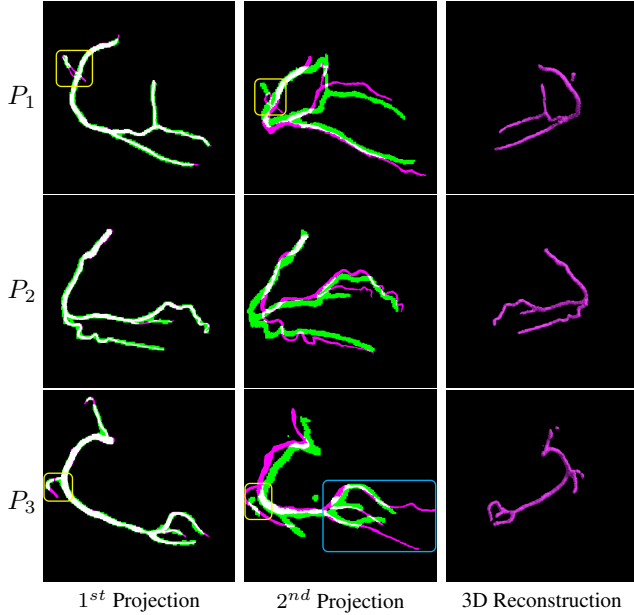


Figure 6. Three qualitative examples. From top to bottom: three patients $P_{1,2,3}$. From left to right: the comparisons between the real ICA data and our reprojections on the first and second projection planes after rigid registration, and our DeepCA model’s 3D reconstruction result. The colour purple represents ICA data, green represents reprojection, and white shows the overlap.

Figure 7 shows two example evaluations of our DeepCA model’s 3D reconstruction on an additional projection plane. It demonstrates that even without involving the information of this projection plane in the input, our model can still reconstruct the accurate vascular structures.

5.3. Ablation Study

We evaluate the significance of different components of our proposed DeepCA model through an ablation study. We evaluate (1) WCGAN-GP (termed as WGP), (2) WCGAN-GP with latent CTLs (termed as +CTLs), and (3) WCGAN-GP with DSConv critic (termed as +DSCC). As demonstrated in Table 2, the combined models consistently provide improved performance. The qualitative results in the supplementary material on real ICA data also illustrate the advantages of each component of our model.

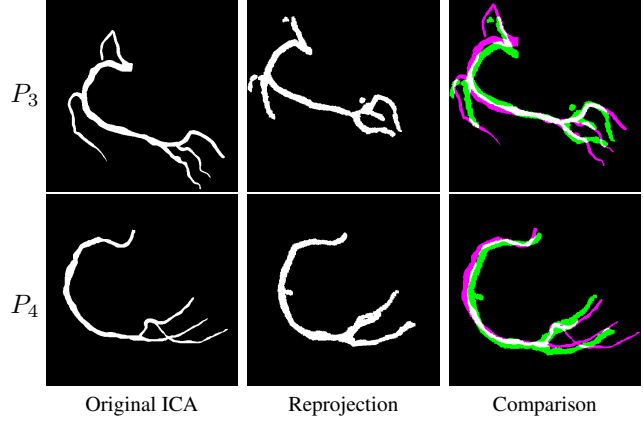


Figure 7. Two example cases of our DeepCA model’s reconstruction on the additional projection plane. From top to bottom: two patients $P_{3,4}$. From left to right: the original ICA data, the reprojection, and the comparison between them after rigid registration. The colour purple represents ICA data, green represents reprojection, and white shows the overlap.

Table 2. Quantitative results of 3 ablation models in terms of $Or(d)$ (%) and $CD(\ell_2)$ (mm). All values represent mean (\pm standard deviation), and the best results are annotated in **bold**.

Model	3D CCTA Test Dataset		
	$Or(1)$	$Or(2)$	$CD(\ell_2)$
WGP	62.06 \pm 10.61	74.38 \pm 10.01	3.43 \pm 1.29
+CTLs	62.87 \pm 11.68	74.39 \pm 10.70	3.24 \pm 1.23
+DSCC	63.46 \pm 10.85	75.14 \pm 10.00	3.24 \pm 1.23
DeepCA	64.21 \pm 10.78	76.25 \pm 9.72	3.22 \pm 1.20

6. Conclusion

In this paper, we propose DeepCA, leveraging the WCGAN with gradient penalty, latent convolutional transformer layers, and a dynamic snake convolutional critic for accurate 3D coronary tree reconstruction. Through simulating projections from CCTA data, we achieve generalisation on real non-simultaneously acquired ICA data. To the best of our knowledge, this is the first study that leverages deep learning in 3D coronary tree reconstruction from two real non-simultaneous ICA projections. The evaluations in this paper provide a baseline for future work in this area. Together with automated coronary vessels segmentation [14, 15], DeepCA can allow end-to-end automated real-time 3D coronary tree reconstruction during cardiac interventions.

Acknowledgments

AB is supported by the Royal Society University Research Fellowship (Grant No. URF\R1\221314).

References

- [1] Abhirup Banerjee, Robin P. Choudhury, and Vicente Grau. Optimized rigid motion correction from multiple non-simultaneous x-ray angiographic projections. In *Pattern Recognition and Machine Intelligence*, pages 61–69, Cham, 2019. Springer International Publishing. 2
- [2] Abhirup Banerjee, Francesca Galassi, Ernesto Zacur, Giovanni Luigi De Maria, Robin P. Choudhury, and Vicente Grau. Point-cloud method for automated 3D coronary tree reconstruction from multiple non-simultaneous angiographic projections. *IEEE Transactions on Medical Imaging*, 39(4):1278–1290, 2020. 2
- [3] Abhirup Banerjee, Rajesh K. Kharbanda, Robin P. Choudhury, and Vicente Grau. Automated motion correction and 3d vessel centerlines reconstruction from non-simultaneous angiographic projections. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*, pages 12–20, Cham, 2019. Springer International Publishing. 2
- [4] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction. *Biomedical Physics & Engineering Express*, 2(5):055010, 2016. 6
- [5] Kit Mills Bransby, Vincenzo Tufaro, Murat Cap, Greg Slabaugh, Christos Bourantas, and Qianni Zhang. 3D coronary vessel reconstruction from bi-plane angiography using graph convolutional networks. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2023. 2, 3
- [6] Alexandre Cafaro, Quentin Spinat, Amaury Leroy, Pauline Maury, Alexandre Munoz, Guillaume Beldjoudi, Charlotte Robert, Eric Deutsch, Vincent Grégoire, Vincent Lepetit, et al. X2Vision: 3D CT reconstruction from biplanar x-rays with deep structure prior. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 699–709. Springer, 2023. 3
- [7] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016. 4
- [8] Serkan Çimen, Ali Gooya, Michael Grass, and Alejandro F Frangi. Reconstruction of coronary arteries from x-ray angiography: A review. *Medical Image Analysis*, 32:46–68, 2016. 1, 2
- [9] Serkan Çimen, Mathias Unberath, Alejandro Frangi, and Andreas Maier. CoronARE: A coronary artery reconstruction challenge. In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, pages 96–104, Cham, 2017. Springer International Publishing. 6
- [10] Francesca Galassi, Mohammad Alkhalil, Regent Lee, Philip Martindale, Rajesh K Kharbanda, Keith M Channon, Vicente Grau, and Robin P Choudhury. 3D reconstruction of coronary arteries from 2D angiographic projections using non-uniform rational basis splines (NURBS) for accurate modelling of coronary stenoses. *PLoS One*, 13(1):e0190650, 2018. 2
- [11] Nathan E Green, S-Y James Chen, Adam R Hansgen, John C Messenger, Bertron M Groves, and John D Carroll. Angiographic views used for percutaneous coronary interventions: A three-dimensional analysis of physician-determined vs. computer-generated views. *Catheterization and Cardiovascular Interventions*, 64(4):451–459, 2005. 1
- [12] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 4
- [13] Ali Hassani, Steven Walton, Nikhil Shah, Abulikemu Abuduweili, Jiachen Li, and Humphrey Shi. Escaping the big data paradigm with compact transformers. *arXiv preprint arXiv:2104.05704*, 2021. 5
- [14] Haorui He, Abhirup Banerjee, Marcel Beetz, Robin P Choudhury, and Vicente Grau. Semi-supervised coronary vessels segmentation from invasive coronary angiography with connectivity-preserving loss function. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022. 8
- [15] Haorui He, Abhirup Banerjee, Robin P. Choudhury, and Vicente Grau. Automated coronary vessels segmentation in x-ray angiography using graph attention network. In *Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers*, pages 209–219, Cham, 2024. Springer Nature Switzerland. 8
- [16] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020. 6, 7
- [17] Atli İbrahim and Osman Serdar Gedik. 3D reconstruction of coronary arteries using deep networks from synthetic x-ray angiogram data. *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, 64(1):1–20, 2022. 2
- [18] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 4
- [19] Kritika Iyer, Brahmajee K Nallamothu, C Alberto Figueroa, and Raj R Nadakuditi. A multi-stage neural network approach for coronary 3D reconstruction from uncalibrated x-ray angiography images. *Scientific Reports*, 2023. 2, 3
- [20] Yoni Kasten, Daniel Doktovsky, and Ilya Kovler. End-to-end convolutional neural network for 3D reconstruction of knee bones from bi-planar x-ray images. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 123–133. Springer, 2020. 3
- [21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

- [22] Mojtaba Lashgari, Robin P. Choudhury, and Abhirup Banerjee. Patient-specific *in silico* 3D coronary model in cardiac catheterisation laboratories. *Frontiers in Cardiovascular Medicine*, 11, 2024. 1
- [23] Meihua Li, Haiquan Yang, and Hiroyuki Kudo. An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections. *Physics in Medicine & Biology*, 47(15):2599, 2002. 2
- [24] Kirsten W. H. Maas, Nicola Pezzotti, Amy J. E. Vermeer, Danny Ruijters, and Anna Vilanova. NeRF for 3D reconstruction from x-ray angiography: Possibilities and limitations. In *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2023. 2, 3
- [25] Sihwa Park, Seongjun Kim, In-Seok Song, and Seung Jun Baek. 3D teeth reconstruction from panoramic radiographs using neural implicit functions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 376–386. Springer, 2023. 3
- [26] Yaolei Qi, Yuting He, Xiaoming Qi, Yuan Zhang, and Guanyu Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6070–6079, October 2023. 5, 6, 7
- [27] Md Aminur Rab Ratul, Kun Yuan, and WonSook Lee. CCX-rayNet: A class conditioned convolutional neural network for biplanar x-rays to CT volume. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1655–1659. IEEE, 2021. 3
- [28] Gülay Yavuz Uluhan and Öğr Üys Osman Serdar Gedik. 3D reconstruction of coronary artery vessels from 2D x-ray angiograms and their pose’s details. In *2022 30th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2022. 2
- [29] Arso M Vukicevic, Serkan Çimen, Nikola Jagic, Gordana Jovicic, Alejandro F Frangi, and Nenad Filipovic. Three-dimensional reconstruction and NURBS-based structured meshing of coronary arteries from the conventional x-ray angiography projection images. *Scientific Reports*, 8(1):1–20, 2018. 2
- [30] Lu Wang, Dong-xue Liang, Xiao-lei Yin, Jing Qiu, Zhiyun Yang, Jun-hui Xing, Jian-zeng Dong, and Zhao-yuan Ma. Weakly-supervised 3D coronary artery reconstruction from two-view angiographic images. *arXiv preprint arXiv:2003.11846*, 2020. 2, 3
- [31] Tao Wang, Wenjun Xia, Jingfeng Lu, and Yi Zhang. A review of deep learning CT reconstruction from incomplete projection data. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 8(2):138–152, 2024. 3
- [32] Yiying Wang, Abhirup Banerjee, and Vicente Grau. NeCA: 3D coronary artery tree reconstruction from two 2D projections by neural implicit representation. *arXiv preprint arXiv:2409.04596*, 2024. 2, 3
- [33] Yiying Wang, Tao Yang, and Weimin Huang. Limited-angle computed tomography reconstruction using combined FDK-based neural network and U-Net. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1572–1575, 2020. 3
- [34] World Health Organisation. Cardiovascular diseases (CVDs), 2021. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)), Last accessed on 2024-06-28. 1
- [35] Xingde Ying, Heng Guo, Kai Ma, Jian Wu, Zhengxin Weng, and Yefeng Zheng. X2CT-GAN: reconstructing CT from biplanar x-rays with generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10619–10628, 2019. 3
- [36] An Zeng, Chunbiao Wu, Guisen Lin, Wen Xie, Jin Hong, Meiping Huang, Jian Zhuang, Shanshan Bi, Dan Pan, Najeeb Ullah, Kaleem Nawaz Khan, Tianchen Wang, Yiyu Shi, Xiaomeng Li, and Xiaowei Xu. 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. 5
- [37] Pinxian Zeng, Luping Zhou, Chen Zu, Xinyi Zeng, Zhengyang Jiao, Xi Wu, Jiliu Zhou, Dinggang Shen, and Yan Wang. 3D CVT-GAN: A 3D convolutional vision transformer-GAN for PET reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 516–526, 2022. 6, 7
- [38] Ruyi Zha, Yanhao Zhang, and Hongdong Li. NAF: Neural attenuation fields for sparse-view CBCT reconstruction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 442–452. Springer, 2022. 3
- [39] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, 2018. 6, 7