Novel View Synthesis of Coronary Arteries from Rotational X-Ray Angiography

by

JAY KSHIRSAGAR

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Applied Science in Electrical & Computer Engineering

School of Electrical Engineering & Computer Science
Faculty of Engineering
University of Ottawa

Supervisor
SHERVIN SHIRMOHAMMADI

Abstract

Coronary artery disease (CAD) is commonly diagnosed and treated using X-ray coronary angiography, which requires multiple angiograms and the administration of contrast dye. While effective, the overuse of contrast dye can lead to adverse effects in patients, particularly those with kidney issues. This thesis explores a novel approach aimed at minimizing the need for multiple angiograms and contrast dye by leveraging neural radiance field (NeRF) algorithms to generate novel views of coronary arteries. By training these algorithms on rotational angiography data, we seek to develop a generalizable model capable of producing detailed coronary artery images without requiring patient-specific retraining, potentially reducing contrast dye usage.

The study systematically evaluates different NeRF architectures. MedNeRF, one of the initial models tested, demonstrated its ability to generate realistic coronary artery images. The right coronary artery (RCA) was easier to model due to its well-defined structure, while the left coronary artery (LCA), which had a larger dataset, was the primary focus of the experiments. Building on this, StyleNeRF, which overcame the limitations of MedNeRF by producing highly detailed and accurate coronary artery images with a Fréchet Inception Distance (FID) score of 75.32, the best among the models tested. StyleNeRF's hierarchical latent space enabled efficient generation of high-resolution novel views, and a web-based tool was developed to allow clinicians to interpolate target views at different angles.

Lastly, dynamic NeRF architectures were explored to account for heart motion and the flow of contrast dye, addressing the dynamic nature of coronary artery imaging. However, the results from dynamic NeRF models, including attempts to merge RoDyNeRF with generative frameworks, were inconclusive, indicating the need for further development in this area.

This research demonstrates the potential of NeRF-based novel view synthesis as a promising technique to reduce the reliance on contrast dye in coronary artery imaging. While the results are promising, challenges remain, particularly in handling dynamic scenes and real-time applications. Future work will focus on advancing generative models, such as latent diffusion and Gaussian splatting, to improve the model’s real-time capabilities and its applicability in clinical settings.

Acknowledgments

Completing this thesis has been a challenging yet deeply fulfilling journey, and it would not have been possible without the support, guidance, and encouragement of many remarkable individuals. First and foremost, I would like to express my heartfelt gratitude to Dr. Shervin Shirmohammadi. His intellectual insight, patient mentorship, and constant encouragement have been invaluable in shaping this research. His passion for knowledge and pursuit of excellence continually inspired me to push my boundaries and strive for improvement.

I also owe sincere thanks to Dr. Bahareh Taji and my colleagues at the University of Ottawa Heart Institute. Their invaluable feedback, expert guidance, and thought-provoking discussions were instrumental in refining the direction of this work and overcoming the challenges along the way. I am equally thankful to clinicians Dr. Pascal, Dr. Derek So, and Dr. A.Y. Chong, whose domain expertise greatly enriched this research. The collaboration with Mr. John McNulty and Mr. AJ Wisniewski from MITRE Corporation was also crucial, and I appreciate their teamwork and dedication.

To my wife, Kiran, thank you for your companionship and for lifting my spirits during tough times. Your constant positivity and guidance made this experience brighter.

Lastly, and most importantly, my deepest appreciation goes to my family. Your unwavering love, patience, and encouragement have been my constant source of strength. A special thank you to Khushi—this thesis is as much a reflection of your support as it is of my efforts.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Coronary Angiography

Coronary artery disease (CAD) is the leading cause of death worldwide, responsible for a significant burden on global health systems [1]. The diagnosis and treatment of CAD often involve Invasive Coronary Angiography (ICA), a procedure that utilizes X-ray imaging to visualize the coronary arteries [2] as shown in Figure **1-1**. ICA is particularly valuable for its ability to guide Percutaneous Coronary Interventions (PCI), a procedure that employs balloons and stents to open blocked coronary arteries. PCI is commonly used to treat both acute coronary syndrome and stable CAD, making ICA an essential tool in the management of this disease.

Figure 1-1 Coronary Angiogram showing Left Coronary Artery

However, ICA is not without its risks. The use of iodinated contrast agents in ICA has been associated with nephrotoxicity, leading to a condition known as contrast-induced acute kidney injury (CI-AKI). CI-AKI is a serious complication that occurs in approximately 15% of patients undergoing coronary angiography, particularly those with pre-existing kidney conditions [3]. The risk factors for CI-AKI include chronic kidney disease, diabetes,

and heart failure, all of which are common in patients with CAD. CI-AKI is the third leading cause of hospital-acquired acute kidney injury and is associated with significant morbidity and mortality [4].

The limitations of Invasive Coronary Angiography (ICA) extend to its inherent two-dimensional (2D) imaging capability, which reduces the complex three-dimensional (3D) structure of the coronary tree to a simplified 2D representation. This reduction can lead to significant issues, such as unrecognized stenoses, underestimation of stenosis severity, and vessel foreshortening. To mitigate these challenges, acquiring images from multiple angles is crucial for a thorough assessment of coronary anatomy, although this approach requires additional contrast and increases radiation exposure, further exacerbating the associated risks.

One advancement aimed at addressing these limitations is Rotational Angiography (RA), also known as rotational X-ray angiography or swing angiography. RA is a powerful imaging technique that generates a series of 2D X-ray images as the X-ray source and detector rotate around the patient, providing a more detailed and comprehensive visualization of the coronary arteries. This method facilitates an in-depth analysis of coronary anatomy and function, improving the detection of pathological changes such as stenoses or occlusions, thereby enhancing diagnostic accuracy and therapeutic outcomes. Despite being available in modern interventional facilities, RA's widespread use is hindered by the complex nature of its acquisition process and the unconventional look of its images. Additionally, the reduction in contrast usage with RA is relatively modest (76 cc vs. 60 cc) [5]. This underscores the ongoing need for technical advancements to minimize the risks associated with ICA.

X-ray coronary angiography, in particular, is crucial for visualizing coronary artery anatomy, identifying stenoses, and assessing, to some extent, the flow of blood through the coronary vessels [1]. By using radiopaque contrast agents, this imaging technique provides essential anatomical information that is critical for making clinical decisions and planning interventions [6].

Rotational Angiography (RA) is a key imaging technique employed in the cardiac catheterization laboratory to visualize coronary arteries from multiple angles (Refer to Figure **1-1**). Unlike conventional coronary angiography (CCA), which provides only two-dimensional representations, RA allows for three-dimensional reconstruction, enabling a more comprehensive assessment of coronary anatomy. This is especially useful in detecting stenoses and other coronary pathologies that may be missed in traditional 2D imaging. The method involves the continuous rotation of the C-arm X-ray system, capturing projections from different angles and reconstructing them into detailed 3D views of the coronary arteries. RA is increasingly favored for its ability to reduce the overlap of vessel structures and foreshortening, improving diagnostic accuracy without significantly increasing radiation or contrast media exposure.

A detailed discussion of the evolution of angiography systems, including conventional coronary angiography, computed tomography coronary angiography, magnetic resonance coronary angiography, and dual-axis rotational angiography (DARCA), will be presented in Chapter 2.

1.2 Use cases of Novel view synthesis of Coronary Arteries

The standard coronary angiogram is a widely used diagnostic tool in interventional cardiology, offering direct visualization of coronary arteries to assess the severity of coronary artery disease. While it has become the gold standard for evaluating coronary lesions, the technique's inherent limitation is its reliance on a limited number of fixed 2D projections. These views, often captured from only a few predetermined angles, can sometimes obscure critical details of the coronary anatomy, especially in cases of complex or overlapping vessels [7]. Moreover, the need to obtain additional angiograms to visualize the coronary arteries from different angles can increase the cumulative radiation dose and contrast medium exposure, leading to heightened risks of radiation-induced injuries and contrast-induced nephropathy, which are significant concerns in clinical practice [8].

Rotational coronary angiography has emerged as a technique to address these limitations by capturing a continuous 3D rotational sweep of the coronary arteries. This method provides a more comprehensive assessment by enabling the acquisition of multiple

projections with a single contrast injection, thus reducing the overall contrast volume and radiation dose required compared to traditional angiography [9]. Despite these advantages, rotational angiography is not without its drawbacks. The technique requires specialized equipment and software, which may not be readily available in all clinical settings. Furthermore, the procedure can be more time-consuming, and the need for patients to hold their breath for extended periods during the image acquisition can lead to discomfort and motion artifacts, potentially compromising image quality [10].

In light of these challenges, there is a growing need for the development of novel view synthesis techniques that can generate 3D reconstructions or alternative 2D views from standard angiograms using advanced computational algorithms. These techniques, often based on machine learning and artificial intelligence, can simulate different viewing angles from a single angiographic dataset, thus providing clinicians with enhanced diagnostic information without the need for additional invasive imaging.

The clinical benefits of novel view synthesis are manifold. In situations such as bifurcation lesions or tortuous arteries, when conventional angiographic perspectives may not be sufficient, synthetic views can provide improved visualization of the morphology of the lesion and assist in more precise placement of the stent [11]. Additionally, these techniques can facilitate more precise pre-procedural planning for complex interventions, such as chronic total occlusion (CTO) revascularization, where understanding the 3D spatial relationships of the coronary anatomy is critical. Furthermore, by reducing the need for multiple invasive imaging procedures, novel view synthesis could lead to cost savings and improved patient outcomes, making it a valuable addition to the interventional cardiologists toolkit [12].

1.3 Contributions

This thesis makes several contributions to the field of coronary artery imaging by leveraging neural radiance field (NeRF) algorithms for novel view synthesis:

- Development of a Generalizable NeRF Model for coronary artery images: The research presents a novel approach to create a generalizable NeRF-based model that can produce high-quality coronary artery images without the need for retraining

for each patient. This offers a significant advantage for clinical settings by reducing the time and computational resources required for personalized imaging.

- High-Resolution Novel View Synthesis: The thesis successfully applied NeRF architectures to produce highly detailed and accurate reconstructions of coronary arteries. This work not only improved the resolution and fidelity of the synthesized images but also introduced a practical solution for generating novel views that can assist clinicians in capturing specific regions of interest efficiently.
- Integration of Real-Time Interpolation Tool: A web-based tool was developed as part of this research, enabling real-time interpolation of coronary artery views. This tool allows clinicians to explore different angles interactively, enhancing the practical usability of NeRF-based models in medical imaging.
- Foundational Work on Dynamic NeRFs: This thesis also laid the groundwork for applying, for the first time, dynamic NeRFs to coronary artery imaging by addressing the challenges posed by heart motion and contrast dye flow. Although this remains an area for future improvement, the research has made important strides toward handling dynamic imaging conditions.

In summary, this thesis provides advancements in coronary artery imaging through the use of novel view synthesis techniques, demonstrating the potential for reducing contrast dye usage and improving imaging efficiency. Work done in the thesis was published in the following paper. It should be noted that this thesis reuses material from the paper with permission.

1. J. Kshirsagar et al., "Generative AI-Assisted Novel View Synthesis of Coronary Arteries for Angiography," 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Eindhoven, Netherlands, 2024, pp. 1-6

1.4 Organization of Thesis

In the next chapter, we review the theoretical background relevant to the topic of this research. This includes an overview of coronary artery disease (CAD) imaging techniques, highlighting both traditional methods like X-ray angiography and more advanced

approaches such as 3D reconstruction and neural networks. We examine the limitations of existing techniques, including the risks associated with contrast dye, and introduce the concept of neural radiance fields (NeRF) as a potential solution for novel view synthesis in coronary artery imaging.

In Chapter 3, we describe the system design and dataset used for our experiments. This includes the details of the rotational angiography data, pre-processing steps, and the setup for training NeRF-based models. We also discuss how the dataset was structured to allow for efficient training and inference, ensuring that the generated views maintain clinical relevance.

Chapter 4 presents our first experimental results, focusing on the application of MedNeRF for novel view synthesis of coronary arteries. This chapter demonstrates how MedNeRF was able to generate detailed images of the coronary arteries and discusses the challenges encountered during the experiments.

In Chapter 5, we introduce StyleNeRF, a more advanced model that produced superior results in terms of image resolution and accuracy. This chapter highlights the encoder-based inference process and discusses the development of a web-based tool that allows for real-time interpolation of coronary artery views from various angles, providing a practical solution for clinical settings.

Chapter 6 explores dynamic NeRF, investigating its potential for handling heart motion and the flow of contrast dye in coronary artery imaging. This chapter outlines the limitations of dynamic NeRF in real-time applications and discusses possible directions for future improvements.

Finally, in Chapter 7, we conclude the thesis with a summary of the key findings and contributions. We also provide insights into future research directions, including the exploration of more advanced generative models such as latent diffusion and Gaussian splatting, which could further enhance the quality and real-time performance of coronary artery imaging in clinical practice.

Chapter 2

Background & Related Work

2.1 Types of Coronary Angiography Systems

The cardiac catheterization laboratory has experienced technological advancements in C-arm-based angiography systems, which have enabled the development of diverse modalities of X-ray coronary angiography. These include conventional coronary angiography, computed tomography coronary angiography, magnetic resonance coronary angiography, and rotational, and dual-axis coronary angiography techniques. These advancements have expanded the diagnostic capabilities and imaging options available for evaluating coronary artery anatomy and pathologies.

1. Conventional Coronary Angiography:

   Conventional Coronary Angiography (CCA) is the standard technique for visualizing coronary arteries. It involves the injection of iodinated contrast media into the coronary arteries, followed by X-ray imaging to capture 2D images of the coronary lumen. While CCA provides high-resolution images that are crucial for detecting stenoses, its limitations include the reduction of complex three-dimensional coronary structures to two-dimensional representations, which can result in diagnostic inaccuracies such as unrecognized stenoses and vessel foreshortening [13], [14]. Furthermore, the reliance on ionizing radiation and nephrotoxic contrast agents poses significant risks to patients, particularly those with pre-existing kidney conditions.

2. Computed Tomography Coronary Angiography:

   Computed Tomography Coronary Angiography (CTCA) is a non-invasive imaging modality that utilizes multi-detector computed tomography (MDCT) to visualize the coronary arteries. Unlike CCA and RA, which require catheterization, CTCA can be performed without the need for invasive procedures. CTCA offers several

advantages, including the ability to visualize both the coronary lumen and vessel walls, thereby detecting not only stenoses but also non-calcified plaques and other features of coronary artery disease. However, its limitations include lower spatial resolution compared to CCA, the need for careful patient selection to avoid radiation exposure in low-risk populations, and the potential for overestimating the severity of stenosis due to artifacts [15].

3. Magnetic Resonance Coronary Angiography:

   Magnetic Resonance Coronary Angiography (MRCA) is another non-invasive imaging technique that uses magnetic resonance imaging (MRI) to visualize the coronary arteries. MRCA does not involve ionizing radiation, making it a safer alternative for patients, particularly those at risk of radiation-induced complications. It also does not require iodinated contrast agents, which reduces the risk of nephrotoxicity [16]. However, MRCA is less commonly used in clinical practice due to its lower spatial resolution compared to CCA and CTCA, as well as technical challenges such as motion artifacts caused by the beating heart and breathing.

4. Rotational Angiography:

   Rotational Angiography (RA) is an advanced imaging technique designed to overcome some of the limitations of CCA. RA involves the continuous rotation of the X-ray source and detector around the patient as illustrated in Figure **2-1**, while acquiring images, resulting in a series of 2D projections from different angles. These images can then be reconstructed into 3D representations of the coronary arteries, providing a more comprehensive assessment of coronary anatomy. RA improves the detection of stenoses and other pathologies by offering multiple views of the coronary tree without significantly increasing the dose of contrast media or radiation exposure.

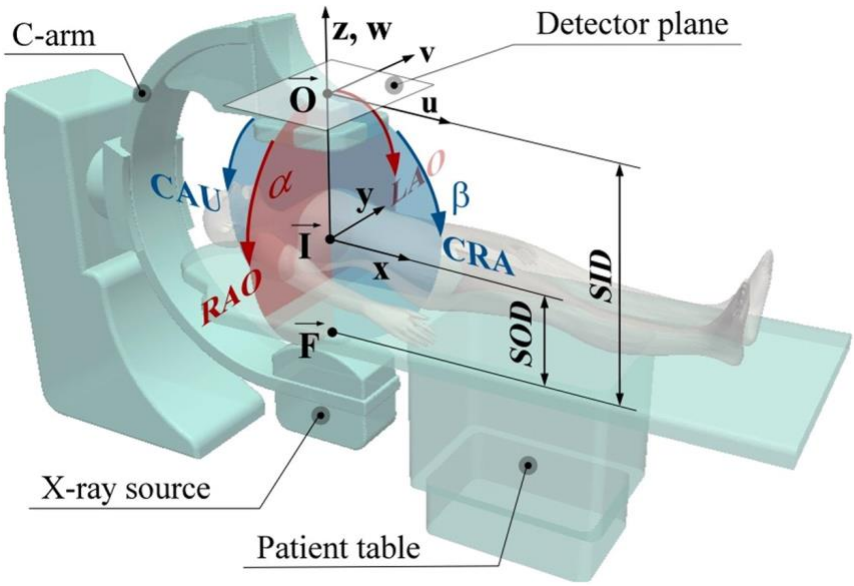


Figure 2-1 Rotational Angiography Setup [17] under CC BY 4.0

In traditional rotational X-ray angiography, the gantry typically rotates from approximately 55 to 60 degrees in the right anterior oblique (RAO) position to a similar angle in the left anterior oblique (LAO) position, often incorporating some cranial (CRA) or caudal (CAU) angulation. Extended rotational X-ray angiography introduces an innovative protocol, expanding the arc from 120 degrees LAO to 60 degrees RAO without adding additional angulation [9]. This method supports tomographic reconstruction algorithms, effectively reconstructing high-contrast coronary arteries filled with contrast. However, it is distinct from cardiac C-arm CBCT, which also offers tomographic reconstruction. Extended rotational X-ray angiography operates faster and requires fewer images than cardiac C-arm CBCT because it focuses on reconstructing high-contrast objects [18]. Despite this, extended rotational X-ray angiography has been incorporated into advanced C-arm CBCT systems. Nonetheless, this technique faces challenges due to its longer acquisition time, such as extended contrast injection [19] and motion artifacts caused by breathing. Dual-axis rotational coronary angiography (DARCA) represents an enhanced version of rotational X-ray angiography, improving patient safety and simplifying the image acquisition process. DARCA integrates CRA and CAU angulation into a single acquisition run [20]. Additionally, DARCA employs optimized C-arm rotation trajectories to minimize vessel overlap and foreshortening, aligning with expert-recommended views [21].

2.2 Traditional 3D reconstruction

3D reconstruction involves creating and visualizing detailed three-dimensional (3D) surfaces of specific anatomical features, including arteries, vessels, organs, and various body parts, as well as atypical formations, such as tumors, lesions, injuries, scars, and cysts. These reconstructions are critical in various medical fields, enabling clinicians to better understand complex structures within the human body and aiding in both diagnosis and treatment planning. For instance, 3D reconstructions are essential in neurosurgery, where precise mapping of brain tumors is necessary to minimize damage to surrounding tissues during surgery [22].

X-rays, CT scans, and MRIs play vital roles in medical imaging by providing distinct perspectives of the human body. X-rays, for example, are particularly effective for visualizing bones and detecting fractures, but their 2D nature limits their utility in understanding complex 3D structures [23]. CT scans, which provide cross-sectional images, are widely used to create detailed 3D models of internal organs, aiding in the diagnosis of conditions such as tumors and vascular diseases [24]. MRI, known for its superior contrast in soft tissues, is invaluable in reconstructing detailed images of the brain, muscles, and joints, making it crucial for diagnosing neurological conditions [25]. These imaging modalities not only assist in diagnosing illnesses but also play a significant role in planning treatments and overseeing different health conditions.

Advancements in medical imaging have greatly facilitated the development of 3D reconstruction techniques. Traditionally, the reconstruction process involves three main steps: segmentation, registration, and surface reconstruction. Segmentation is identifying and isolating relevant anatomical structures from imaging data. This step is often performed manually, requiring radiologists to outline the structures of interest, such as tumors or blood vessels, which is a time-consuming and labor-intensive task. Registration involves aligning the segmented images from different angles or different imaging modalities to ensure that they correspond correctly in the 3D space, a process that is critical for accurate reconstruction [26]. Surface reconstruction converts the registered and segmented data into a continuous, smooth 3D surface that represents the anatomical

structure. This step involves complex mathematical algorithms to ensure the accuracy and realism of the reconstructed model [27].

However, traditional reconstructions often require more time and may not perform as well as supervised Deep Learning (DL) methods. For instance, manual segmentation is prone to inter-operator variability, meaning different radiologists might segment the same image differently, leading to inconsistencies in the final 3D model. This limitation has led to the exploration of DL-based architectures, which have shown promise in automating the segmentation process. These architectures, such as convolutional neural networks (CNNs), are trained on large datasets to recognize patterns and features within medical images, enabling them to perform segmentation and reconstruction tasks with greater speed and accuracy than traditional methods [28]. However, DL-based methods require a large number of annotated samples to effectively learn from data and generalize it. Unfortunately, such large annotated datasets are often not available in medical imaging due to the time-consuming and expensive process of data annotation, as well as privacy concerns [29].

2.3 Novel View Synthesis

It's essential to distinguish between 3D reconstruction and novel view synthesis, two techniques often discussed in the context of NeRF. 3D reconstruction involves creating a complete 3D model of an object or scene from multiple images or scans, as discussed in the previous section. On the other hand, novel view synthesis focuses on generating new perspectives or views of a scene that were not part of the original input data. While 3D reconstruction seeks to build a detailed and accurate representation of the entire scene, novel view synthesis aims to simulate what the scene would look like from different angles, relying heavily on the underlying model's ability to interpolate and extrapolate based on existing data [30].

A key approach that has gained prominence in generating synthetic data for novel view synthesis is the use of Generative Adversarial Networks (GANs) [31]. GANs consist of two neural networks, a generator and a discriminator, which are trained simultaneously. The generator aims to create realistic data (such as images or views) from random input,

while the discriminator attempts to differentiate between real and generated data. This adversarial training process enables GANs to learn complex data distributions, making them particularly powerful for generating realistic images or scenes from various viewpoints. GANs have been widely used in data augmentation, allowing models to be trained on synthetic data when real-world data is scarce or difficult to obtain, especially in fields like medical imaging.

Wasserstein Generative Adversarial Networks (WGANs) [32], a variant of GANs, have demonstrated impressive potential in producing realistic images, addressing common issues with standard GANs, such as instability and mode collapse, which occur when the generator produces limited variations of images. WGANs have been particularly successful in generating synthetic medical data, including multi-sequence brain MRIs, and have shown promise in improving model performance by augmenting training datasets. This capability makes GANs an essential tool for novel view synthesis, where generating plausible new views or perspectives from learned data is crucial.

Neural Radiance Fields (NeRFs) [30], a more advanced technique, excel in novel view synthesis by generating high-quality images from novel viewpoints. NeRF leverages volumetric scene representations and optimizes for view consistency, allowing it to synthesize new perspectives from a sparse set of input views. While GANs generate new data directly from learned distributions, NeRFs focus on accurately rendering novel views by optimizing for radiance values and depth from multiple angles. Although NeRFs have been more commonly applied in computer vision tasks, their potential for 3D reconstruction in medical imaging, particularly when adapted to specific requirements, is promising.

In medical imaging, synthetic data plays a critical role in overcoming the limitations of traditional 3D reconstruction by providing diverse, labeled datasets for training AI models. In a study comparing the training of vessel segmentation models on real and synthetic images, models trained on real images achieved an average Area Under the Curve (AUC) of 0.887, while those trained on synthetic images showed a comparable average AUC of 0.841. Notably, models trained with synthetic images demonstrated greater sensitivity

when allowing a false positive rate greater than 0.35, highlighting synthetic data's potential to enhance model performance, especially in scenarios where real data may be limited [33].

Moreover, the authors in [34] propose a Variational Autoencoder (VAE)-based architecture that can generate annotated cardiac MRIs, showing that including synthetic images during model training improves the Dice score on segmentation tasks by 12%. This highlights the synergy between synthetic data generation techniques such as GANs and VAEs and advanced models like NeRFs, as both approaches contribute to generating accurate novel views and enhancing medical imaging tasks.

2.4 Neural Radiance Fields (NeRF)

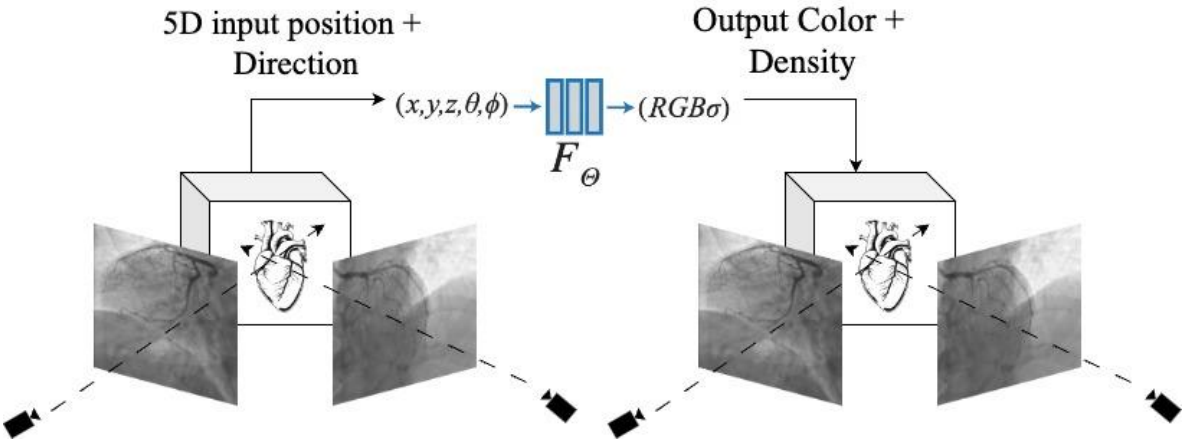


Figure 2-2 Neural Radiance Field Architecture

Figure **2-2** illustrates the architecture of Neural Radiance Fields (NeRF), which is an approach to representing a 3D scene using a 5D function. This function maps a 3D spatial location and a 2D viewing direction to corresponding RGB color values and volume density. The neural network is trained on a sparse set of images of a scene with known camera positions, allowing it to predict the emitted or reflected light at any point when viewed from a specific angle and the density of that point. Following are the key NeRF concepts:

- Scene Representation

  NeRF represents a 3D scene using a continuous function F_Θ that maps spatial location and viewing direction to color and volume density. Formally, the function is expressed as:

$$\boldsymbol{F_{\Theta}}: (\boldsymbol{x}, \boldsymbol{d}) \rightarrow \boldsymbol{z} \qquad (1)$$

Here, Θ denotes the neural network parameters, $x = (x, y, z) \in R^3$ represents the spatial location, and $d = (\theta, \phi) \in S^2$ corresponds to the viewing direction on the unit sphere. The output is a 4D vector consisting of the RGB color $c = (r, g, b)$ and the volume density σ. This can be mathematically defined as:

$$F_{\Theta}: R^3 \times S^2 \rightarrow R^4 \qquad (2)$$

The neural network typically consists of 8–10 fully connected layers with ReLU activation functions. Initially, the spatial coordinates are processed to generate a feature vector, which is then concatenated with the viewing direction to compute the final color and density.

- Volumetric Rendering

  As depicted in Figure **2-2**, the network samples multiple points along each ray that passes through a pixel into the scene. For each sample, the network outputs the corresponding color and density values. The final color of the pixel is computed through volumetric rendering, where the integration of colors and densities along the ray simulates how light accumulates as it travels through the scene, considering occlusion and light absorption.

  For a given ray $r(t) = o + td$,where o is the ray origin and d its direction, the expected color $C(r)$ is computed as:

$$C(r) = \int_0^{\infty} T(t)\sigma\big(r(t)\big)c(r(t), d)\ dt \qquad (3)$$

Here,$T(t)$ represents the accumulated transmittance along the ray up to distance t , and is defined as:

$$T(t) = \exp\left(-\int_0^t \sigma\big(r(s)\big)\ ds\right) \qquad (4)$$

This integral is approximated using numerical quadrature, with discrete samples taken along the ray to compute the color.

- Positional Encoding

  To capture fine details, both 3D coordinates and viewing directions undergo a positional encoding before being input into the network. This transformation maps the inputs into a higher-dimensional space, enabling the network to better model high-frequency details within the scene. NeRF uses positional encoding γ. For a scalar input x , the encoding is defined as:

$$\gamma(x) = (\sin(2^0\pi x), \cos(2^0\pi x), \dots, \sin(2^{L-1}\pi x), \cos(2^{L-1}\pi x)) \quad (5)$$

  The encoding is applied elementwise to both 3D coordinates and viewing directions, resulting in a higher-dimensional input vector for the network.

- Camera Parameters

  Accurate camera parameters are essential for synthesizing novel views of a 3D scene in NeRF. These parameters define the camera's position, orientation, and intrinsic properties during training. They ensure the correct relationship between the 2D images and the 3D scene, which is crucial for view synthesis.

  Extrinsic Parameters: These describe the camera's position and orientation in 3D space. The extrinsic parameters include a rotation matrix R and a translation vector t . The rotation matrix ($R \in R^{3\times3}$) specifies the orientation of the camera with respect to the world coordinate system, while the translation vector ($t \in R^3$) represents the camera's position.

  Mathematically, the relationship between a 3D point X_{world} in the world coordinate system and its corresponding point X_{camera} in the camera coordinate system is:

$$X_{\text{camera}} = RX_{\text{world}} + t \quad (6)$$

Intrinsic Parameters: These describe the internal characteristics of the camera, such as the focal length, principal point, and skew coefficient. The intrinsic parameters are encapsulated in the camera matrix ($K \in R^{3\times3}$), which transforms 3D points from the camera coordinate system into 2D image points. The intrinsic matrix K is defined as:

$$K = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \qquad (7)$$

where (f_x) and (f_y) represent the focal lengths along the x and y axes, (c_x, c_y) are the coordinates of the principal point (optical center), and s is the skew coefficient for non-rectangular pixels.

Accurate camera parameters are critical because they determine how rays are cast from the camera through each pixel into the 3D scene. In Section 3.2 we talk about extracting these camera parapeters from our data. NeRF synthesizes new views by integrating the color and density of points sampled along these rays. Incorrect camera parameters result in distorted or misaligned synthesized views, leading to poor reconstruction quality.

Camera parameters can be extracted through Structure-from-Motion (SfM) techniques, which use multiple 2D images to recover 3D structure and camera parameters. SfM identifies key points across images, matches them, and optimizes both the camera parameters and 3D points to minimize reprojection errors. Errors in camera parameters can introduce artifacts like blurriness, ghosting, or incorrect depth, reducing the fidelity of NeRF's output.

2.4.1 Generative Neural Radiance Fields

GRAF (Generative Radiance Fields for 3D-Aware Image Synthesis) [35] utilizes radiance fields to generate images that remain consistent when viewed from different angles, unlike older methods that use voxel grids, which can be limited in detail. Trained using 2D images without specific camera positions, GRAF employs an adversarial learning framework to produce high-quality images. Although GRAF currently excels at handling simple scenes

with single objects, it demonstrates that radiance fields are a powerful tool for generating 3D-aware images. pixelNeRF [36] takes a different approach by conditioning NeRFs on one or a few input images. This allows it to learn to generate a 3D scene from just a small number of views, a significant improvement over traditional NeRF methods that require many views and substantial computational resources. Trained on multiple scenes, pixelNeRF can generate new views of an object or scene not present in the original images, making it highly versatile and useful for tasks like 3D reconstruction from a single image. π-GAN (Periodic Implicit Generative Adversarial Networks) [37] further advances 3D-aware image synthesis by utilizing a more advanced network architecture that creates consistent 3D representations with fine details. π-GAN allows for precise control over camera viewpoints and can generate images at different scales, even from angles unseen during training. This capability makes it a robust tool for creating high-quality 3D images with detailed textures and structures. GIRAFFE (Generative Image Representation with Compositional Neural Feature Fields) [38] introduces a compositional approach to 3D-aware image synthesis. Unlike methods that model scenes as a whole, GIRAFFE decomposes scenes into objects and background, representing each component with its own neural feature field. This allows for independent control over the shape, appearance, and position of each object within a scene. By combining NeRF with a GAN framework, GIRAFFE enables efficient rendering and manipulation of complex scenes, providing greater flexibility and control in generating 3D-aware images. StyleNeRF [41] represents a significant advancement by integrating the style-based generator architecture of StyleGAN2 [39] with NeRF for high-resolution, 3D-aware image synthesis. By leveraging the strengths of StyleGAN2 in generating photorealistic images and the 3D representation capabilities of NeRF, StyleNeRF produces images with high fidelity and consistent multi-view appearances. It enables precise control over style and geometry, allowing manipulation of generated images in terms of both content and style attributes. Furthermore, StyleNeRF demonstrates strong performance in generating complex scenes and objects with fine details while being efficient in training and rendering. This makes it a valuable tool for applications requiring high-quality 3D-aware image synthesis, and it is utilized in our experiments elaborated in Chapter 4.

These advancements in NeRF technology mark significant steps forward in creating more realistic and consistent 3D images, offering better control and quality even with limited input data.

2.4.2 Dynamic Neural Radiance Fields

While generative neural radiance fields (NeRFs) have significantly advanced 3D-aware image synthesis by allowing the creation of high-quality, view-consistent images from static scenes, they are inherently limited to static environments where objects do not change over time. This limitation poses a significant challenge in real-world applications where scenes are often dynamic, involving moving objects or changing perspectives. To address these shortcomings, researchers have developed dynamic NeRF models that extend the capabilities of traditional generative NeRFs, enabling the modeling and reconstruction of scenes that evolve over time.

Dynamic NeRFs are crucial for capturing the complexities of the real world, where both rigid and non-rigid motions are common. For instance, traditional NeRFs can model a static object from multiple angles, but they struggle to handle scenarios where the object or the environment changes—such as a person moving their head, objects shifting in a scene, or even entire environments transforming over time. Dynamic NeRFs fill this gap by introducing mechanisms to account for these changes, allowing for more accurate and realistic representations of dynamic scenes. These models are particularly important for applications in virtual reality, augmented reality, and film production, where the ability to render and interact with dynamic scenes is essential.

Here we explore various dynamic NeRF approaches that have been developed to overcome the limitations of static NeRFs, each offering unique methods for handling the challenges associated with modeling time-varying scenes.

Building on the concept of modeling dynamic scenes, D-NeRF (Neural Radiance Fields for Dynamic Scenes)[40] introduces time as an additional input to the NeRF model, enabling it to handle scenes where objects are in motion. D-NeRF splits the learning process into two stages: one that encodes the scene into a canonical configuration and another that maps this canonical space to the scene's deformed state at a specific time. This

dual-stage approach allows D-NeRF to generate new views of the scene while controlling both the camera position and the time variable, effectively capturing the dynamic nature of the scene. The method excels in synthesizing high-quality views of scenes with various types of deformations, from rigid objects to complex human movements, making it a robust tool for dynamic scene reconstruction.

RoDynRF (Robust Dynamic Radiance Fields) [41] further enhance the ability to reconstruct dynamic scenes by addressing common issues related to camera pose estimation. Traditional methods rely on Structure from Motion (SfM) algorithms to estimate camera poses, but these can fail in dynamic scenes with challenging conditions, such as poorly textured surfaces or fast-moving objects. This new approach jointly estimates both static and dynamic radiance fields along with the camera parameters, improving robustness against the shortcomings of SfM. The model is capable of reconstructing dynamic radiance fields from a wide range of challenging videos without needing pre-determined camera poses, making it highly adaptable to real-world scenarios. Extensive experiments have shown that this method outperforms previous state-of-the-art techniques, particularly in scenarios where camera motion and object dynamics are complex.

2.4.3 Neural Radiance Fields for Medical Imaging

Recent studies have explored the application of NeRF in the medical imaging field, demonstrating its potential to synthesize novel views from X-rays, CT scans, and MRIs, which are crucial for non-invasive diagnostics and treatment planning. In [42], the authors introduced MedNeRF, a model that adapts the generative NeRF approach to render CT projections from limited X-ray views, even in single-view scenarios. MedNeRF synthesizes realistic images and models the variation in anatomical structures with viewpoint changes without requiring 3D supervision. This capability is enabled by a novel discriminator architecture, which enhances the model's performance on CT scans by providing a more detailed feedback signal, ensuring that the generated images maintain anatomical accuracy.

Another significant contribution is Neural Adaptive Tomography (NeAT) [43], which is designed for tomographic settings where scenes are semi-transparent and volumetric rather than opaque. NeAT is the first adaptive, hierarchical neural rendering pipeline tailored for tomography, addressing the unique challenges posed by the semi-transparency of volumetric scenes. This method improves the accuracy and quality of tomographic reconstructions, especially when dealing with limited input data, a common scenario in medical imaging.

In contrast, [44] presents Spars-view CBCT Reconstruction with Neural Attenuation Fields (SNAF), a model that focuses on creating superior reconstructions using only 20 input views of Cone-Beam Computed Tomography (CBCT). SNAF's approach is particularly relevant in clinical settings where it is often challenging to acquire many views due to time, cost, or patient safety concerns. By utilizing a minimal number of input views, SNAF demonstrates the ability to produce high-quality reconstructions, making it a promising tool for improving the accessibility and effectiveness of CBCT in medical diagnostics.

A recent study [45] has demonstrated the potential of using Neural Radiance Fields (NeRF) for 3D X-ray angiography reconstruction, achieving promising results with sparse and limited-angle projections. NeRF's ability to capture detailed vascular structures under challenging conditions highlights its effectiveness compared to traditional methods, which often struggle with issues like vessel sparsity, overlap, and visibility. However, there is still a significant gap in research focused on generating novel views during dynamic cardiac phases, such as systole and diastole, which are critical for assessing heart function and diagnosing conditions like coronary artery disease (CAD).

While NeRF shows promise in addressing challenges like sparse data and projection constraints in coronary angiography, further research is needed to overcome limitations such as background overlap. Particularly, applying NeRF to dynamic imaging during the cardiac cycle could significantly enhance the accuracy and utility of 3D models, offering better insights into heart function and potentially improving clinical outcomes. This area holds considerable potential for advancing cardiovascular imaging and diagnostics.

Chapter 3

System Design

3.1 Data Preparation

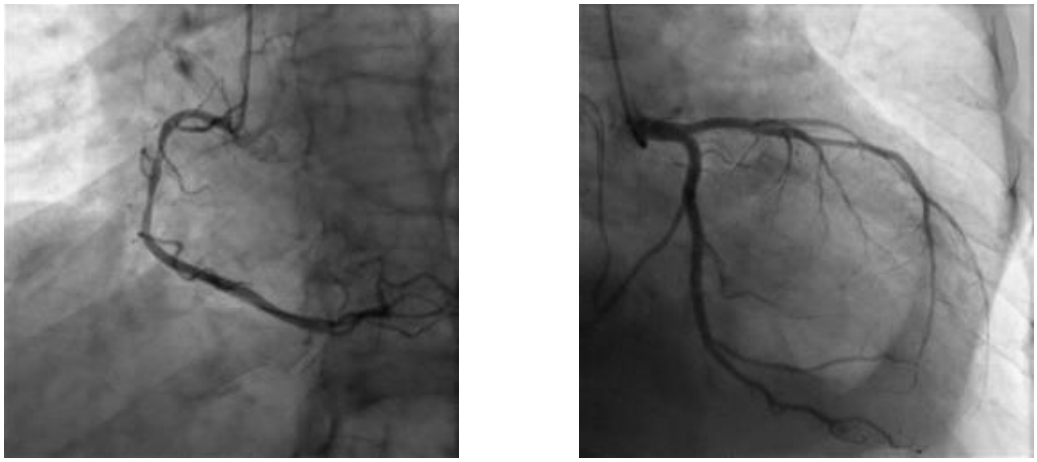
The dataset was derived from a comprehensive collection of angiogram scans obtained through rotational angiography, focusing on the left coronary artery (LCA) and right coronary artery (RCA). The data collection process was performed under ethical approval number 20230622-01H, ensuring compliance with all relevant privacy and ethical guidelines.

The raw data was sourced from the Picture Archiving and Communication System (PACS) at UOHI, where Digital Imaging and Communication in Medicine (DICOM) files were retrieved. To protect patient confidentiality, all identifying information was removed, and a generic patient ID was assigned to each file. This process of anonymization was crucial for maintaining the privacy of individuals while still allowing for the use of detailed medical data.

The final dataset consists of 800 of angiogram scans, including 700 LCA scans and 100 RCA scans. Each scan was processed to extract frame-wise images that provide detailed views of the coronary arteries. The dataset includes specific images of the right coronary artery, as well as the left coronary artery, with their respective branches.

In Figure **3-1**.(a) The RCA originates from the right aortic sinus and travels along the right atrioventricular groove. It primarily supplies oxygenated blood to the right atrium, right ventricle, sinoatrial (SA) node, and atrioventricular (AV) node. Additionally, the RCA gives rise to the posterior descending artery in most individuals, which supplies the interventricular septum. Understanding the anatomy of the RCA is crucial for diagnosing and treating conditions such as right ventricular infarctions and arrhythmias related to the SA and AV nodes.

In Figure **3-1** (b) The LCA emerges from the left aortic sinus and quickly bifurcates into two major branches: the left anterior descending (LAD) artery and the circumflex artery. The LAD runs down the anterior interventricular groove, supplying the front portion of the left ventricle and the interventricular septum, while the circumflex artery extends into the left atrium and the lateral and posterior walls of the left ventricle. The LCA is essential for providing blood to a significant portion of the heart muscle, and blockages in its branches are common causes of myocardial infarction.



(a) (b)

Figure 3-1: (a) Right Coronary Artery and (b) Left Coronary Artery

3.1.1 DICOM Tags

The DICOM (Digital Imaging and Communications in Medicine) standard was employed in the collection of angiogram scans to ensure a consistent format for medical imaging data. DICOM is widely recognized as the global standard for transmitting, storing, and managing medical imaging information. It standardizes the representation of imaging data, enabling interoperability across various systems and facilitating accurate diagnosis and analysis.

Each DICOM file contains a set of metadata attributes known as DICOM tags. These tags describe various aspects of the scan, including patient information, technical imaging parameters, equipment settings, and study details. The metadata plays a vital role in documenting the circumstances under which the scan was acquired, helping to ensure that essential data required for diagnosis, review, and further processing is retained. The

consistent format provided by DICOM ensures that the information can be interpreted across diverse platforms and systems.

For the purposes of this study, specific DICOM tags were carefully selected and preserved from the angiogram scans. These tags were chosen to retain critical information about the imaging process while ensuring that patient privacy was protected. The DICOM standard provides the ability to anonymize certain patient data, such as the Patient ID, while retaining essential technical parameters required for detailed analysis.

Table 1 provides a summary of the key DICOM tags extracted from the dataset. These tags contain essential information about each scan, including the imaging modality, study description, and various acquisition parameters such as pixel spacing, frame time, and x-ray exposure settings. These technical details are necessary for performing detailed analysis and generating new synthetic views of the coronary arteries.

The retention of these DICOM tags is crucial for the detailed characterization of each scan. The metadata provides critical information for understanding the technical conditions under which the scan was performed, which is vital for tasks such as reconstructing 3D models or synthesizing new views. For example, information such as the pixel spacing, x-ray tube current, and distance from the source to the detector can influence the accuracy of geometric reconstructions.

In addition to aiding analysis, the metadata also plays a significant role in the organization and retrieval of data within a database. The DICOM tags serve as identifiers, allowing researchers to efficiently search and categorize the images based on specific criteria such as study date, modality, or institution. This organized structure is essential when working with large datasets, as it enables quick access to relevant scans and associated metadata.

The following table lists the key DICOM tags extracted from the angiogram dataset, along with their values. These tags provide a comprehensive overview of the imaging parameters and technical settings used during the scans, offering insight into the acquisition process and facilitating subsequent analysis.

Table 1 DICOM Tags

Tag	Value
Patient ID	1001
Study Date	20231103
Modality	XA
Series Description	CardiacSwing 15 fps LCA CRA 35 6s
Image Dimensions	512 x 512
Pixel Spacing	[0.40124997907318, 0.40124997907318]
PositionerPrimaryAngle	
PositionerSecondaryAngle	
PositionerPrimaryAngleIncrement	
PositionerSecondaryAngleIncrement	
KVP	75.461
Frame Time	66.657
Exposure Time	492
Number of Frames	88
Institution Name	OTTAWA HEART INSTITUTE
X-Ray Tube Current	704
Distance Source to Detector	1233.0
Distance Source to Patient	765.0
Acquired Image Area Dose Product	24.26

This table highlights key parameters that are essential for interpreting and analyzing the angiogram scans. For instance, the pixel spacing value indicates the physical distance represented by each pixel in the image, which is critical for accurate spatial measurements. The x-ray tube current and acquired image area dose product provide insight into the exposure settings, which affect image quality and patient safety.

3.2 Camera Parameters Extraction

In this section, we detail the process of extracting camera parameters from DICOM files. These parameters such as the camera's position, orientation, and intrinsic properties are essential for transforming the 2D medical images into a coherent 3D representation, enabling accurate reconstructions of anatomical structures, like the coronary arteries, within the novel view synthesis framework.

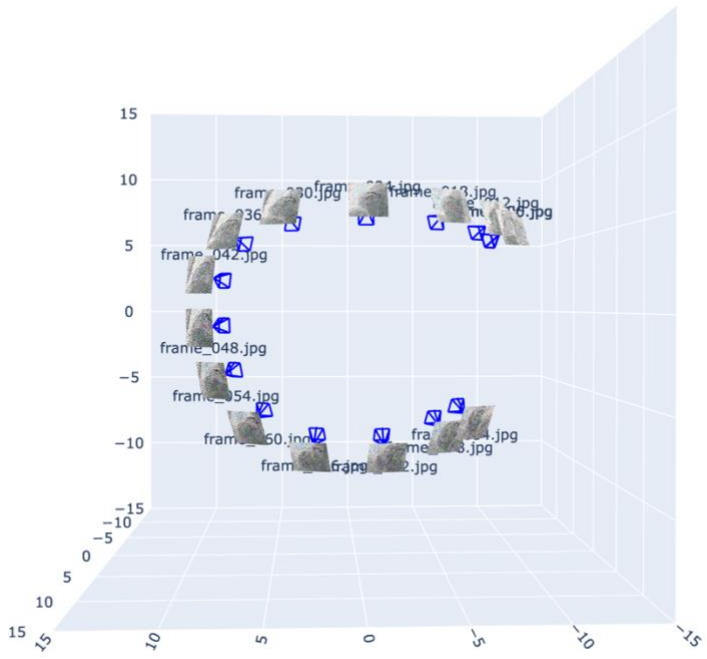


Figure 3-2 Rotational Angiography camera path

Figure 3-3 visually depicts the coordinate system and the orientation of the camera relative to the patient. The image highlights the key axes (X, Y, Z) and their respective orientations, which are crucial for interpreting the geometric conventions of the DICOM standard. These conventions ensure consistent interpretation of image data across all scans.

DICOM files follow strict geometrical conventions that define the orientation of the imaging system relative to the patient's body. Understanding these conventions is critical for accurately interpreting the image data and reconstructing the 3D environment. The conventions followed in the DICOM standard for medical imaging are as follows:

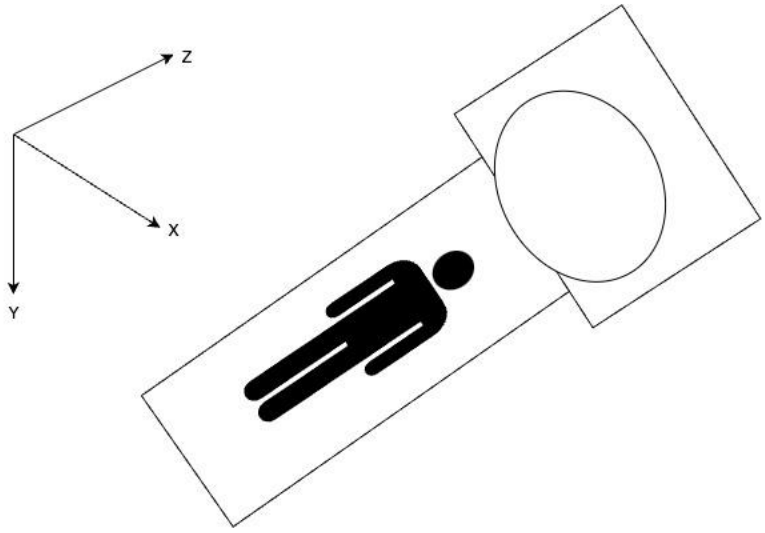


Figure 3-3 DICOM Coordinate Geometry

The X-axis increases toward the left of the patient, while the Y-axis increases towards the posterior (back) of the patient. The Z-axis increases towards the superior direction (head of the patient, as shown in Figure 3-3). In addition to these spatial orientations, angular measurements in the DICOM system adhere to specific definitions. These geometrical and angular conventions are essential for interpreting image data and determining the camera's orientation within the 3D coordinate system. They are crucial for extracting camera parameters, which directly influence the accuracy of 3D reconstructions.

The process of extracting camera parameters from DICOM files involves retrieving key metadata related to imaging geometry. These parameters are vital for ensuring accurate 3D scene reconstruction as they define both the intrinsic and extrinsic properties of the camera during image acquisition. Among the most critical camera parameters is the Source-to-Detector Distance, which is analogous to the focal length in traditional camera systems. This parameter is retrieved from the DICOM tag DistanceSourceToDetector and represents the distance between the X-ray source and the detector. In our dataset, this distance is recorded as 1233 mm. The source-to-detector distance plays a key role in determining the magnification and scale of the captured image.

Another important parameter is the Detector Size, which is calculated using pixel spacing (obtained from the ImagerPixelSpacing tag) and the number of pixels along each dimension (from the Rows and Columns tags). This detector size is essential for computing

the field of view (FOV), which defines the extent of the scene captured by the imaging system. Accurate knowledge of the detector size ensures that the 3D scene reconstruction retains the correct spatial proportions.

The Field of View (FOV) is another critical parameter that is calculated based on the detector size and the source-to-detector distance. The vertical and horizontal FOVs are computed using the following equations:

- Vertical FOV (fovy):

$$\text{fovy} = 2 \times \arctan\left(\frac{\text{detector\_size\_y}}{2 \times \text{DistanceSourceToDetector}}\right) \qquad (8)$$

- Horizontal FOV (fovx):

$$\text{fovx} = 2 \times \arctan\left(\frac{\text{detector\_size\_x}}{2 \times \text{DistanceSourceToDetector}}\right) \qquad (9)$$

These FOV values are integral to determining the angular extent of the scene visible to the camera, which directly influences the accuracy of the 3D scene reconstruction by defining the region of space captured in each image.

The angular positions of the imaging device during acquisition are essential for characterizing the motion and orientation of the camera. These angles are extracted from the DICOM metadata. The Positioner Primary Angle (RL) captures the right-left orientation of the camera, and determining how far the camera is rotated along the right-left axis is vital. The Positioner Secondary Angle (CC) captures the craniocaudal orientation, defining the rotation of the camera along the craniocaudal axis. These angular measurements, along with their increments (which are extracted from the arrays of angle increments in the DICOM metadata), provide a detailed description of the camera's orientation and motion throughout the acquisition process. This detailed orientation data is crucial for rendering novel views from varying perspectives.

To accurately reconstruct the 3D scene from 2D medical images, it is essential to compute a transformation matrix that represents the camera’s position and orientation in 3D space.

3.3 Evaluation Metrics

In the context of evaluating the performance of generative models and image reconstruction techniques, several quantitative metrics are commonly employed to assess the quality of generated or reconstructed images. Among these, Fréchet Inception Distance (FID), Kernel Inception Distance (KID), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) are pivotal. These metrics offer complementary insights into the similarity between generated images and their real counterparts, providing both statistical and perceptual evaluations.

- Fréchet Inception Distance (FID)

Fréchet Inception Distance (FID) is a widely adopted metric used to quantify the similarity between two sets of images—one set produced by a generative model and the other comprising authentic images. FID compares the feature distributions of these image sets using a pre-trained neural network, often Inception-v3, which extracts high-level features from the images. Specifically, the metric computes the difference between the mean and covariance of the features for the generated and real images, resulting in a score that reflects how closely the generated images resemble the real ones.

Mathematically, FID is expressed as:

$$\mathrm{FID} = |\mu_r - \mu_g|^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{1/2}\right) \qquad (10)$$

where μ_r and μ_g represent the mean feature vectors of the real and generated image distributions, respectively, and Σ_r and Σ_g denote the covariance matrices of these distributions. The FID score captures both the difference in means and the overall spread (covariance) of the feature distributions, providing a comprehensive measure of similarity.

A lower FID score indicates that the generated images have a feature distribution more closely aligned with that of the real images, thus signifying higher quality. In generative modeling, reducing the FID score is often a primary objective, as it directly correlates with the model's ability to produce realistic images.

- Kernel Inception Distance (KID)

Kernel Inception Distance (KID) is another metric used to evaluate the similarity between generated and real images, with a methodology closely related to FID. However, KID employs kernel-based techniques to compare the distributions of the two image sets. Unlike FID, which relies on the Fréchet distance and a Gaussian approximation of the feature distributions, KID uses the squared Maximum Mean Discrepancy (MMD) to compare the distributions. This kernel-based approach avoids the assumption that the distributions are Gaussian, potentially offering greater flexibility in certain scenarios.

The KID score is computed as follows:

$$\text{KID} = \frac{1}{m(m-1)}\sum_{i\neq j} k(x_i, x_j) + \frac{1}{n(n-1)}\sum_{i\neq j} k(y_i, y_j) - \frac{2}{mn}\sum_{i,j} k(x_i, y_j) \qquad (11)$$

where x_i and y_i represent features from the real and generated images, respectively, and $k(x, y)$ is a kernel function, often a Gaussian kernel. Like FID, a lower KID score suggests that the generated images are of higher quality, more closely resembling real images. One notable advantage of KID over FID is that it provides unbiased estimates, even with a small number of samples, which can be beneficial when working with limited data.

- Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for assessing image reconstruction quality. It quantifies the ratio between the maximum possible power of a signal and the power of noise that corrupts the signal. PSNR is calculated in decibels (dB) and is generally used to measure the quality of reconstructed images by comparing them to reference images. A higher PSNR value signifies better image fidelity, as it implies that the reconstruction is less affected by noise and closer to the original image.

The PSNR is defined as:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \qquad (12)$$

where (MAX) is the maximum possible pixel value of the image (e.g., 255 for an 8-bit image), and (MSE) is the mean squared error between the reference and reconstructed images. PSNR is particularly effective in scenarios where the noise is additive and independent of the image content, offering a straightforward measure of the reconstruction error. However, while PSNR focuses on pixel-wise error, it may not always correlate well with perceptual image quality, making it more suitable for technical evaluations rather than perceptual assessments.

- Structural Similarity Index (SSIM)

The Structural Similarity Index (SSIM) is a perceptual metric that evaluates the visual quality of an image by comparing three key characteristics: luminance, contrast, and structure. Unlike PSNR, which focuses on error magnitude, SSIM assesses how similar two images are in terms of their visual structure, offering a more human-centric measure of image quality. SSIM values range between -1 and 1, with values closer to 1 indicating greater similarity to the reference image and, therefore, higher image quality.

The SSIM between two images (x) and (y) is computed as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (13)$$

where μ_x and μ_y are the local means of the images, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance between the images, and $C_1$and C_2 are constants to stabilize the division. By assessing these three factors, SSIM offers a robust evaluation of image quality that aligns more closely with human visual perception, particularly when small perceptual differences, rather than large numerical errors, are critical.

Together, FID, KID, PSNR, and SSIM provide a comprehensive set of tools for evaluating the quality of images produced by generative models and image reconstruction algorithms. While FID and KID focus on statistical similarities between real and generated images, PSNR and SSIM offer insights into reconstruction quality, with PSNR focusing on error magnitude and SSIM emphasizing perceptual similarity. By employing these metrics, researchers can rigorously evaluate the performance of generative and reconstruction models, ensuring that the produced images are both statistically and perceptually similar to the real or reference images.

Chapter 4

Novel View Synthesis via MedNeRF

4.1 Introduction

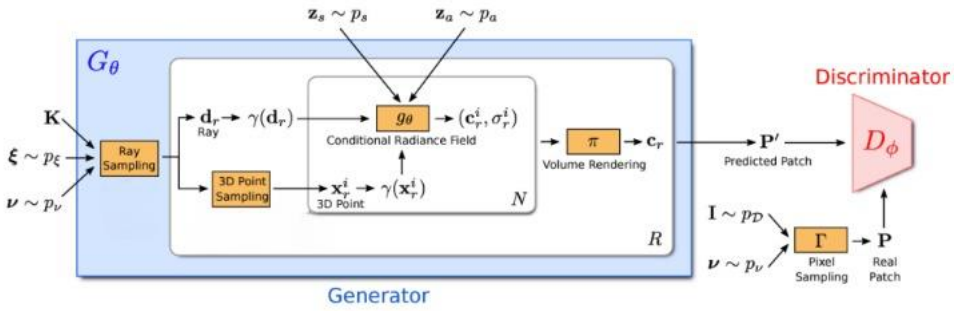


Figure 4-1 MedNeRF Architecture, reused from [35] with Author's permission / original car graphics removed

NeRF's application in the medical field was initially constrained by its lack of generalizability in results. However, the introduction of the Generative Radiance Fields (GRAF) marked a significant advancement, integrating a generative model that enhanced the resolution of synthesized images and the separation of camera perspectives from scene characteristics. Nevertheless, GRAF's broad applicability did not entirely meet the specific needs of medical imaging, which demands high anatomical precision and the ability to work with small datasets. In response to these challenges, MedNeRF was developed as a specialized adaptation of GRAF for medical imaging purposes.

As a result, MedNeRF provides an efficient solution for working with sparse and minimally invasive input data, such as single-view X-rays. Key components of the MedNeRF architecture include also shown in Figure **4-1**:

1. Generator (G): The generator predicts an image patch based on the camera matrix, sampled camera pose, pixel sampling pattern, and shape/appearance codes. An important aspect of MedNeRF is its ability to generate images with different shapes and appearances in a single model by introducing randomly sampled data of different medical instances during training.
2. Discriminator (D): The discriminator compares predicted image patches to real

ones extracted from real images at multiple scales. This comparison is used to train the model and make it "learn" from the disparities in predictions. MedNeRF introduces a new discriminator architecture that incorporates self-supervised learning. Training on a pretext task is designed to learn useful global and local features.

3. Volumetric Rendering: Once the model has been trained, it can reconstruct a complete set of X-ray projections within a full vertical rotation from a single-view X-ray. This feature of MedNeRF is based on modifying the generation objective by adding a distortion Mean Square Error (MSE) loss that balances blurriness and accuracy.

The architecture makes use of an implicit linear network that can disentangle 3D anatomy identity and the attenuation response of different medical instances. The architectural framework under consideration exhibits certain limitations when confronted with dynamic data, such as rotational angiography. It omits any temporal characteristics of the data, thereby failing to account for the vasculature deformation caused by heartbeats and non-cardiac body movements.

4.2 Implementation

Three experiments, described in the next subsections, were designed to evaluate the model's performance across different datasets and training iterations, focusing on the generation quality assessed by several evaluation matrices metioned in section 3.3

4.2.1 Experiment A

In this experiment, we aimed to assess the individual impact of Left Coronary Artery (LCA) and Right Coronary Artery (RCA) images on the training and performance of our generative model. By training the model separately on datasets of LCA and RCA images, we sought to understand how the anatomical differences and dataset sizes affect the model's ability to learn and synthesize realistic coronary artery images.

For the LCA, we utilized a dataset comprising 760 scans, training the model over 300,000 iterations. The training allowed the model to capture the complex and intricate vasculature

characteristic of the LCA. The synthetic images generated from various angles demonstrated the model's capability to represent the detailed branching patterns and fine structures of the LCA, showcasing its proficiency in handling complex anatomical features.

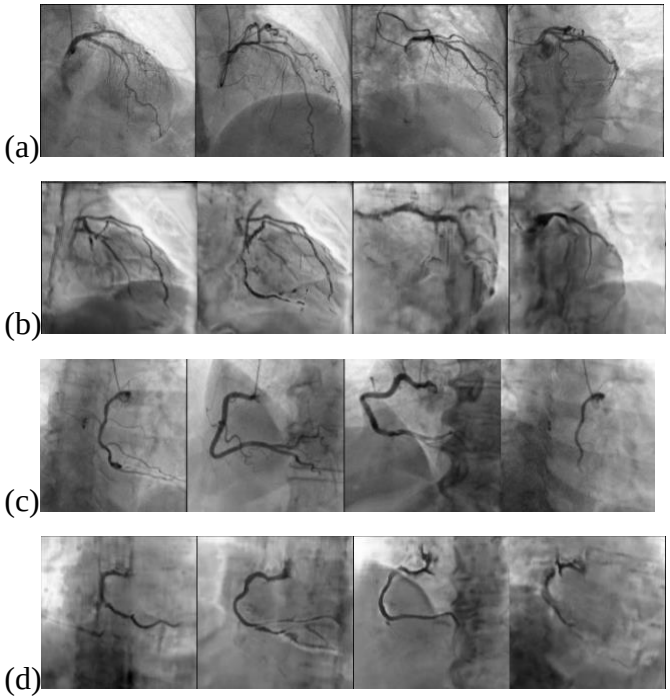


Figure 4-2 Qualitative results: (a) LCA ground truth, (b) LCA generated, (c) RCA ground truth, and (d) RCA generated.

In contrast, we trained a separate model on a smaller dataset of 100 scans of the RCA, which is known for having more distinct and well-defined arterial structures compared to the LCA [46]. Despite the reduced amount of training data, the model effectively learned to generate synthetic images with precisely defined vasculature. The results suggest that the RCA's simpler and more consistent anatomical features make it easier for the model to learn and replicate, even with less data and fewer training iterations.

To further analyze the individual impact of both LCA and RCA images on model performance, we conducted an additional experiment using a balanced dataset. We trained the model on approximately 95 scans from each artery type for 100,000 iterations. This balanced approach allowed us to directly compare how the model learns from equal amounts of data representing both artery types. The findings indicated that while the model could generate realistic images for both LCAs and RCAs, the quality and precision were

influenced by the inherent anatomical complexity and the amount of training data available for each artery type.

These experiments demonstrated that RCAs are easier for the model to generate accurately, even with less training data, due to their more distinct and well-defined structures. On the other hand, LCAs, with their complex and intricate vasculature, require more extensive data and training iterations to achieve comparable levels of detail and realism in the synthetic images (Refer to Figure **4-2**). Understanding these differences is crucial for optimizing model training strategies, particularly in medical imaging applications where data availability and anatomical variability can significantly impact model performance.

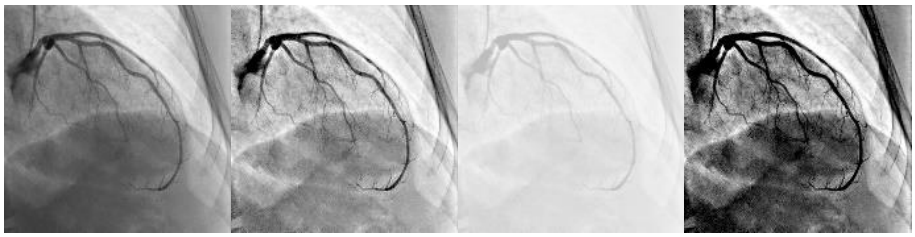
By analyzing the individual contributions of LCA and RCA images, we gained valuable insights into the model's learning dynamics and its capacity to generalize across different anatomical structures. These findings inform future work on improving generative models for medical imaging, emphasizing the importance of dataset composition and the need to tailor training processes according to the specific characteristics of the anatomical features being modeled.

4.2.2 Experiment B

This experiment aims to achieve better qualitative results by applying preprocessing steps to the training dataset. Image preprocessing is a crucial step for image enhancement of angiographic images. Angiograms are noisy, blurred, and full of occlusions. Additionally, motion from breathing and the beating nature of the heart add more complexity to these 2D scans. Various studies have highlighted the importance of preprocessing in improving the visibility and accuracy of angiographic analysis. For instance, techniques such as Gaussian filtering and anisotropic diffusion have been employed to reduce noise while preserving edge details, which are critical for accurate diagnosis [47]. Furthermore, contrast enhancement methods like CLAHE (Contrast Limited Adaptive Histogram Equalization) have been shown to enhance the visibility of vascular structures, making them more discernible in clinical settings [48].

In addition to traditional methods, more advanced approaches such as machine learning-based denoising techniques have been developed to tackle the complex noise patterns in angiograms. These methods leverage convolutional neural networks (CNNs) to learn and

apply noise reduction filters dynamically, resulting in cleaner and more interpretable images. However, these advanced methods often require large datasets and significant computational resources, which may limit their applicability in certain settings [49].



(a) (b) (c) (d)

Figure 4-3 Preprocessed Images: (a) actual image, (b) MSRCP, (c) MSRCR, and (d) Auto-MSRCR

Hence, we explored image processing techniques like Multiscale Retinex with Color Restoration (MSRCR), Multi-scale Retinex with Chromacity Preservation (MSRCP) [50], and Auto-MSRCR [51] methods that remove artifacts from the image and denoise it [47]. Recent studies have indicated that MSRCR, when combined with wavelet-based denoising, can further enhance image quality by effectively suppressing noise while preserving important diagnostic features. Figure **4-3** compares these methods. However, there were no major visual improvements in the generated results using preprocessed images, which may be attributed to the inherent limitations of the preprocessing algorithms used or the specific characteristics of the dataset employed in this study [52].

4.2.3 Experiment C

In this experiment, we aimed to optimize the volume rendering of high-resolution images using the MedNeRF architecture. However, rendering high-resolution images with MedNeRF proved to be computationally expensive due to the intensive memory and processing power required for handling large image dimensions.

To address these computational challenges, we experimented with reducing the resolution of Left Coronary Artery (LCA) scans from 512×512 pixels to 128×128 pixels. The high-resolution images at 512×512 pixels demanded high GPU memory, which posed constraints on the training process and limited our ability to efficiently leverage the volumetric rendering capabilities of MedNeRF. By downscaling the images, we

significantly reduced the computational load, making it feasible to train the model within the available GPU memory constraints.

We trained the MedNeRF model on the reduced-resolution LCA scans for 200,000 iterations. This number of iterations was selected to ensure that the model had sufficient exposure to the data to learn the underlying anatomical features despite the lower resolution. The reduction in image size allowed us to expedite the training process while still capturing essential structural details of the coronary arteries.

Our primary goal was to overcome the GPU memory limitations without substantially compromising the quality and accuracy of the rendered images. By optimizing the model training with lower-resolution inputs, we aimed to maintain a balance between computational efficiency and the fidelity of the volumetric renderings. The use of MedNeRF's volumetric rendering approach on reduced-resolution images enabled us to generate synthetic views of the LCA from various angles, demonstrating the model's ability to reconstruct 3D anatomical structures effectively.

The results of this experiment indicated that reducing the image resolution is a viable strategy for managing computational resources while still benefiting from the advanced rendering capabilities of MedNeRF. Although there is an inherent trade-off between image resolution and detail, the 128×128 pixel images retained sufficient anatomical information for the purposes of our study. This approach allowed us to improve the efficiency of the training process and overcome hardware limitations, ultimately enhancing the practicality of using MedNeRF for high-resolution medical image synthesis.

4.3 Analysis

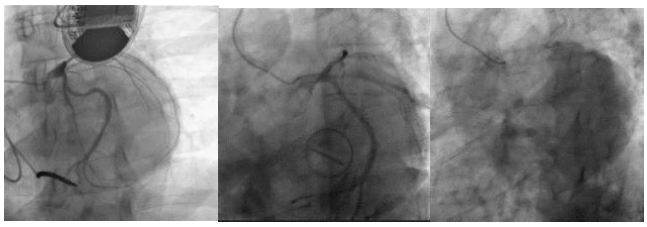
In our study, we conducted a series of experiments to evaluate the capability of our model in generating realistic vasculature images of the heart, specifically focusing on the left coronary artery (LCA) and the right coronary artery (RCA). The aim was to assess how different factors such as artery complexity, image preprocessing, and data dimensions affect the model's performance in synthesizing accurate representations of coronary vasculature.

The results from Experiment A, presented in Figure **4-2**, demonstrate that RCA-dominant

views are easier for the model to generate compared to LCA views. This is primarily due to the RCA's more concise and less complex anatomical structure. The RCA typically has fewer branches and a more straightforward path than the LCA, which makes it less challenging for the model to learn and replicate its features.

The outcome of Experiment B denied our hypothesis that image processing will improve the results.

These experiments collectively demonstrate the model's potential to generate vasculature images that resemble the anatomy of the heart. However, due to inherent challenges with this data type such as anatomical variability, imaging artifacts, and motion distortions (Refer to Figure **4-4**) there are likely to be anatomical abnormalities in the generated images. These issues can manifest as missing vessel segments, distortions, or inaccuracies in vessel branching patterns.



(a) (b) (c)

Figure 4-4 (a, b) surgical instruments blocking arteries (c) No contrast dye.

An analysis of the results presented in Table I suggests that the model used in Experiment C outperformed the previous experiments, despite being trained on reduced data dimensions and with less training time. In this experiment, we reduced the resolution of the LCA scans from 512×512 pixels to 128×128 pixels to overcome GPU memory constraints and improve computational efficiency.

Table 2 Quantitative metrics comparison

	Iterations	FID	KID
LCA	300,000	116.52	0.088
RCA	200,000	132.23	0.098
Lowres. LCA	200,000	110.76	0.076

The fine-tuned model in Experiment C achieved the lowest Fréchet Inception Distance (FID) score of 110.76 and the lowest Kernel Inception Distance (KID) of 0.076 among all experiments. Lower FID and KID values indicate better quality and diversity of the generated images compared to the real images. Moreover, this model achieved a Structural Similarity Index Measure (SSIM) of 0.77 and a Peak Signal-to-Noise Ratio (PSNR) of 30.1, which are higher than those obtained in previous experiments. These metrics reflect a higher similarity and reconstruction quality between the generated images and the ground truth images.

The improved performance, despite reduced image resolution and fewer training iterations, suggests that optimizing data dimensions can enhance model efficiency without compromising image quality. This finding is significant for practical applications where computational resources are limited.

One of the key challenges in working with rotational angiograms is the sparsity of information due to limited angular coverage. MedNeRF was initially trained on views from a full 360-degree vertical rotation along one axis. However, in clinical practice, rotational angiograms as depicted in Figure **4-5** follow a predefined trajectory along two axes around the patient and do not perform a complete 360-degree rotation due to physical and safety constraints. The limited angular range hinders reconstruction accuracy and can result in incomplete representations of the coronary arteries. Certain views are missing, which forces the model to interpolate unseen perspectives, potentially leading to inaccuracies. Additionally, the beating nature of the heart introduces motion artifacts. The heart's

movement during angiography causes deformations and distortions in the images, making it difficult for the model to learn consistent structural patterns of the vasculature.

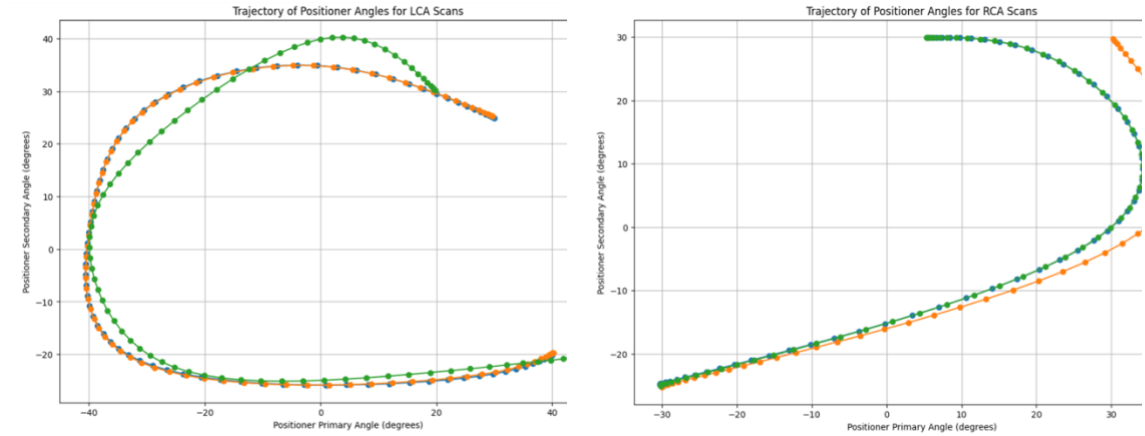
The quality of the reconstructed images is further affected by several factors. Inconsistent contrast dye distribution at the start of angiogram scans can lead to incomplete capture of the vasculature, as the dye has not yet fully permeated the arteries. Obstructions from overlapping structures such as vasculature, ribs, and other anatomical features can obstruct the view of the coronary arteries, introducing noise and occlusions in the images.

Moreover, the presence of non-cardiac artifacts poses a significant challenge. Since the data is collected from real patients, images may include stents, pacemakers, surgical wires, and other artifacts, which can interfere with the model's learning process. Anatomical variations in coronary artery patterns also contribute to the complexity. Coronary artery dominance varies among individuals, with approximately 80% of people having a right-dominant configuration where the RCA supplies the posterior descending artery, about 10% having a left-dominant configuration, and the remaining 10% having a codominant configuration where both the RCA and LCA contribute. These variations make it challenging to achieve balanced training data that adequately represents all possible anatomical configurations.

4.4 Conclusion

Our experiments highlight the model's potential to generate vasculature images that closely resemble the heart's anatomy. The findings indicate that the RCA is easier for the model to generate due to its simpler and more consistent structure, even with smaller datasets and fewer training iterations. The importance of image preprocessing is not evident in the case of NeRf for X-ray imaging. Optimizing data dimensions by reducing image resolution can overcome computational constraints and improve model performance without sacrificing output quality, as evidenced by the superior results in Experiment C.

Despite these advancements, challenges such as limited angular coverage, motion artifacts, imaging obstructions, and anatomical variability continue to affect reconstruction accuracy. Addressing these issues is crucial for further improving model performance.



(a) (b)

Figure 4-5 Swing angiogram trajectory: (a) Angular path of LCA scans, and (b) Angular paths of RCA scans.

The appearance of the reconstructed images is affected by inconsistent contrast dyes in the scans, as well as obstructions caused by overlapping vasculature and ribs.

To further improve the model's accuracy and reliability, future work should focus on enhancing the resolution of training data to capture finer anatomical details more effectively, a strategy we explored in Chapter 4. Additionally, incorporating methods to address the dynamic and temporal challenges posed by the beating heart and rotational angiograms is critical for more accurate reconstructions. In Chapter 5, we discuss our attempts to overcome these dynamic challenges, emphasizing the importance of motion correction and temporal consistency in improving the model's ability to generate accurate vasculature images.

Chapter 5

Novel View Synthesis via StyleNeRF

5.1 Introduction

In the previous chapter, we explored novel view synthesis through the MedNeRF architecture, demonstrating its effectiveness in generating views from medical imaging datasets. MedNeRF excels in handling structured, domain-specific data, particularly in tasks where low to medium-resolution outputs are sufficient. However, a significant limitation of MedNeRF lies in its inability to effectively generate high-resolution images—a requirement crucial to the accurate visualization of intricate anatomical structures in medical scans. This constraint led us to seek alternative architectures capable of leveraging our high-resolution images more effectively.

In this chapter, we introduce StyleNeRF as a solution to overcome these limitations. StyleNeRF's architecture enables the synthesis of high-quality images with fine-grained control over texture and resolution, making it particularly well-suited for coronary artery imaging, where precision and clarity are essential. By utilizing StyleNeRF's unique hierarchical latent space, we aim to generate realistic and detailed novel views from high-resolution medical imagery, providing superior results compared to MedNeRF.

The architecture of StyleNeRF follows a similar design to StyleGAN2, incorporating both a mapping network and a synthesis network. The mapping network samples latent codes from a Gaussian distribution, which are then processed and broadcasted to the synthesis network. The synthesis network leverages NeRF++ [53], a backbone that divides the scene into a foreground NeRF (within a unit sphere) and a background NeRF (using an inverted sphere parameterization). This separation allows StyleNeRF to focus on the vascular structures while effectively managing the background, enhancing the quality of the generated images.

Progressive training, starting from low to high resolution, allows for stable training across increasing resolutions, ensuring that high-resolution details are captured without

overwhelming the model during early training stages. Additionally, StyleNeRF applies a NeRF path regularization to ensure 3D consistency in the synthesized views.

5.2 Implementation

The left coronary artery dataset used for StyleNeRF consists of high-resolution images derived from rotational angiography scans as mentioned in 3.1. Unlike traditional RGB images, X-ray images present unique challenges due to their monochromatic nature. This required careful data handling to ensure that the fine vascular structures, such as the left circumflex and left marginal arteries, were adequately represented in the training data. By utilizing the NeRF++ backbone, we could disintegrate the background and foreground effectively, which helped isolate coronary arteries from surrounding structures. This was a key improvement over MedNeRF, which struggled with merging vessel and background pixels.

To address 3D consistency, NeRF path regularization played a critical role by enforcing accurate depth and structure relationships between different views. The model was trained with u and v angular ranges derived from the rotational angiography data, improving its ability to generate consistent multi-angle views. These ranges, defined as -20 to 30 degrees in v and -40 to 40 degrees in u, were critical in ensuring the synthesized images matched the real-world perspectives seen in angiograms.

Training StyleNeRF required significant computational resources due to the high-resolution nature of the dataset. The model was trained on a multi-GPU setup at the Ottawa Heart Institute, utilizing GPUs with sufficient memory to handle the large images. Key hyperparameters such as learning rate, batch size, and the number of training epochs were fine-tuned to balance training speed and model accuracy. Progressive training allowed the model to capture more detailed structures without encountering convergence issues at higher resolutions.

StyleNeRF's progressive training strategy proved essential for generating high-resolution images of coronary arteries. As shown in Figure **5-1**, at certain iterations, marked by peaks, resolution upsampling occurred. These peaks correspond to moments when the model shifted from one resolution level to the next, starting from low-resolution images and

gradually increasing their size throughout the training process. Each increase in resolution initially caused a spike in the FID score, signaling a temporary difficulty for the model to adjust to the new level of detail. However, after these jumps, the model successfully stabilized, refining its ability to capture finer anatomical structures, such as the intricate branches of the coronary arteries. This stepwise resolution scaling was key to ensuring that StyleNeRF incrementally learned and improved, adding more detailed textures and depth to the images at each stage. Eventually, the model effectively handled high-resolution data, accurately generating sharp, detailed representations of the coronary arteries, as demonstrated by the subsequent reduction in FID scores after each upsampling phase.

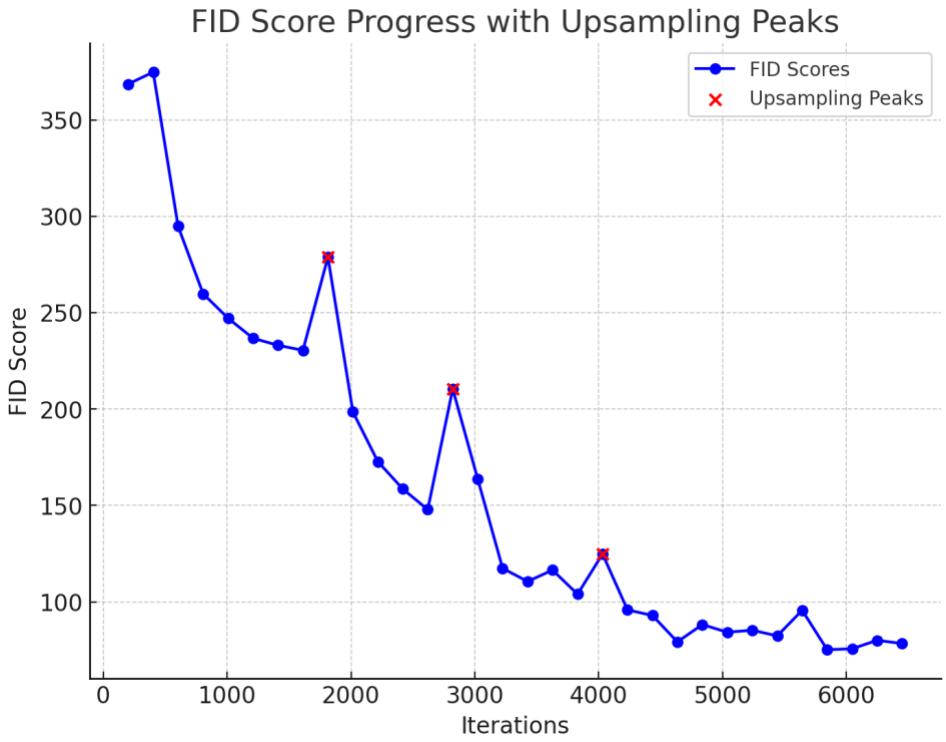


Figure 5-1 FID Progress with Upsampling peaks

As illustrated in Figure **5-2**, the encoder within the StyleNeRF architecture plays a critical role in transforming high-dimensional input data (real images) into meaningful latent representations. These latent codes are key to the model's ability to generate novel views of coronary arteries. In the figure, the top path represents how random latent codes pass through a trained model to generate an output image. The bottom path shows how an input image is processed by the encoder to calculate latent codes, which are then fed into the

synthesis network to produce a new image. This workflow highlights how the encoder serves as a bridge between real angiographic images and their latent representations, enabling accurate and high-resolution view generation.

The encoder, is trained using several loss functions, including smooth L1 loss for the latent code matching and VGG loss for perceptual quality. The encoder learns to map input angiograms to latent codes, which are passed through the synthesis network to generate novel views. During training, the random latent codes and calculated latent codes are compared, ensuring that the generated images retain structural accuracy when reconstructed.

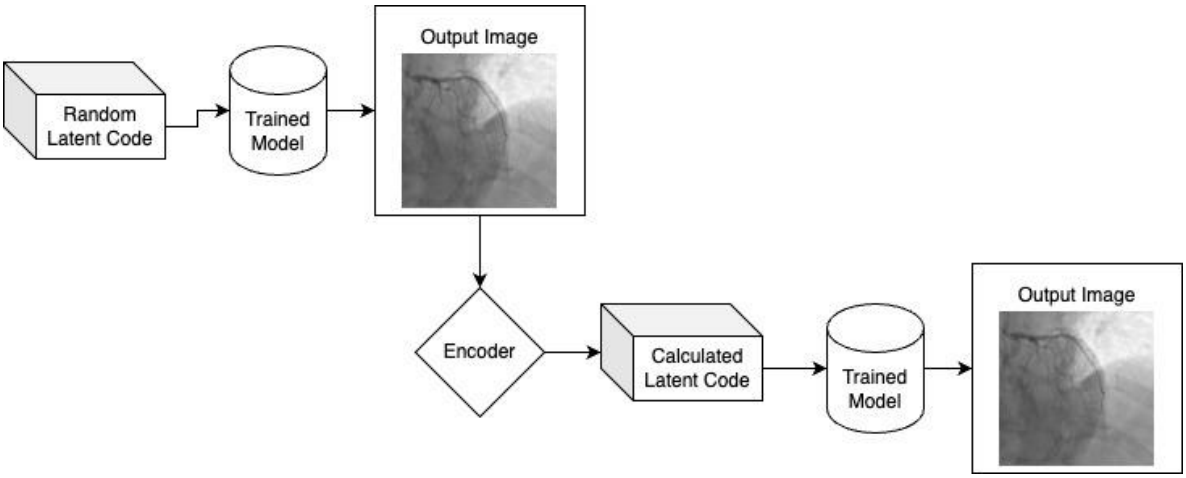


Figure 5-2 Encoder Training

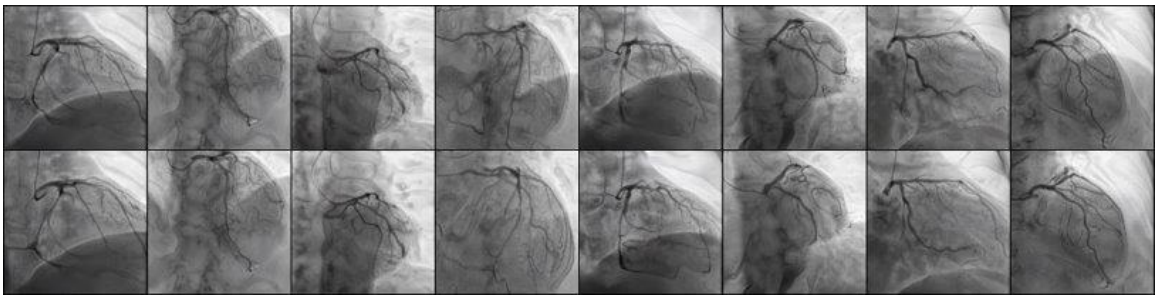


Figure 5-3 Encoder validation

5.3 Analysis

In both quantitative and qualitative evaluations, StyleNeRF outperformed MedNeRF. While MedNeRF was sufficient for generating low to medium-resolution outputs, it struggled with fine vascular structures, often leading to blurred or merged artery branches. In contrast, StyleNeRF excelled at rendering complex anatomical structures, such as the left marginal and left circumflex arteries, providing significantly clearer and more accurate images.

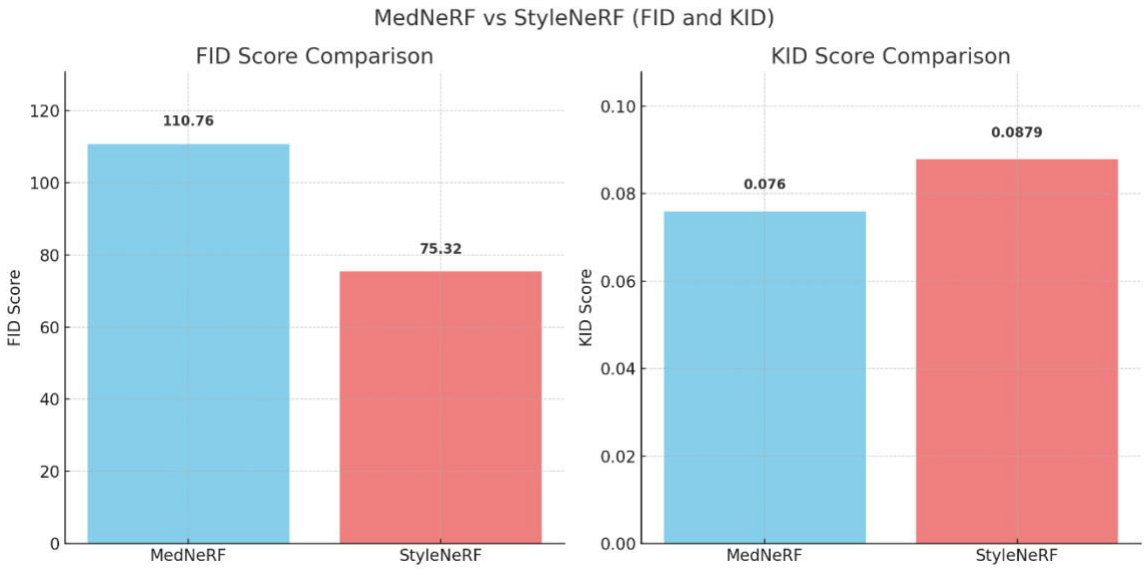


Figure 5-4 Quantitative Score comparison between MedNeRF and StyleNeRF

Quantitatively, StyleNeRF achieved an FID score of 75.32, compared to MedNeRF's score of 110.76. This significant improvement is clearly depicted in Figure **5-4**, where the FID and KID scores for both models are compared side by side. StyleNeRF's lower FID score, a direct indicator of its superior image quality, demonstrates its ability to capture fine details and preserve anatomical structures more effectively than MedNeRF.

While StyleNeRF had a slightly higher KID score 0.0879 compared to MedNeRF's 0.076, this can be explained by the model's focus on generating finer details in high-resolution images. KID, which is more sensitive to pixel-level differences, likely picked up small local inconsistencies that arose from StyleNeRF's hierarchical training approach. These pixel-level deviations, although minor, resulted in a higher KID score, but they did not detract from the overall quality and accuracy of the images, as evidenced by the

significantly lower FID score. The increase in KID reflects a trade-off between pixel-level consistency and the model's ability to generate clearer, more detailed structures in high resolution.

StyleNeRF's progressive training allowed for the generation of highly detailed, high-resolution outputs, a key improvement over MedNeRF. The model's ability to generate finer details was validated using both an encoder and an inversion script as shown in Figure **5-3**. These validation techniques ensured that the synthesized images accurately represented real artery structures, providing a higher level of confidence in the model's outputs. The comparison of FID and KID scores serves as a testament to StyleNeRF's effectiveness in handling high-resolution medical images, ultimately leading to more precise and reliable representations of coronary arteries.

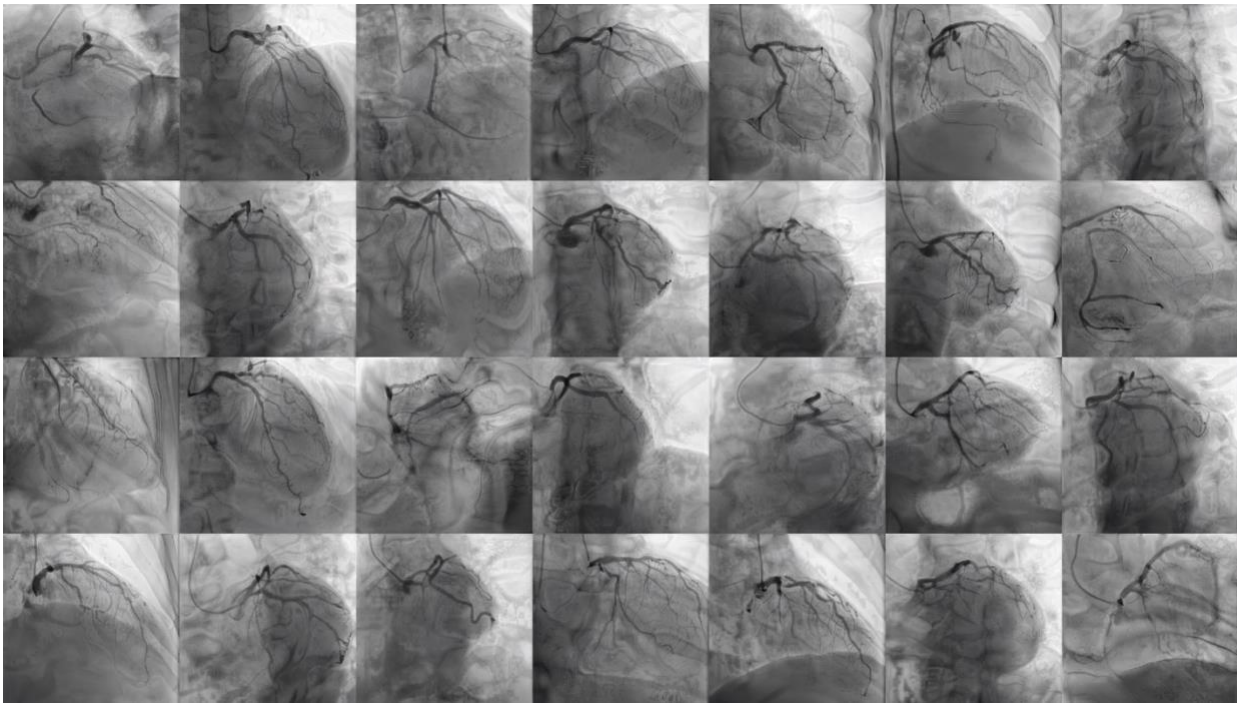


Figure 5-5 High-Resolution Coronary Arteries Images generated by StyleNeRF

Figure 5-5 shows a sample high-resolution output generated by StyleNeRF, clearly capturing the complex branching and fine textures of the coronary arteries.

StyleNeRF’s incorporation of 3D consistency through path regularization has significantly enhanced the realism and structural accuracy of the synthesized coronary artery views, even when observed from different angles. This capability is best demonstrated through a

web app-based inference viewer, which allows for real-time interpolation between various input angles of the coronary arteries. Figure **5-6** showcases a screenshot of this inference web app, where users can interactively explore different perspectives of the arteries, helping to visualize critical areas and supporting both diagnosis and treatment planning.

The web app provides a user-friendly interface, where clinicians can input specific angles or smoothly transition between views based on predefined rotational angiography data. The app leverages the latent codes generated by StyleNeRF's encoder, allowing for the interpolation of images in real-time. This functionality is especially useful for creating novel views from input data, as it enables the user to seamlessly visualize the coronary arteries from angles that may not have been captured during the original angiogram. As shown in Figure **5-6**, users can adjust the parameters such as view angles, offsets, and zoom levels, and the tool instantly reflects these changes by generating corresponding high-resolution artery images.

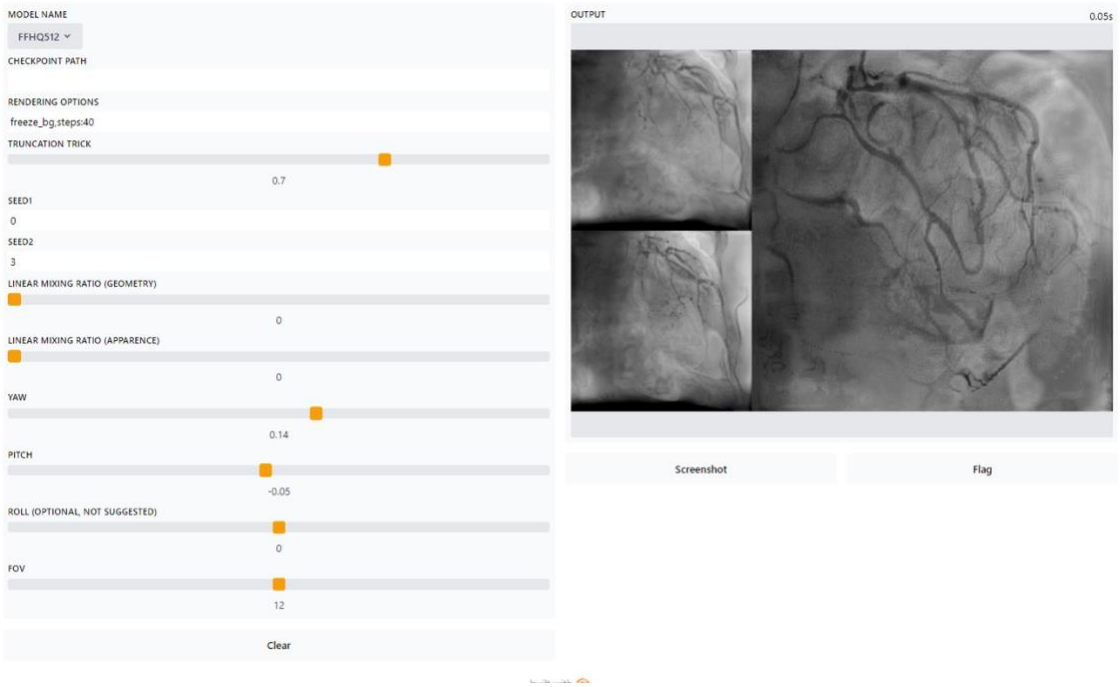


Figure 5-6 Web Interface for Angular Interpolation

One feature of the app is the ability to select any angle within the range of the rotational angiography data, allowing clinicians to obtain views that best suit their needs. By interpolating between input angles, the app ensures smooth transitions between different perspectives, providing a cohesive 3D visualization of the coronary structures. The right-

hand side of Figure displays the resulting coronary images in high clarity, offering a clear view of the artery's structure and branching patterns. The bottom panel of the interface even includes options for taking screenshots or flagging specific views for further analysis, adding to the tool's utility in clinical settings.

This web-based inference tool, as illustrated in Figure **5-6**, significantly enhances the clinical value of StyleNeRF, transforming it from a standalone model into a practical application. Clinicians can utilize this tool for procedural planning, exploring arterial structures without the need for additional imaging. Currently, the tool generates views using random latent codes, which allows for the creation of novel views based on the encoded latent space. Although this is highly useful for generating multiple perspectives, it does not yet allow the user to input a specific target image to guide the view generation. However, the ability to generate high-resolution images from previously unseen angles reduces the need for repeated invasive angiography scans, making the tool both time-efficient and safer for patients.

In future work, this tool could be further refined to support input of target images for specific patient cases. This would allow clinicians to provide a real image as input, and the model would generate novel views based on the given input, ensuring even more precise control over the output images. By incorporating this functionality, the web app would enhance its real-time interpolation feature, ensuring that coronary artery visualization is not only detailed but also customizable to individual clinical needs. This enhancement would make the tool an even more powerful asset for diagnosis and treatment planning, offering a seamless transition from random latent-based generation to target-image-guided synthesis for personalized medical imaging.

Despite the improved resolution, StyleNeRF faced challenges in handling very fine structures, particularly when interpolating angles outside of the rotational angiography path. In these cases, the model generated inconsistent images, as it struggled to infer accurate details from sparse angular data. Additionally, the model occasionally overfitted on predominant views, leading to inconsistencies in less frequent viewing angles.

5.4 Conclusion

The implementation of StyleNeRF in this study demonstrates its superiority over MedNeRF in generating high-resolution, detailed views of coronary arteries. StyleNeRF's progressive training, combined with its advanced architecture, allows for the synthesis of highly accurate and realistic images, addressing many of the limitations observed in MedNeRF. The model was particularly successful in capturing fine anatomical structures, such as small artery branches, and maintaining 3D consistency across different views.

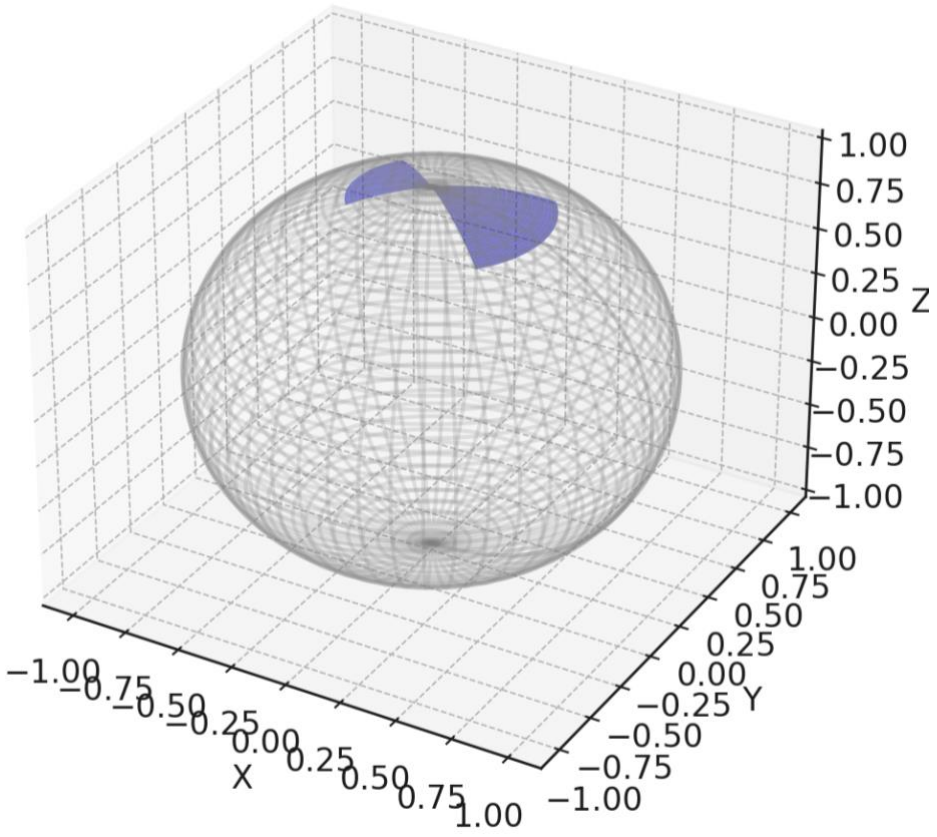


Figure 5-7 Angular Range of Rotational angiograms visualized on a sphere.

However, challenges remain, particularly in generating consistent views when interpolating outside the predefined angular range as shown in Figure **5-7** and handling dynamic aspects of the data, such as heart motion and contrast dye flow. Future work will focus on incorporating dynamic components into the architecture, through the integration of merging with dynamic NeRF architectures to capture temporal changes more effectively.

Chapter 6

Novel View Synthesis via Dynamic NeRFs

6.1 Introduction

NeRFs have revolutionized the field of 3D scene representation and rendering by modeling scenes as continuous functions of spatial coordinates and viewing directions, mapping them to radiance and volume density. This approach enables the synthesis of photorealistic images from novel viewpoints with remarkable fidelity. Building upon this foundation, generative NeRF models such as GRAF, StyleNeRF, and GRAM have extended these capabilities to 3D-aware image synthesis. These models integrate generative adversarial networks (GANs) with NeRFs to produce high-resolution, 3D-consistent images from unstructured 2D datasets, often disentangling camera parameters and scene attributes to allow for explicit control over style and viewpoint.

However, a significant limitation of these generative models is their focus on static scenes. They are designed to synthesize images of objects or environments that do not change over time, which restricts their applicability in scenarios involving motion or temporal dynamics. In contrast, many real-world applications involve dynamic scenes where the content evolves over time due to motion, deformation, or other temporal factors.

To address this gap, dynamic neural radiance fields have been proposed, notably in works like D-NeRF and subsequent methods. Dynamic NeRFs extend the original NeRF framework by incorporating time as an additional input dimension. This inclusion allows the model to capture both spatial geometry and temporal evolution, effectively learning a spatiotemporal radiance field. By doing so, dynamic NeRFs can reconstruct and render novel views of scenes that exhibit rigid body movements, articulated motions, and non-rigid deformations.

Recent developments in the field have introduced more robust dynamic models, such as RodyNeRF (Robust Dynamic Neural Radiance Fields). RodyNeRF focuses on addressing challenges posed by noisy or incomplete temporal data, leveraging advanced techniques to improve the robustness of dynamic scene reconstruction. This method incorporates data

augmentation and regularization strategies to handle variability in dynamic sequences, allowing for better performance in environments with occlusions or inconsistent temporal information. By improving temporal stability and reconstruction quality, robust dynamic models like RodyNeRF are particularly valuable for applications where temporal dynamics play a crucial role.

The key difference between generative NeRFs and dynamic NeRFs lies in their handling of temporal information. While generative NeRFs aim to produce diverse images of static scenes with variations in style or viewpoint, dynamic NeRFs are specifically designed to model scenes where the content itself changes over time. This capability is crucial for accurately representing and rendering dynamic phenomena, which static models cannot capture.

The importance of dynamic and robust NeRFs becomes particularly evident in medical imaging applications such as rotational angiography. Rotational angiography involves acquiring a series of X-ray images while rotating around the patient, resulting in a 3D visualization of blood vessels. The data obtained from this procedure contains significant dynamic and temporal attributes due to physiological movements like cardiac motion and blood flow. Traditional static NeRF models are insufficient for this type of data because they cannot account for the temporal changes inherent in the vascular structures.

By employing dynamic and robust NeRFs like D-NeRF and RodyNeRF, it becomes possible to model these temporal dynamics accurately. The inclusion of time as an input allows the model to capture the continuous motion of blood through the vessels and the pulsatile nature of the cardiovascular system. This leads to more accurate reconstructions of the vascular anatomy, which is essential for diagnostic purposes and surgical planning. Moreover, robust dynamic NeRFs can facilitate the visualization of blood flow patterns and the assessment of vascular anomalies, providing clinicians with valuable insights that are not accessible through static imaging techniques.

Dynmaic neural radiance fields represent a significant advancement over generative NeRFs by enabling the modeling and rendering of dynamic scenes with temporal consistency. Their application to rotational angiography demonstrates their potential to enhance medical imaging by accurately capturing the complex temporal dynamics of physiological

processes. For our experiments, we have tested both D-NeRF and RodyNeRF (Robust Dynamic Neural Radiance Fields) to capture the temporal dynamics inherent in angiographic data.

6.2 Implementation

In this section, we explore the application of D-NeRF and RoDyNeRF models. Our goal is to evaluate their capabilities and limitations when applied to multiple dynamic scenes and to investigate the feasibility of creating a generalizable generative model through model merging. We conducted extensive experiments using our dataset to assess these models and attempted to merge RoDyNeRF with a generative framework inspired by GRAF (Generative Radiance Fields) to overcome inherent limitations.

D-NeRF extends the original Neural Radiance Fields (NeRF) framework to handle dynamic scenes by introducing time as an additional input dimension. It models the radiance field as a continuous function of spatial coordinates, enabling the reconstruction of scenes that change over time from monocular video data. This approach allows for the synthesis of novel views of dynamic scenes by learning a representation that accounts for temporal variations.

The D-NeRF model utilizes a Multi-Layer Perceptron (MLP) that accepts 4D inputs comprising spatial coordinates and time, and outputs corresponding color and density values. To capture high-frequency details essential for realistic rendering, positional encoding is applied to both spatial and temporal dimensions. The training process involves minimizing the difference between the rendered images and the ground truth frames from the video, effectively teaching the model to represent the dynamic scene accurately.

For our experiments, we employed individual rotational scan sequences from our dataset, each capturing a dynamic scene from multiple viewpoints over time. The model was trained using standard settings as proposed in the original D-NeRF paper, with necessary adjustments to accommodate the specific characteristics of our dataset. Training was conducted on a per-scene basis, requiring the model to learn the dynamics of each scene individually.

In dynamic coronary artery imaging, accurately synchronizing the X-ray pulses with the cardiac cycle is essential for capturing high-quality, temporally coherent images. To achieve this, an ECG-based time-stamping system was implemented. The goal was to align the X-ray pulses with the heart's electrical activity by using data extracted from the DICOM files. This time-stamping process enables the synchronization of X-ray frames with specific phases of the cardiac cycle, improving the accuracy of dynamic imaging.

The electrocardiogram (ECG) signal provides vital information about the heart's electrical activity, particularly the R-peaks, which correspond to the contraction of the heart ventricles. By extracting the ECG curve data from the DICOM files, we were able to map the timing of the X-ray pulses to these R-peaks, thereby gating each X-ray frame to a specific point in the cardiac cycle. The timing of each X-ray pulse was calculated based on the pulse width and frame intervals.

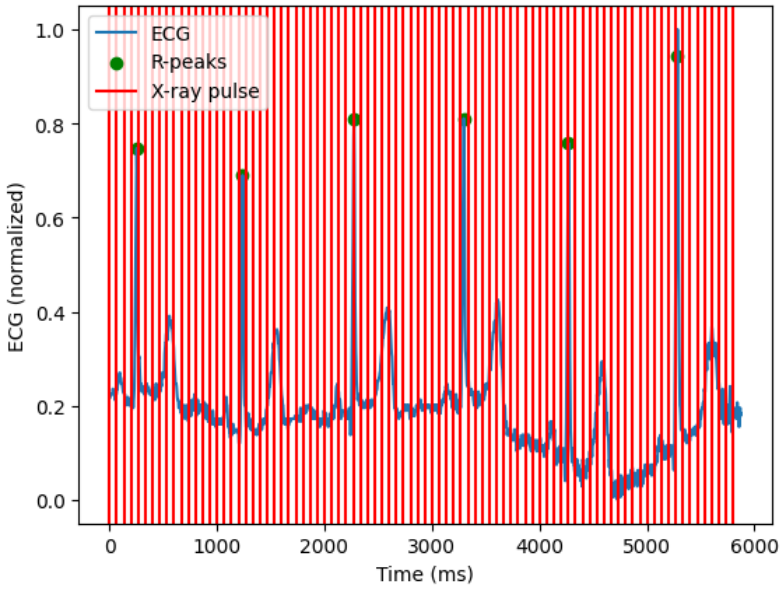


Figure 6-1 ECG Signals from DICOM File

Figure **6-1** illustrates the relationship between the X-ray pulses and the ECG signal. The ECG waveform (blue) is shown alongside vertical red lines that indicate the timing of the X-ray pulses, while the green markers denote the detected R-peaks. This visualization demonstrates how the X-ray pulses are aligned with specific cardiac events, enabling us to capture images at precise moments in the cardiac cycle. By ensuring that the X-ray frames

are synchronized with the heart's electrical activity, we can generate temporally consistent images that account for the dynamic movement of the coronary arteries.

To further refine the synchronization between the X-ray pulses and the heart's motion, we calculated the R-R interval percentage for each X-ray frame. The R-R interval is the time between consecutive R-peaks, representing one complete cardiac cycle. By determining the percentage of the R-R interval corresponding to each X-ray frame, we were able to map the frames to their respective positions within the cardiac cycle.

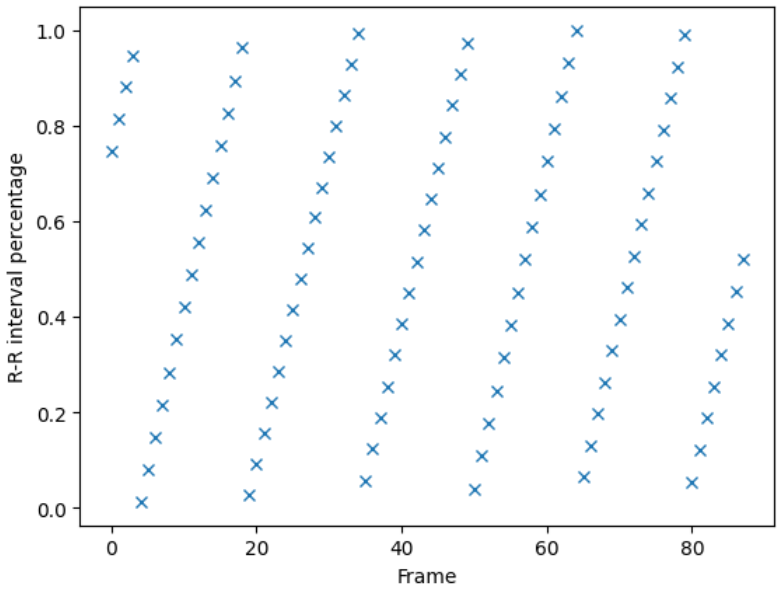


Figure 6-2 Frame wise R-R Intervals

Figure **6-2** shows the calculated R-R interval percentages for each frame, demonstrating the cyclical nature of the heart's motion. This mapping ensures that each X-ray frame is associated with a specific phase of the heart's contraction or relaxation, allowing for more accurate modeling of the coronary arteries during dynamic imaging. By interpolating between the R-R intervals and frame times, the model can represent the subtle changes in artery structure that occur throughout the cardiac cycle.

The ECG-based time-stamping system was integrated into the Dynamic NeRF framework, enabling the generation of 3D coronary artery models that account for both the spatial positioning of the X-ray frames and the temporal dynamics of the heart. Each X-ray frame was tagged with its corresponding R-R interval percentage, which allowed the Dynamic

NeRF model to interpolate between frames and generate temporally coherent reconstructions of the coronary arteries.

By aligning the frames with the cardiac cycle, this system improves the model's ability to capture real-time changes in the coronary arteries, such as the movement of blood flow and the shifting positions of artery walls. The integration of time-stamped X-ray data into the Dynamic NeRF framework represents a significant advancement in the accuracy and realism of dynamic coronary artery imaging.

D-NeRF demonstrated success in reconstructing high-fidelity representations of individual angiography scans. It effectively captured motion and temporal variations, producing realistic renderings of the input videos. However, a significant limitation became evident: the model was inherently scene-specific and required retraining for each new video sequence. This lack of generalization across different scenes highlighted the model's inability to function as a generative model capable of handling multiple dynamic inputs.

RoDyNeRF, or Robust Dynamic Neural Radiance Fields, builds upon D-NeRF by enhancing robustness to real-world challenges such as noise, occlusions, and incomplete data. It introduces mechanisms designed to handle imperfect data commonly encountered outside controlled environments, making it more suitable for practical applications where data quality cannot be guaranteed.

The RoDyNeRF model incorporates an enhanced architecture with additional layers and modules specifically designed to process incomplete and noisy inputs. It employs spatial and temporal regularization techniques to maintain consistency across frames, ensuring coherent temporal dynamics in the reconstructed scenes. Advanced optimization methods, including modified loss functions that account for data imperfections, are used to stabilize training and improve robustness.

We utilized the same individual video sequences from our dataset, Training involved modified loss functions as outlined in the RoDyNeRF framework, allowing the model to handle the introduced imperfections. This setup aimed to evaluate the model's ability to maintain reconstruction quality in less-than-ideal conditions.

RoDyNeRF demonstrated improved performance in handling noisy and incomplete data compared to D-NeRF. The model maintained high-quality reconstructions despite these challenges, effectively addressing some practical limitations of dynamic scene reconstruction. However, like D-NeRF, RoDyNeRF remained limited to single-scene modeling and required retraining for each new dynamic scene. This limitation underscored the inherent design of dynamic NeRFs as scene-specific models.

Our experiments with D-NeRF and RoDyNeRF highlighted a fundamental limitation: both models are inherently designed to handle only one dynamic scene from a specific video input. These models encode scene-specific details into the network's weights during training, making them incapable of generalizing to new, unseen scenes without retraining. This fixed mapping from input coordinates and time to color and density for a particular scene makes them unsuitable for applications requiring a general model capable of handling multiple dynamic scenes. Additionally, retraining for each new scene is computationally intensive and impractical for real-time or large-scale applications, posing significant challenges for scalability.

To overcome these limitations, we proposed merging RoDyNeRF with a generative model framework inspired by GRAF, as depicted in the merged architecture in Figure **6-3**. The architecture introduces a latent space representation that allows the model to learn a distribution over multiple scenes, enabling the generation of novel instances. In this merged model, RoDyNeRF’s robustness in handling dynamic scenes and noisy data is enhanced by GRAF's generative capabilities, which allow the model to generalize across different dynamic scenes.

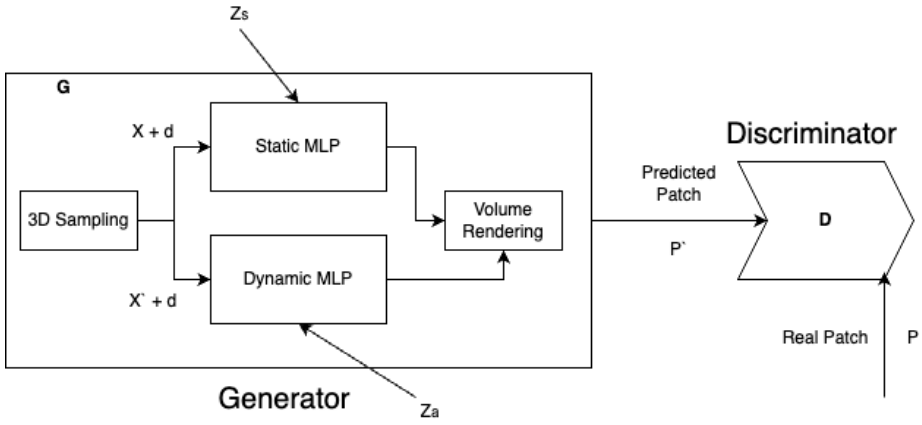


Figure 6-3 Combined Architecture (Dynamic + Generative NeRF)

Figure **6-3** shows how we incorporated a per-scene latent vector z_s, which captures the unique characteristics of each dynamic scene. This vector is sampled from a prior distribution and conditions the radiance field generation in the NeRF. As shown in the image, the generator architecture consists of 3D sampling, which takes spatial coordinates and temporal input (denoted as ($x + d$)) along with the latent vector (z_s). The model passes these inputs through two Multi-Layer Perceptrons (MLPs), one Static MLP and one Dynamic MLP, which together ensure both static and dynamic features of the scene are captured. The final output is processed by Volume Rendering to generate the 3D scene, which is then evaluated by the Discriminator against the real patch to distinguish between the predicted patch (P') and real patch (P).

We modified the NeRF network of RoDynRF to accept the latent vector (z_s) alongside the spatial coordinates and time, effectively conditioning the scene representation on both the physical input and the latent code. A mapping network projects the latent vector into a higher-dimensional space before conditioning the NeRF, similar to the techniques used in generative models to capture complex variations across different data instances. This approach allows the model to adapt to different scenes using a shared latent space rather than being constrained to a single scene.

For training, we used the same dataset which we used for Generative models. The training objective combined reconstruction loss and adversarial loss to encourage the model to generate realistic and diverse dynamic scenes. The Discriminator, as shown in Figure **6-3**, helps to ensure that the predicted scene patches closely resemble the real scene patches, refining the model's output and improving its realism.

This approach ensured the model maintained temporal coherence within each scene while learning across multiple scenes. The merged architecture provides a more scalable and generalizable framework for dynamic scene reconstruction, as it can now learn from and generate multiple dynamic scenes without requiring retraining for each one.

6.3 Analysis and Conclusion

Despite the inability of the proposed model to produce coherent visual outputs, quantitative analysis using FID scores sheds light on the challenges encountered. The FID scores for the generated outputs consistently exceeded 350 across all test cases, significantly higher than the threshold for acceptable generative quality in comparable tasks. These results highlight the model's failure to capture meaningful distributions and corroborate the observed noise in the generated images. By comparison, the standalone RoDyNeRF model, which was not integrated with a generative framework, produced visually coherent results without this degradation in quality.

To contextualize the performance of the merged architecture, comparisons were made against the individual components: the standalone RoDyNeRF and the generative model framework. RoDyNeRF demonstrated effective novel view synthesis within single-scene settings but lacked the versatility for generalization across dynamic scenes. The generative framework alone was able to produce diverse outputs (as explained in Chapter 4 & Chapter 5) but lacked the temporal and spatial consistency required for dynamic scene modeling. The merged architecture, unfortunately, failed to leverage the strengths of either component. Instead, it amplified their limitations, resulting in outputs with poor quality and significantly higher FID scores.

The integration of RoDyNeRF with a generative framework introduced considerable challenges, particularly in representing dynamic scene variations within the latent space. The merged model failed to disentangle meaningful scene-specific features, leading to mode collapse and incoherent outputs. Despite experimenting with adjustments to the loss functions, including varying the weights of adversarial and reconstruction terms, the model's performance did not improve. This indicates a deeper issue in the architectural design, which necessitates a more structured approach to combining these frameworks.

Resource constraints further compounded these challenges. The memory-intensive architecture restricted training to small batch sizes and low resolutions, directly impacting the model's capacity to learn generalizable representations. Additionally, the complexity of incorporating adversarial loss into the framework resulted in significant training instability. The competing objectives of generating novel views and reconstructing input scenes created conflicting gradients, preventing effective optimization. Multiple attempts to stabilize training, including adjusting learning rates and experimenting with alternative adversarial loss functions, failed to resolve these issues, and the model struggled to converge to meaningful solutions.

The high FID scores and lack of coherence in outputs reflect the inherent difficulty of integrating generative frameworks with Dynamic NeRF architectures. Existing architectures are inherently limited in their ability to generalize across dynamic scenes, and their latent spaces lack the capacity to represent diverse scene variations effectively. The complexity introduced by adversarial training, coupled with the constraints of limited computational resources, further exacerbated these challenges. These limitations emphasize the need for innovative architectural solutions and more robust training strategies.

While the results did not meet expectations, the insights gained from this work provide valuable guidance for addressing these challenges in future research. Potential directions include incorporating mechanisms to disentangle scene-specific features within the latent space, developing optimized training frameworks such as curriculum learning or latent space pretraining, and designing lightweight architectures to reduce computational overhead. Additionally, expanding the dataset to include more diverse dynamic scenes would enhance the model's capacity for generalization.

Although this attempt fell short of achieving its goals, it highlights critical areas for improvement and lays the groundwork for future advancements in dynamic scene novel view synthesis. By addressing the challenges identified in this work, future research can contribute significantly to developing generalizable and efficient Dynamic NeRF architectures.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

This research set out to address a critical challenge in coronary artery imaging: the reliance on multiple angiograms and contrast dye, which poses risks to patients when used in excess. Our primary goal was to develop a method that could reduce the number of angiograms by generating high-quality novel views of coronary arteries from existing rotational angiography data. The vision was to create a generalizable model that, once trained, could be applied across different patients without requiring retraining, thus saving time and reducing the need for additional imaging procedures. This would enable clinicians to obtain specific imaging angles and efficiently capture regions of interest, all while minimizing the patient's exposure to contrast dye.

Existing approaches in the literature typically focus on building 3D reconstructions for individual patients, often requiring extensive retraining and model adjustments for each new case. These methods are impractical for real-time clinical applications due to the computational complexity and time required. In contrast, this research took a novel approach by exploring the potential of neural radiance fields (NeRFs) to generate generalizable, high-resolution coronary artery images that could work across different patients. The key focus was to determine which NeRF architecture MedNeRF, StyleNeRF, or dynamic NeRF could produce the most clinically relevant and accurate images.

Among the architectures explored, MedNeRF showed early promise, particularly in generating images of the right coronary artery (RCA), a structure that is easier to model due to its more consistent geometry. Despite limited data and training iterations, MedNeRF managed to produce accurate images, demonstrating its capability in less computationally intensive scenarios. However, the best performance was achieved with StyleNeRF, which consistently produced high-resolution, detailed images of coronary arteries, with a Fréchet Inception Distance (FID) score of 75.32, indicating a strong potential for clinical use. The encoder-based inference in StyleNeRF further enhanced its practicality, enabling efficient

novel view generation with minimal manual input, a key factor in reducing procedural time in clinical settings.

In tackling the dynamic nature of coronary artery imaging, where heart motion and contrast dye flow are significant factors, we experimented with dynamic NeRF. Despite incorporating time-stamped data, such as electrocardiogram (ECG) signals, the dynamic NeRF did not produce satisfactory results in handling real-time motion. Integrating static and dynamic models proved to be complex, and the results highlighted the need for further advancements in this area to address the full range of challenges posed by dynamic scenes.

Overall, this research makes a significant contribution by demonstrating the potential of NeRF architectures, particularly StyleNeRF, in generating high-resolution, accurate novel views of coronary arteries. However, challenges remain in optimizing the model's performance for dynamic imaging and reducing its computational demands. Future work should focus on addressing these limitations, with particular attention to improving real-time capabilities, handling dynamic motion more effectively, and exploring ways to further streamline the computational requirements for broader clinical adoption.

7.2 Future Scope

Looking ahead, there are several directions for future research that could help overcome the limitations we encountered in this study. First, we plan to investigate more advanced generative models, such as latent diffusion models[54], [55], [56] and Gaussian splatting models[57], [58], [59]. These models are emerging as state-of-the-art in image synthesis and could potentially handle the dynamic aspects of coronary artery imaging more effectively than the NeRF-based approaches we have tested so far. By exploring these advanced models, we aim to improve the ability to capture the movement of the heart and contrast dye, which is essential for accurate real-time imaging in clinical practice.

Additionally, transitioning this research from a controlled experimental environment into real-world clinical use will require significant optimization, particularly in reducing the processing time required to generate novel views. Currently, the need for high-performance GPU setups limits the practical application of these methods in clinical settings, where processing must be both fast and accessible. We will work on improving the computational

efficiency of the models, exploring methods to reduce the model's reliance on large, expensive GPU clusters while maintaining the quality of the generated images.

Moreover, as part of the future work, we aim to reach key clinical benchmarks that will allow us to begin testing our system in a clinical environment. This involves collaborating with clinicians to integrate the tool into their workflow, ensuring that it meets the practical needs of real-time coronary imaging. One promising avenue is the development of a software tool that can generate novel views of coronary arteries with minimal input from the user, making it more accessible for use in a busy clinical setting. By incorporating feedback from clinical experts, we hope to create a tool that not only enhances imaging accuracy but also reduces the burden on both clinicians and patients by minimizing the need for repeated angiograms.

Finally, we plan to address the limitations of the dynamic NeRF model by refining its ability to capture the heart's motion. This could involve combining different model architectures or developing a hybrid approach that balances the strengths of both static and dynamic models. Successfully overcoming these challenges would significantly improve the utility of our model, making it a viable option for real-time, dynamic coronary artery imaging.

In summary, while our work has made major strides in improving coronary artery imaging, the journey toward full clinical integration is ongoing. With continued development, we believe that the methods presented in this research have the potential to become a powerful tool in the diagnosis and treatment of coronary artery disease.

Chapter 8

References

[1] Writing Committee Members *et al.*, "2021 ACC/AHA/SCAI Guideline for Coronary Artery Revascularization: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines," *J. Am. Coll. Cardiol.*, vol. 79, no. 2, pp. e21–e129, Jan. 2022, doi: 10.1016/j.jacc.2021.09.006.

[2] A. Timmis *et al.*, "European Society of Cardiology: Cardiovascular Disease Statistics 2019," *Eur. Heart J.*, vol. 41, no. 1, pp. 12–85, Jan. 2020, doi: 10.1093/eurheartj/ehz859.

[3] Y. Yang *et al.*, "Contrast-induced acute kidney injury and adverse clinical outcomes risk in acute coronary syndrome patients undergoing percutaneous coronary intervention: a meta-analysis," *BMC Nephrol.*, vol. 19, no. 1, p. 374, Dec. 2018, doi: 10.1186/s12882-018-1161-5.

[4] V. Pistolesi *et al.*, "Contrast medium induced acute kidney injury: a narrative review," *J. Nephrol.*, vol. 31, no. 6, pp. 797–812, Dec. 2018, doi: 10.1007/s40620-018-0498-y.

[5] R. S. O. Giuberti *et al.*, "A randomized trial comparing dual axis rotational versus conventional coronary angiography in a population with a high prevalence of coronary artery disease," *J. Intervent. Cardiol.*, vol. 27, no. 5, pp. 456–464, Oct. 2014, doi: 10.1111/joic.12148.

[6] N. E. Green, S.-Y. J. Chen, J. C. Messenger, B. M. Groves, and J. D. Carroll, "Three-dimensional vascular angiography," *Curr. Probl. Cardiol.*, vol. 29, no. 3, pp. 104–142, Mar. 2004, doi: 10.1016/j.cpcardiol.2004.01.002.

[7] J. H. Lee, D. Han, I. Danad, B. ó Hartaigh, F. Y. Lin, and J. K. Min, "Multimodality Imaging in Coronary Artery Disease: Focus on Computed Tomography," *J. Cardiovasc. Ultrasound*, vol. 24, no. 1, pp. 7–17, Mar. 2016, doi: 10.4250/jcu.2016.24.1.7.

[8] P. J. Scanlon *et al.*, "ACC/AHA guidelines for coronary angiography. A report of the American College of Cardiology/American Heart Association Task Force on practice

guidelines (Committee on Coronary Angiography). Developed in collaboration with the Society for Cardiac Angiography and Interventions," *J. Am. Coll. Cardiol.*, vol. 33, no. 6, pp. 1756–1824, May 1999, doi: 10.1016/s0735-1097(99)00126-6.

[9] P. A. Hudson *et al.*, "A Novel Dual-Axis Rotational Coronary Angiography Evaluation of Coronary Artery Disease — Case Presentation and Review," *Clin. Cardiol.*, vol. 33, no. 7, pp. E16–E19, 2010, doi: 10.1002/clc.20651.

[10] W. Shen *et al.*, "Impact of respiratory motion artifact on coronary image quality of one beat coronary CT angiography," *J. X-Ray Sci. Technol.*, vol. 29, no. 2, pp. 287–296, 2021, doi: 10.3233/XST-200812.

[11] B. Balaney *et al.*, "Improved visualization of the coronary arteries using motion correction during vasodilator stress CT myocardial perfusion imaging," *Eur. J. Radiol.*, vol. 114, pp. 1–5, May 2019, doi: 10.1016/j.ejrad.2019.02.010.

[12] J. Narula *et al.*, "SCCT 2021 Expert Consensus Document on Coronary Computed Tomographic Angiography: A Report of the Society of Cardiovascular Computed Tomography," *J. Cardiovasc. Comput. Tomogr.*, vol. 15, no. 3, pp. 192–217, 2021, doi: 10.1016/j.jcct.2020.11.001.

[13] K. Nieman and F. Cademartiri, "CT Angiography for the Detection of Coronary Artery Stenosis," in *CT of the Heart: Principles and Applications*, U. J. Schoepf, Ed., Totowa, NJ: Humana Press, 2005, pp. 321–332. doi: 10.1385/1-59259-818-8:321.

[14] C. A. Karlo, S. Leschka, P. Stolzmann, N. Glaser-Gallion, S. Wildermuth, and H. Alkadhi, "A systematic approach for analysis, interpretation, and reporting of coronary CTA studies," *Insights Imaging*, vol. 3, no. 3, Art. no. 3, Jun. 2012, doi: 10.1007/s13244-012-0167-y.

[15] A. Farzaneh-Far, M. Steigner, and R. Y. Kwong, "Applications and limitations of cardiac computed tomography in the evaluation of coronary artery disease," *Coron. Artery Dis.*, vol. 24, no. 7, p. 606, Nov. 2013, doi: 10.1097/MCA.0000000000000027.

[16] A. Chiribiri, R. M. Botnar, and E. Nagel, "Magnetic Resonance Coronary Angiography: Where Are We Today?," *Curr. Cardiol. Rep.*, vol. 15, no. 2, p. 328, Jan. 2013, doi: 10.1007/s11886-012-0328-0.

[17] A. M. Vukicevic, S. Çimen, N. Jagic, G. Jovicic, A. F. Frangi, and N. Filipovic, "Three-dimensional reconstruction and NURBS-based structured meshing of

coronary arteries from the conventional X-ray angiography projection images," *Sci. Rep.*, vol. 8, no. 1, p. 1711, Jan. 2018, doi: 10.1038/s41598-018-19440-9.

[18] S. V. Raman *et al.*, "Rotational X-ray coronary angiography," *Catheter. Cardiovasc. Interv.*, vol. 63, no. 2, pp. 201–207, 2004, doi: 10.1002/ccd.20130.

[19] A. J. Klein *et al.*, "Safety and efficacy of dual-axis rotational coronary angiography vs. standard coronary angiography," *Catheter. Cardiovasc. Interv. Off. J. Soc. Card. Angiogr. Interv.*, vol. 77, no. 6, pp. 820–827, May 2011, doi: 10.1002/ccd.22804.

[20] A. J. P. Klein and J. A. Garcia, "Rotational coronary angiography," *Cardiol. Clin.*, vol. 27, no. 3, pp. 395–405, Aug. 2009, doi: 10.1016/j.ccl.2009.03.002.

[21] J. A. Garcia *et al.*, "Determination of optimal viewing regions for X-ray coronary angiography based on a quantitative analysis of 3D reconstructed models," *Int. J. Cardiovasc. Imaging*, vol. 25, no. 5, pp. 455–462, Jun. 2009, doi: 10.1007/s10554-008-9402-5.

[22] A. S. Panayides *et al.*, "AI in Medical Imaging Informatics: Current Challenges and Future Directions," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 1837–1857, Jul. 2020, doi: 10.1109/JBHI.2020.2991043.

[23] X. Ou *et al.*, "Recent Development in X-Ray Imaging Technology: Future and Challenges," *Research*, vol. 2021, p. 9892152, Dec. 2021, doi: 10.34133/2021/9892152.

[24] J. Hsieh and T. Flohr, "Computed tomography recent history and future perspectives," *J. Med. Imaging*, vol. 8, no. 5, p. 052109, Sep. 2021, doi: 10.1117/1.JMI.8.5.052109.

[25] S. Saleh Elessawy, M. Abd El-Ghaffar Borg, M. Abd El-Salam Mohamed, G. Elsayed Elhawary, and E. Mohamed Abd El-Salam, "The role of MRI in the evaluation of muscle diseases," *Egypt. J. Radiol. Nucl. Med.*, vol. 44, no. 3, pp. 607–615, Sep. 2013, doi: 10.1016/j.ejrnm.2013.04.002.

[26] M. Sarmah, A. Neelima, and H. R. Singh, "Survey of methods and principles in three-dimensional reconstruction from two-dimensional medical images," *Vis. Comput. Ind. Biomed. Art*, vol. 6, no. 1, p. 15, Jul. 2023, doi: 10.1186/s42492-023-00142-7.

[27] B. Preim and C. Botha, "Chapter 6 - Surface Rendering," in *Visual Computing for Medicine (Second Edition)*, B. Preim and C. Botha, Eds., Boston: Morgan Kaufmann, 2014, pp. 229–267. doi: 10.1016/B978-0-12-415873-3.00006-7.

[28] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges,” *J. Digit. Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019, doi: 10.1007/s10278-019-00227-x.

[29] S. K. Zhou *et al.*, “A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises,” *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021, doi: 10.1109/JPROC.2021.3054390.

[30] 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,” Aug. 03, 2020, *arXiv*: arXiv:2003.08934. doi: 10.48550/arXiv.2003.08934.

[31] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” arXiv.org. Accessed: Sep. 12, 2024. [Online]. Available: https://arxiv.org/abs/1406.2661v1

[32] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” Dec. 06, 2017, *arXiv*: arXiv:1701.07875. doi: 10.48550/arXiv.1701.07875.

[33] P. Costa *et al.*, “End-to-End Adversarial Retinal Image Synthesis,” *IEEE Trans. Med. Imaging*, vol. 37, no. 3, pp. 781–791, Mar. 2018, doi: 10.1109/TMI.2017.2759102.

[34] T. Zhou, S. Tulsiani, W. Sun, J. Malik, and A. A. Efros, “View Synthesis by Appearance Flow,” arXiv.org. Accessed: Aug. 17, 2024. [Online]. Available: https://arxiv.org/abs/1605.03557v3

[35] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, “GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis,” Mar. 30, 2021, *arXiv*: arXiv:2007.02442. Accessed: Aug. 21, 2024. [Online]. Available: http://arxiv.org/abs/2007.02442

[36] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, “pixelNeRF: Neural Radiance Fields from One or Few Images,” May 30, 2021, *arXiv*: arXiv:2012.02190. doi: 10.48550/arXiv.2012.02190.

[37] E. R. Chan, M. Monteiro, P. Kellnhofer, J. Wu, and G. Wetzstein, “pi-GAN: Periodic Implicit Generative Adversarial Networks for 3D-Aware Image Synthesis,” Apr. 05, 2021, *arXiv*: arXiv:2012.00926. Accessed: Aug. 21, 2024. [Online]. Available: http://arxiv.org/abs/2012.00926

[38] M. Niemeyer and A. Geiger, “GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields,” in *2021 IEEE/CVF Conference on Computer*

Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, Jun. 2021, pp. 11448–11459. doi: 10.1109/CVPR46437.2021.01129.

[39] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," Mar. 23, 2020, *arXiv*: arXiv:1912.04958. doi: 10.48550/arXiv.1912.04958.

[40] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-NeRF: Neural Radiance Fields for Dynamic Scenes," Nov. 27, 2020, *arXiv*: arXiv:2011.13961. doi: 10.48550/arXiv.2011.13961.

[41] Y.-L. Liu *et al.*, "Robust Dynamic Radiance Fields," arXiv.org. Accessed: Jul. 18, 2024. [Online]. Available: https://arxiv.org/abs/2301.02239v2

[42] A. Corona-Figueroa, J. Frawley, S. Bond-Taylor, S. Bethapudi, H. P. H. Shum, and C. G. Willcocks, "MedNeRF: Medical Neural Radiance Fields for Reconstructing 3D-aware CT-Projections from a Single X-ray," Apr. 08, 2022, *arXiv*: arXiv:2202.01020. doi: 10.48550/arXiv.2202.01020.

[43] D. Rückert, Y. Wang, R. Li, R. Idoughi, and W. Heidrich, "NeAT: neural adaptive tomography," *ACM Trans. Graph.*, vol. 41, no. 4, p. 55:1-55:13, Jul. 2022, doi: 10.1145/3528223.3530121.

[44] Y. Fang *et al.*, "SNAF: Sparse-view CBCT Reconstruction with Neural Attenuation Fields," Nov. 30, 2022, *arXiv*: arXiv:2211.17048. doi: 10.48550/arXiv.2211.17048.

[45] K. Maas, N. Pezzotti, A. Vermeer, D. Ruijters, and A. Vilanova, "NeRF for 3D Reconstruction from X-ray Angiography: Possibilities and Limitations," p. 12 pages, 2023, doi: 10.2312/VCBM.20231210.

[46] C. K. Eschen *et al.*, "Classification of Left and Right Coronary Arteries in Coronary Angiographies Using Deep Learning," *Electronics*, vol. 11, no. 13, Art. no. 13, Jan. 2022, doi: 10.3390/electronics11132087.

[47] Z. Ding, J. C. Gore, and A. W. Anderson, "Reduction of noise in diffusion tensor images using anisotropic smoothing," *Magn. Reson. Med.*, vol. 53, no. 2, pp. 485–490, Feb. 2005, doi: 10.1002/mrm.20339.

[48] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, USA: Academic Press Professional, Inc., 1994, pp. 474–485.

[49] S. Lefkimmiatis, “Universal Denoising Networks : A Novel CNN Architecture for Image Denoising,” Mar. 24, 2018, *arXiv*: arXiv:1711.07807. doi: 10.48550/arXiv.1711.07807.

[50] V. Bhateja, A. Yadav, D. Singh, and B. K. Chauhan, “Multi-scale Retinex with Chromacity Preservation (MSRCP)-Based Contrast Enhancement of Microscopy Images,” in *Smart Intelligent Computing and Applications, Volume 2*, S. C. Satapathy, V. Bhateja, M. N. Favorskaya, and T. Adilakshmi, Eds., in Smart Innovation, Systems and Technologies. Singapore: Springer Nature, 2022, pp. 313–321. doi: 10.1007/978-981-16-9705-0_31.

[51] S. Parthasarathy and P. Sankaran, “An automated multi Scale Retinex with Color Restoration for image enhancement,” in *2012 National Conference on Communications (NCC)*, Feb. 2012, pp. 1–5. doi: 10.1109/NCC.2012.6176791.

[52] B. Fadida-Specktor, “Preprocessing Prediction of Advanced Algorithms for Medical Imaging,” *J. Digit. Imaging*, vol. 31, no. 1, pp. 42–50, Feb. 2018, doi: 10.1007/s10278-017-9999-9.

[53] K. Zhang, G. Riegler, N. Snavely, and V. Koltun, “NeRF++: Analyzing and Improving Neural Radiance Fields,” Oct. 21, 2020, *arXiv*: arXiv:2010.07492. doi: 10.48550/arXiv.2010.07492.

[54] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” Dec. 16, 2020, *arXiv*: arXiv:2006.11239. doi: 10.48550/arXiv.2006.11239.

[55] Y. Zhong *et al.*, “MedDiffusion: Boosting Health Risk Prediction via Diffusion-based Data Augmentation,” Oct. 05, 2023, *arXiv*: arXiv:2310.02520. doi: 10.48550/arXiv.2310.02520.

[56] R. Shi *et al.*, “Zero123++: a Single Image to Consistent Multi-view Diffusion Base Model,” Oct. 23, 2023, *arXiv*: arXiv:2310.15110. doi: 10.48550/arXiv.2310.15110.

[57] “3D Gaussian Splatting for Real-Time Radiance Field Rendering.” Accessed: Sep. 19, 2024. [Online]. Available: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[58] W.-H. Chu, L. Ke, and K. Fragkiadaki, “DreamScene4D: Dynamic Multi-Object Scene Generation from Monocular Videos,” May 23, 2024, *arXiv*: arXiv:2405.02280. Accessed: Sep. 19, 2024. [Online]. Available: http://arxiv.org/abs/2405.02280

[59] P. Ungermann, A. Ettenhofer, M. Nießner, and B. Roessle, “Robust 3D Gaussian Splatting for Novel View Synthesis in Presence of Distractors,” Aug. 21, 2024, *arXiv*: arXiv:2408.11697. doi: 10.48550/arXiv.2408.11697.