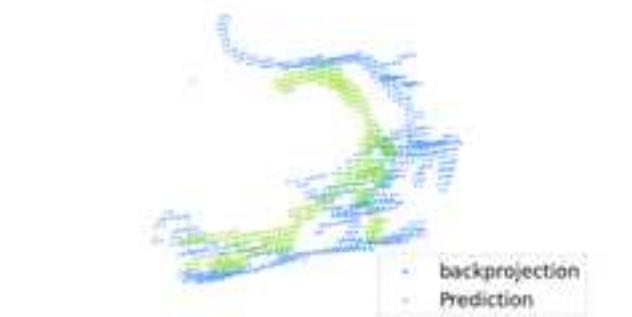
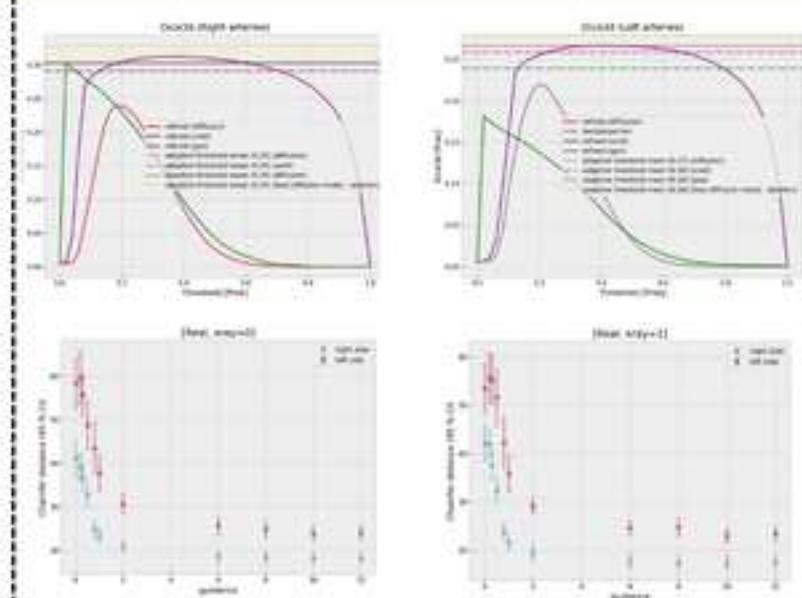
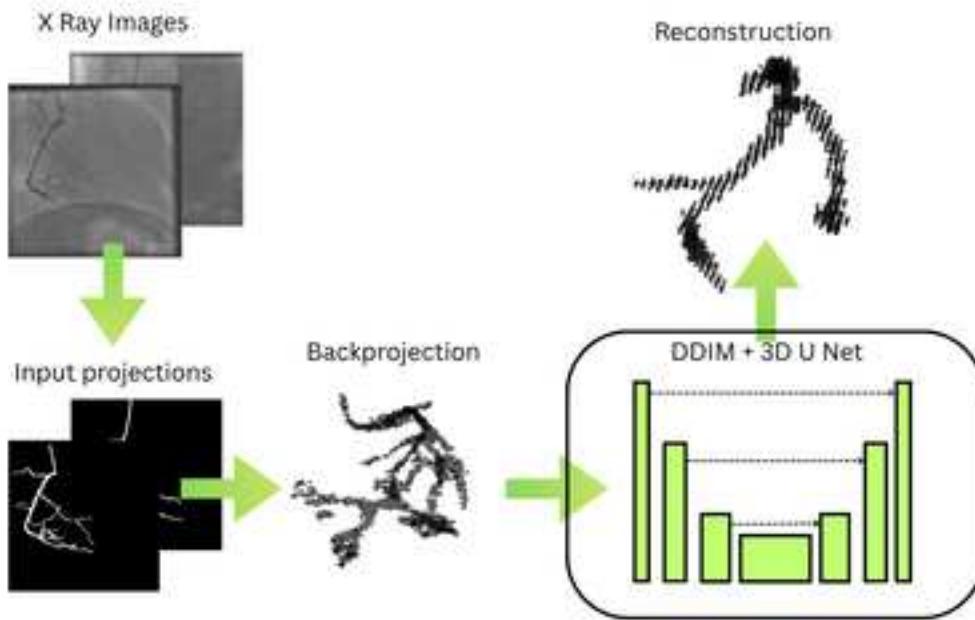


Applied Soft Computing Journal

XRA-Diff: Guided diffusion as a refinement for backprojected coronary angiography X-ray images --Manuscript Draft--

Manuscript Number:	ASOC-D-25-12568
Article Type:	Full Length Article
Keywords:	diffusion; x-ray coronary angiography; 3D reconstruction; deep learning; biomedical engineering
Corresponding Author:	Dominik Bernard Lau Gdańsk University of Technology Gdańsk, Pomorskie POLAND
First Author:	Dominik Lau
Order of Authors:	Dominik Lau Hubert Malinowski Jerzy Szyjut Tomasz Dziubich
Abstract:	In this paper a novel coronary artery 3D reconstruction method was introduced. The approach first backprojects two input x-ray images to retrieve first approximation of three-dimensional coordinates. Afterwards the point cloud is refined using a classifier free guidance diffusion model - we use two separate models for left and right arteries. The sampled result is then denoised and returned as algorithm output. The method was trained on data synthesized with ImageCAS dataset. Tests were performed on the mentioned dataset as well as on 100 samples of clinical data. A few variants of the approach were also presented together with the performance metrics. For ImageCAS 3D metrics were computed and for the clinical data, due to the lack of ground truth, reprojection was used for performance assessment. The method achieves state-of-the-art both on ImageCAS and on the clinical data and is a generative alternative to other approaches to the task

XRA-Diff: Guided diffusion as a refinement for backprojected coronary angiography X-ray images

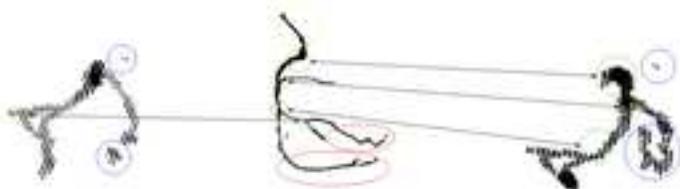


Method description

1. Backprojection is created from two X ray images.
2. Two separate models are used for right side vs left side reconstruction.
3. The backprojection is used for conditioning a diffusion model.
4. DDIM sampling with 10 steps is performed to obtain a refined backprojection.

Results

- Diffusion-based method for 3D coronary artery reconstruction achieves state-of-the-art performance on synthetic and clinical data.
- The method improves the visual features of the reconstructed arteries.



- A novel diffusion-based method for 3D coronary artery reconstruction is introduced
- The approach uses classifier-free guidance with separate models for left and right arteries
- The method is trained on synthetic data generated from the ImageCAS dataset
- Evaluation is performed on both ImageCAS and clinical dataset
- The method achieves state-of-the-art on both synthetic and clinical data

Dominik Lau

Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology
ul. Gabriela Narutowicza 11/12
80-233 Gdańsk, Poland
dominik.lau@pg.edu.pl

Editors of Applied Soft Computing

Dear Editors

I am pleased to submit our manuscript entitled “*XRA-Diff: Guided diffusion as a refinement for backprojected coronary angiography X-ray images*” for consideration as a Full Article in *Applied Soft Computing*.

In this work, we present a novel method for 3D coronary artery reconstruction from two angiographic X-ray images. The approach first generates an initial 3D point cloud via backprojection, which is then refined using a classifier-free guidance diffusion model, trained on synthetic projections generated from the ImageCAS dataset. We evaluate the method on both the synthetic dataset and clinical data, demonstrating state-of-the-art performance compared to existing approaches. Qualitative examples illustrate the robustness of our model, which provides a generative alternative to conventional reconstruction methods, particularly in scenarios where clinical ground truth is limited.

We believe this manuscript will appeal to the readership of *Applied Soft Computing*, as it presents a novel computational approach for coronary artery reconstruction with potential applications in clinical practice.

The authors declare no conflicts of interest. This manuscript is original, has not been published elsewhere, and is not under consideration by any other journal. We warrant that the article is the author’s original work.

Thank you for your consideration. We look forward to your feedback.

Sincerely,

Dominik Lau

Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology

[Click here to view linked References](#)

XRA-Diff: Guided diffusion as a refinement for backprojected coronary angiography X-ray images

Dominik Bernard Lau^{DOI a,*}, Hubert Malinowski^{DOI a}, Jerzy Szyjut^{DOI a}, and Tomasz Dziubich^{DOI a}

^aFaculty of Electronics, Telecommunications and Informatics, Department of Computer Architecture, Gdańsk University of Technology, Gabriela Narutowicza 11/12, 80-233 Gdańsk, Poland

*Correspondence: dominik.lau@pg.edu.pl

November 19, 2025

In this paper a novel coronary artery 3D reconstruction method was introduced. The approach first backprojects two input x-ray images to retrieve first approximation of three-dimensional coordinates. Afterwards the point cloud is refined using a classifier free guidance diffusion model - we use two separate models for left and right arteries. The sampled result is then denoised and returned as algorithm output. The method was trained on data synthesized with ImageCAS dataset. Tests were performed on the mentioned dataset as well as on 100 samples of clinical data. A few variants of the approach were also presented together with the performance metrics. For ImageCAS 3D metrics were computed and for the clinical data, due to the lack of ground truth, reprojection was used for performance assessment. The method achieves state-of-the-art both on ImageCAS and on the clinical data and is a generative alternative to other approaches to the task.

Keywords: diffusion, x-ray coronary angiography, 3D reconstruction, deep learning, biomedical engineering

1 Introduction

1.1 Justification

Cardiovascular diseases are the leading cause of death in the EU, making for 32.7 [%] of all deaths [1], out of which 32.4 [%] is attributed to Ischaemic Heart Disease [2]. Obstructive coronary artery disease - a leading cause of IHD, often resulting from the abnormal build-up of cholesterol plaque - is diagnosed by the gold-standard procedure of X-ray coronary angiography (XRA). Approximately 4 million angiograms are performed each year in Europe and the United States [3] therefore it is critical to strive for the optimization of diagnoses making process. However, a considerable amount of 3D information of the coronary arteries is lost - the examination creates a series of, inherently 2D, X-ray images on which coronary arteries, having been injected a dose of contrast material, are visible [4]. An obvious way of addressing the limitations posed by 2D representation would be to find 3D reconstruction best representing the depicted arteries, reducing the influence of subjective impressions on experts' conclusions as to whether a clinical treatment should be started for an individual or not [5]. Unfortunately, not only is the task ill-posed, as it tries to restore information that has been essentially lost in the acquisition process, but also, the spatial notions that may have been left on the images are greatly limited, requiring the use of multiple projection angles for successful reconstruction. To add to the complexity, no public dataset containing both ground-truth 3D volumes and x-rays is available, as it would require performing both the angiography and computerized tomography on an individual [6], which is rarely clinically justified. This lack of data jeopardizes deep learning-based approaches to the task.

1.2 Related works

1.2.1 3D reconstruction

Many approaches to 3D reconstruction of heart arteries from XRA have been explored. Galassi et al. [7] and Vukicevic et al. [8] presented methods based on extracting vessel centerlines from projections and matching corresponding points between them. However, these approaches require indicating endpoints by an expert and are sensitive to motions. Cong et al. [9] proposed a deformable model-based method for reconstructing the 3D geometry of coronary arteries from multiple 2D angiographic projections. In this approach, the parametric snake evolves toward the real vascular centerline in 3D space based on the integrated internal energy and composited external energy.

State-of-the-art results are achieved using deep learning methods. Fu et al. [10] introduced 3DGR-CAR. In the first stage, a U-Net network is used to estimate voxel depth from a single projection. In the second stage, positional parameters are

obtained from the image using U-Net. Wang et al. [11] proposed NeCA, approach presenting an application of NeRF (Neural Radiance Fields) [12] to the domain. The NeCA model utilizes a multiresolution hash encoder to efficiently represent the 3D vessel structure from just two 2D projections. As a self-supervised method, NeCA is optimized for each data point individually, meaning it does not transfer knowledge between different cases. This leads to a long reconstruction time for each coronary tree. Furthermore, Wang et al. [13] presented DeepCA - reconstruction algorithm based on Wasserstein conditional generative adversarial network (WCGAN) with gradient penalty [14] having a backprojection at input. However, above method was validated only on right coronary artery. Black-box approach was also explored by Iyer et al. [6] hinting at a limited scalability of the vanilla convolutional neural networks towards the task. Brandby et al. [15] presented 3DAngioNet, a reconstruction algorithm based on Graph Convolutional Networks (GCNs) that utilizes bi-plane angiography. However, this method requires an expert to define the start and end points of the segment of interest. Thus, it cannot be used as a fully-automated method. In Tab. 1 a summary of key traits of the above methods was presented.

Approach	Type	Keypoints	Automated
Galassi et al. [7]	Centerline projection method	Eight-connectivity seed-fill algorithm, Non-Uniform Rational B-Spline (NURBS)	semi-automated
Vukicevic et al. [8]	Centerline projection method	Partial matching, Non-Uniform Rational B-Spline (NURBS)	semi-automated
Cong et al. [9]	Deformable model	Parametric snake, internal and external energies	semi-automated
Fu et al. [10]	Deep learning	3D Gaussian Representation, U-Net network	fully-automated
Wang et al. [11]	Deep learning	NeRF-based, self-supervised, multiresolution hash encoder, differentiable cone-beam projector	fully-automated
Wang et al. [13]	Deep learning	Wasserstein conditional generative adversarial network, gradient penalty, backprojection	fully-automated
Brandby et al. [15]	Deep learning	EfficientB3-UNet to segmentation and projection geometries, GCN to deformation	semi-automated

Table 1: Existing approaches to 3D reconstruction summary.

1.2.2 Diffusion

Diffusion models are a model family that learn the decoder given a fixed encoder [16]. The encoder maps ground truth and a timestep to z_t i.e. it represents $q(z_t/x)$ given a latent representation z_t . The decoder then attempts to learn $p(z_{t-1}/z_t)$ which inverts the encoding process. Finally, for the reconstruction task, $z_0 = \mathbf{x}_{\text{diffusion}}$ is equivalent to the reconstructed volume. Such convergence of a diffusion model producing realistic data after denoising requires an access to a large dataset [17].

As the main purpose of the diffusion model is to generate new data through denoising, additional conditioning and guidance must be performed to ensure that the condition $\mathbf{x}_{\text{diffusion}} \approx \mathbf{x}_{gt}$ is fulfilled. The former is usually achieved through the injection of latent conditioning vectors to the attention layers or concatenation to input. The latter can be performed by a separate classifier or through a classifier-free guidance [18], which is used in the paper. In the approach the condition is being dropped in forward propagation during the training with $\text{Prob}(\text{drop condition}) = p_{-c}$, then during the sampling step (that is the computation of $\hat{\mathbf{z}}_t \approx \mathbf{z}_t$), a linear combination of unconditional and conditional predictions is used

$$\hat{\mathbf{z}}_t = \hat{\mathbf{z}}_{t:\text{unconditional}} + \gamma(\hat{\mathbf{z}}_{t:\text{conditional}} - \hat{\mathbf{z}}_{t:\text{unconditional}})$$

where γ is a guidance scale.

The diffusion models have already been applied in cardiology, albeit to different tasks - Guo et al. proposed VesselDiffusion [19] utilizing 3D point cloud generation diffusion network conditioned on 2D images to synthesize realistically looking coronary vessel systems. Zhao et al. [20] used conditional diffusion for the generation of realistic pairs of projections with keypoints matching. Li et al. [21] used diffusion for noise-to-box process, each box matching a candidate for stenosis location. There have also been approaches to three-dimensional reconstruction via diffusion among different domains. Jeong, Yoo and Chun proposed DX2CT [22], a diffusion model for the reconstruction of biplanar or monoplanar x-ray images to volumes, which was then benchmarked on the Lung Image Database Consortium (LIDC) datasets [23] while Yoon et al. [24] validated the proposed diffusion-based approach on wrist X-rays.

1.2.3 Segmentation

Although the study assumes the input data is presegmented (binary segmentation is out of the scope of this work), this section reviews recent findings that have been made in that direction. For the selected segment such as one of the main segments: LCX, LAD, RCA or a limited subset of the most common segments (frequency of particular segments following SYNTAX taxonomy [25] being discussed by the ARCADE dataset authors [26]) multiple methods have been shown to achieve high alignment with ground truth. One of such methods proposed by Zhang et al. [27] uses modules for providing context (main vs secondary vessel segments), boundary perception and contrast enhancement reaching an impressive Dice score greater than 95%. Another approach named PSPNet [28] (a variation on U Net) was reported to achieve recall of $\approx 94\%$ for a task of highlighting all of the displayed (and clinically significant) arteries. In summary, the posed problem of binary segmentation is widely addressed by accurate approaches.

1.3 Goal of the study

This study contributes a novel approach to the domain of coronary artery three-dimensional reconstruction, that is the usage of guided diffusion models, to authors knowledge, not applied to XRA. The above method is examined in terms of clinical feasibility and the performance is benchmarked against the state-of-the-art methods. Furthermore, an analysis of the impact of hyperparameters such as denoising steps and guidance strength is performed, followed by ablation study. As we prioritize research transparency, the source code of methods used in the paper together with the experiments and metric measurements were published under github.com/cvlab-ai/xra-diff alongside an archive containing trained model weights used in the study in [Zenodo record 17413997](https://zenodo.org/record/17413997).

Statement of contribution

This work advances the field of three-dimensional coronary artery reconstruction from X-ray angiography (XRA) through the following contributions:

- **Novel approach:** Introduction of a novel diffusion-based method for 3D reconstruction of coronary arteries from XRA modality.
- **Performance benchmark:** Establishment of a new state-of-the-art of the domain.
- **Comprehensive quantitative analysis:** Extensive comparison of the method with the selected baselines.
- **Qualitative demonstration:** Evaluation of the quality of the most interesting parts of reconstructed coronary trees.

2 Materials and methods

2.1 Dataset synthesis

To construct a training dataset ImageCAS was used [29]. It is a dataset of 1000 volumes of coronary arteries from computerized tomography. To simulate the targeted modality, that is XRA, Open Discretization Library (ODL) operators were used [30], specifically cone beam camera model. For each volume two projections were acquired according to the parameters present in Tab. 2. The angles were intended to mimic the ones used in clinical practice [31]. Left and right artery cases were generated separately.

right side	1. proj.	2. proj.	left side	1. proj.	2. proj.
SID [mm]	$\in [970, 1010)$	$\in [1050, 1070)$	SID [mm]	$\in [1030, 1200)$	$\in [960, 1160)$
SOD [mm]	$\in [745, 785)$	$\in [742, 788)$	SOD [mm]	$\in [740, 785)$	$\in [737, 852)$
primary angle (α) [°]	$\in [0, 50)$	$\in [-18, -42)$	primary angle (α) [°]	$\in [-18, 18)$	$\in [-47, -22)$
secondary angle (β) [°]	$\in [-20, 20)$	$\in [-18, 18)$	secondary angle (β) [°]	$= 0$	$\in [-10, 45)$
common					
		image resolution	512 × 512		
		grid resolution	128 × 128 × 128		
		grid spacing	$\in [0.8, 1.1)$		
		image spacing	$\in [0.2769, 0.2789)$		

Table 2: Parameters for each projection.

The X-ray acquisition model is a geometry with circular acquisition curve, symmetry axis perpendicular to source and detector circles. The following holds for the aforementioned circles radii

$$r_{source} = SOD \quad (1)$$

$$r_{detector} = SID - SOD \quad (2)$$

further, the isocenter of the DICOM coordinate system lies in the origin of such geometry. Finally, heart artery voxels are projected onto image planes via Radon transform [32].

2.2 Clinical dataset

The dataset utilized in this study was derived from the University Clinical Center in Gdańsk. It comprised a total of 1,227 angiographic images obtained from anonymized examinations of 114 patients. The mean age of the cohort was 64.6 years, with a median age of 65 years. Male participants accounted for 58% of the study population, with ages ranging from 34 to 87 years, while females represented 40% of the cohort, aged between 69 and 91 years. The gender of two patients was not specified. Data acquisition was performed using a single angiography system — Siemens AXIOM-Artis through the years 2015–2019. The labelling process was executed using AngioTagger software throughout the years 2019 -2021 and performed by 3 specialists in interventional cardiology with at least 3 years of experience. Each image was accompanied by a metadata file in JSON format, which included both general information and detailed technical parameters. The dataset comprises binarized (segmented) X-ray images of individual coronary vessel segments annotated in accordance with the SYNTAX Score standards [33], accompanied by the corresponding DICOM metadata [34], from which parameters relevant to the reconstruction algorithm were extracted, i.e.:

- SID, SOD
- acquisition angles (primary and secondary)
- imager pixel spacing

From the acquired data 51 pairs of X-ray images were selected for each side (102 samples total), to account for missing or corrupted labels.

2.3 Preprocessing

As mentioned earlier, image binary segmentation must be performed, that is for each pixel of the image denoted $im(x, y)$

$$\begin{aligned} im_{bin}(x, y) &= 1; (x, y) \in vessel \\ im_{bin}(x, y) &= 0; \text{ otherwise} \end{aligned}$$

Methods to achieve such representation were mentioned in section 1.2.3. For both input x-ray images ($xray_0$ and $xray_1$) a backprojection is performed and an intersection of the two backprojections is found

$$x_{backprojected} = x_{backprojected:0} \cap x_{backprojected:1} \quad (3)$$

the triplet $(x_{backprojected}, xray_0, xray_1)$ makes for the deep learning model input.

2.4 Architecture

The approach first splits the data into two categories based on xrays - left and right arteries. For that a convolutional neural network such as ResNet can be used [35]. Since the problem is well posed and the side is easily distinguishable (our non-formal tests on Res-Net reached the accuracy of 98.33 [%] on already mentioned ARCADE dataset), it will not be discussed any further.

Once the artery side is established, there are two separate diffusion models at work, each trained only on the particular side. In both cases the diffusion model noise predictor (both for left and right arteries) consists of three encoders - one for each x-ray image (projection encoder) and one for the backprojection. For x-ray image encoders regular 2D CNNs were used with three convolutions and a global average pooling to 64 features, then passed through fully-connected layer to obtain a 1×1024 feature vector for each image. The backprojection encoder is a 3D CNN following the exact same approach as above (three convolutions and a global average pooling). The encoded feature vectors are summed and passed as a conditioning to a 3D U-Net, injected via cross-attention layers. Aside from the regular gaussian noise (and a timestamp) that is the primary input of the diffusion model the backprojection is also concatenated as a second input channel (where the first channel represents the aforementioned noise). 3D U-Net implementation from Hugging Face diffusers package [36] was used. The module-wise summary of the architecture was presented in Tab. 3 (generated with PyTorch Lightning [37]). Finally, the approach uses

DDIM sampler [38] with CFG and a customizable number of denoising steps n . For the default version of our approach $\gamma = 7.5$ and $n = 10$. In summary, at each timestep the method computes

$$\begin{aligned} emb_0 &= CNN_{2D:0}(xray_0) \\ emb_1 &= CNN_{2D:1}(xray_1) \\ emb_2 &= CNN_{3D}(x_{backprojected}) \\ cond &= emb_0 + emb_1 + emb_2 \\ z(t) &= UNet3D(z(t+1), x_{backprojected}, cond) \end{aligned}$$

The rationale supporting such approach is the fact that semi-supervised methods have been proven to work in the domain [11]. Furthermore the splitting of data into two models (left and right arteries treated separately) was proven to be successful in other tasks [39]. Lastly, backprojected data is a noisy version of reconstructed arteries which aligns with what diffusion models are doing at heart - denoising that is and using that as model input makes it partially independent from projection angles which increases robustness as the model does not need to be trained for every possible angle. Entire flow was demonstrated in Fig. 1.

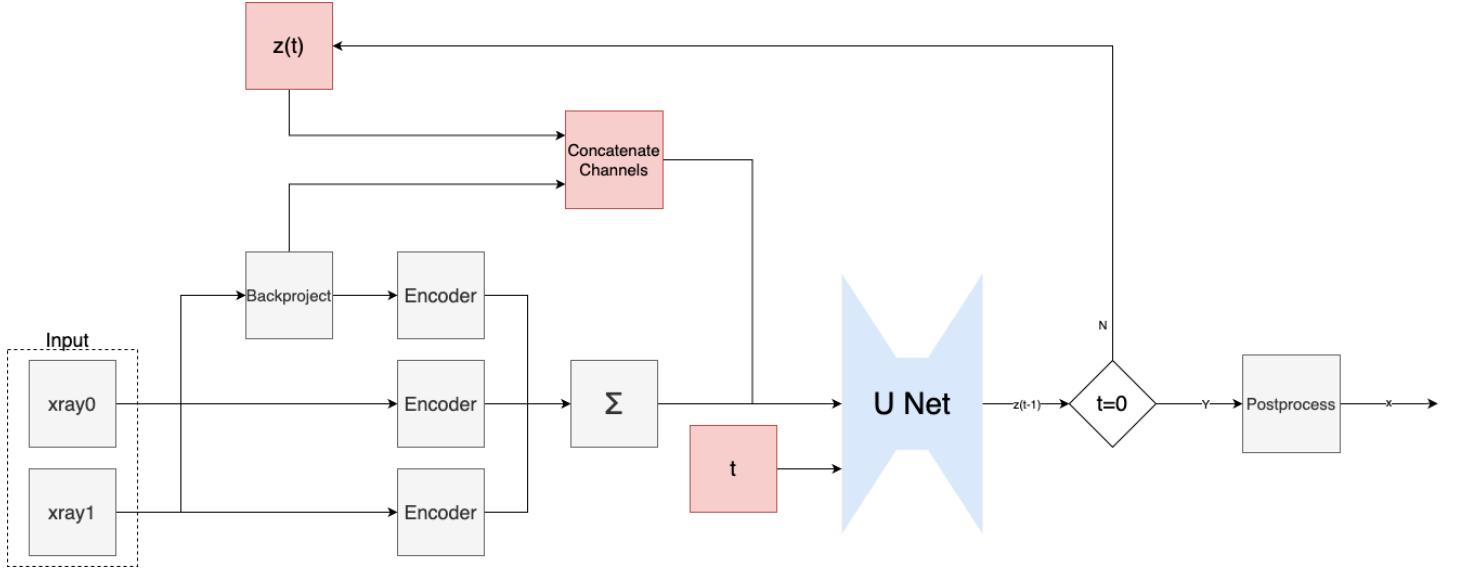


Figure 1: Data flow diagram of the proposed approach (basic version).

	Name	Type	Params
0	unet3d	UNet3DConditionModel	811 M
1	unet3d.conv_in	Conv2d	608
2	unet3d.time_proj	Timesteps	0
3	unet3d.time_embedding	TimestepEmbedding	20.7 K
4	unet3d.transformer_in	TransformerTemporalModel	5.3 M
5	unet3d.down_blocks	ModuleList	269 M
6	unet3d.up_blocks	ModuleList	427 M
7	unet3d.mid_block	UNetMidBlock3DCrossAttn	109 M
8	unet3d.conv_norm_out	GroupNorm	64
9	unet3d.conv_act	SiLU	0
10	unet3d.conv_out	Conv2d	289
11	bp_encoder	BackprojectionEncoder	289 K
12	bp_encoder.conv1	Conv3d	1.8 K
13	bp_encoder.conv2	Conv3d	110 K
14	bp_encoder.conv3	Conv3d	110 K
15	bp_encoder.pool	AdaptiveAvgPool3d	0
16	bp_encoder.fc	Linear	66.6 K
17	bp_encoder.norm1	BatchNorm3d	128
18	bp_encoder.norm2	BatchNorm3d	128
19	bp_encoder.activation	GELU	0
20	p0_encoder	ProjectionEncoder	141 K
21	p0_encoder.conv1	Conv2d	640
22	p0_encoder.conv2	Conv2d	36.9 K
23	p0_encoder.conv3	Conv2d	36.9 K
24	p0_encoder.pool	AdaptiveAvgPool2d	0
25	p0_encoder.fc	Linear	66.6 K
26	p0_encoder.activation	GELU	0
27	p1_encoder	ProjectionEncoder	141 K
28	p1_encoder.conv1	Conv2d	640
29	p1_encoder.conv2	Conv2d	36.9 K
30	p1_encoder.conv3	Conv2d	36.9 K
31	p1_encoder.pool	AdaptiveAvgPool2d	0
32	p1_encoder.fc	Linear	66.6 K
33	p1_encoder.activation	GELU	0

Table 3: Details of the model architecture. Output of ModelSummary utility from PyTorch Lightning package (maximum depth=2).

Due to the resource-wise infeasibility of the search strategies for U Net hyperparameters, those used in the paper were chosen by an educated guess.

2.5 Postprocessing

Adaptive thresholding is a technique that accounts for the varying intensity of diffusion denosing process output voxel values. In our approach we have used threshold th_δ such that

$$|\{v \in \mathbf{x}_{\text{diffusion}} | v \geq th_\delta\}| / |\mathbf{x}_{\text{diffusion}}| = \delta \quad (4)$$

$\delta = 0.99$ was used. Furthermore

$$v \in \mathbf{x}_{\text{diffusion}} \text{ s. t. } |N(v)| < \gamma \quad (5)$$

were removed where $N(v)$ is the Moore neighborhood of the voxel v . $\gamma = 10$ was selected. Thus the reconstruction

$$\hat{\mathbf{x}} = \{v \in \mathbf{x}_{\text{diffusion}} | |N(v)| < \gamma \wedge v \geq th_\delta\} \quad (6)$$

is produced.

2.6 Training

The models (left, right) were trained on synthesized data from section 2.1 of which 80% made training data and 20% were held out for testing. The training process utilized four H100 GPUs and the progress was tracked through ClearML and Tensorboard with $PSNR$ and validation/training losses. Adam optimizer was used with $lr = 10^{-5}$, batch size = 2 (VRAM limitations). Drop probability (used for CFG) was set to $p_{\sim c} = 0.1$

2.7 Experiments methodology

The quantitative experiments following this section were performed on a virtual machine with four H100 GPUs. For the synthetic dataset (that is produced from ImageCAS as described in section 2.1) two metrics were taken into account - Dice Similarity Coefficient for topology similarity evaluation and Chamfer distance for point cloud density quantification, both defined below

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (7)$$

$$d_{Chamfer}(\mathbb{P}, \mathbb{Q}) = \frac{1}{|\mathbb{P}|} \sum_{p \in \mathbb{P}} \min_{q \in \mathbb{Q}} \{d(p, q)\} + \frac{1}{|\mathbb{Q}|} \sum_{q \in \mathbb{Q}} \min_{p \in \mathbb{P}} \{d(p, q)\} \quad (8)$$

where TP, FP, FN are taken from a confusion matrix for the validation set. Moreover,

$$O(t) = \frac{2TP_{d < t}}{2TP_{d < t} + FP_{d < t} + FN_{d < t}} \quad (9)$$

Where $d < t$ means points in t voxel radius are considered the same. In later parts we compute $O(1)$ and $O(2)$ [13]. The definitions hold regardless of whether $\mathbb{P}, \mathbb{Q} \subset \mathbb{R}^3$ or $\mathbb{P}, \mathbb{Q} \subset \mathbb{R}^2$. Also peak signal-to-noise ratio (PSNR) was used.

$$PSNR(\hat{\mathbf{x}}, \mathbf{x}_{gt}) = 10 * \log_{10} \frac{\max\{\hat{\mathbf{x}}\}^2}{MSE(\hat{\mathbf{x}}, \mathbf{x}_{gt})} [dB] \quad (10)$$

Furthermore, for a basic performance assessment method, runtime τ was measured.

While the above metrics are straightforward to compute when ground truth is available (which is the case for synthetic data), clinical data, typically missing the target 3D volume resulting in the angiography images, required us to use the following approach:

1. Reconstruct the data based on input xrays $xray_0, xray_1$ i.e. obtain $\hat{\mathbf{x}}$
2. project $\hat{\mathbf{x}} \subset \mathbb{R}^3$ onto $xray_0$ & $xray_1$ planes which results in $\hat{\mathbf{x}}_0^p, \hat{\mathbf{x}}_1^p \subset \mathbb{R}^2$
3. compute the metrics defined above on those projections, that is $f(\hat{\mathbf{x}}_j^p, xray_j), j \in \{0, 1\}$ where f is substituted for DSC and $d_{Chamfer}$

That way reprojection metrics are obtained. It ought to be emphasized that for such approach the 2D reprojection metrics are maximized for backprojection, which only vaguely matches the real 3D point cloud of the arteries, therefore the looseness of the metric as a form of reconstruction quality assessment must be taken into account. Furthermore, keeping in mind the fact that the reconstruction outputs a static volume (that is not moving like a real artery tree would) the reconstruction for non-simultaneous projections will always have the reprojection metrics lowered due to the shape and position change in-between the X-Ray images. To mitigate that, erosion on 3x3 neighborhood was performed before DSC was computed. Lastly, to account for the limited dataset size, bootstrap method was used for confidence intervals computation for each metric.

For state-of-the-art comparison, the following baselines were chosen:

- the backprojection by itself (i.e. without our refinement)
- regular 3D U Net for direct reconstruction (with about the same parameter count as the one used in diffusion)
- Generative Adversarial Network (already explored in [13])

The question about synthetic dataset faithfulness to real data should be made. There are a few factors that are not taken into account when training on non-clinical data:

- heart motion
- acquisition artifacts

- isocenter shifts

In order to simulate those translation vectors $\mathbf{t} \in [-20, 20] \text{ [px} \times [-20, 20] \text{ [px]}$ and scaling coefficients $s_x, s_y \in [0.8, 1.2]$ were selected at random and associated transformations were applied to the second image. Then a separate experiment was carried out and measures taken. The parameters simulate modest variability during acquisition with vessels being roughly 2-3 mm in girth [40] and 20 pixels equivalent to 6 mm assuming an imager pixel spacing $\approx 0.3 \text{ [mm/px]}$.

It must be reiterated, that no diffusion-based method was suggested in the domain, thus none can be used in the comparison. Later section contains ablation study (see: section 3.2) of the model with respect to a selected subset of hyperparameters to account for that lack of reference.

3 Results and discussion

3.1 State-of-the-art

In this section general experiments were performed to determine the overall performance of the method. The basic variant is the one described in section 2.4, while the best variant details for each side are enclosed in a later part of the article (section 3.2).

As seen in Tab. 4 for both sides of the arteries our method in its best variant is significantly better than the rest of the methods in terms of DSC and $O(1) / O(2)$ which would suggest its increased correspondence to ground truth in terms of topology and morphological traits of the vessels. On the other side $d_{Chamfer:3d}^{synth}$ is for the right worse than 3D U Net - it can be concluded, that point cloud-locally classic method is better. The same cannot be stated for the left side, as the mean value for 3D U-Net stays in the best variant's confidence interval, thus the difference is not that significant. Chamfer distance between the reconstructed and ground truth volumes is the worst for GAN, which still reaches relatively high $DSC/O(1)/O(2)$. The basic variants of the method, while fulfilling the role of refinement of backprojections toward better 3D representation do not offer statistical improvements from using a simple 3D U-Net (in terms of the mean 3D metrics). Noticeably, there is a decrease in metric values between right and left arteries, which is a predictable result due to the left side increased tortuosity and anatomical variance. Further than the results above, Tab. 5 contains the complementary 2D reprojection metrics. $DSC^{synth, xray_0}$ and $DSC^{synth, xray_1}$ are significantly better for the right side, while the improvement is not as well pronounced for the left side.

To compare not only the clinical usability of the models, but also the robustness, in Tab. 7 reprojection metrics for real data were presented. The metric values are significantly lower, which is caused by different data distribution, nevertheless the diffusion model variants perform the best for most of the cases. It should also be clear, that the method itself has an increased robustness as the dependency on specific acquisition angles of x ray images is reduced - it is only present in conditioning, while the backprojection construction (that is an input to the models) is handled by an unparametric algorithm (i.e. inverse Radon preprocessing described in section 2.3).

Lastly the supplementary metrics should be taken into account when assessing the performance. As seen in Tab. 6, the diffusion inference process is several orders of magnitude slower than the conventional supervised architectures. Surprisingly, GAN has by far the highest $PSNR$ among the studied methods, meaning the model is concentrating points in the correct area, being the best in doing so, but missing significant local features (captured by Chamfer distance) and morphological traits (captured by Dice).

In the abovementioned tables containing results for experiments on synthetic datasets a row "Ours (Motion simulation)" was included. As can be seen there is a significant decrease in terms of 3D metric values when motion-related transforms are included, however the detrimental effects do not strongly affect $O_{3D}(2)$ as the metric amortizes the slight motions introduced. The rest of the scores are lowered, which means the decrease is caused by the shifts of the arteries in 3D. For 2D reprojection metrics the discrepancy between values for each projection is noticeable - as the transforms were applied to the second projection (i.e. $xray_1$) the measurements taken with respect to that image are significantly lower. The slight increase in the runtime of the method was caused by the measure including time to apply transformations to the second image.

Model (right)	$d_{Chamfer:3d}^{synth}$ [vox]	DSC_{3D}^{synth} [%]	$O_{3D}(1)$ [%]	$O_{3D}(2)$ [%]
Backprojection	3.25 (3.14-3.36)	0.10 (0.10-0.11)	0.22 (0.21-0.23)	0.36 (0.34-0.37)
GAN	1.93 (1.82-2.04)	0.30 (0.29-0.32)	0.54 (0.51-0.56)	0.70 (0.68-0.71)
3D U Net	1.35 (1.28-1.43)	0.29 (0.28-0.30)	0.51 (0.49-0.52)	0.70 (0.68-0.72)
Ours (Basic variant)	1.59 (1.51-1.67)	0.30 (0.29-0.31)	0.57 (0.55-0.58)	0.76 (0.75-0.78)
Ours (Best variant)	1.49 (1.41-1.59)	0.33 (0.32-0.34)	0.59 (0.58-0.61)	0.78 (0.77-0.79)
Ours (Motion simulation)	2.10 (2.01-2.20)	0.20 (0.19-0.21)	0.42 (0.40-0.44)	0.63 (0.61-0.65)
Model (left)				
Backprojection	4.10 (3.98-4.21)	0.09 (0.09-0.10)	0.19 (0.18-0.20)	0.30 (0.28-0.31)
GAN	2.61 (2.45-2.79)	0.24 (0.22-0.26)	0.44 (0.41-0.46)	0.60 (0.57-0.62)
3D U Net	1.50 (1.42-1.60)	0.26 (0.25-0.27)	0.46 (0.44-0.48)	0.65 (0.62-0.67)
Ours (Basic variant)	1.61 (1.54-1.68)	0.27 (0.26-0.28)	0.51 (0.49-0.52)	0.70 (0.68-0.71)
Ours (Best variant)	1.54 (1.48-1.61)	0.28 (0.27-0.29)	0.51 (0.50-0.52)	0.70 (0.68-0.71)
Ours (Motion simulation)	2.18 (2.09-2.28)	0.18 (0.18-0.19)	0.38 (0.36-0.39)	0.57 (0.55-0.59)

Table 4: Mean 3D metric values with 95% confidence intervals compared to state-of-the-art on synthetic data.

Model (right)	$d_{Chamfer:2d}^{synth, xray_0}$ [px]	$d_{Chamfer:2d}^{synth, xray_1}$ [px]	$DSC_{2D}^{synth, xray_0}$ [%]	$DSC_{2D}^{synth, xray_1}$ [%]
GAN	10.21 (9.48-10.97)	8.45 (7.91-9.01)	0.44 (0.42-0.46)	0.53 (0.52-0.54)
3D U Net	8.87 (8.11-9.65)	5.83 (5.39-6.29)	0.48 (0.46-0.50)	0.59 (0.58-0.61)
Ours (Basic variant)	10.45 (9.84-11.10)	6.72 (6.30-7.17)	0.42 (0.41-0.44)	0.56 (0.55-0.58)
Ours (Best variant)	8.70 (8.10-9.34)	6.28 (5.69-6.91)	0.50 (0.48-0.51)	0.62 (0.61-0.63)
Ours (Motion simulation)	11.53 (10.87-12.24)	17.83 (17.50-18.19)	0.40 (0.39-0.42)	0.29 (0.28-0.29)
Model (left)				
GAN	19.75 (14.28-26.11)	21.62 (16.21-27.86)	0.46 (0.44-0.48)	0.39 (0.37-0.41)
3D U Net	13.47 (8.34-19.62)	15.99 (10.89-21.97)	0.59 (0.57-0.61)	0.52 (0.49-0.54)
Ours (Basic variant)	15.05 (11.79-18.63)	16.90 (13.72-20.36)	0.57 (0.55-0.58)	0.49 (0.47-0.50)
Ours (Best variant)	14.45 (11.14-18.12)	15.98 (12.73-19.54)	0.60 (0.59-0.62)	0.53 (0.52-0.55)
Ours (Motion simulation)	15.50 (12.45-18.84)	24.88 (22.01-27.96)	0.55 (0.54-0.56)	0.30 (0.29-0.30)

Table 5: Mean reprojection metric values with 95% confidence intervals compared to state-of-the-art on synthetic data.

Model (right)	$PSNR^{synth}$ [dB]	τ [s]
GAN	39.63 (39.33-39.91)	0.01
3D U Net	33.90 (33.71-34.08)	0.01
Ours (Basic variant)	28.45 (28.19-28.71)	2.71
Ours (Best variant)	32.92 (32.68-33.16)	2.71
Ours (Motion simulation)	27.84 (27.59-28.09)	2.86
Model (left)		
GAN	39.04 (38.58-39.54)	0.01
3D U Net	31.13 (30.97-31.29)	0.01
Ours (Basic variant)	27.46 (27.19-27.71)	2.70
Ours (Best variant)	28.89 (28.63-29.14)	2.71
Ours (Motion simulation)	26.87 (26.59-27.16)	2.87

Table 6: Supplementary metrics on synthetic data with 95 % CI. As the runtime τ has very little variability, it is shown without the intervals.

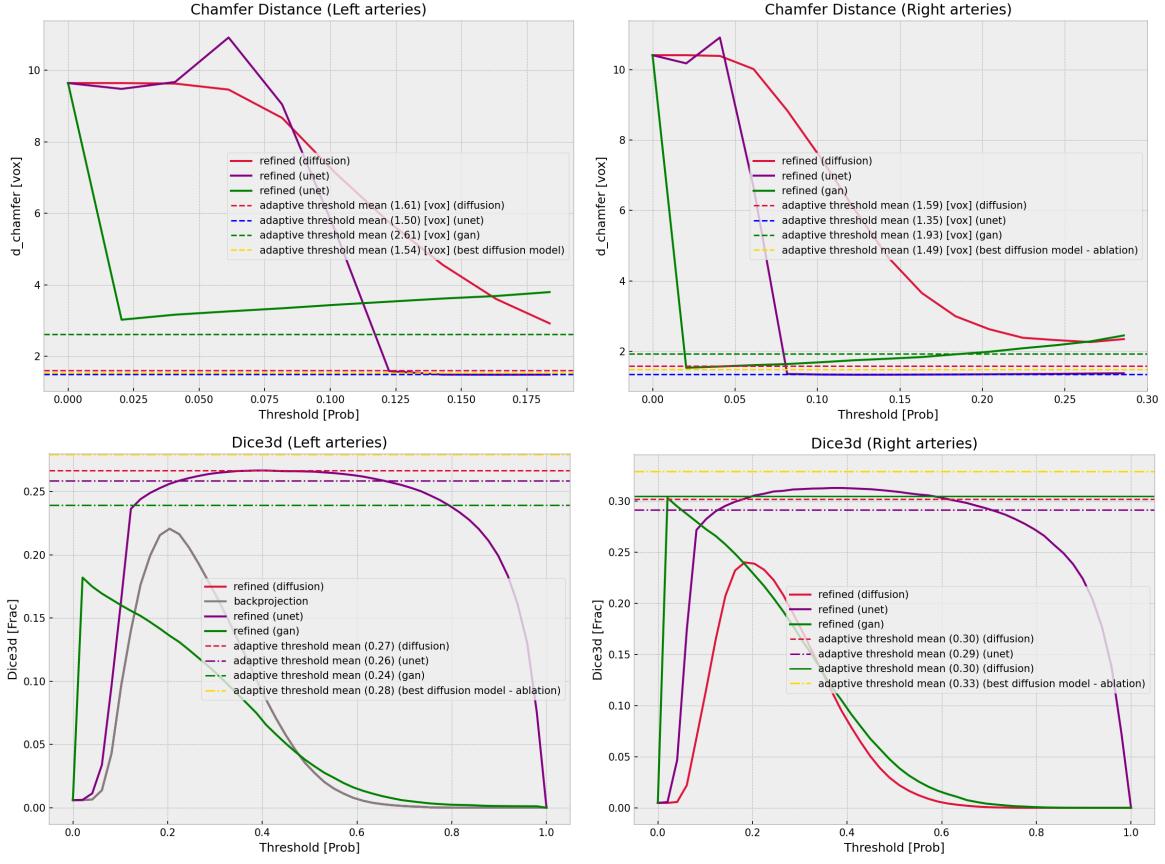


Figure 2: Baseline-to-ours $d_{\text{Chamfer:3d}}^{\text{synth}}$ and $DSC_{\text{3D}}^{\text{synth}}$ comparison with respect to threshold. Lines parallel to X axis mark the mean metric value when adaptive threshold is used.

Model (right)	$d_{\text{Chamfer:2d}}^{\text{real, xray}_0}$ [px]	$d_{\text{Chamfer:2d}}^{\text{real, xray}_1}$ [px]	$DSC_{\text{2D}}^{\text{real, xray}_0}$ [%]	$DSC_{\text{2D}}^{\text{real, xray}_1}$ [%]
GAN	16.75 (15.72-17.81)	16.64 (15.30-18.12)	0.27 (0.25-0.28)	0.28 (0.27-0.29)
3D Unet	17.02 (15.25-18.87)	17.20 (14.92-19.80)	0.29 (0.27-0.31)	0.32 (0.29-0.34)
Ours (basic variant)	18.64 (17.38-19.96)	17.69 (16.24-19.18)	0.27 (0.25-0.28)	0.29 (0.28-0.31)
Ours (best variant)	14.42 (13.29-15.64)	13.75 (12.39-15.21)	0.32 (0.31-0.34)	0.35 (0.34-0.37)
Model (left)				
GAN	32.51 (30.42-34.63)	32.63 (30.46-34.93)	0.20 (0.19-0.21)	0.21 (0.19-0.22)
3D Unet	20.50 (17.92-23.85)	18.26 (16.38-20.32)	0.31 (0.29-0.33)	0.33 (0.31-0.35)
Ours (basic variant)	25.08 (23.45-26.73)	24.62 (23.11-26.12)	0.30 (0.28-0.31)	0.31 (0.30-0.32)
Ours (best variant)	20.78 (19.18-22.47)	19.05 (17.87-20.24)	0.32 (0.30-0.33)	0.33 (0.32-0.34)

Table 7: Mean reprojection metrics with 95 % CI on clinical data.

3.2 Ablation study

Further tests were performed to determine how the inclusion and exclusion of certain parameters or techniques (e.g. conditioning) affects the method performance. Furthermore, for the same reason the impact of backprojection results quality was subjected to ablation. The results were presented in Tab. 8, 9.

Model (right)	$d_{Chamfer:3d}^{synth}$ [vox]	DSC_{3D}^{synth} [%]	$O_{3D}^{synth}(1)$	$O_{3D}^{synth}(2)$
basic ($P = 2$, projections, side-specific)	1.59 (1.51-1.67)	0.30 (0.29-0.31)	0.57 (0.55-0.58)	0.76 (0.75-0.78)
no projections	1.75 (1.68-1.84)	0.26 (0.25-0.27)	0.51 (0.50-0.53)	0.72 (0.70-0.73)
side-agnostic (BEST!)	1.49 (1.41-1.59)	0.33 (0.32-0.34)	0.59 (0.58-0.61)	0.78 (0.77-0.79)
$P = 3$	2.70 (2.56-2.85)	0.20 (0.19-0.21)	0.39 (0.38-0.41)	0.57 (0.55-0.59)
$P = 4$	3.48 (3.28-3.68)	0.17 (0.17-0.18)	0.35 (0.33-0.36)	0.50 (0.49-0.52)
Model (left)				
basic	1.61 (1.54-1.68)	0.27 (0.26-0.28)	0.51 (0.49-0.52)	0.70 (0.68-0.71)
no projections (BEST!)	1.54 (1.48-1.61)	0.28 (0.27-0.29)	0.51 (0.50-0.52)	0.70 (0.68-0.71)
side-agnostic	1.61 (1.53-1.70)	0.29 (0.28-0.30)	0.53 (0.51-0.54)	0.71 (0.70-0.73)
$P = 3$	2.39 (2.27-2.53)	0.21 (0.20-0.21)	0.41 (0.39-0.42)	0.58 (0.56-0.60)
$P = 4$	2.91 (2.75-3.09)	0.22 (0.21-0.23)	0.40 (0.38-0.42)	0.55 (0.53-0.57)

Table 8: Ablation study - different versions of the model in question were presented above, information on which parameter changed is contained in the first column, if the parameter is not mentioned it takes the default value. In both categories the "side-agnostic" row is the results for the model performing (i.e. being trained and tested) on both sides at once. "no projections" means it is not present in conditions.

Model (right)	$d_{Chamfer:2d}^{real, xray_0}$ [px]	$DSC_{2D}^{real, xray_0}$ [%]	$d_{Chamfer:2d}^{real, xray_1}$ [px]	$DSC_{2D}^{real, xray_1}$ [%]
basic	18.64 (17.38-19.96)	0.27 (0.25-0.28)	17.69 (16.24-19.18)	0.29 (0.28-0.31)
no projections	19.36 (17.91-20.87)	0.27 (0.25-0.29)	18.77 (17.22-20.35)	0.29 (0.27-0.31)
side-agnostic	14.42 (13.29-15.64)	0.32 (0.31-0.34)	13.75 (12.39-15.21)	0.35 (0.34-0.37)
Model (left)				
basic	25.08 (23.45-26.73)	0.30 (0.28-0.31)	24.62 (23.11-26.12)	0.31 (0.30-0.32)
no projections	20.78 (19.18-22.47)	0.32 (0.30-0.33)	19.05 (17.87-20.24)	0.33 (0.32-0.34)
side-agnostic	16.28 (15.10-17.51)	0.36 (0.35-0.37)	14.81 (13.56-16.09)	0.37 (0.36-0.39)

Table 9: Ablation study - reprojection metrics on clinical data.

Side-agnostic: The "side-agnostic" row shows how the model behaves without the left/right split (also trained on both of these sides at once). For the right arteries the results clearly stand out as the best meaning the benefit from the reduction of complexity of the problem that is gained by just reconstructing the right arteries is outweighed by the gain from increasing the dataset size. For the left side such benefit does not seem to occur for the synthetic data evaluation, while there is a noticeable improvement in the robustness (measured by the clinical reprojection metrics). All in all, the assumption that splitting the training data into left and right would simplify the task seems to have been disproven, as in both cases the side-agnostic model either works better or shows no significant metric value decrease. **No projections:** determines how including the projections in the conditioning (i.e. projection encoders described in detail in section 2.4) affects the performance. For synthetic data, not including projections in condition significantly boosts $d_{Chamfer:3d}^{synth}$, thus it is assumed to be the "best" (as most of the left side variants offer roughly the same performance) for the sake of state-of-the-art comparison in section 3.1. An assumption that including projections in conditioning decreases robustness is partially true - such behavior occurs for the left side, whereas for the right side the inclusion of projection encoders seems beneficial as there is a major decrease in metric values when it is not provided. **Number of projections:** to determine how the model behaves if there are more projections improving the backprojection quality (without training the model again) the same experiments were run again. Still only two projections were used for conditioning. For a given P , now input backprojection is acquired from the following identity in place of eq.(3):

$$x_{backprojected} = \bigcap_{j=0}^{P-1} x_{backprojected:j} \quad (11)$$

Fig. 3 depicts qualitative results when reconstruction is run on a single sample with $P \in \{2 \dots 4\}$. Against the intuition for $P = 4$ the method produces the most tortuous and differing volume, the result looks the best for $P = 2$. The subjective impressions are supported by the quantitative results depicted in Tab. 8 (rows starting with $P = \dots$), where the detrimental effect is largely reflected on metrics.

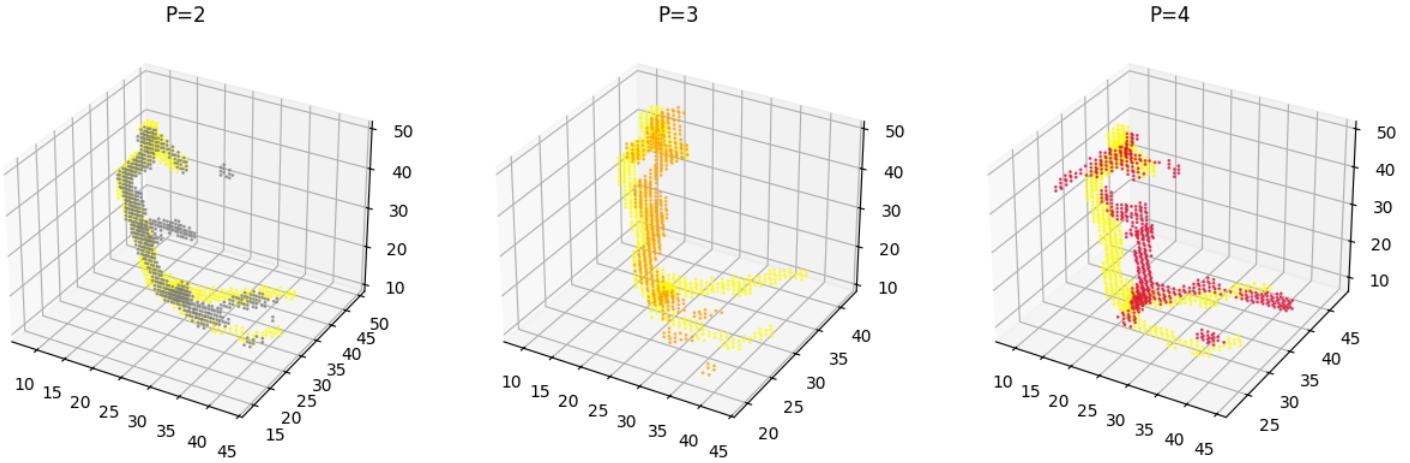
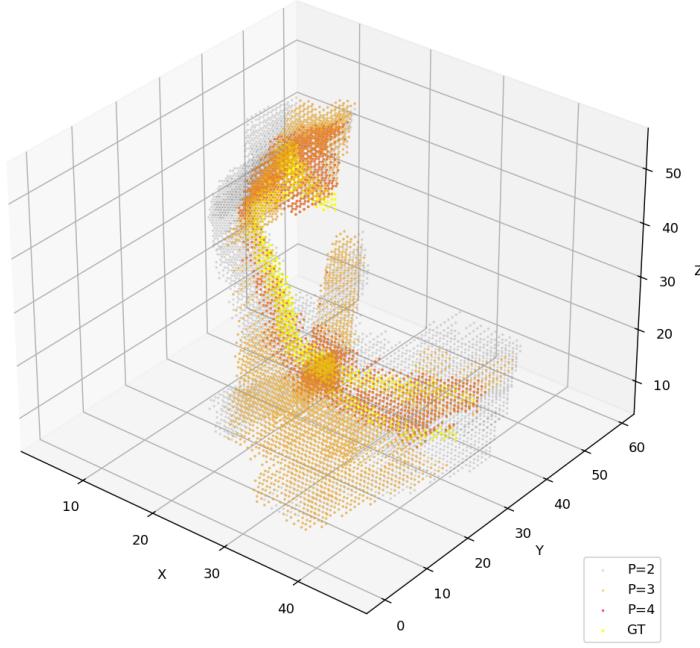


Figure 3: Above: Backprojection convergence towards ground truth with a growing number of projections used. Below: Reconstruction results from each backprojection.

Guidance scale: Fig. 4 shows how guidance strength γ affects the test results. As can be seen there is a steep decrease for $\gamma \in (0, 2]$, for $\gamma > 2$ the performance seems mostly unaffected by the changing variable, which is depicted as a saturation - plateau on each plot. Again we see worsened performance for the left side, although for higher guidance scales Chamfer distance is the same for both sides. Peak performance on synthetic dataset is reached for $\gamma = 2$, however the reprojection metrics suggest differently with measured values the highest for $\gamma \in \{6, 8, 10, 12\}$.

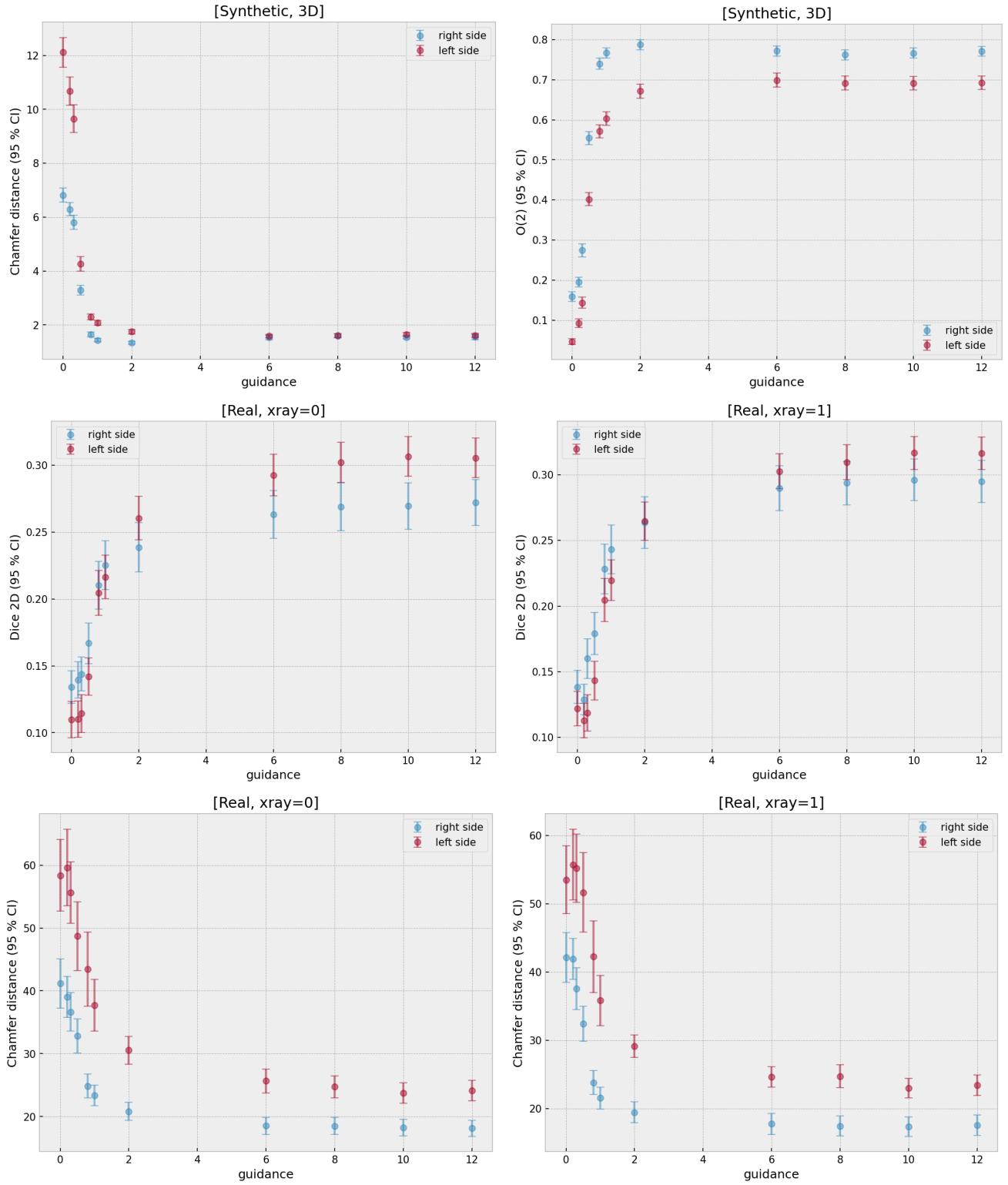


Figure 4: Chamfer distance decrease and $O(2)$ increase with respect to guidance scale for both sides, with 95 [%] confidence interval whiskers on synthetic dataset (above). Reprojection Dice and Chamfer distance on clinical dataset for both x-ray images with 95 [%] CI (below).

Sampling steps: The experiments were also run with different number of sampling steps used in the inference and the results summarized in Fig. 5. Contrary to the expectations, more steps does not necessarily mean a better reconstruction quality. In fact, it is the opposite. For more sampling steps the metric values seem to worsen, moreover the standard deviation of the results grows - increasing the number of sampling steps means a finer spacing between timesteps. For $n = 5$ the performance seems to be the best, however it does not differ much from $n = 10$ (which is a value initially assumed). On the boxplots a relation between left and right side of the arteries can be once again noticed that is the model producing

worse reconstructions for the left side. Interestingly for some of the cases (e.g. Dice2d for $xray_0$) the standard deviation for the left arteries is much lower. Last thing worth noting is the outliers appearing only on one side of the distribution - for $Ot(2)$ it is for the lower values of the metric and for $d_{Chamfer}$ is for the larger distances.

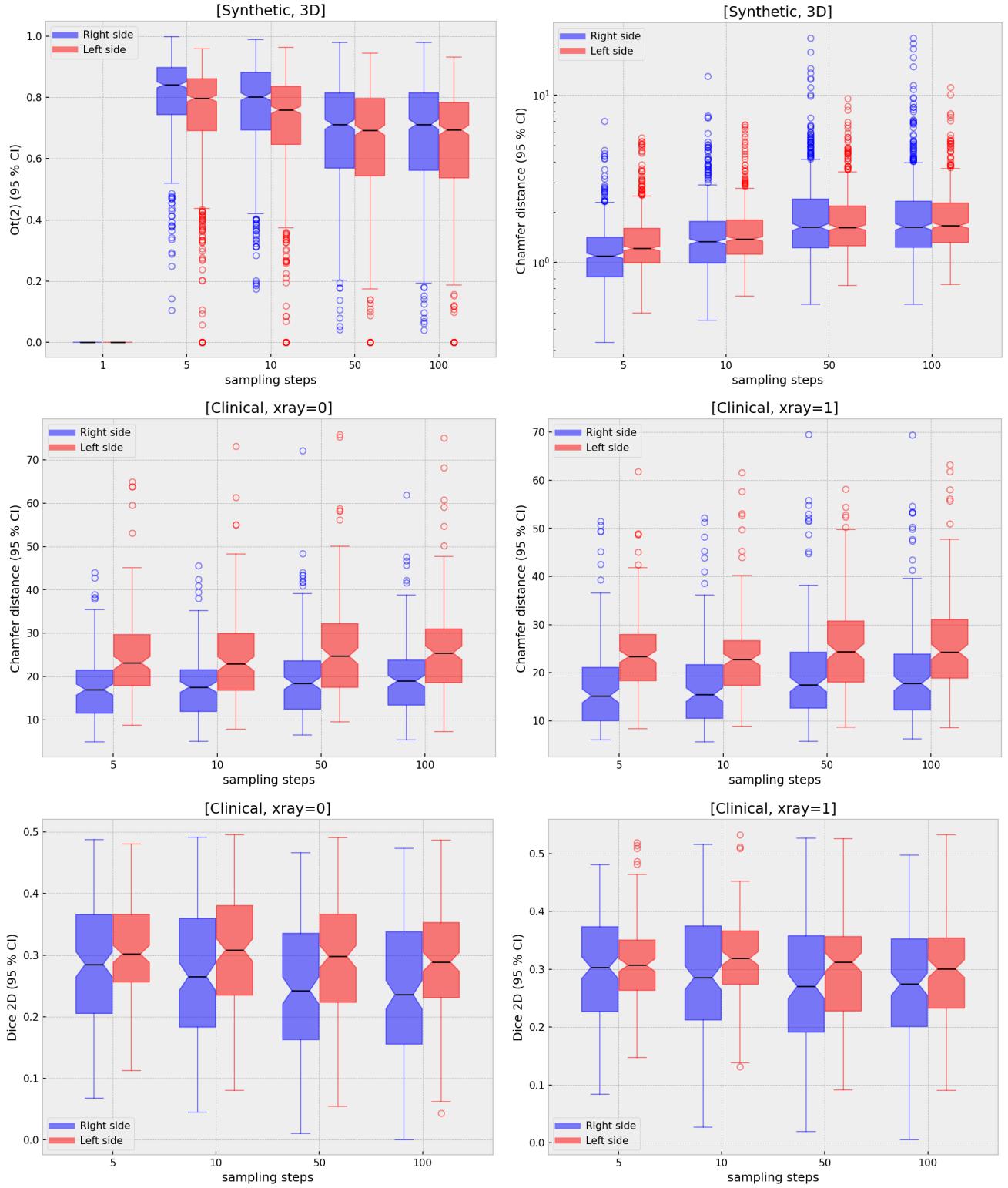


Figure 5: Boxplots presenting the effect of varying sampling steps on different metric values.

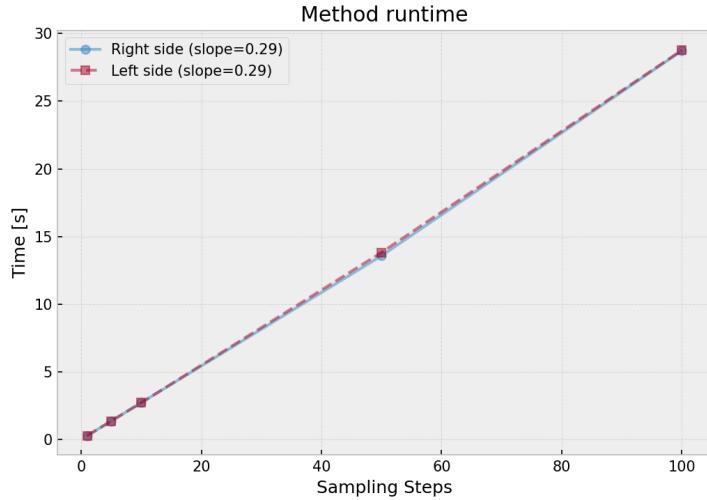


Figure 6: Method runtime with growing number of sampling steps.

3.3 Qualitative result demonstration

While the quantitative metrics are important for performance assessment, diffusion models have the advantage of capturing the underlying data distribution, allowing them to generate anatomically realistic and diverse reconstructions that better reflect structural improving qualitative impressions of the reconstruction. For the qualitative assessment a few key matters were discussed. Firstly, a few example results were shown in Fig. 7. The reconstruction is of varying quality, in some cases being noisy, in other missing parts of the volume.

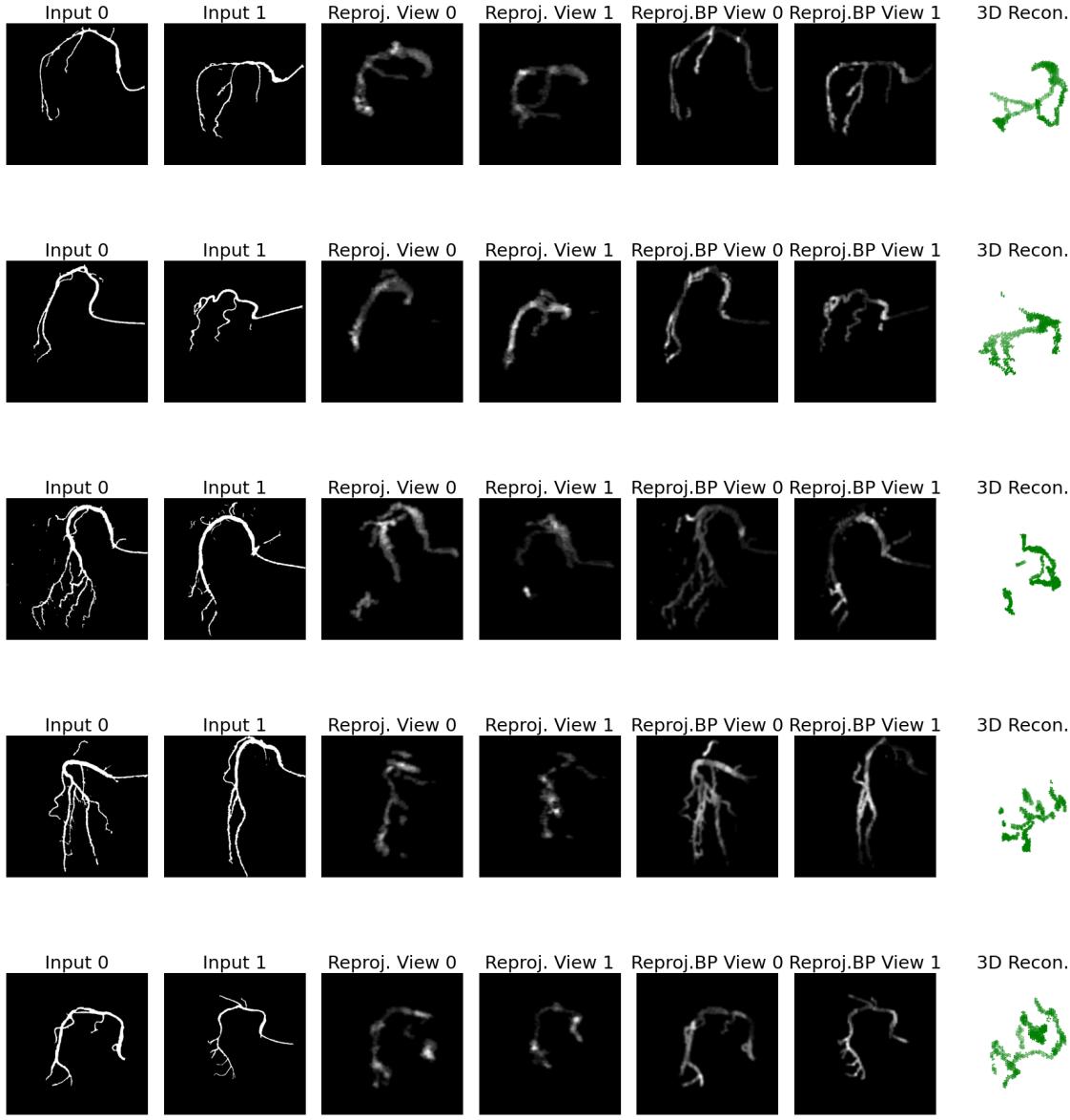


Figure 7: Example results of the model for 5 different inputs.

In Fig. 8 a comparison of diffusion model results with one of the baselines (U Net 3D) was shown. It can clearly be seen that, not captured enough by quantitative metrics, there is a significant difference between the morphological accuracy of the two models. U-Net fails to capture the proper shape of segments while diffusion is more coherent in that aspect. The overall noisiness of the produced volume is prominent in U-Net, furthermore in multiple cases it fails to reconstruct the bifurcations and the continuity of the segments is violated. The diffusion model reduces the overall noisiness and reflects the natural vessel flow.

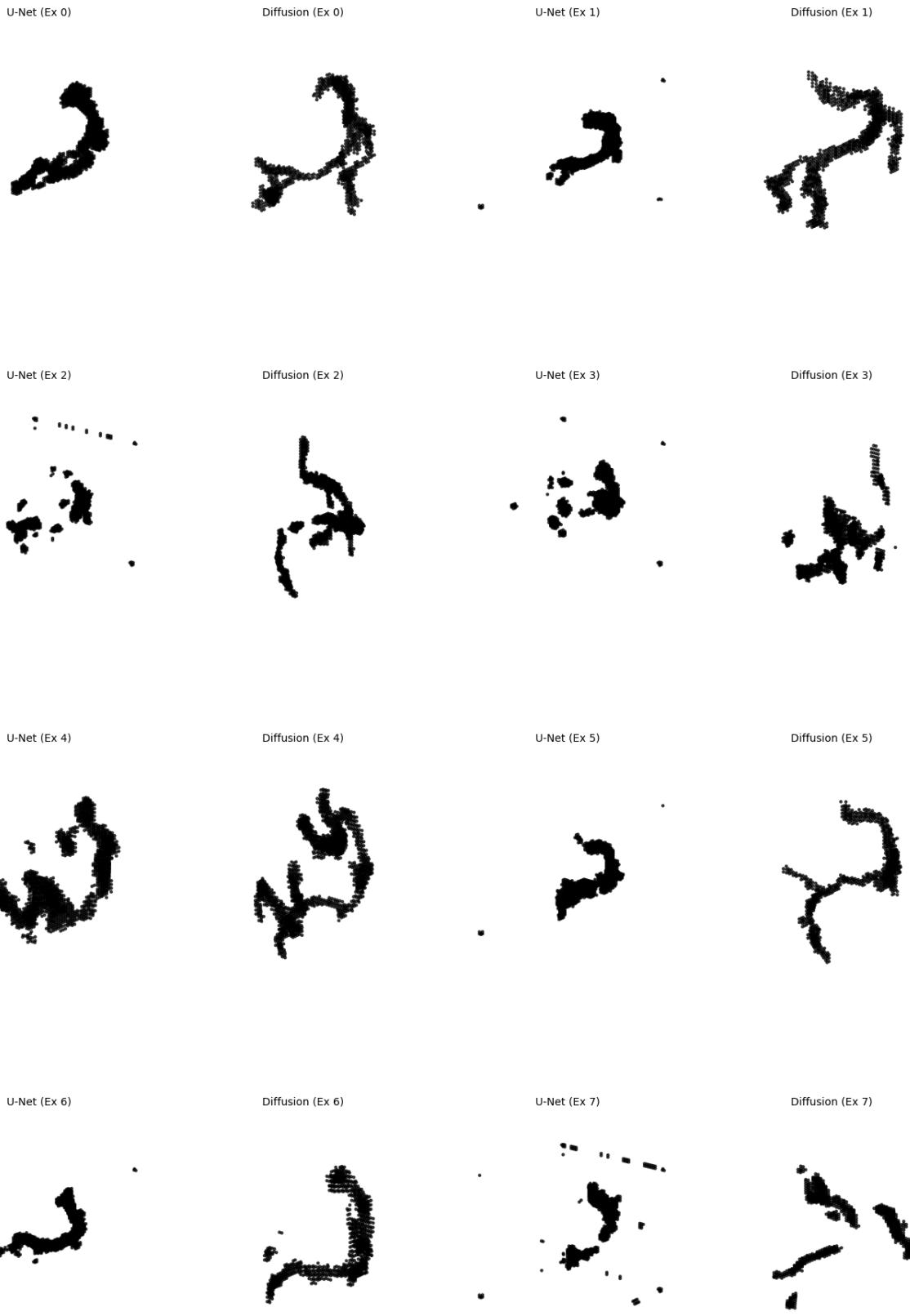


Figure 8: Diffusion model to baseline (U-Net) output comparison.

Fig. 9 contains a comparison between a single input projection and the resulting volume. As shown, the bifurcation points (the division of a branch into two child branches) are reconstructed well, there are also lesions and other changes to the volume radius visible, which is promising towards the increased utility in clinical practice. There are however many inaccuracies - parts of the vessels missing, especially the voxels that are the farthest from the main segment, and superfluous points in

some areas. The observations might imply that there is a limited sensitivity for reconstructing the fine-grained features of the vessels, which might be attributed to the synthetic dataset limitations (lack of such features in the training set overall) and to the resolution that was assumed in this work.

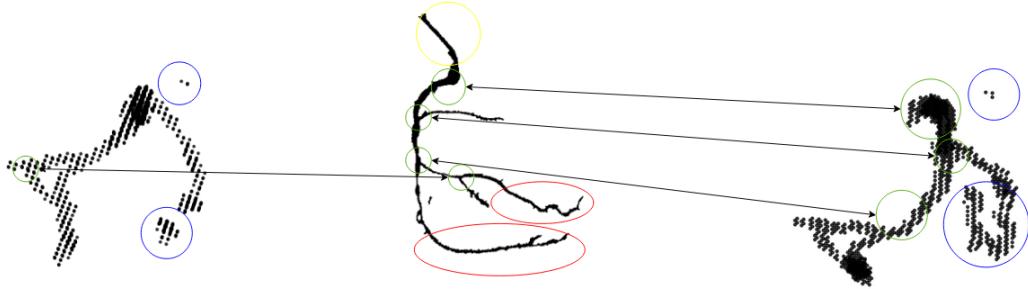


Figure 9: Comparison between input projection (in the middle) and output from two different perspectives (to the left and right). Blue circles mark points that are not present in input data, but are present in the reconstruction. Red circles mark parts that are missing from reconstruction. Green circles mark matching areas of the volume.

Another matter to consider is the fact that diffusion models inference is stochastic starting from $\mathbf{x}_{noise} \sim N(0, 1)$ therefore there is a variety of results that can be generated for the same input data, which is demonstrated in Fig. 10, where there is a variability visible between reconstructions of the same volume. While the volumes share a similar core there are differences in terms of the finer details. This randomness might result in a limited reliability or the need to rerun the model inference. Last, but not least, we encourage the reader to examine the animations depicting the sampling process that were attached in source code repository at github.com/cvlab-ai/xra-diff/media. Moreover, a final visualization present in Fig. 11 presents the reconstruction overlayed on backprojection, demonstrating the improvement in the quality of reconstruction from the starting point, which is the backprojected arteries. Such presentation might be beneficial for the radiologists' understanding of the results.



Figure 10: Subsequent outputs generated in inverse diffusion process for the same input data.

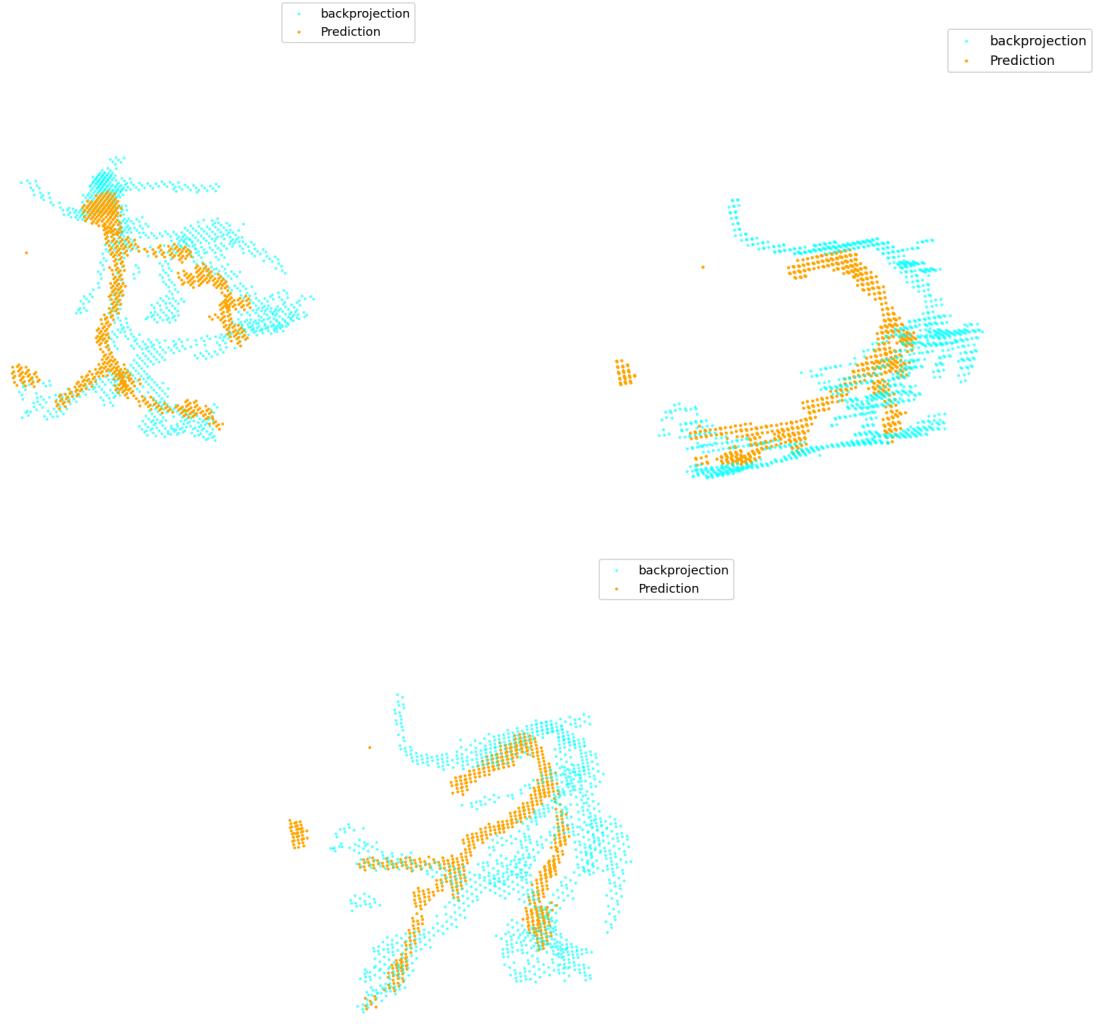


Figure 11: Example of a reconstructed right artery overlayed on backprojection results from three different points of view.

4 Conclusions

In the paper we proved the usability of diffusion models for the task of refining backprojection to resemble coronary arteries better, both on the clinical and synthetic data. Moreover, our approach reached state of the art among the deep learning-based approaches in the domain, significantly improving the results for the right side. For the left side arteries an improvement, although not a major one, was also achieved. We emphasized however the limited robustness of the used metrics especially when lacking ground truth 3D data, which was the case clinical examples.

We have experimented with modifications of the basic version of the method for the influence of guidance scale, conditioning and sampling strategies demonstrating potential improvement of the output quality. For the right arteries reconstruction, a method trained on both left and right arteries performed the best, while for the left arteries, a method skipping the conditioning with 2D projections produced the best outcomes. The method was proven to be partially affected by guidance scale choice as the most prominent changes happened in $\gamma \in (0, 2)$ interval reaching the maximum at $\gamma \approx 2$. Furthermore, results have shown a counter intuitive performance decrease of the model when utilizing more projections for improved backprojection quality. Moreover, our initial assumption that an universal model trained for the reconstruction of both sides of the arteries will have a worse performance was invalidated as the "side-agnostic" model has shown a performance matching or exceeding the basic variant.

In the final part of the work, we have shown example outputs proving the qualitative results of the method better reflect the topological features of the coronary arteries compared to baselines. We have highlighted the multiple correspondences between input projections and output reconstructions, at the same time pinpointing to locations at which major artifacts were visible.

Limitations Diffusion models, as mentioned earlier, are data-extensive. In our case, only 800 training samples were used for each side, which might still leave much room for improvement provided a bigger dataset is supplied - this hypothesis remains untested due to the lack of such public data. Furthermore, as shown in the paper, the inference takes orders of magnitude longer for diffusion compared to U Net, which might become a bottleneck in high-load production environments. Moreover, the postprocessing method used was primitive, therefore requiring potential improvements. Finally, as shown, the model might fail to converge for out-of-distribution data points, although its robustness has been proven to an extent (using reprojection metrics).

Future works: The authors will focus on deployment of the presented method in a clinical environment and the usability assessment by experts. Also a study on postprocessing methods will be conducted. At the moment authors are also preparing a multi-label dataset from XRA for a release at a later date.

Glossary

- XRA - X-ray Angiography
- SID - Source-Image distance
- SOD - Source-Object distance
- DICOM - "Digital Imaging and Communications in Medicine" format
- AngloTagger - a software for labelling data from coronary angiography

Contributions

D. B. Lau: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft
H. Malinowski: Investigation, Software, Validation **J. Szyjut:** Investigation, Software **T. Dziubich:** Conceptualization, Review & Editing, Supervision, Validation, Data Curation.

Data availability

ImageCAS is publically available at kaggle/datasets/xiaoweixumedicalai/imagecas. The clinical dataset is available on request made to the corresponding author.

Informed consent was obtained from all subjects involved in the study.

Funding sources

The research was supported in part by project “Cloud Artificial Intelligence Service Engineering (CAISE) platform to create universal and smart services for various application areas”, No. KPOD.05.10-IW.10-0005/24, as part of the European IPCEI-CIS program, financed by NRRP (National Recovery and Resilience Plan) funds.

Acknowledgements

Authors would like to thank **CI TASK** for granting access to resources and **CAISE** - a cloud computing platform that allowed for the efficient training and reduced the experiments run time (notably four H100 GPUs).

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Cardiovascular diseases statistics — ec.europa.eu. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cardiovascular_diseases_statistics. [Accessed 15-10-2025].
- [2] Causes of death statistics — ec.europa.eu. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Causes_of_death_statistics. [Accessed 15-10-2025].
- [3] A. Besteemir, Z. Apaydin, and A.Y. Kılınç. Analysis of coronary angiography and revascularization rates made over 5 years in public institutions in turkeye. *The Anatolian Journal of Cardiology*, 27:529–533, 2023.
- [4] 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.
- [5] Wei Wu, Usama M. Oguz, Akshat Banga, Shijia Zhao, Anjani Kumar Thota, Vinay Kumar Gadamidi, Charu Hasini Vasa, Khaled M. Harmouch, Abdallah Naser, Xiarepati Tielawaerdi, and Yiannis S. Chatzizisis. 3d reconstruction of coronary artery bifurcations from intravascular ultrasound and angiography. *Scientific Reports*, 13(1), August 2023.
- [6] 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*, 13(1), October 2023.
- [7] 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):1–23, 01 2018.
- [8] 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):1711, January 2018.
- [9] Weijian Cong, Jian Yang, Danni Ai, Yang Chen, Yue Liu, and Yongtian Wang. Quantitative analysis of deformable model-based 3-d reconstruction of coronary artery from multiple angiograms. *IEEE Transactions on Biomedical Engineering*, 62(8):2079–2090, August 2015.
- [10] Xueming Fu, Yingtai Li, Fenghe Tang, Jun Li, Mingyue Zhao, Gao-Jun Teng, and S. Kevin Zhou. 3dgr-car: Coronary artery reconstruction from ultra-sparse 2d x-ray views with a 3d gaussians representation. 2024.
- [11] Yiying Wang, Abhirup Banerjee, and Vicente Grau. Nec: 3d coronary artery tree reconstruction from two 2d projections via neural implicit representation. *Bioengineering*, 11(12), 2024.
- [12] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM*, 65(1):99–106, December 2021.
- [13] Yiying Wang, Abhirup Banerjee, Robin P. Choudhury, and Vicente Grau. Deepca: Deep learning-based 3d coronary artery tree reconstruction from two 2d non-simultaneous x-ray angiography projections. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pages 337–346, February 2025.
- [14] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans, 2017.
- [15] 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. 2023.
- [16] Simon J.D. Prince. *Understanding Deep Learning*. The MIT Press, 2023.
- [17] Xiyu Wang, Baijong Lin, Daochang Liu, Ying-Cong Chen, and Chang Xu. Bridging data gaps in diffusion models with adversarial noise-based transfer learning. In *Proceedings of the 41st International Conference on Machine Learning, ICML'24*. JMLR.org, 2024.
- [18] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance (preprint). 2022.
- [19] Zhanqiang Guo, Zimeng Tan, Jianjiang Feng, and Jie Zhou. Vesseldiffusion: 3d vascular structure generation based on diffusion model. *IEEE Transactions on Medical Imaging*, 44(9):3845–3857, 2025.
- [20] Lin Zhao, Xin Yu, Yikang Liu, Xiao Chen, Eric Z. Chen, Terrence Chen, and Shanhui Sun. Leveraging diffusion model and image foundation model for improved correspondence matching in coronary angiography (preprint). 2025.

- [21] Xinyu Li, Danni Ai, Hong Song, Jingfan Fan, Tianyu Fu, Deqiang Xiao, Yining Wang, and Jian Yang. Stqd-det: Spatio-temporal quantum diffusion model for real-time coronary stenosis detection in x-ray angiography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9908–9920, 2024.
- [22] Yun Su Jeong, Hye Bin Yoo, and Il Yong Chun. Dx2ct: Diffusion model for 3d ct reconstruction from bi or mono-planar 2d x-ray(s). In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025.
- [23] Samuel G. Armato et al. The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, 38(2):915–931, January 2011.
- [24] Siyeop Yoon, Jay Sairam Pratap, Wen-Chih Liu, Matthew Tivnan, Hui Ren, Abhiram Bhashyam, Quanzheng Li, Neal Chen, and Xiang Li. High-resolution 3d ct synthesis from bidirectional x-ray images using 3d diffusion model. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, 2024.
- [25] SyntaxScore. Definitions. <https://syntaxscore.org>. Accessed: 2025-06-08.
- [26] Maxim Popov, Akmaral Amanturdieva, Nuren Zhaksylyk, Alsabir Alkanov, Adilbek Saniyazbekov, Temirgali Aimyshev, Eldar Ismailov, Ablay Bulegenov, Arystan Kuzhukeyev, Aizhan Kulambayeva, Almat Kalzhanov, Nurzhan Temenov, Alexey Kolesnikov, Orazbek Sakhov, and Siamac Fazli. Dataset for automatic region-based coronary artery disease diagnostics using x-ray angiography images. *Scientific Data*, 11(1), January 2024.
- [27] Hongwei Zhang, Zhifan Gao, Dong Zhang, William Kongto Hau, and Heye Zhang. Progressive perception learning for main coronary segmentation in x-ray angiography. *IEEE Transactions on Medical Imaging*, 42(3):864–879, March 2023.
- [28] Xiliang Zhu, Zhaoyun Cheng, Sheng Wang, Xianjie Chen, and Guoqing Lu. Coronary angiography image segmentation based on pspnet. *Computer Methods and Programs in Biomedicine*, 200:105897, 2021.
- [29] 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.
- [30] Jonas Adler, Holger Kohr, Axel Ringh, Julian Moosmann, sbanert, Matthias J Ehrhardt, Gregory R Lee, niinimaki, bgris, Olivier Verdier, Johan Karlsson, zickert, Willem Jan Palenstijn, Ozan Öktem, Chong Chen, Hector Andrade Loarca, and Michael Lohmann. odlgroup/odl: ODL 0.7.0, 2018.
- [31] Steve Ramcharita Peregrine Green, Paul Frobisher. Optimal angiographic views for invasive coronary angiography: a guide for trainees. *British Journal of Cardiology*, 2016.
- [32] Peter Aundal Toft. *The Radon Transform - Theory and Implementation*. PhD thesis, November 1996.
- [33] Franz-Josef Neumann, Miguel Sousa-Uva, Anders Ahlsson, Fernando Alfonso, Adrian P Banning, Umberto Benedetto, Robert A Byrne, Jean-Philippe Collet, Volkmar Falk, Stuart J Head, Peter Jüni, Adnan Kastrati, Akos Koller, Steen D Kristensen, Josef Niebauer, Dimitrios J Richter, Petar M Seferović, Dirk Sibbing, Giulio G Stefanini, Stephan Windecker, Rashmi Yadav, Michael O Zembala, and ESC Scientific Document Group. 2018 esc/eacts guidelines on myocardial revascularization. *European Heart Journal*, 40(2):87–165, 08 2018.
- [34] Digital Imaging and Communications in Medicine (DICOM). Standard, National Electrical Manufacturers Association (NEMA), Rosslyn, VA, USA, 2025.
- [35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [36] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, Steven Liu, William Berman, Yiyi Xu, and Thomas Wolf. Diffusers: State-of-the-art diffusion models.
- [37] William Falcon and The PyTorch Lightning team. PyTorch Lightning, March 2019.
- [38] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. 2022.
- [39] Jonghoe Ku, Yong-Hee Lee, Junsup Shin, In Kyu Lee, and Hyun-Woo Kim. Mpseg : Multi-phase strategy for coronary artery segmentation (preprint). 2023.
- [40] Katherine Lal, Rebecca Gosling, Mina Ghobrial, Gareth J Williams, Vignesh Rammohan, D Rod Hose, Patricia V Lawford, Andrew Narracott, John Fenner, Julian P Gunn, and Paul D Morris. Operator-dependent variability of angiography-derived fractional flow reserve and the implications for treatment. *European Heart Journal - Digital Health*, 2(2):263–270, February 2021.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: