

# NeRF-CA: Dynamic Reconstruction of X-Ray Coronary Angiography With Extremely Sparse-Views

Kirsten W.H. Maas , Danny Ruijters , Anna Vilanova , Member, IEEE, and Nicola Pezzotti 

**Abstract**—Dynamic three-dimensional (4D) reconstruction from two-dimensional X-ray coronary angiography (CA) remains a significant clinical problem. Existing CA reconstruction methods often require extensive user interaction or large training datasets. Recently, Neural Radiance Field (NeRF) has successfully reconstructed high-fidelity scenes in natural and medical contexts without these requirements. However, challenges such as sparse-views, intra-scan motion, and complex vessel morphology hinder its direct application to CA data. We introduce NeRF-CA, a first step toward a fully automatic 4D CA reconstruction that achieves reconstructions from sparse coronary angiograms. To the best of our knowledge, we are the first to address the challenges of sparse-views and cardiac motion by decoupling the scene into the moving coronary artery and the static background, effectively translating the problem of motion into a strength. NeRF-CA serves as a first stepping stone for solving the 4D CA reconstruction problem, achieving adequate 4D reconstructions from as few as four angiograms, as required by clinical practice, while significantly outperforming state-of-the-art sparse-view X-ray NeRF. We validate our approach quantitatively and qualitatively using representative 4D phantom datasets and ablation studies.

**Index Terms**—4D reconstruction, X-ray coronary angiography, neural radiance field, sparse-view, static and dynamic decomposition.

## I. INTRODUCTION

X-RAYS are fundamentally limited by providing a two-dimensional (2D) representation of a three-dimensional (3D) structure, resulting in a lack of depth perception. This problem is especially apparent in X-ray coronary angiography (CA) interventions, where dynamic sequences of 2D X-ray projections or coronary angiograms represent complex dynamic 3D (4D) vessel structures. On top of that, in standard CA interventions, an extremely limited set of these angiogram views are acquired per patient, which we refer to as sparse-view [1]. These

interventions could benefit from 4D reconstructions to enhance the 4D perception of the vessel structure for catheter navigation applications [2]. Nonetheless, accurate 4D reconstruction from CA data poses a challenge due to inherent characteristics from the X-ray system, complex vessel characteristics, and intra-scan motion [3]. The challenges of the X-ray system include extremely sparse-view X-ray projections and system geometry inaccuracies. The vessel characteristics include structure sparsity, vessel overlap, and poor vessel visibility due to background obstruction, related to overlapping background structures or blood vessel contrast inhomogeneity. Moreover, coronary angiography data exhibit intra-scan motion due to cardiac and respiratory motion. Overall, 4D reconstruction of CA data remains a challenge.

Several CA reconstruction methods have been proposed, including traditional and machine learning methods [3], [4], [5]. Both methods can obtain satisfactory reconstructions from extremely sparse-view sequences. However, they rely on time-consuming user segmentations or large amounts of training data, consisting of many views per patient, which are not available due to the clinical sparse-view setting of CA. These dependencies hinder the adaptation of these methods in clinical practice. While some methods ignore the dynamics, others assume cyclic cardiac motion through an acquired electrocardiogram (ECG) signal but still lead to motion artifacts [3].

Neural Radiance Field (NeRF) is a promising deep learning technique for 3D reconstruction that can reconstruct high-fidelity 3D natural scenes from 2D projections [6]. Unlike machine learning methods that learn a prior, NeRF learns an individual scene given a set of 2D images at hand. As such, it does not rely on segmentations or large amounts of training data, showing the potential to overcome the limitations of prior CA reconstruction techniques. Recent work explored the potential of NeRF in X-ray angiography, showing that limitations still lie in reconstruction in a sparse-view setting [7].

Meanwhile, many extended NeRF works have been proposed to reconstruct in the sparse-view setting [8]. Regularization-based techniques for natural scenes have been shown to reconstruct from only 3 views [9]. Techniques designed for natural scenes rely on properties such as depth and sparsity, which are absent in the CA setting, making them unsuitable. Concurrently, NeRF-based techniques proposed for medical scenes do not reach the extremely sparse view requirements for CA scenes, requiring a minimum of 15 projections for adequate performance [10]. NeRF methods, such as 3DGR-CAR [11] and NeCA [12], address sparse-views for CA scenes. However,

Received 29 August 2024; revised 4 June 2025; accepted 6 June 2025. Date of publication 12 June 2025; date of current version 5 September 2025. Recommended for acceptance by T. Ropinski. (Corresponding author: Kirsten W.H. Maas.)

Kirsten W.H. Maas and Anna Vilanova are with the Department of Mathematics & Computer Science, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands (e-mail: k.w.h.maas@tue.nl; a.vilanova@tue.nl).

Danny Ruijters is with the Philips Healthcare, 5684PC Best, The Netherlands, and also with the Department of Electrical Engineering, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands (e-mail: danny.ruijters@philips.com).

Nicola Pezzotti is with the ASML, 5504DR Veldhoven, The Netherlands.

To accelerate research in this domain, we made our codebase public: <https://github.com/kirstenmaas/NeRF-CA>.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TVCG.2025.3579162>, provided by the authors. Digital Object Identifier 10.1109/TVCG.2025.3579162

1077-2626 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

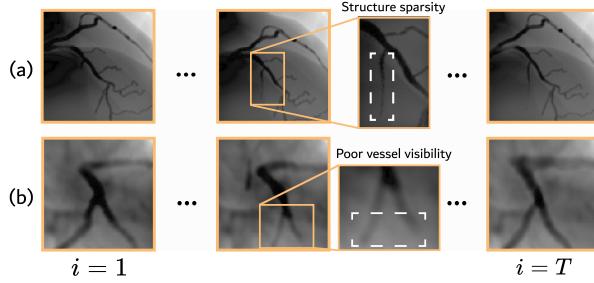


Fig. 1. One coronary angiogram sequence for the (a) XCAT [16] and (b) MAGIX [17] datasets with cardiac phases  $i \in \{1, \dots, T\}$ .

they rely on accurate segmentation or extensive training data and ignore the inherent motion of CA data. On the other hand, Liu et al. [13] recently presented static and dynamic scene decomposition for modeling vessel dynamics in digital subtraction angiography (DSA) separating foreground and background. Although similar to CA, DSA does not suffer from poor vessel visibility, and the technique proposed by Liu et al. [13] is not adequate for sparse-view settings since it requires many training projections.

We propose NeRF-CA, the first step towards a 4D X-ray Coronary Angiography (CA) reconstruction method addressing the challenges of sparse-views and the cardiac motion characteristics of CA acquisitions. We present 4D vessel reconstructions from an extremely sparse amount of synthetic coronary angiogram sequences exhibiting cardiac motion. We build on the extended work of NeRF for static and dynamic natural scene decomposition [14] and techniques for reconstruction in a sparse-view setting [9], [15]. Specifically, we utilize the rapid motion of the coronary artery to decouple the scene into a dynamic coronary artery component and a static background component. The novelty of our work lies in developing a combination of self-supervised decomposition with regularization techniques to enable sparse-view reconstructions for CA scenes. Specifically, we propose regularization techniques for the dynamic coronary artery component, enforcing properties such as scene sparsity and minimal occlusion. Together, we demonstrate vessel reconstruction capabilities from as few as four coronary angiogram views, matched to clinical workflows. We validate the effect of these constraints through ablation studies. Similar to other related works [4], [11], we demonstrate our method on two 4D phantom datasets [16], [17], simulating synthetic coronary angiograms. These synthetic coronary angiograms allow us to address the challenges of sparse-views and cardiac motion in CA data, having a ground-truth as realistic as possible. Meanwhile, we can perform the necessary quantitative validation, as we can acquire ground-truth sequences from any viewpoint. We compare our method with X-ray sparse-view NeRF reconstruction methods, significantly outperforming them in the extremely sparse-view X-ray coronary angiography setting. This provides the initial validation of the method in an ideal setting, showing the potential for future developments and applications to real data.

In summary, the main contribution is a methodology to address cardiac motion and sparse-view characteristics for 4D reconstruction of CA data. This serves as a first step towards a solution for clinical data. Unlike previous methods, our method does not rely on manual user interactions or large datasets. We decouple the coronary angiography scene in a dynamic coronary artery and static background by uniquely combining scene decomposition with regularization constraints fitting the sparse nature of the blood vessel structure. Moreover, we perform experiments to compare our method to existing state-of-the-art NeRF-based methods while validating the relevance of our choices through ablation studies.

## II. BACKGROUND

In this section, we will discuss the clinical background of coronary X-ray angiography and the technical background of Neural Radiance Fields (NeRFs), which are relevant to understanding our work.

### A. X-Ray Coronary Angiography

X-ray coronary angiography (CA) is the most common imaging modality to diagnose and treat coronary artery disease. Although it has fundamental limitations due to its 2D representation of a 3D anatomy, it still outperforms 3D imaging modalities. The benefits are its spatial and temporal resolution [18]. This work focuses on CA imaging acquired by a single-plane C-arm angiography system, which is still the gold standard for CA interventions [19].

A coronary angiogram sequence consists of multiple frames showing a contrast-enhanced coronary artery experiencing cardiac and respiratory motion. In single-plane CA, these sequences are obtained from a single acquisition angle or view. The terms "sequence" and "view" will be used interchangeably throughout this paper, depending on whether the focus is on the temporal or directional aspect, respectively. In Fig. 1, an example of one coronary angiogram sequence is shown for two datasets, where each frame represents a single cardiac phase  $i$  within one cardiac cycle. As can be seen, the imaging contains sparse blood vessel structures that may overlap, as highlighted in Fig. 1(a). Moreover, the imaging suffers from poor vessel visibility through overlapping background structures or inhomogeneous contrast, which is highlighted in Fig. 1(b). In a clinical setting, an extremely sparse amount of angiogram views (i.e., 2-4 per patient) is captured to minimize X-ray exposure. C-arm systems can be rotated using 3D Euler angles  $\theta$  and  $\phi$  to obtain varying viewing points [3]. The coronary arteries experience cardiac and respiratory motion while acquiring the angiogram sequences. Cardiac motion dominates the dynamics of the coronary arteries [20]. This motion aligns with the beating of the heart, which is assumed to be synchronous with the simultaneously recorded electrocardiogram (ECG). Therefore, individual frames of the coronary angiogram can be assigned an estimated cardiac phase within each cardiac cycle, as shown in Fig. 1. On the other hand, respiratory motion is correlated with the breathing rate, which is typically not measured clinically. As a result, this motion is

more complex to estimate [20] and is assumed to be residual for 3D reconstruction tasks [3].

4D reconstructions in the context of CA can aid both the diagnosis and treatment of coronary artery disease. An example of a clinical application of 4D reconstructions is called coronary roadmap, in which the reconstruction is overlayed over the X-ray to aid catheter navigation through the blood vessels [2]. This overlay is created by re-projecting the reconstruction based on the C-arm angle of interest, similar to the novel view application of NeRFs.

### B. Neural Radiance Field

A Neural Radiance Field (NeRF) represents a single scene through a multilayer perceptron (MLP), a fully connected neural network [6]. Given a 3D coordinate  $\mathbf{x} = (x, y, z)$  and a viewing direction  $\mathbf{d} = (\theta, \phi)$ , the network outputs a density  $\sigma$  and color  $\mathbf{c} = (r, g, b)$  to represent natural scenes. NeRF therefore learns a continuous function  $F_{\Theta}(\mathbf{x}, \mathbf{d}) = (\sigma, \mathbf{c})$ , where  $\Theta$  represents the learned weights of the MLP. A NeRF needs to be re-optimized for each scene.

As MLPs have difficulty learning high frequencies, NeRF employs positional encoding  $\gamma$  on the input parameters to allow for high-frequency details in the reconstructions [6]. Early works used sinusoidal functions with different frequencies to map the input to a higher dimensional space [21].

NeRF renders pixels through the learned density  $\sigma$  and color  $\mathbf{c}$  by discretizing the volume rendering equation [22]. Every pixel is associated with a ray  $\mathbf{r}$ , which is defined as  $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ , with camera origin  $\mathbf{o}$ , ray direction  $\mathbf{d}$  and  $t$  the position along the ray. The obtained predicted pixel color  $\hat{\mathbf{C}}(\mathbf{r})$  is compared to the ground truth pixel color  $\mathbf{C}(\mathbf{r})$  of a given image to optimize the model utilizing mean squared error (MSE) through the photometric loss. We refer to the work of Mildenhall et al. [6] for further details on the method.

## III. RELATED WORK

This section will discuss X-ray coronary angiography (CA) reconstruction and Neural Radiance Field techniques related to the work in this paper.

### A. X-Ray Coronary Angiography Reconstruction Techniques

A wide range of X-ray coronary angiography (CA) reconstruction techniques have been proposed, which can be categorized into model-based, tomographic-based, and machine learning-based [3], [4]. We will discuss these techniques and their strategies to handle cardiac motion and sparse-views.

Model-based techniques extract the vessels from the CA imaging to reconstruct them as binary trees. These methods achieve accurate reconstruction with sparse clinical views but rely on time-consuming manual input [18]. Recently, machine learning approaches have been proposed for automatic segmentation [23], [24] or direct 3D reconstruction [4], [5], [25]. However, these methods require large amounts of accurately segmented training data, including multiple views per patient, which is unfeasible in clinical settings due to the need to minimize X-ray exposure and the time-consuming nature of

segmentation. Although automatically obtained segmentations offer a potential solution, 3D reconstruction methods that depend on them are highly sensitive to segmentation quality. As stated by several works [5], [23], [24], inaccuracies, such as missing vessels or inconsistencies across 2D views, can propagate through the reconstruction process, resulting in noisy or incomplete 3D reconstructions. Tomographic techniques reconstruct attenuation volumes from raw X-ray imaging without needing manual segmentation. However, they require many views, limiting their use in single-plane CA. Methods based on NeRF have the potential to overcome these limitations, and will be discussed in Section III-B. Our work will build on these methods in the context of CA.

Traditional 3D reconstruction methods use various strategies to mitigate motion artifacts, primarily focusing on cardiac motion and treating respiratory motion as residual [3]. For the cardiac motion, they use ECG signals for frame extraction through gating or motion compensation. Gating selects frames around the same cardiac phase, while motion compensation aligns all frames to the same phase.

Our work focuses on solving the challenge of cardiac motion within the sparse-view setting, not considering respiratory motion and other real-world data aspects, like noise or patient motion. NeRF-CA serves as a first step towards a solution for the full clinical problem, given that a method has yet to be proposed that addresses cardiac motion and sparse-views without the reliance on manual segmentations or large datasets, where both factors have so far hindered the development of a clinically applicable solution to the reconstruction problem. Incorporating respiratory motion and other real-world data aspects is out of scope for this paper, such that the cardiac motion and sparse-view challenges can be isolated and evaluated independently. In case of potential results, future work should address the remaining data aspects.

### B. Neural Radiance Fields

NeRFs have been widely adopted for reconstructing natural scenes due to their ability to reconstruct high-fidelity details. However, challenges persist in reconstructing natural scenes from sparse-views, especially when dynamic [26]. On top of that, CA scenes present unique challenges such as sparse structures and poor vessel visibility [27], [28].

Many works have addressed the sparse-view problem, which can be categorized into prior-based and regularization-based methods [8]. Whereas prior-based methods train a general model on large datasets, regularization-based methods introduce additional scene-specific regularization terms. Given the lack of data in single-plane CA, as discussed in Section III-A, prior-based methods are unsuitable in our context. Regularization-based methods instead rely on scene-specific properties [9], [15], [29], [30]. However, these methods are yet to be adapted to medical scenes, which pose unique properties such as poor vessel visibility and therefore complicating foreground and background distinction. For medical scenes, reconstruction in a sparse-view setting also poses a significant challenge [31], [32]. Prior-based [33], [34], [35] or X-ray regularization-based [10], [32], [36] strategies have been proposed to address this in static

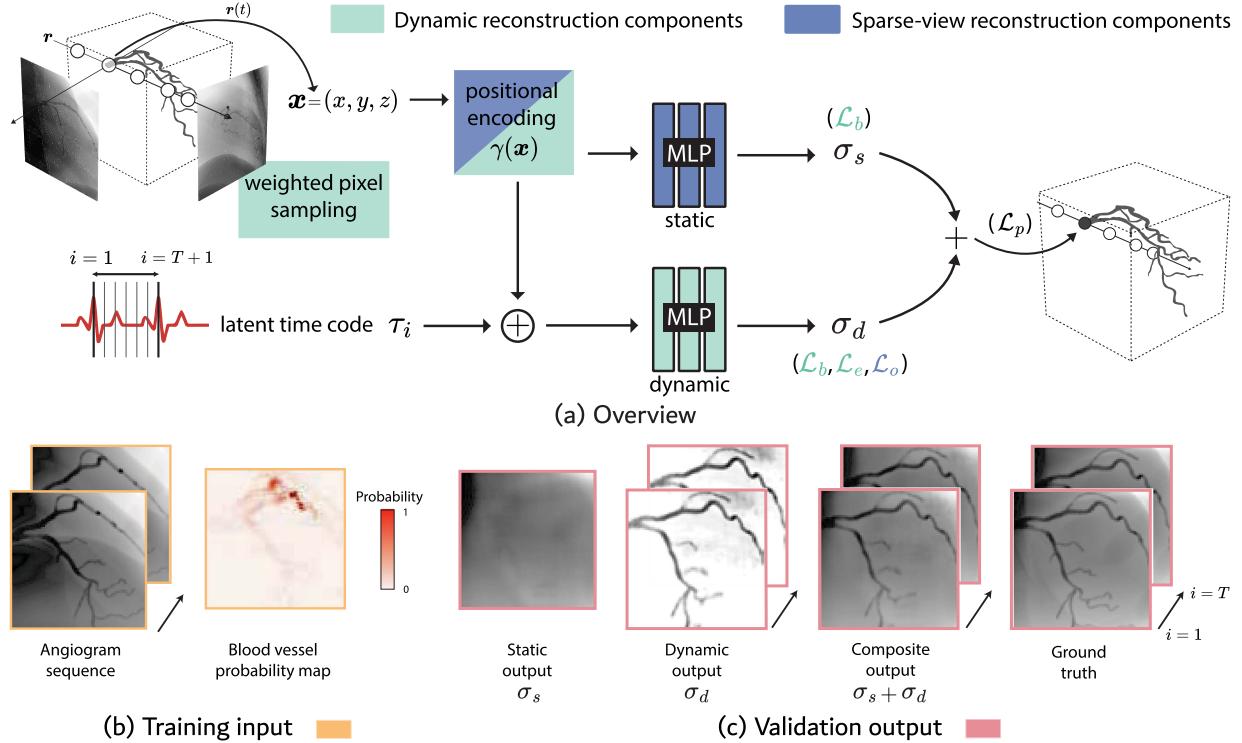


Fig. 2. Overview of the NeRF-CA method (a). The full loss function  $\mathcal{L}_f$  (1) consists of (2) Photometric Loss  $\mathcal{L}_p$ , (3) Static vs. Dynamic Factorization Loss  $\mathcal{L}_b$ , (4) Dynamic Entropy Loss  $\mathcal{L}_e$ , and (5) Dynamic Occlusion Loss  $\mathcal{L}_o$ . We also depict the (b) training input and (c) validation output, in yellow and pink, respectively.

scenes, reaching a limit in performance with 15 projections [10]. However, this number of projections is still inadequate for our clinical context and does not account for sparse structures in dynamic scenes. In our work, we are inspired by the regularization-based methods utilized for natural scenes. We propose regularization-based methods that fit blood vessel properties, effectively translating regularization-based techniques to the medical imaging setting.

NeRF methods have also been proposed for blood vessel reconstruction, addressing sparse-views and dynamics independently. For CA scenes, several NeRF-based techniques have been proposed, including 3DGR-CAR [11] and NeCA [12]. These methods address the sparse-view problem, enabling reconstructions from as few as two projections. However, they depend on accurate segmentation or extensive training data and overlook the inherent dynamics of CA data. For vessels that are not prone to extensive motion, digital subtraction angiography (DSA) can be used to remove all background structures from the 2D projection images. TiaVox [37] and Liu et al. [13] successfully reconstruct these sparse structure scenes from only 30 DSA projections. Liu et al. [13] model blood flow dynamics by decoupling the scene into a static background and dynamic blood vessel foreground. Static and dynamic scene decomposition is one of the approaches to model dynamics with NeRFs, where scenes can be decoupled in a static and dynamic component [26]. However, these DSA scenes do not experience poor vessel visibility, recently highlighted as a significant challenge in

sparse-view X-ray angiography reconstruction with NeRFs [7]. Moreover, CA scenes exhibit even more significant dynamics in foreground and background than DSA scenes.

In this work, we decouple the static and dynamic scenes and extend this strategy to address the properties of CA scenes. To the best of our knowledge, NeRF-CA is the first to effectively address the challenges of sparse-views and dynamics of CA data. Inspired by the natural scene decomposition technique D<sup>2</sup> NeRF [14], we decouple the CA scene into a dynamic coronary artery component and a background component. We uniquely leverage this static and dynamic decomposition by imposing regularization on each learned decoupled component. This regularization is inspired by successful techniques from the natural imaging domain, including structure sparsity [15], scene smoothness [9], and edge occlusion minimization [9]. However, their extension to the CA setting is not straightforward, hence a new methodology is needed. NeRF-CA allows 4D reconstruction from an extremely sparse number of angiogram views without user interaction, which is currently not possible with state-of-the-art methods.

#### IV. NERF-CA

In this section, we describe our method, NeRF-CA. Fig. 2(a) shows an overview of our method. We categorize the components of our method according to their function. Specifically, we propose dynamic reconstruction components, shown in green in

Fig. 2(a), and sparse-view reconstruction components, shown in purple in Fig. 2(a). Overall, the X-ray scene is expressed as an absorption-only model with attenuation coefficients  $\sigma$ , as will be further described in Section IV-B. Fig. 2(b), (c) depicts the model training input and validation output in yellow and pink, respectively. We will utilize these colors throughout the paper to distinguish the training (yellow) and validation (pink) views.

The basis of our method is decoupling the scene into a static and dynamic scene. The static model represents the background, and the dynamic model represents the dynamic coronary artery structure. The static and dynamic models predict their attenuation coefficients  $\sigma_s$  and  $\sigma_d$ , respectively. These coefficients are composited to obtain the scene from which the predicted pixel intensities are calculated. Fig. 2(c) depicts the outputs of our methods, specifically, the static, dynamic, and composite outputs for one validation view. Whereas the static model solely relies on the positionally-encoded 3D coordinate as input, the dynamic models also receive a per-cardiac phase latent code  $\tau_i$ . This latent code represents a cardiac phase along the cardiac cycle derived from the simultaneously recorded ECG signal. More details on the scene decomposition can be found in Section IV-C.

By decomposing the scene, we can individually impose regularizers for the static and dynamic scenes, allowing the final dynamic scene to model the coronary artery. In Section IV-A, we provide the overall objective and loss function of our method while introducing these regularizers. Sections IV-D and IV-E describe the regularizers for dynamic and sparse-view reconstruction, respectively. A core aspect of our approach is weighted pixel sampling, which guides the model to distinguish the sparse blood vessel structures from the background. This is achieved using a blood vessel probability map, which estimates the likelihood of each pixel containing a blood vessel. This probability map is shown in Fig. 2(b) and detailed in Section IV-D.

#### A. Overall Objective

Our contribution lies in developing a unique combination of dynamic and sparse-view reconstruction, enabled by an ECG-timed temporal integration of the cardiac sequence and a combination of several regularization techniques targeted to tackle the unique challenges presented by CA data. Fig. 2(c) shows examples of the predicted coronary angiogram views and the ground truth views.

The full loss of NeRF-CA  $\mathcal{L}_f$  can be summarized as

$$\begin{aligned} \mathcal{L}_f(\mathbf{r}, \tau_i) = & \mathcal{L}_p(\mathbf{r}, \tau_i) + \lambda_b \mathcal{L}_b(\mathbf{r}, \tau_i) \\ & + \lambda_e \mathcal{L}_e(\mathbf{r}, \tau_i) + \lambda_o \mathcal{L}_o(\mathbf{r}, \tau_i), \end{aligned} \quad (1)$$

where  $\mathcal{L}_p$  represents the photometric loss,  $\mathcal{L}_b$  represents the static vs. dynamic factorization loss with weight  $\lambda_b$ ,  $\mathcal{L}_e$  represents the dynamic entropy loss with weight  $\lambda_e$ , and  $\mathcal{L}_o$  represents the dynamic occlusion loss with weight  $\lambda_o$ . Each of these components will be introduced in the next paragraphs. The final model optimization is performed by sampling rays  $\mathbf{r}$  from a total set of rays  $\mathcal{R}$  defined by the ground truth pixels. Each pixel has a cardiac phase  $i$  assigned, which is denoted by a time latent code  $\tau_i$ , which will be detailed in Section IV-C.

For dynamic reconstruction, we decouple the static background from the dynamic coronary artery structure, as detailed in Section IV-D. We combine static and dynamic decomposition with a weighted pixel sampling strategy and two regularizers to accurately decouple the coronary artery from the background. Our method is self-supervised through the standard photometric loss  $\mathcal{L}_p$ , computed based on the composition of the static and dynamic components. Weighted pixel sampling is proposed to separate the sparse blood vessel structure from the background at the image level. The static vs. dynamic factorization regularizer  $\mathcal{L}_b$  is enforced to separate the static and dynamic scene on a 3D point level [14]. In addition, we enforce entropy  $\mathcal{L}_e$  in the decoupled dynamic scene to achieve the sparse structure of the blood vessel [15].

For sparse-view reconstruction, we reconstruct from an extremely sparse number of angiogram views. The sparse-view reconstruction components, consisting of windowed positional encoding and dynamic occlusion loss, will be detailed in Section IV-E. Windowed positional encoding is a proven strategy to enforce smoothness [9]. We employ this strategy in the decoupled static and dynamic scenes to enforce smoothness in both scenes, avoiding degenerate high-frequency solutions. The occlusion loss  $\mathcal{L}_o$  punishes high occlusion areas close to the camera, an expected faulty behavior when applying NeRFs for sparse-view reconstruction [9]. We demonstrate that imposing such a loss can benefit our sparse dynamic scene.

#### B. Neural Radiance Field for X-Ray

X-ray scenes can be described by the absorption-only model, assuming the Beer-Lambert law [22]. Similarly to other works applying NeRFs for X-ray scenes (e.g., [31]), we approximate the scene by  $F_\Theta(\mathbf{x}) = \sigma$ , where  $\sigma$  represents the attenuation coefficient in the X-ray scene. The predicted pixel intensity  $\hat{I}(\mathbf{r})$  of an X-ray image is then described as:

$$\hat{I}(\mathbf{r}) = I_0 \exp\left(-\int_{t_n}^{t_f} \sigma(\mathbf{r}(t)) dt\right),$$

where  $t$  represents the position along the ray, and  $t_n$  and  $t_f$  represent the near and far thresholds along the ray. The X-ray source is assumed to emit a constant initial intensity  $I_0$  independent of the viewing direction.

#### C. Static and Dynamic Scene Decomposition

The base of our method is the static and dynamic decoupling of the scene, inspired by the work of D<sup>2</sup> NeRF and Liu et al. [13], [14], and CA reconstruction works that separate the blood vessel from the background during reconstruction [38]. Specifically, we decouple our scene in a static component  $F_{\Theta_s}$  and dynamic component  $G_{\Theta_d}$ . Two MLPs represent them with weights  $\Theta_s$  and  $\Theta_d$ , respectively, as shown in Fig. 2(a).

$$\begin{aligned} F_{\Theta_s}(\gamma(\mathbf{x})) &= \sigma_s \\ G_{\Theta_d}(\gamma(\mathbf{x}), \tau_i) &= \sigma_d \end{aligned}$$

The static component  $F_{\Theta_s}$  predicts a static attenuation coefficient  $\sigma_s$  from a positionally-encoded 3D coordinate  $\gamma(\mathbf{x})$ ,

similar to the originally proposed NeRF [6]. We use a coarse-to-fine windowed positional encoding  $\gamma$  in our work, which will be further described in Section IV-E. Unlike the static component, the dynamic component  $G_{\Theta_d}$  is also conditioned on a latent code per cardiac phase  $\tau_i$  as input and predicts a dynamic attenuation coefficient  $\sigma_d$ . This latent code  $\tau_i$  is a learned vector representing a discrete cardiac phase  $i$ . Specifically, each cardiac cycle, represented by a fixed interval of the ECG as shown in Fig. 2(a), is divided into  $T$  equal parts. The parameter  $T$ , where  $i \in \{1, \dots, T\} \bmod T$ , is pre-defined based on the frame rate of the angiogram sequence. In other words, we assume that the frame time matches the cardiac phase  $i$ . A key difference in our approach lies in how we define  $\tau_i$ . While D<sup>2</sup> NeRF [14] and Liu et al. [13] represent  $\tau_i$  as per-frame time, we instead represent  $\tau_i$  as a cardiac phase in the cycle. This approach allows us to match cardiac phases across viewpoints.

The static and dynamic attenuation coefficient for each 3D coordinate  $\mathbf{x}$  are composed as  $\sigma = \sigma_s + \sigma_d$  to obtain the predicted pixel intensity  $\hat{I}(\mathbf{r}, \tau_i)$  at cardiac phase  $i$

$$\hat{I}(\mathbf{r}, \tau_i) = I_0 \exp \left( - \int_{t_n}^{t_f} (\sigma_s(\mathbf{r}(t)) + \sigma_d(\mathbf{r}(t), \tau_i)) dt \right).$$

The photometric loss  $\mathcal{L}_p$  is computed from the predicted pixel intensity and the ground truth pixel intensity. This photometric loss  $\mathcal{L}_p$  ensures that the predicted composite frame at cardiac phase  $i$  is similar to the ground truth frame at cardiac phase  $i$  from the angiogram sequence. Specifically, for every ray  $\mathbf{r}$  at cardiac phase  $i$ , we optimize

$$\mathcal{L}_p(\mathbf{r}, \tau_i) = (\hat{I}(\mathbf{r}, \tau_i) - I(\mathbf{r}, \tau_i))^2. \quad (2)$$

#### D. Dynamic Reconstruction Components

The function of the static and dynamic decoupling in our work is to separate the scene in the background and the dynamic coronary artery in a self-supervised manner. However, the coronary artery is not the only dynamic element in our CA scene. For example, the heart and ribs also experience motion. Moreover, the sparsity of the structure of the blood vessel also poses a challenge in the separation. Specifically, we propose weighted pixel sampling to address the blood vessel's separation from the image's background. We adopt static vs. dynamic factorization loss to correctly separate static and dynamic scenes on a position level [14]. Lastly, to force our dynamic scene to only model the blood vessel, we impose a sparsity structure constraint on our dynamic scene [15]. An alternative to our self-supervised approach is to directly supervise the dynamic output of our model using a coronary artery segmentation mask, such as one generated by a Frangi filter [39], which we will ablate in Section V-E. Fig. 2(c) shows an example of our dynamic reconstruction. As can be seen, the dynamic output mostly consists of the coronary artery structure, while the static output represents the background. Next, we will provide more details on the strategies to achieve these results.

**Weighted Pixel Sampling:** To optimize a NeRF model, rays are sampled for every pixel of every frame, as was described in Section IV-C. However, the blood vessel structure occupies

only a few pixels in every coronary angiogram frame. When all pixels are sampled uniformly, this structure may be missed and, therefore, difficult to separate from the background. Therefore, we uniquely introduce weighted pixel sampling to reconstruct the blood vessels rather than the background. The weights are computed based on the assumption that the blood vessel structure is of high contrast and motion. We calculate a sampling probability map based on the variance of each pixel value across the frames of one angiogram sequence. Fig. 2(b) provides an example of a probability map for one angiogram sequence. The blood vessel structure is distinguished from the background with higher sampling probability, indicated by the darker red areas. During training, we sample a pre-defined fraction of rays  $V$  from this map's thresholded set of high-variance pixels. The remaining pixels are sampled randomly from all possible pixels.

**Static vs. Dynamic Factorization Loss:** A position in space should be occupied by the static or dynamic scene but never by both. The photometric loss, detailed in (2), does not guarantee this behavior and, therefore, a correct separation of static and dynamic scenes. To enforce this separation, we adopt the Static vs. Dynamic Factorization regularizer [14]. This regularizer defines a spatial ratio of dynamic and static attenuation coefficients as

$$w(\mathbf{x}, \tau_i) = \frac{\sigma_d(\mathbf{x}, \tau_i)}{\sigma_d(\mathbf{x}, \tau_i) + \sigma_s(\mathbf{x})}.$$

This ratio is penalized when deviating from a categorical distribution  $\{0, 1\}$  through a binary entropy loss  $\mathcal{L}_b$ :

$$\mathcal{L}_b(\mathbf{r}, \tau_i) = \int_{t_n}^{t_f} H_b(w(\mathbf{r}(t), \tau_i)) dt \quad (3)$$

$$H_b(x) = -(x \cdot \log(x) + (1-x) \cdot \log(1-x))$$

D<sup>2</sup> NeRF uses a skewness term on top of the binary entropy loss to favor static explanations, assuming scenes are mostly static. However, since our dynamic component is sparse and occupies only a few points in the 3D space, applying skewness leads to degenerate solutions where the dynamic component is entirely removed. We omit the skewness term and instead achieve structure sparsity uniquely through a dynamic entropy loss, as detailed below.

**Dynamic Entropy Loss:** In the final reconstruction, the part of the scene occupied by the coronary artery should be small. Previous works for the reconstruction of CA scenes have accomplished this by enforcing a small number of points of the scene to have non-zero attenuation values [40]. However, we would also like to force a sparsity constraint on our blood vessel structure. We utilize ray entropy to enforce sparsity specifically on the dynamic component of our scene. As a result, our composite scene favors most points to be static while modeling the dynamic component as a sparse structure, accomplishing an ideal static vs. dynamic factorization. Previous works have defined ray entropy to avoid cloud-like artifacts in NeRF scenes [14], [15]. Based on these works, we define the ray density  $p_{\mathbf{r}, \tau_i}$  as

$$p_{\mathbf{r}, \tau_i}(t) = \frac{\sigma_d(\mathbf{r}(t), \tau_i)}{\int_{t_n}^{t_f} \sigma_d(\mathbf{r}(s), \tau_i) ds}.$$

The dynamic entropy loss function  $\mathcal{L}_e$  is then defined as

$$\mathcal{L}_e(\mathbf{r}, \tau_i) = - \int_{t_n}^{t_f} p_{\mathbf{r}, \tau_i}(t) \log(p_{\mathbf{r}, \tau_i}(t)) dt. \quad (4)$$

To avoid artifacts from appearing in empty spaces, we disregard non-hitting rays based on their accumulated density, similar to previous work [15]. However, given that we want to achieve sparseness for the areas that have blood vessels, we always compute entropy for rays likely to hit a blood vessel, avoiding these rays to near zero. We derive this likelihood from the blood vessel probability map, which we also utilize for weighted pixel sampling as described in the respective paragraph and shown in Fig. 2(b).

#### E. Sparse-View Reconstruction Components

The main goal of our work is to reconstruct a coronary artery structure from an extremely sparse amount of views. While the dynamic reconstruction components provide a decoupled reconstruction of the coronary artery, degenerate solutions may still occur when limiting the number of views for training. We, therefore, introduce two components for sparse-view reconstruction. Namely, we incorporate windowed positional encoding  $\gamma(\mathbf{x})$  to achieve a smooth CA scene. Moreover, we also impose edge occlusion regularization on the dynamic scene to avoid common degenerate solutions where artifacts appear near the camera.

*Windowed Positional Encoding:* A key element in reconstructing high-fidelity scenes with NeRFs is the positional encoding [21]. While these resulting high-frequency inputs allow for fast reconstruction of high-frequency components, they are prone to bias the model to generate high-frequency artifacts, especially in a sparse-view setting [9]. Several works in NeRF have introduced a frequency regularization method, which linearly increases the frequency of the positional encoding over time [9], [41], [42], enforcing smoothness in the scene. Recently, FreeNeRF [9] has shown that their windowed positional encoding is especially effective in avoiding degenerate solutions in an extremely sparse-view setting. This windowed positional encoding is applied on both the static and dynamic component of our scene, which enforces smoothness in each decoupled component. The positional encoding is defined as

$$\gamma_\alpha(\mathbf{x}) = (\mathbf{x}, \dots, \alpha_L(n) \sin(2^L \pi \mathbf{x}), \alpha_L(n) \cos(2^L \pi \mathbf{x})). \quad (5)$$

The weight  $\alpha_L(n) = \text{clamp}(\frac{nL}{N}, 0, 1)$  is set based on the current training iteration  $n$ , and the hyperparameter  $N$  which denotes the number of iterations needed to reach the total number of frequency bands  $L$ .

*Dynamic Occlusion Loss:* While windowed positional encoding is effective, this strategy still allows for degenerate solutions in the sparse-view setting. Specifically, FreeNeRF [9] demonstrated that NeRF models may generate a solution where high-frequency artifacts appear near the camera when presented with an extremely sparse amount of training views. These solutions are correct for the training views but do not correctly generalize to new viewpoints. To avoid this behavior, FreeNeRF [9] proposes an occlusion regularization loss, punishing high occlusion areas close to the camera. While this loss may be effective in

natural scenes with objects in the center, our composite X-ray scene consists of a few non-zero attenuation points. As a result, applying this loss directly to the composite scene would not avoid degenerate outcomes. In contrast, our learned dynamic scene, representing the coronary artery, is expected to be nearly empty, with the coronary artery positioned near the center due to the X-ray system setup. Therefore, on top of our self-supervised decomposition approach, we apply occlusion regularization exclusively to the learned dynamic scene.

We define the occlusion regularization loss  $\mathcal{L}_o$  as

$$\mathcal{L}_o(\mathbf{r}, \tau_i) = \int_{t_n}^{t_f} (\sigma_d(\mathbf{r}(t), \tau_i) \cdot M(t)) dt, \quad (6)$$

where  $M(t)$  is a binary mask applied respectively to the distance  $t$  of a point along the ray starting from the X-ray source. This mask evaluates to 1 if the point's distance  $t$  is lower than the distance threshold  $D$ , otherwise it evaluates to 0. Note that FreeNeRF [9] does not penalize based on a distance  $t$  but rather on the first  $C$  points starting from the X-ray source. We rather penalize based on a distance threshold  $D$ , as this is more applicable to a C-arm system setup where physical distances between the X-ray source and patient are known. Specifically, we enforce the points with a distance smaller than  $D$  to be outside the patient and, therefore, nearly empty in density, which matches the goal of the occlusion regularization.

(1) provides an overview of NeRF-CA's full loss function. Rather than utilizing constant weights for the hyperparameters  $\lambda_b$ ,  $\lambda_e$ , and  $\lambda_o$ , we linearly increase them over time with a pre-determined number of delay steps. This approach is fitting due to the coarse-to-fine transition of the windowed positional encoding, as imposing weights too early in a low-frequency setting may lead to degenerate solutions.

## V. EXPERIMENTS

In this section, we present experiments to evaluate and validate NeRF-CA. We present the 4D phantom datasets that allow us to evaluate quantitatively. Our method has multiple components and parameters that we evaluate through extensive ablation studies. Furthermore, we compare our method to existing state-of-the-art methods to show the strengths and weaknesses of our approach.

#### A. Datasets

For the evaluation, we use parameterized 4D synthetic phantoms previously used in a similar evaluation context [43]. These phantoms allow us to generate realistic ground-truth X-ray sequences from the anatomical cardiac vessel structure from any viewpoint at any cardiac phase. Utilizing these phantoms, we can effectively focus our evaluation on the aspects NeRF-CA specifically addresses, i.e., sparse-views and cardiac motion. Meanwhile, these phantoms also exhibit varying aspects of vessel sparsity and poor vessel visibility, allowing us to validate our model under these conditions. These phantoms intentionally exclude complex characteristics of real clinical CA data, such as blood vessel contrast inhomogeneity, lack of cardiac synchronization among views, and respiratory motion, to ensure the

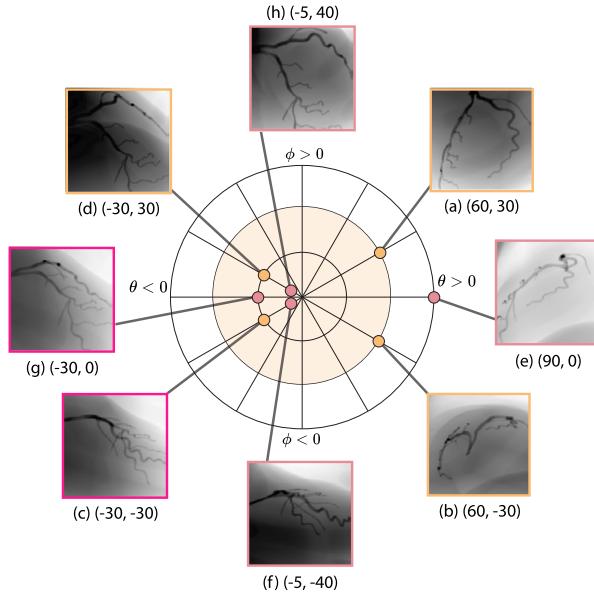


Fig. 3. Optimal viewing angles for the left coronary artery expressed in Euler angles ( $\theta, \phi$ ). The angles are indicated by circles displayed on a flattened sphere. Our 4 training views (a-d) and 4 validation views (e-h) are displayed in yellow and pink circles, respectively. The light yellow area represents the limited angle range of  $60^\circ$ , from which we obtain training views beyond the 4 projection setting.

evaluation remains focused on the aspects NeRF-CA is designed to address. Using clinical data with these factors, as described in Section II-A, would introduce confounding variables, limiting the evaluation's value. Additionally, real clinical CA data offers limited flexibility for ablation studies; for example, the typical clinical setting provides a maximum of 4 views per patient. Moreover, clinical CA data lacks 3D or 4D ground-truth volumes needed for generating or evaluating 2D CA sequences.

We use the 4D XCAT phantom [16] and a 4D coronary computed tomography angiography (CCTA) scan, called MAGIX [17]. Both datasets depict the left coronary artery (LCA), which is more complex in geometry than the right coronary artery and, therefore, more clinically interesting for 3D reconstruction applications. For both datasets, we simulate homogeneous contrast injection using higher attenuation values in segmented areas, similar to previous work [7]. Fig. 1 shows examples of one angiogram sequence for these 4D datasets. In these datasets, one angiogram consists of  $T = 10$  frames, each corresponding to a discrete cardiac phase  $i$ , assuming minimal respiratory motion. As can be seen from Fig. 1, the 4D XCAT dataset [16] includes a detailed LCA structure, fitting for evaluating the effect of vessel sparsity on our reconstruction quality. Moreover, the 4D XCAT phantom allows for the simulation of cardiac motion according to the ECG-based cardiac phase. We sample the XCAT phantom with an isotropic resolution of  $0.5 \text{ mm}^3$ . The 4D CCTA dataset MAGIX, obtained from the OsiriX platform [17], consists of 10 consecutive scans that represent different cardiac phases across the cardiac cycle. As contrast is injected throughout the body during CCTA scans,

each scan also includes varying contrast in the background, which allows us to evaluate the effect of the dynamic background on our reconstructions. Due to the CCTA's lower resolution of  $0.4 \times 0.4 \times 2 \text{ mm}^3$ , the obtained LCA structure is coarse. To generate projections from the 4D volumes, we utilize the tomographic toolbox TIGRE [44]. We adjust the geometry settings to simulate the C-arm setting of CA data, where the focus is on the main branches while deeper branches may be out of view.

In clinical practice, the LCA is typically visualized with 8 literature-based distinctive angiographic views [45], [46] to minimize overlap and foreshortening [47]. Fig. 3 shows these views in Euler angles ( $\theta, \phi$ ) distributed along a flattened sphere. The image examples are from the XCAT dataset. For 4D reconstruction, selecting viewing angles in a sparse setting is crucial. We, therefore, base our training and validation sets on these clinical optimal views. We will train our method using 4 to 70 projections. For 4 projections, we choose 4 well-distributed optimal angles, indicated by the yellow circles in the figure and images (a-d). The remaining views, indicated by pink circles and images (e-h), are used to validate all cases. For more projections, we uniformly sample training views within a limited range of  $60^\circ$ , indicated by the light yellow area. This area covers the range of most optimal viewing angles while avoiding views too close to the validation set.

### B. Evaluation Set-Up

We perform both qualitative and quantitative evaluations in our work. The qualitative approach is suitable to overcome known problems in assessing medical imaging with pure quantitative metrics [48]. For our application, where topological accuracy is most relevant, this is especially valuable [7]. We evaluate quantitatively by computing the re-projection error for 2D novel view synthesis, fitting the roadmap application where the 4D reconstruction would be re-projected to the C-arm angle of interest, as discussed in Section II-A.

The ultimate goal of our work is to evaluate the reconstruction of the coronary artery structure rather than direct volume-to-volume comparison. Specifically, we are interested in the topology of the structure, for which, to our knowledge, no appropriate quantitative measures have been proposed [7]. Therefore, we mostly rely on qualitative analysis for the structure itself. Besides, we also perform binary comparisons by computing Dice, a well-known metric for blood vessel segmentation tasks [49]. We only consider the positive or blood vessel structure class when computing Dice. This approach compensates for the imbalance of these classes due to the sparsity of the structure in the image. To compute a binary Dice value, we generate novel views by applying maximum intensity projection (MIP) from the dynamic attenuation values  $\sigma_d$ , assuming that the blood vessel structure has the highest attenuation values and is purely dynamic. This means that we strictly consider the dynamic coronary artery component for our Dice computation. We obtain the final binarized images from thresholding the MIP views based on the 4D ground-truth values of the coronary artery structure. When comparing to static models, we compute the MIP from the composite  $\sigma$  attenuation values. We also evaluate the overall 2D

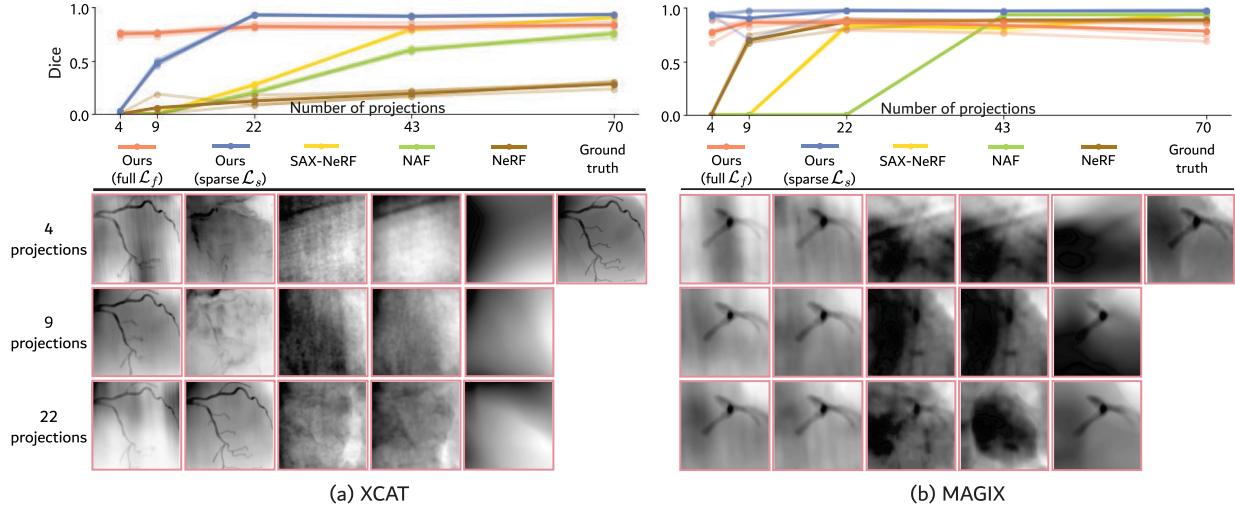


Fig. 4. Dice scores and qualitative results for the comparison to existing methods across varying number of training projections, including Ours (full), Ours (sparse), SAX-NeRF [10], NAF [36], and NeRF [6] for the (a) XCAT and (b) MAGIX datasets. The low-opacity lines represent the scores of the four validation views, while the bold line shows the mean score across these views.

CA image prediction with common image quality metrics peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM). Besides quality metrics, we also report the computation time.

### C. Implementation Details

The models were implemented in Pytorch and trained on an RTX A5000 GPU. We utilize the same hyperparameters for both datasets. The models have an MLP architecture of 4 hidden layers with 128 neurons with ReLU activations, commonly used in NeRF applications [6]. We utilize a Softplus activation to obtain a positive attenuation value output. We train our models with an Adam optimizer with a linearly decaying learning rate from  $1 \times 10^{-3}$  to  $1 \times 10^{-5}$  in 150000 steps. The number of training iterations is 200000 with a ray batch size of 1024. We train with images of size  $200 \times 200$  pixels and sample 500 points along each ray.

The hyperparameters for our method components were chosen based on a grid search. Although many different hyperparameters are involved, we show they can be consistently used across different datasets, highlighting the robustness of our method. We impose a maximum frequency band for the windowed positional encoding, as shown in (5), of  $L = 12$  over  $N = 150000$  iterations. We start the coarse-to-fine transition at a frequency band of 1, therefore activating the first frequency bands ( $\sin(\pi\mathbf{x}), \cos(\pi\mathbf{x})$ ) from the start of training. This has shown to be more advantageous for separating foreground and background. A fraction of  $V = 0.5$  of the ray batch size is sampled for the purpose of weighted pixel sampling. The static vs. dynamic factorization loss weight  $\lambda_b$  increases from  $1 \times 10^{-12}$  to  $1 \times 10^{-10}$ , with a delay of 40000 iterations. The dynamic entropy loss weight  $\lambda_e$  linearly increases from  $1 \times 10^{-12}$  to  $1 \times 10^{-10}$  without a delay, where only rays are considered with a minimum accumulated density of  $1 \times 10^{-4}$ . The dynamic

occlusion loss weight  $\lambda_o$  increases from  $1 \times 10^{-8}$  to  $1 \times 10^{-5}$  linearly, delayed by 40000 iterations with a distance threshold of  $D = 0.2$ .

### D. Comparison to Existing Methods

In this section, we compare our method to state-of-the-art 3D reconstruction methods. As discussed in Section III-A, coronary X-ray angiography techniques require manual interactions or large training datasets. As our method does not require these settings, we compare it to state-of-the-art NeRF X-ray reconstruction methods, which obtain static reconstructions from one dataset directly. We particularly focus on state-of-the-art methods for sparse-view reconstruction, according to the main goal of our work. We compare our method to SAX-NeRF [10] and NAF [36], as well as the original NeRF method as a baseline [6]. Lastly, we include our method with only the sparse-view reconstruction components, referred to as sparse method  $\mathcal{L}_s$ , as a representative of a state-of-the-art sparse-view natural scene NeRF-based reconstruction method [9]. This sparse method is represented by a part of our full loss  $\mathcal{L}_f$  (see (1)) as loss  $\mathcal{L}_s$ , where  $\mathcal{L}_s(\mathbf{r}, \tau_i) = \mathcal{L}_p(\mathbf{r}, \tau_i)$ . From our proposed components, only the photometric loss  $\mathcal{L}_p$  (see (2)) and the windowed positional encoding  $\gamma$  (see (5)) is applied. We utilize one static MLP trained individually for every single timestep. We note that, unlike the natural scene method [9], we do not apply the dynamic occlusion loss  $\mathcal{L}_o$  (see (6)). This is because the assumption of edge emptiness, as described in Section IV-D, does not apply to our static non-empty scene.

As our work focuses on the reconstruction of the coronary artery structure, we report the Dice score for this evaluation. We refer to Section V-B for details on the Dice score computation from the predicted 4D attenuation volumes. Fig. 4 shows the Dice scores for the (a) XCAT and (b) MAGIX dataset for our full

method  $\mathcal{L}_f$ , our sparse method  $\mathcal{L}_s$ , SAX-NeRF [10], NAF [36], and NeRF [6]. The low-opacity lines represent the scores of the four validation views, introduced in Fig. 3, while the bold line shows the mean score across these views. The figure also displays qualitative examples for the settings of 4, 9, and 22 projections. SAX-NeRF, NAF, and NeRF both show low Dice scores in the setting of 4 and 9 projections for both datasets. This low score is consistent across the four validation views, as can be seen by the consistent low-opacity lines. In the qualitative examples of these methods, we can see that the coronary artery structure is either missing or obstructed by background noise. In this setting, our sparse method  $\mathcal{L}_s$  performs well on the MAGIX dataset but fails to reconstruct a topologically accurate structure for the XCAT dataset. Notably, our full method  $\mathcal{L}_f$  significantly outperforms the other methods in this setting. It reconstructs topologically accurate structures even with as few as 4 training projections. Only starting from 22 and 43 for the MAGIX and XCAT datasets, respectively, SAX-NeRF, NAF, and NeRF start to produce accurate reconstructions. Although our full method  $\mathcal{L}_f$  loses slight quality compared to these methods, our sparse method  $\mathcal{L}_s$  still significantly outperforms them in this setting. All in all, we significantly outperform state-of-the-art sparse-view NeRF reconstruction methods in the CA domain, with our full method  $\mathcal{L}_f$  in the extremely sparse setting and our sparse method  $\mathcal{L}_s$  beyond this setting.

### E. Ablation Studies

We evaluate the multiple components and parameters of our work through ablation studies. First, we examine the dynamic and sparse-view reconstruction components individually, as introduced in Fig. 2(a). Next, we analyze the consistency in quality for the dynamic reconstruction across the cardiac phases. Lastly, we evaluate the effect of weighted pixel sampling and the dynamic entropy and occlusion loss components. Furthermore, we validate our choice for self-supervision over a segmentation-supervised approach, which can be found in the supplementary material. Specifically, we demonstrate that the self-supervised approach achieves higher accuracy, as it does not depend on the automatic error-prone segmentation methods that produce inconsistencies in supervision.

*Dynamic and Sparse-view Reconstruction Components:* In Section IV and Fig. 2(a), we illustrate how NeRF-CA effectively combines dynamic and sparse-view reconstruction components. In this section, we discuss the outcomes of our full method  $\mathcal{L}_f$  compared to these separate components, denoted as our sparse method and our dynamic method, respectively. The definition of our sparse method with loss  $\mathcal{L}_s$  is given in Section V-D. We define our dynamic method, as shown in Fig. 2(a) in green, as a part of our full loss  $\mathcal{L}_f$  (see (1)) as loss  $\mathcal{L}_d$

$$\mathcal{L}_d(\mathbf{r}, \tau_i) = \mathcal{L}_p(\mathbf{r}, \tau_i) + \lambda_b \mathcal{L}_b(\mathbf{r}, \tau_i) + \lambda_e \mathcal{L}_e(\mathbf{r}, \tau_i),$$

where we also apply the windowed positional encoding  $\gamma$  (see (5)). We exclude the dynamic occlusion loss  $\mathcal{L}_o$  (see (1)) to verify its role in sparse-view reconstruction. Fig. 5 shows the quantitative results for the XCAT and MAGIX dataset, where (a-c) show the Dice, PSNR, and SSIM scores for the XCAT

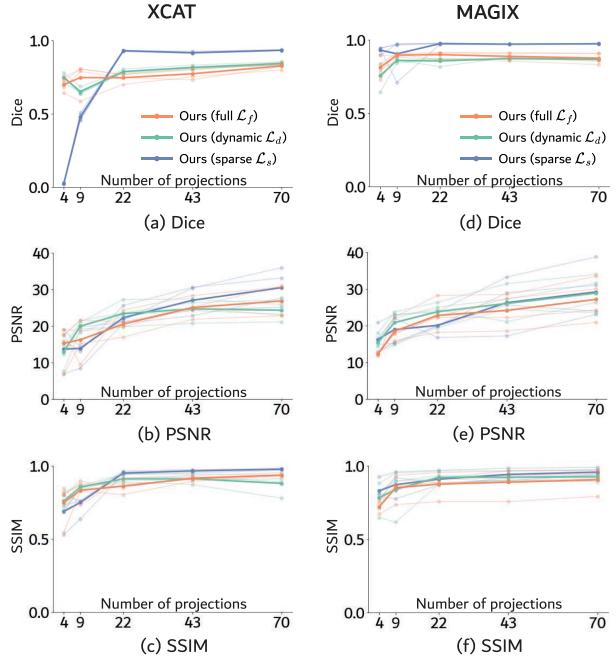


Fig. 5. Quantitative results in Dice, PSNR, and SSIM for the XCAT and MAGIX datasets across different numbers of training projections. XCAT: (a) Dice, (b) PSNR, and (c) SSIM. MAGIX: (d) Dice, (e) PSNR, and (f) SSIM. We display measures for our full  $\mathcal{L}_f$ , dynamic  $\mathcal{L}_d$ , and sparse  $\mathcal{L}_s$  methods. The low-opacity lines represent the scores of the four validation views, while the bold line shows the mean score across these views.

dataset and (d-f) show these scores for the MAGIX dataset. We display these scores across various projection counts for our full  $\mathcal{L}_f$  (in orange), sparse  $\mathcal{L}_s$  (in blue), and dynamic  $\mathcal{L}_d$  (in green) methods. The low-opacity lines represent the scores of the four validation views averaged over all timesteps, while the bold line shows the mean score across these views.

First, we evaluate the reconstruction of only the coronary artery structure using Dice scores, as it addresses the main goal of our work. Our full method  $\mathcal{L}_f$  (orange) achieves high Dice scores in the extremely sparse-view settings of 4 and 9 projections. These scores are consistent across the four validation views, as shown by the low-opacity lines. For the XCAT dataset, our full method  $\mathcal{L}_f$  significantly outperforms our sparse method  $\mathcal{L}_s$  (blue) and slightly outperforms our dynamic method  $\mathcal{L}_d$  (green), as can be seen from Fig. 5(a). The scores across the MAGIX dataset are more similar, as seen in Fig. 5(d), showing that all three of our methods perform well on this dataset, although the sparse method  $\mathcal{L}_s$  slightly outperforms our other methods. Fig. 6 shows the qualitative examples for the 4 training projections setting. We show the static output, dynamic output, composite output, and ground truth images. By comparing the different methods across the rows, we can observe that our full method  $\mathcal{L}_f$  (bottom row, orange) achieves the best separation of the dynamic coronary artery from the static background, removing the occluding background from the novel views. For the XCAT dataset (Fig. 6(a)), the reconstructed structure is topologically

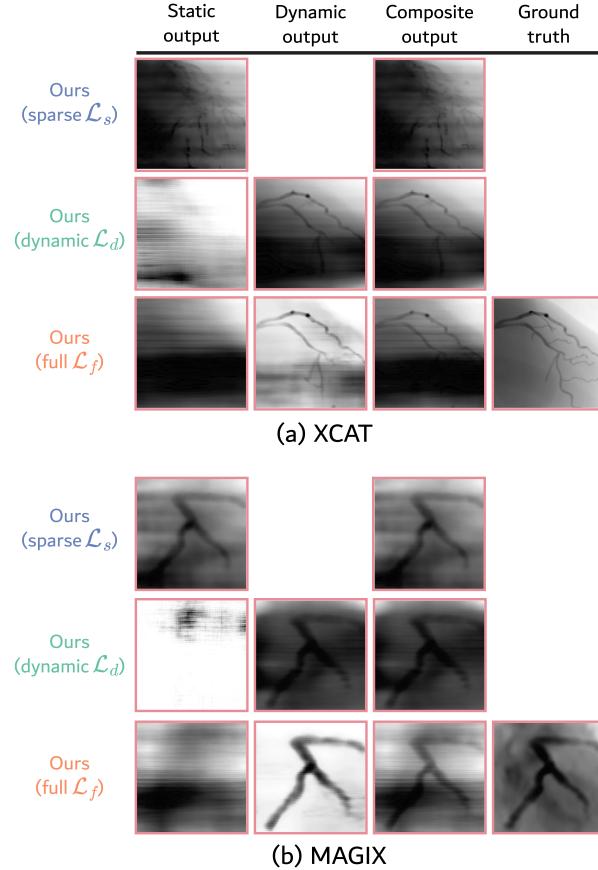


Fig. 6. Qualitative results for the (a) XCAT and (b) MAGIX datasets for our full  $\mathcal{L}_f$ , dynamic  $\mathcal{L}_d$ , and sparse  $\mathcal{L}_s$  methods for the setting of 4 training projections. We show the static, dynamic, and composite output, as well as the ground truth projection.

accurate, with the primary branches clearly defined, which fits our clinical applications. For the MAGIX dataset (Fig. 6(b)), our models achieve a high-fidelity structure without loss of quality, likely due to the easier coarser nature of the structure. We also report scores beyond the extremely sparse-view setting, as seen in Fig. 5(a), (d). For many projections, our full  $\mathcal{L}_f$ , dynamic  $\mathcal{L}_d$ , and sparse  $\mathcal{L}_s$  methods all perform well in reconstructing the coronary artery structure. However, our full  $\mathcal{L}_f$  and dynamic  $\mathcal{L}_d$  methods lose quality with more training projections, whereas our sparse method  $\mathcal{L}_s$  does not as can be seen from the flat line. Our full method  $\mathcal{L}_f$  and dynamic  $\mathcal{L}_d$  method impose regularization constraints on the separation of foreground and background, which may explain the score loss with more projections. Although our work focuses on extremely sparse-view settings, the sparse method  $\mathcal{L}_s$  might be more suitable for reconstructions with more projections.

We also report the overall CA image comparison scores in PSNR and SSIM in Fig. 5 and the overall running time. The drop in PSNR scores (Fig. 6(b), (e)) indicates that our full method is susceptible to background noise. Meanwhile, the SSIM scores (Fig. 6(c), (f)) indicate that the overall performance

is still accurate, even in the extremely sparse setting of 4 and 9 projections for both datasets. The sparse method  $\mathcal{L}_s$  (blue) is more accurate in background reconstruction when presented with more projections, highlighting its potential outside the extremely sparse-view setting. We do observe variation in scores among the validation views as can be seen from the low-opacity lines in Fig. 5(b), (c), (e), and (f), which is likely due to poor vessel visibility, which can be observed in the qualitative examples in Fig. 6. However, separating the structure from the background mitigates this issue. Therefore, our full method  $\mathcal{L}_f$  remains effective, whereas our sparse  $\mathcal{L}_s$  and dynamic  $\mathcal{L}_d$  methods may be less applicable due to poor vessel visibility. Since the Dice score best reflects our application, which is reconstructing the blood vessel structure, we will only report this score for the remaining cases. The running time of our full method  $\mathcal{L}_f$ , dynamic method  $\mathcal{L}_d$ , and sparse method  $\mathcal{L}_f$  are 5.5, 5.5 and 3 hours, respectively.

Overall, our full method  $\mathcal{L}_f$  effectively decouples the scene into a dynamic coronary artery structure and a static background, performing well in extremely sparse-view settings and showing high potential for coronary angiography applications. Beyond the extremely sparse-view setting, our sparse method  $\mathcal{L}_s$  outperforms our full method  $\mathcal{L}_f$ , highlighting its possible applications when more training projections are available.

*Dynamic Scene Reconstruction:* In our work, we decouple the scene into a static and dynamic component while imposing several regularizers to force the dynamic component to solely model the coronary artery structure. We, therefore, assume that the CA scene is mostly static aside from the coronary artery structure. As a result, any remaining background dynamics, such as contrast inhomogeneity in the MAGIX CCTA dataset, will be modeled as static, which may lead to background noise and inconsistency in output across the cardiac phases. In this section, we evaluate whether our full method  $\mathcal{L}_f$  maintains quality across the dynamic cardiac phases.

Fig. 7 shows how our method compares to the ground truth for our two datasets, (a) XCAT and (b) MAGIX, in a setting of 4 angiogram views. We display the scores for the four validation views as low opacity lines, and the mean score computed from these views as a bold line. For the (a) XCAT dataset, the line graphs indicate that the scores of our full  $\mathcal{L}_f$  method are consistent across the cardiac phases, which can be derived from the constant slope of the line. The qualitative examples, displaying the dynamic output and MIP of the ground truth, show that the slightly lower quality of the full method  $\mathcal{L}_f$  is due to the missing sparse branches. These sparse branches are less important in our application, where topology is most relevant. Therefore, our full method  $\mathcal{L}_f$  maintains quality across the dynamic cardiac phases. For the (b) MAGIX dataset, we observe more variation in Dice score across timesteps and among the different validation views. We can observe that this variation is due to background noise appearing in the dynamic scene. The MAGIX dataset consists of dynamic contrast flow across the cardiac phases, resulting in these parts being modeled as dynamic by our model. However, the background noise is primarily outside of the coronary artery structure, not affecting the reconstruction quality of the structure itself. All in all, our model consistently maintains quality for

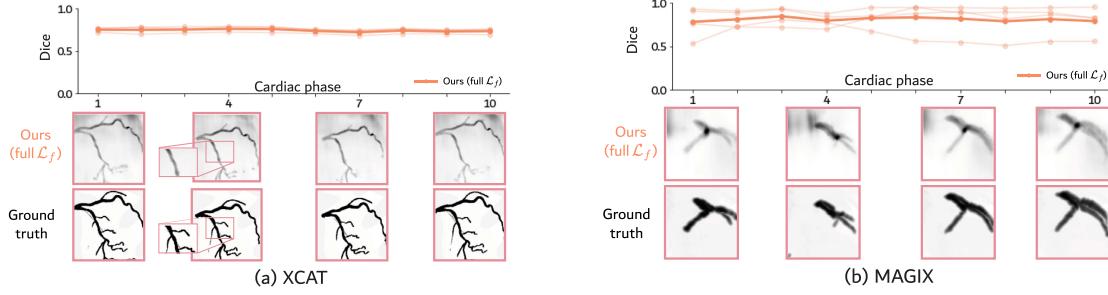


Fig. 7. Dice scores for our full method  $\mathcal{L}_f$  over the cardiac phases for the (a) XCAT and (b) MAGIX datasets. The model is trained with 4 angiogram views. The low-opacity lines represent the scores of the four validation views, whereas the bold line represents the mean score across these views.

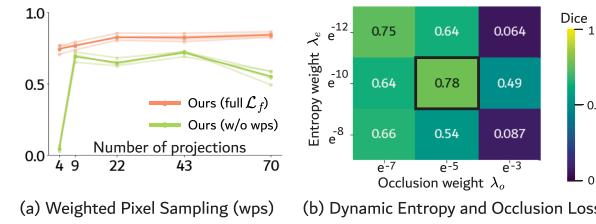


Fig. 8. Dice scores for the XCAT dataset for the (a) Weighted Pixel Sampling (wps) and (b) Dynamic Entropy and Occlusion Loss ablation studies. (a) Scores for our full method (full  $\mathcal{L}_f$ ) and our method without weighted pixel sampling (w/o wps). (b) Scores across hyperparameter combinations of the dynamic entropy weight  $\lambda_e$  and dynamic occlusion weight  $\lambda_o$ , where the black box indicates our hyperparameter setting.

reconstructing the coronary artery structure, fitting our clinical application.

**Weighted Pixel Sampling:** We incorporate weighted pixel sampling in our method to enhance the reconstruction of blood vessel structures. This approach targets sampling in the heuristically determined regions of the blood vessels, enhancing the reconstruction of these sparse structures, as described in Section IV-D. Fig. 8(a) displays the Dice scores for our full method  $\mathcal{L}_f$  and our method without weighted pixel sampling (w/o wps) over the number of training projections for the XCAT dataset. The results indicate that weighted pixel sampling (in orange in Fig. 8(a)) generally improves the score across all projection counts, significantly boosting performance in the extremely sparse setting of 4 projections from a nearly zero score to a high Dice score with accurate structure reconstruction. For the MAGIX dataset, we also observe a slight overall score increase with weighted pixel sampling, as shown in the supplementary material.

**Dynamic Entropy and Occlusion Loss:** In this section, we evaluate the the dynamic entropy  $\mathcal{L}_e$  and dynamic occlusion  $\mathcal{L}_o$  losses, which are described in Section IV-D and Section IV-E, respectively. Since both losses affect the dynamic model, we analyze their combined effectiveness, as they are likely interrelated. Fig. 8(b) shows the Dice scores for the XCAT dataset across the dynamic entropy  $\lambda_e$  and dynamic occlusion  $\lambda_o$  weights for the 4 training projection settings. The optimal weights for our model are indicated by the black square with an entropy weight  $\lambda_e$  of  $e^{-10}$  and the occlusion weight  $\lambda_o$  of  $e^{-5}$ , leading to a Dice

score of 0.78. We observe that the most optimal Dice score is obtained by combining both losses, demonstrating that both are essential for optimal results. For the MAGIX dataset, we observe similar patterns for the same hyperparameters, as shown in the supplementary material.

## VI. DISCUSSION AND CONCLUSION

We propose NeRF-CA, the first step toward a 4D Neural Radiance Field reconstruction method for X-ray coronary angiography in an extremely sparse-view setting. Our main contribution is the methodological aspects to solve the components of cardiac motion and sparse-views of the complex CA 4D reconstruction problem, given that literature has yet to propose a solution to the full clinical problem. Specifically, we propose a method that uniquely integrates scene decomposition with tailored regularization techniques to accurately decouple the CA scene in a static background and dynamic coronary artery structure. We demonstrate the effectiveness of our method both in the dynamic and sparse-view reconstruction settings. We have compared NeRF-CA to existing NeRF-based techniques for sparse-view 3D reconstruction, significantly outperforming them for CA blood vessel reconstruction. Our ablation studies have also demonstrated the need for the individual components proposed in our method. Notably, we achieve useful 4D reconstructions with four synthetic angiogram views, which are designed to capture realistic characteristics of CA data, showing potential for clinical adaptation.

Despite these advantages, we also highlight a few limitations of NeRF-CA. NeRF-CA does not address all challenges in a clinical setting. Motion-related challenges that are not addressed include respiratory motion, contrast inhomogeneity, and synchronization across viewpoints. Specifically, our method assumes synchronous motion with the ECG signal across viewpoints, which cannot be assumed in clinical data due to these additional challenges. Future work could take inspiration from the various CA-based works that propose specific techniques to deal with these motion-related challenges in CA data [3]. Nevertheless, integrating these techniques into NeRF-CA is complex, as they typically rely on deformable models [50] or image registration [18], which require human interactions, such as segmentations. In contrast, NeRF-CA implicitly models motion without segmentations, making direct integration non-trivial

and beyond the scope of this work. Furthermore, we validate our method using synthetic data, which has limitations such as simplified motion and homogeneous vessel contrast. However, this approach enabled a quantitative evaluation of NeRF-CA and its components. Given a further developed method which addresses all challenges in real CA data, testing on real data would be a necessary step. Additionally, geometry inaccuracies from C-arm calibration in clinical settings, unlike the perfect camera geometry assumed in our work, could impact performance and require further investigation. Finally, improving the running time of our method from hours to minutes is crucial for integration into clinical workflows. Future work could explore techniques like volumetric representations [36] or 3D Gaussian Splatting [51], which have shown promise in accelerating NeRF-based medical reconstructions. However, integrating these techniques is not straightforward, given that these methods do not represent the scene as a continuous space, which we rely on in our approach, both for static-dynamic decomposition and sparse-view modelling. Therefore, these existing frameworks would require large adaptations and are beyond the scope of this work.

In this work, we proposed NeRF-CA, the first step toward a 4D reconstruction method for CA data in the extremely sparse-view setting without user interaction. We demonstrate adequate blood vessel reconstructions from as few as four angiogram viewswith synthetic data, showing the potential for clinical applications and outperforming state-of-the-art NeRF-based 3D reconstruction techniques.

#### ACKNOWLEDGMENT

This research was performed within the MEDUSA project as part of the Eindhoven MedTech Innovation Center collaboration between Eindhoven University of Technology, Philips Medical Systems, and Catharina Ziekenhuis Eindhoven.

#### REFERENCES

- [1] N. E. Green, S.-Y. J. Chen, J. C. Messenger, B. M. Groves, and J. D. Carroll, “Three-dimensional vascular angiography,” *Curr. Problems Cardiol.*, vol. 29, no. 3, pp. 104–142, 2004.
- [2] K. Piayda et al., “Dynamic coronary roadmapping during percutaneous coronary intervention: A feasibility study,” *Eur. J. Med. Res.*, vol. 23, pp. 1–7, 2018.
- [3] S. Çimen, A. Gooya, M. Grass, and A. F. Frangi, “Reconstruction of coronary arteries from X-ray angiography: A review,” *Med. Image Anal.*, vol. 32, pp. 46–68, 2016.
- [4] K. Iyer, B. K. Nallamothu, C. A. Figueroa, and R. R. Nadakuditi, “A multi-stage neural network approach for coronary 3D reconstruction from uncalibrated X-ray angiography images,” *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 17603.
- [5] Y. Zhu, Y. Wang, C. Di, H. Liu, F. Liao, and S. Ma, “Sparse and transferable three-dimensional dynamic vascular reconstruction for instantaneous diagnosis,” *Nature Mach. Intell.*, vol. 7, pp. 730–742, 2025.
- [6] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing scenes as neural radiance fields for view synthesis,” *Commun. ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [7] K. W. H. Maas, N. Pezzotti, A. J. E. Vermeer, D. Ruijters, and A. Vilanova, “NeRF for 3D reconstruction from X-ray angiography: Possibilities and limitations,” in *Proc. VCBM 2023: Eurographics Workshop Vis. Comput. Biol. Med.*, 2023, pp. 29–40.
- [8] K. Gao, Y. Gao, H. He, D. Lu, L. Xu, and J. Li, “NeRF: Neural radiance field in 3D vision, a comprehensive review,” 2022, arXiv: 2210.00379.
- [9] J. Yang, M. Pavone, and Y. Wang, “FreeNeRF: Improving few-shot neural rendering with free frequency regularization,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 8254–8263.
- [10] Y. Cai, J. Wang, A. Yuille, Z. Zhou, and A. Wang, “Structure-aware sparse-view X-ray 3D reconstruction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 11174–11183.
- [11] X. Fu et al., “3DGR-CAR: Coronary artery reconstruction from ultra-sparse 2D X-ray views with a 3D gaussians representation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2024, pp. 14–24.
- [12] Y. Wang, A. Banerjee, and V. Grau, “NECA: 3D coronary artery tree reconstruction from two 2D projections via neural implicit representation,” *Bioengineering*, vol. 11, no. 12, 2024, Art. no. 1227.
- [13] Z. Liu et al., “3D vessel reconstruction from sparse-view dynamic DSA images via vessel probability guided attenuation learning,” 2024, arXiv: 2405.10705.
- [14] T. Wu, F. Zhong, A. Tagliasacchi, F. Cole, and C. Oztireli, “D<sup>2</sup>NeRF: Self-supervised decoupling of dynamic and static objects from a monocular video,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 32653–32666.
- [15] M. Kim, S. Seo, and B. Han, “InfoNeRF: Ray entropy minimization for few-shot neural volume rendering,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12912–12921.
- [16] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, “4D XCAT phantom for multimodality imaging research,” *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, 2010.
- [17] A. Rosset, L. Spadola, and O. Ratib, “Osirix: An open-source software for navigating in multidimensional dicom images,” *J. Digit. Imag.*, vol. 17, pp. 205–216, 2004.
- [18] S. J. Chen and D. Schäfer, “Three-dimensional coronary visualization, part 1: Modeling,” *Cardiol. Clin.*, vol. 27, no. 3, pp. 433–452, 2009.
- [19] J. Knutti et al., “2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes: The task force for the diagnosis and management of chronic coronary syndromes of the European society of cardiology (ESC),” *Eur. Heart J.*, vol. 41, no. 3, pp. 407–477, 2020.
- [20] G. Shechter, J. R. Resar, and E. R. McVeigh, “Displacement and velocity of the coronary arteries: Cardiac and respiratory motion,” *IEEE Trans. Med. Imag.*, vol. 25, no. 3, pp. 369–375, Mar. 2006.
- [21] M. Tancik et al., “Fourier features let networks learn high frequency functions in low dimensional domains,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 7537–7547.
- [22] N. Max, “Optical models for direct volume rendering,” *IEEE Trans. Vis. Comput. Graphics*, vol. 1, no. 2, pp. 99–108, Jun. 1995.
- [23] M. Hwang et al., “A simple method for automatic 3D reconstruction of coronary arteries from X-ray angiography,” *Front. Physiol.*, vol. 12, 2021, Art. no. 724216.
- [24] D. Bappy, A. Hong, E. Choi, J.-O. Park, and C.-S. Kim, “Automated three-dimensional vessel reconstruction based on deep segmentation and bi-plane angiographic projections,” *Computerized Med. Imag. Graph.*, vol. 92, 2021, Art. no. 101956.
- [25] H. Zhao et al., “Self-supervised learning enables 3D digital subtraction angiography reconstruction from ultra-sparse 2D projection views: A multicenter study,” *Cell Rep. Med.*, vol. 3, no. 10, 2022, Art. no. 100775.
- [26] R. Yunus et al., “Recent trends in 3D reconstruction of general non-rigid scenes,” in *Computer Graphics Forum*. New York, NY, USA: Wiley, 2024, Art. no. e15062.
- [27] X. Wang, S. Hu, H. Fan, H. Zhu, and X. Li, “Neural radiance fields in medical imaging: Challenges and next steps,” 2024, arXiv: 2402.17797.
- [28] A. Molaei et al., “Implicit neural representation in medical imaging: A comparative survey,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 2381–2391.
- [29] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. Sajjadi, A. Geiger, and N. Radwan, “RegNeRF: Regularizing neural radiance fields for view synthesis from sparse inputs,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5480–5490.
- [30] A. Jain, M. Tancik, and P. Abbeel, “Putting NeRF on a diet: Semantically consistent few-shot view synthesis,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 5885–5894.
- [31] D. Rückert, Y. Wang, R. Li, R. Idoughi, and W. Heidrich, “NeAT: Neural adaptive tomography,” *ACM Trans. Graph.*, vol. 41, no. 4, pp. 1–13, 2022.
- [32] G. Zang, R. Idoughi, R. Li, P. Wonka, and W. Heidrich, “IntraTomo: Self-supervised learning-based tomography via sinogram synthesis and prediction,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1960–1970.
- [33] Y. Lin, Z. Luo, W. Zhao, and X. Li, “Learning deep intensity field for extremely sparse-view CBCT reconstruction,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2023, pp. 13–23.

- [34] Y. Fang et al., “SNAF: Sparse-view CBCT reconstruction with neural attenuation fields,” 2022, arXiv: 2211.17048.
- [35] J. Kshirsagar et al., “Generative ai-assisted novel view synthesis of coronary arteries for angiography,” in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, 2024, pp. 1–6.
- [36] R. Zha, Y. Zhang, and H. Li, “NAF: Neural attenuation fields for sparse-view CBCT reconstruction,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2022, pp. 442–452.
- [37] Z. Zhou et al., “Tiavox: Time-aware attenuation voxels for sparse-view 4D DSA reconstruction,” 2023, arXiv: 2309.02318.
- [38] B. Liu, F. Zhou, and X. Bai, “Improved c-arm cardiac cone beam ct based on alternate reconstruction and segmentation,” *Biomed. Signal Process. Control*, vol. 13, pp. 113–122, 2014.
- [39] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Proc. Med. Image Comput. Comput.-Assisted Interv.: 1st Int. Conf.*, Cambridge, MA, USA, 1998, pp. 130–137.
- [40] M. Li, H. Kudo, J. Hu, and R. H. Johnson, “Improved iterative algorithm for sparse object reconstruction and its performance evaluation with micro-CT data,” *IEEE Trans. Nucl. Sci.*, vol. 51, no. 3, pp. 659–666, Jun. 2004.
- [41] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, “BARF: Bundle-adjusting neural radiance fields,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 5741–5751.
- [42] K. Park et al., “Nerfies: Deformable neural radiance fields,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 5865–5874.
- [43] C. Rohkohl, G. Lauritsch, A. Keil, and J. Hornegger, “CAVAREV—an open platform for evaluating 3D and 4D cardiac vasculature reconstruction,” *Phys. Med. Biol.*, vol. 55, no. 10, 2010, Art. no. 2905.
- [44] A. Biguri, M. Dosanjh, S. Hancock, and M. Soleimani, “Tigre: A matlab-GPU toolbox for CBCT image reconstruction,” *Biomed. Phys. Eng. Exp.*, vol. 2, no. 5, 2016, Art. no. 055010.
- [45] N. E. Green, S.-Y. J. Chen, A. R. Hansgen, J. C. Messenger, B. M. Groves, and J. D. Carroll, “Angiographic views used for percutaneous coronary interventions: A three-dimensional analysis of physician-determined vs. computer-generated views,” *Catheterization Cardiovasc. Interv.*, vol. 64, no. 4, pp. 451–459, 2005.
- [46] P. Green, P. Frobisher, and S. Ramcharitar, “Optimal angiographic views for invasive coronary angiography: A guide for trainees,” *Brit. J. Cardiol.*, vol. 23, pp. 110–113, 2016.
- [47] C. D. Mario and N. Sutaria, “Coronary angiography in the angioplasty ERA: Projections with a meaning,” *Heart*, vol. 91, no. 7, pp. 968–976, 2005.
- [48] S. Kastryulin, J. Zakirov, N. Pezzotti, and D. V. Dylov, “Image quality assessment for magnetic resonance imaging,” *IEEE Access*, vol. 11, pp. 14154–14168, 2023.
- [49] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, “Blood vessel segmentation algorithms—Review of methods, datasets and evaluation metrics,” *Comput. Methods Programs Biomed.*, vol. 158, pp. 71–91, 2018.
- [50] N. Baka, B. Lelieveldt, C. Schultz, W. Niessen, and T. van Walsum, “Respiratory motion estimation in X-ray angiography for improved guidance during coronary interventions,” *Phys. Med. Biol.*, vol. 60, no. 9, 2015, Art. no. 3617.
- [51] Y. Cai et al., “Radiative Gaussian splatting for efficient X-ray novel view synthesis,” 2024, arXiv: 2403.04116.