UNIVERSITY OF CALIFORNIA
SANTA CRUZ

LEARNED GRIDLESS REPRESENTATIONS OF CONE BEAM COMPUTED TOMOGRAPHY SCANS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

by

Emmanouil Nikolakakis

August 2026

The Dissertation of Emmanouil Nikolakakis
is approved:

Prof. J. Eshraghian, Chair

Prof. Razvan Marinescu

Prof. Mircea Teodorescu

Prof. Roberto Manduchi

Donald Smith
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Learned Gridless Representations of Cone Beam Computed Tomography Scans

by

Emmanouil Nikolakakis

Medical image representation has long been dominated by voxel-grid matrices. While their inherent structure and order work efficiently for various linear transformations and provide a seamless visualization method on monitors, they fail to preserve the topology of the scan and to encode sparse information in a memory-efficient way.

The recent emergence of machine learning-based **continuous coordinate-based scene representations** such as neural radiance fields and Gaussian splatting has provided alternative representation techniques. These approaches overfit the weights of a model by iterative differentiable rendering and have been shown to be more compact than grid representations. Moreover, they are then able to perform novel view synthesis from any given camera pose.

Off-grid representations translate directly to Cone Beam Computed Tomography sparse-view acquisitions, where streaking and quantum noise artifacts are dominant. Using differentiable rendering, a continuous representation is achieved, with interpolation providing a path to recover some of the lost signal.

In this dissertation, we apply a variety of methodologies, including Gaussian splatting, implicit occupancy fields, and Neural Attenuation Fields regularized with an anatomic prior, to Cone Beam Computed Tomography reconstruction, and evaluate their performance across a range of anatomic datasets. Our models show that learned gridless representations achieve substantial memory reduction, recover signal under extreme view sparsity, and preserve scene topology.

To my family, my friends, my girl, my colleagues, my advisors...all of those who supported me always and unconditionally.

Acknowledgments

Choosing to leave my job in France and cross the Atlantic to do a PhD in Santa Cruz will go down as one of the most important decisions of my life. Being at this point typing my thesis seems unreal, and yet, the past five years have been exactly what I hoped for. And I owe it to all the people who have been by my side and encouraged me at all points.

I would like to thank all the professors who have helped me develop academically over the past five years. First and foremost, Razvan, who trusted me from the very start and guided me as I joined his group. A lot of the ideas that we had since we first met and had a few discussions over coffee made it to medical imaging conferences and into my final dissertation. I will always be grateful for the motivation and the energy he sparked within me, in particular during periods of stress and frustration. Next, I'd like to thank Julie Digne from CNRS-Lyon, with whom I worked on two projects and whose lab I visited during my fourth year. I learned an incredible amount from her and I am hoping to collaborate again with her in the future. Moreover, I want to thank Shiva Abbaszadeh, with whom I worked the first few years of my PhD. It was under her guidance that I transitioned from a full-time developer to a researcher. Finally, I want to thank professors Jason Eshraghian, Mircea Teodorescu, and Roberto Manduchi for agreeing to read and validate my dissertation and providing very useful feedback following my submission to candidacy.

My family has always walked the academic path. After my mother, father, and sister, I am now also about to complete my PhD. Yet, I would not have made it here without any of them. Thank you Γιάννη, Σουζάνα, Φία, I am honored to be a member of our family. I also want to thank my brother-in-law, professor Δεμερτζής for all the discussions we had and the guidance he provided me with. Finally, I want to thank my niece, Μελίνα, one of the most beautiful people to enter my life in the past few years.

I started the PhD thinking that I would focus on work and not distract myself with an active social life and a large friend group. Never have I ever been so wrong. Thank you, Daniel, Αχιλλέα, Απόστολε, McKenna, Hari, Juanita, Χρήστο, Βίκυ, Ian, Mugen, Aditya, Ehsan, Amin, Zack, and everyone from the "Crossover" chat, the "Cruisin' in Santa Cruz" chat, and the Banana Lab. You all know who you are, and you made my life brighter every single day. Of course, I can't forget my friends in Greece who also continuously supported me. Μάνο, Φοίβο, Κωνσταντή, Θεόφιλε, Γιώργο, σας ευχαριστώ μέσα από την καρδιά μου. To my childhood friend...Αλέκο, I will never forget you.

Lastly, I want to thank Heena. Everything has been so much better and easier from the moment I met you. Thank you for making me believe in myself more than ever before. I am the luckiest to have you.

List of Abbreviations

3DGS 3D Gaussian Splatting

ADAM Adaptive Moment Estimation

ASD-POCS Adaptive Steepest Descent – Projection Onto Convex Sets

CBCT Cone Beam Computed Tomography

CGLS Conjugate Gradient Least Squares

CT Computed Tomography

DAS Data Acquisition System

DICOM Digital Imaging and Communications in Medicine

DL Deep Learning

DRR Digitally Reconstructed Radiograph

FBP Filtered BackProjection

FDK Feldkamp–Davis–Kress

FOV Field of View

HU Hounsfield Units

INR Implicit Neural Representation

ISR Implicit Scene Representation

LAC Linear Attenuation Coefficient

LPIPS Learned Perceptual Image Patch Similarity

MC Marching Cubes

ML Machine Learning

MLP Multilayer Perceptron

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

NAF Neural Attenuation Fields

NeRF Neural Radiance Fields

NIFTI Neuroimaging Informatics Technology Initiative

NVS Novel View Synthesis

PET Positron Emission Tomography

PPMI Parkinson's Progression Markers Initiative

PSNR Peak Signal-to-Noise Ratio

SART Simultaneous Algebraic Reconstruction Technique

SDF Signed Distance Field

SH Spherical Harmonics

SSIM Structural Similarity Index Measure

TV Total Variation

UDF Unsigned Distance Field

VGG Visual Geometry Group

Chapter 1

Introduction

1.1 The Necessity for fast sparse signal reconstructions in medical imaging

This dissertation is focused on the problem of accurately reconstructing **Cone Beam Computed Tomography (CBCT)** scans resulting from a low-dose acquisition. We will now discuss the motivation and emerging works that resulted in learned gridless representations being implemented as a reconstruction and representation methodology. Moreover, this chapter also presents a brief overview of the history and significance of medical imaging, the introduction of CBCT, and the reason behind low-dose acquisition protocols. Finally, the discussion turns to existing work that demonstrated the suitability of learned gridless representations for CBCT 3D scene reconstruction and representation from sparse datasets to a denoised and continuous 3D scan.

1.1.1 The importance of Medical Imaging

Information is the key for a clinician to provide a timely and accurate diagnosis for a patient. Throughout human history, doctors evaluated symptoms and physical exams, but everything changed in 1895 when X-rays were introduced by Wilhelm Röntgen [Tub96]. The ability to scan the human body was revolutionary and only the beginning of an array of modalities within medical imaging, with Ultrasound, Positron Emission Tomography (PET), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) being just a few of the most popular scanners. Information captured can be split into two main categories, the first being anatomical (CT, MRI) and the second behavioral (PET). In anatomical imaging, the scanner reconstructs the internal structure of the patient's body, while in behavioral imaging, the scanner captures functional signals such as blood flow or the synapses that are firing in the brain. The variety of options and the ability to combine different types of imaging that resulted in hybrid scans have since greatly improved the diagnostic quality and the treatment effectiveness that patients receive.

To quantify the quality of life improvement since the deployment of scanners, we can investigate the impact on some of the most deadly or common diseases. An ischemic stroke can result in 32,000 dead brain cells each second [Sav06]. With current CT technology, surveys have shown that 64.8% of patients receive a CT scan within 25 minutes of their arrival at the hospital [PGZ+26]. Breast cancer, the most diagnosed type of cancer in the world, has dropped by 40 % in the U.S. according to the American Association for Cancer Research, resulting in 490,000 averted deaths [Ame24]. Over 6.7 million Americans have heart failure [BAA+25], while studies have shown that the mortality of patients with advanced heart failure drops from 88.8 % to 71.8 % when following echocardiography-guided therapies [SSP+22]. The Alzheimer's Association has stated that 6.9 million Americans have Alzheimer's disease, with the number

expected to double by 2060. Amyloid PET has demonstrated the ability to provide an accurate diagnosis to help with early treatment, with research showing that when not used, misdiagnosis increases by 30% [SBG+26]. Finally, according to the World Health Organization, COVID-19 has caused over 7 million deaths since its appearance. Even so, early research studies employing CT showed the scanner's ability to diagnose and quantify the severity of the disease by identifying and analyzing relevant biomarkers [FAAF+21]. All mentioned achievements highlight the importance of fast and accurate medical image reconstructions.

1.1.2 Cone Beam Computed Tomography and low-dose acquisitions

Computed tomography is the most commonly used imaging modality. Its reduced cost of operation, wide availability, and short scan duration make it a popular choice. The value measured is the Linear Attenuation Coefficient of the material in the field of view. This value is acquired due to the original intensity of the X-ray emitted from the X-ray source being reduced as it traverses through the patient's body as a result of the photoelectric effect. The final intensity or energy of the emitted X-ray is then measured by the detector panel. The acquisition of values in CT is done in the **sinogram** domain, which will be explained in detail further in the dissertation. The transformation of information between the image space and the sinogram space is performed using the Radon and Inverse Radon transforms.

A conventional CT acquisition rotates the gantry through a large number of angular positions K with a small angular increment $\Delta\theta$ between consecutive acquisitions, so that the source S and detector panel d sample the patient densely and near-uniformly around the field of view. Reducing the radiation dose administered to the patient typically means reducing K, which increases $\Delta\theta$ and leaves large angular gaps unsampled, as

illustrated in Figure 1.1. This sparse-view setting is the central challenge that motivates the reconstruction methodologies presented in this dissertation.

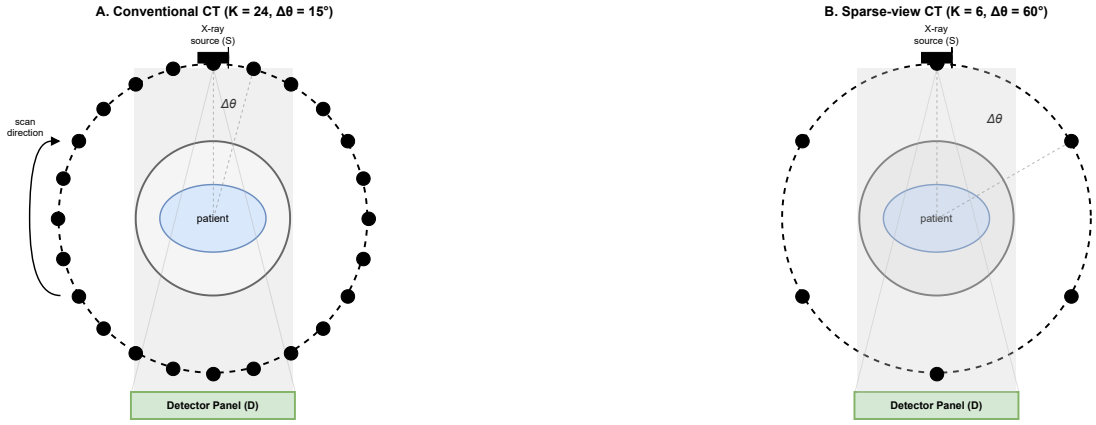


Figure 1.1: Comparison of a conventional, densely sampled CT acquisition (A) against a sparse-view acquisition (B). Notation, defined in full in Table 2.2 (Chapter 2): the X-ray source S rotates around the patient at gantry angle θ and radiates toward the detector panel d, capturing K radiographs P_θ at an angular increment $\Delta\theta = 360°/K$. Reducing K (and thus increasing $\Delta\theta$) reduces the radiation dose but leaves the scanned volume under-sampled.

It is well established that ionizing radiation administered to patients during CBCT scans can be detrimental to overall health since it interferes with cell DNA by breaking chemical bonds [PMT+16b]. Particularly vulnerable populations include the elderly, children, and pregnant women. Therefore, there are constant efforts to reduce the total amount of X-ray radiation per scan, achieved in one of two ways: 1) reducing the total amount of X-ray beam intensity emitted in each projection, or 2) reducing K while preserving the X-ray beam intensity per projection, i.e., the sparse-view acquisition described above.

Both strategies introduce distinct advantages and disadvantages. In the case of sparse-view acquisitions, the total amount of projections is reduced; however, the visual quality and resolution of each projection is preserved. This results in an ill-posed acquisition where not all regions within the FOV are covered equally, introducing the

geometric streaking artifacts illustrated in Figure 1.1. On the other hand, lowering the photon count per projection has its own characteristic artifacts. While ensuring that the information acquired about the 3D scene is uniform, fewer acquired photons result in quantum or Poisson noise [ZGS14]. Therefore, sparse-view acquisition is an ill-posed problem that would benefit from implicit interpolation, while low-dose per projection protocols are more vulnerable to low signal-to-noise ratios and require Poisson-noise-specific denoising operations.

This dissertation will be focused on CBCT, introduced originally in 1996 in Europe [SGCB+21] as an alternative to the traditional CT. This methodology enhances the contrast of bone and dental structures while reducing the contrast for tissues and blood vessels. This makes it particularly useful for imaging skeletal damage across the body. Overall, it is a shorter scan – only lasting 10–20 seconds – with a smaller total radiation dose administered to the patient [PMT+16b, EHR07, HML+14]. It is a cheaper option while also being less prone to motion artifacts. Figure 1.2 compares CBCT against other common medical imaging modalities along these dimensions.

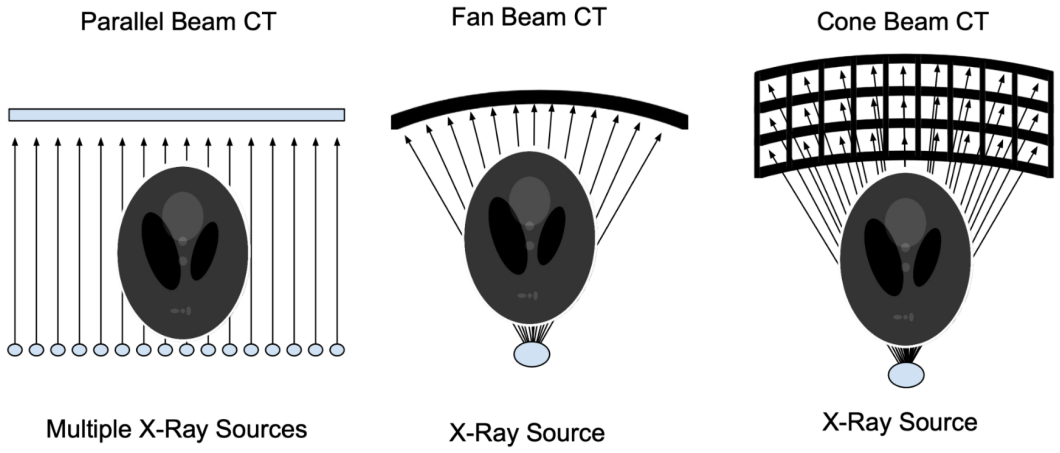


Figure 1.2: Comparison of Cone Beam Computed Tomography (CBCT) against other common medical imaging modalities, highlighting differences in acquisition geometry, dose, cost, and clinical applicability.

1.1.3 Traditional reconstruction methodologies and machine learning

Medical Imaging reconstruction was originally seen as a mathematical problem requiring an analytic solution to solve a system of equations, while simultaneously accounting for the geometry of the scanner and the physics constraints in the system matrix design.

The most basic solution is the **Filtered BackProjection (FBP)** [KS01], which is a simple inverse Radon transform with a ramp filter that is applied to eliminate overall Gaussian blur in the image and enhance high-frequency details. The Feldkamp–Davis–Kress (FDK) [FDK84] algorithm was then introduced as an adaptation of FBP to 3D CBCT image reconstruction. FDK and FBP are considered low computational cost and quick reconstruction methodologies, which are, however, prone to streaking artifacts. Iterative solutions such as the Simultaneous Algebraic Reconstruction Technique (SART) [AK84], Adaptive Steepest Descent – Projection Onto Convex Sets (ASD-POCS) [SP08], and Conjugate Gradient Least Squares (CGLS) [HS+52] were the next breakthrough, where the solution is a continuous optimization of an image based on the prior understanding of the image characteristics. The optimization depends on an objective function that is maximized by adapting the image through every iteration.

Most recently, multiple Machine Learning (ML) models have been introduced that directly translate signal from the sinogram domain to image space. These models are also compensating for noise and artifacts either by implicitly mapping a noisy sinogram to a ground truth clean image or by directly incorporating physics, geometrical, and statistical awareness into the loss function.

1.1.4 The emergence of learned off-grid representations

A recent field of work in Computational Geometry and Computer Vision has involved encoding a single 3D scene in the weights of a small neural network model or other primitives such as Gaussians. The approach of learning to represent a 3D scene leveraging differentiable rendering has been gaining a lot of attention due to its ability to use a single methodology that is guaranteed to work well in out-of-distribution data. Since it is a continuous coordinate representation, it also allows for using interpolation to recover missing information, with the main application being novel view synthesis (NVS) that allows real-time rendering. In addition, their representation is much more sparse than voxel grids, which allows faster reconstruction with a memory-efficient representation. The most popular methodologies include **Neural Radiance Fields (NeRF)** and **3D Gaussian Splatting (3DGS)**, which have been dominating the 3D scene representation field of research. Other implicit neural representation (INR) methodologies have also gained attention, with a bigger emphasis on computational geometry such as Signed Distance Fields (SDF) and Occupancy Fields. More discussion on the fundamentals and the popular methodologies of gridless representations is provided in Chapter 2.

1.1.5 Adapting off-grid representations for low-dose acquisitions

The benefits of differentiable rendering and continuous off-grid representation are directly applicable to the medical imaging field. In particular, NVS using input 2D images as training data translates directly to sparse radiograph acquisitions in CT, where differentiable rendering can be used to synthesize novel radiographs. With the use of this methodology, prevalent low-dose CBCT artifacts, including geometric and statistical streak artifacts, are eliminated. Moreover, the resulting format consumes less memory than alternative 3D voxel grid file formats such as Digital Imaging and Communications

in Medicine (DICOM) or Neuroimaging Informatics Technology Initiative (NIFTI). Indeed, gridless methodologies have recently offered an alternative methodology for reconstruction directly from sinograms and radiographs to 3D representations, without the artifacts from traditional analytic and iterative approaches or the long training times, dataset requirements, and out-of-distribution underperformance of conventional Deep Learning (DL) models. These advantages resulted in a motivation to investigate a variety of learned gridless methodologies for radiographs to 3D scene reconstruction and analyze their results and limitations while comparing them to each other and previous reconstruction methods.

1.2 Contributions

This motivation and the questions about the applicability of gridless representations in medical imaging are the driving force behind this dissertation. Our work explores various learned methodologies based on differentiable rendering [MST+21] that can be used to overfit a 3D CBCT scan. Specifically, we are focusing on implementations based on 3DGS [KKLD23], **Neural Attenuation Fields (NAF)** [ZZL22], and implicit SDF [GYH+20]. We choose CBCT due to its ability to reliably represent complex anatomical structures with well-defined hard boundaries. Moreover, the angular radiograph acquisition process translates directly to existing methodologies used in other Computer Vision applications with camera poses and captured images as inputs to generate the 3D scene representation.

For all three presented works, I functioned as the project lead with my main responsibilities including managing the research team, determining the motivation of the research, performing literature review, and identifying methodologies that could be translated from computer vision and computational geometry to the medical imaging

field. Additionally, I was fully responsible for writing the corresponding papers and presenting them at conferences and guest research talks at other venues. My PI, Professor Marinescu, provided guidance in all steps of project development and paper submission and made sure the research direction was on the right track. For RibPull and ReNAF, Dr. Digne from Université Claude Bernard Lyon, CNRS, co-advised me and helped in particular with literature review, manuscript submission, and methodology or result evaluation. Finally, also for RibPull, PhD candidate Amine Ouasfi from Université de Rennes, INRIA, developed the SparseOcc [OB24] implicit surface reconstruction methodology that was adapted for ribcages.

A note about our results

As described in detail during the dissertation, our work leverages methods, tools, and datasets introduced by other researchers. These include the original authors of 3DGS [KKLD23], NAF [ZZL22], and SparseOcc for learned off-grid methodologies, Laplace-Based Contraction [CTO+10] for skeletonization, Parkinson's Progression Markers Initiative (PPMI) for accessing reconstructed images in voxel grid format, and, finally, TIGRE [BDHS16] and Plastimatch [SLW+10] for digitally reconstructed radiographs from 3D DICOM or NIFTI images.

1.2.1 GaSpCT: Gaussian Splatting for Novel Brain CBCT Projection View Synthesis

Our first work explores the use of 3DGS as a methodology for reconstructing and representing a CBCT 3D scene, with the main motivation being novel view synthesis to compensate for the information loss resulting from sparse-view acquisition. While exploring the usage of NeRF for CBCT, the original 3DGS for real-time rendering was

released, presenting a viable alternative that is also less prone to floaters and trains faster.

As a result, we published one of the first applications of 3DGS in CBCT for brain scans. The results demonstrated our model's ability to *outperform* other representations in reconstruction fidelity and also proved that 3DGS functioned well as an alternative reconstruction methodology to traditional iterative and analytic algorithms. While this was a strong start in off-grid learned representations for medical imaging, our next question was whether a point cloud representation would offer advantages not present in the voxel grid format.

Specific Contributions per Author: My contributions include using 3DGS as a representation and reconstruction methodology for CBCT. The entirety of the methodology, including how to extract the camera pose from the DICOM metadata, the adaptation of the loss function, and the geometrical initialization, were my ideas. Also the choice of the models to compare against and the ablation studies were done by me. Utkarsh Gupta developed the ellipsoid initialization and helped with running MedNeRF. Justin Bui and Jonathan Vengosh helped with fine-tuning all hyperparameters, running MipNeRF, and making Python scripts for the camera parameter extraction from DICOM data. I wrote all the text for the research paper. Professor Marinescu supervised the whole project as my primary advisor.

This result is based on:

Emmanouil Nikolakakis, Utkarsh Gupta, Jonathan Vengosh, Justin Bui & Razvan Marinescu. GaSpCT: Gaussian Splatting for Novel Brain CBCT Projection View Synthesis. *Medical Imaging 2024: Image Processing, SPIE* [NGV+24].

Full details can be seen in Chapter 3.

1.2.2 RibPull: Implicit Occupancy Fields and Medial Axis Extraction for CT Ribcage Scans

Following the question of what advantages are obtained with the use of off-grid representations such as point clouds, our literature review identified Neural Skeletons [CD23] as a perfect example of the geometric operations that medical imaging could benefit from. Extracting the skeleton or medial axis of an anatomical bone structure could help identify fractures or other irregularities. After discussion, ribcages were identified as interesting geometric objects to evaluate off-grid learned representations and skeletonization due to their curvature complexity, interpretable visualization, and the known problem of individual rib labeling. We also decided to use SDFs, similar to Neural Skeletons.

While there was research already performed in skeletonization of ribcages [YGW+21, JGW+23], these works did not include a continuous coordinate-based representation of the object. Moreover, their proposed skeletonization methods (shortest-path-based skeletonization [TBA11], voxel-based TEASAR [SBB+00a], and point-based L1 median skeletonization [HWCO+13a]) were not reliable and robust under sparse acquisition. Therefore, we suggested RibPull, which leverages an implicit binary occupancy field (SparseOcc) as a continuous off-grid representation and a Laplacian-based contraction methodology for medial axis extraction. Our results demonstrate that the SparseOcc worked better than other methodologies to encode and preserve the topology of the ribcages, while Laplacian-based contraction proved to be robust in all test cases.

Specific Contributions per Author: My contributions include the original idea of leveraging implicit SDF representations and, later on, implicit occupancy fields for medical applications. I also used RibSeg to get the segmented ribcages that we used as a dataset and identified Laplacian-Based Contraction as the ideal methodology for

skeletonization. Amine Ouasfi, the original author of the implicit occupancy field used, helped us adapt the model for medical imaging. He also wrote the relevant section on the paper. Dr. Digne supervised the computational geometry aspect of the project, helping with investigating different methodologies to implicitly and accurately capture the ribcages. Professor Marinescu supervised the whole project as my primary advisor.

This result is based on:

Emmanouil Nikolakakis, Amine Ouasfi, Julie Digne & Razvan Marinescu. RibPull: Implicit Occupancy Fields and Medial Axis Extraction for CT Ribcage Scans. *Medical Imaging 2025: Image Processing, SPIE* [NODM25].

Full details can be seen in Chapter 4.

1.2.3 ReNAF: Regularized Neural Attenuation Fields with 3D Reconstruction Priors for Sparse-View CT

The final work presented in this dissertation explores the usage of anatomical priors to generate 3D volumes of higher accuracy with less training time. An additional goal was to be robust under extremely sparse-view acquisitions, where other methodologies collapsed. While a few methods such as ρ-NeRF and NeAS had similar goals, none of them were available in public repositories.

With these targets in mind, we introduced Regularized Neural Attenuation Fields (ReNAF), a model that leveraged physics-aware anatomical priors provided by a traditional reconstruction algorithm. We investigated a variety of algorithms, including FDK, SART, CGLS, and ASD-POCS, with the latter being considered the best candidate. Generating such a prior only increased the total reconstruction time by a few minutes, while the attenuation field model converged much sooner than alternative methodologies and

remained robust under sparse-view reconstructions with as few as 10 radiographs used as input. We also developed ρ-NeRF and NeAS based on the description provided in the papers for thorough comparisons and evaluations.

Specific Contributions per Author: I worked closely with Dr Digne to research open problems in the field of implicit scene reconstruction in CBCT with anatomical priors. We determined the overall methodology proposed together. All development and validation of the model and ρ-NeRF was done by me. Daksh Shah was responsible for developing and validating NeAS. Julia Wong generated anatomical priors using TIGRE, as will be explained in Chapter 5. I was responsible for writing the paper with Dr. Digne and Professor Marinescu supervising and reviewing the process. Professor Marinescu supervised the whole project as my primary advisor.

This result is based on:

Emmanouil Nikolakakis, Daksh K. Shah, Julia Wong, Julie Digne & Razvan Marinescu. ReNAF: Regularized Neural Attenuation Fields with 3D Reconstruction Priors for Sparse-View CT. *Ready to be submitted* [NSW+26].

Full details can be seen in Chapter 5.

Chapter 2

Background and Related Work

2.1 Cone-Beam Computed Tomography

CT images the **linear attenuation coefficient (LAC)** of the materials within the field of view (FOV) of the scanner.

The synthesis of LAC follows the **Beer-Lambert law** [KS01]:

$$I = I_0 \cdot e^{-\int_{\mathbf{r}}^{\mathbf{r}+\ell} \mu(\mathbf{x})\, d\mathbf{x}} \tag{2.1}$$

where I_0 is the original X-ray intensity, I is the intensity measured at the detector, and $\mu(\mathbf{x})$ is the LAC along the ray path dx.

In practice, the ray path between $\mathbf{r}$ and $\mathbf{r} + \ell$ traverses a sequence of materials with different attenuation properties. Figure 2.1 illustrates this by discretizing the ray into a small number of homogeneous layers, each with its own attenuation coefficient μ_k, before taking the continuum limit that recovers Equation (2.1). The following section explains in detail the instrumentation design in a typical CT scanner to match the physics behind CT to the scanner components and signal acquisition regime.

$$I(x) = I_0\, e^{-\mu x} \qquad I(L) = I_0\, e^{-\int_0^L \mu(x)\, dx}$$

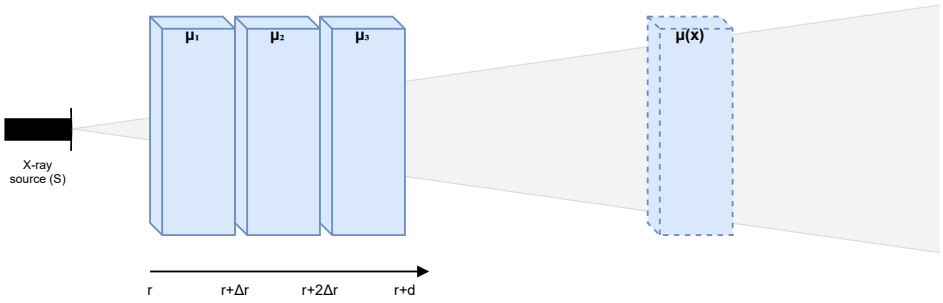


Figure 2.1: Attenuation of the X-ray beam along the ray path. Left: a discretized approximation, where the ray traverses homogeneous layers of thickness Δr, each with its own homogeneous attenuation coefficient μ_1, μ_2, μ_3, following the discretized form $I = I_0 \cdot e^{-\mu_k \Delta r}$ shown above the layers. Right: the continuum limit, where the layers are replaced by the continuous attenuation field $\mu(\mathbf{x})$ along the ray, integrated from $\mathbf{r}$ to $\mathbf{r} + \ell$ and recovering the Beer–Lambert law shown above the panel (Equation (2.1)). I_0, I, and μ are defined earlier in this section.

2.1.1 System Design - Scanner Components

A common CT scanner contains the following components [HY23]:

- A **generator** that provides a high-voltage supply up to 150 kilovolts. The level of voltage determines the maximum intensity of the X-rays emitted by the source.
- A **patient table** where the patient rests during the scan. The table is considered a structural artifact in CT image reconstructions, and methodologies have been explored for its automatic removal.
- An **image processor** which converts the raw intensity measurements that are the output of the Data Acquisition System (DAS) to the cross-sectional set of images.
- A **console** or control unit where an operator inserts the patient information and scanning parameters per desired protocol. These include the duration, energy

intensities, and slice thickness.

- The main **gantry**, also known as the scanning unit, contains the electronic components, including the X-ray source, the detector panel, and the DAS. A detailed analysis of the electronics behind the X-ray emission and photon acquisition will be provided in the following subsection.

2.1.2 System Design - Electronics

The electronics can be described in two parts: the X-ray source, which emits photons at a desired intensity, and the detector panel with the DAS, which detects the attenuated photons and converts the raw electronic signal into photon intensities used to acquire the final LAC values.

The X-ray tube converts electricity into X-ray photons using a cathode and anode sealed inside an evacuated envelope. The tungsten cathode filament releases electrons by thermionic emission, and the tube potential accelerates them toward the tungsten anode, where on impact they produce X-rays by two mechanisms: 1) characteristic radiation, from an incident electron ejecting an inner-shell electron that is replaced by an outer-shell electron emitting a photon, and 2) bremsstrahlung, from an electron being deflected by a target nucleus and radiating its lost kinetic energy [PNBSB20]. At the end of the process, only 1% of the input energy results in X-ray emission, with the remaining 99% being lost to heat.

A related extension to the single-energy source just described is dual-energy CT, motivated by the ill-posedness of single-energy scanning. A single energy system would measure the LAC of each voxel as a single value for the given energy level, yet attenuation depends on both a material's density and its atomic composition. Therefore, different materials such as calcium and iodinated contrast can result in identical values. Dual-

energy CT solves this issue by imaging the same anatomy at two distinct energy spectra (for example, 80 kV and 140 kV). Because the photoelectric effect dominates at low energies while Compton scattering dominates at higher energies, a dual-energy system yields a second independent value per voxel, enough to solve for material composition rather than total attenuation alone [JSLR24].

Once the photon arrives at the detection panel, it first meets the scintillator, a dense ceramic material such as gadolinium oxysulfide, cadmium tungstate, or gadolinium oxide. These materials are chosen due to their high density and high atomic number, which make them effective at absorbing X-ray photons and stopping them within the material [DWL+13, WPP+18]. When a photon is absorbed, its energy is deposited in the scintillator and released again as visible light, with more deposited X-ray energy producing brighter light [vdBvSB+23, WPP+18]. Directly behind the scintillator sits a silicon photodiode, which converts this visible light into a small electrical current whose size is proportional to the amount of light, and therefore to the intensity of the X-rays that reached that detector element [vdBvSB+23, DWL+13]. This current is then passed to the DAS, which amplifies the weak analog signal and digitizes it, producing one intensity reading per detector element [DWL+13]. To turn these intensity readings into attenuation values, each reading I is compared against the unattenuated beam intensity I_0 obtained from reference detectors or an air calibration, and the logarithm of the ratio $\ln(I_0/I)$ gives the line integral of attenuation along that ray path, which forms the input to reconstruction following the Beer-Lambert equation (2.1).

2.2 CT Reconstruction Fundamentals

Every signal acquisition system is based on the following equation:

$$y = A \cdot x + b \tag{2.2}$$

Here, x is the ground truth of the information, A is a representation of the system matrix or forward model that converts the ground truth into the captured signal (in the same or a different domain), and b represents additive noise introduced during the scan. Every reconstruction algorithm will then aim to recover a close representation of x [FDK84, AK84, SP08].

In the context of CBCT, x represents the patient, the system matrix A captures all scanner information such as system geometry, voxel-to-detector mapping, per-detector offset, and gain calibration [TSS16], b represents primarily electronic readout noise or dark current shot noise [ZGS14], and y is the captured information in the sinogram domain. Every point in the FOV results in a single sinusoid in the sinogram domain. The final sinogram is a superposition of all the sinusoids representing each point with any value in the FOV. Whereas a sinogram is built from 1D projections of a 2D slice, a radiograph is the 2D projection of a 3D scene. Essentially, it is the same idea with one extra spatial dimension per view.

Figure 2.2 illustrates this relationship directly. A single point at (x_0, y_0) in image space traces a single sinusoid $s(\theta) = x_0 \cos\theta + y_0 \sin\theta$ in the sinogram, where θ is the projection angle and s is the detector position at which the ray through that point is recorded. When the FOV instead contains 100 points, each one contributes its own sinusoid, and the observed sinogram is the superposition of all 100 curves. A real anatomical image, such as the Shepp–Logan phantom, can be viewed as the continuum limit of this process: a continuous attenuation field $\mu(x, y)$ is equivalent to an uncountable set of points, so its sinogram is the superposition of the sinusoids contributed by every point in the FOV, producing the smooth wave pattern shown in the bottom row of the

figure.

The final value of the reconstructed LAC is measured in the **Hounsfield Scale**. This unit was introduced to encode the LAC values in a form that is easily handled by 12-bit processing [JOT+19]. The formula that converts from LAC to Hounsfield Units is given in Equation (2.3):

$$\mathrm{HU} = 1000 \cdot \frac{\mu - \mu_{\mathrm{water}}}{\mu_{\mathrm{water}} - \mu_{\mathrm{air}}} \tag{2.3}$$

where μ is the LAC of the material being converted, and μ_{water} and μ_{air} are the LAC of water and air, respectively, at the effective energy of the scan. By construction, this normalizes water to 0 HU and air to -1000 HU. Table 2.1 lists representative LAC and Hounsfield values for a few common materials.

Table 2.1: Representative LAC and Hounsfield Unit values for common materials.

Material	**LAC μ (cm^{-1})**	**Hounsfield Units (HU)**
Air	0.0000	-1000
Water	0.1928	0
Soft Tissue	0.2005	+40
Bone	0.3856	+1000

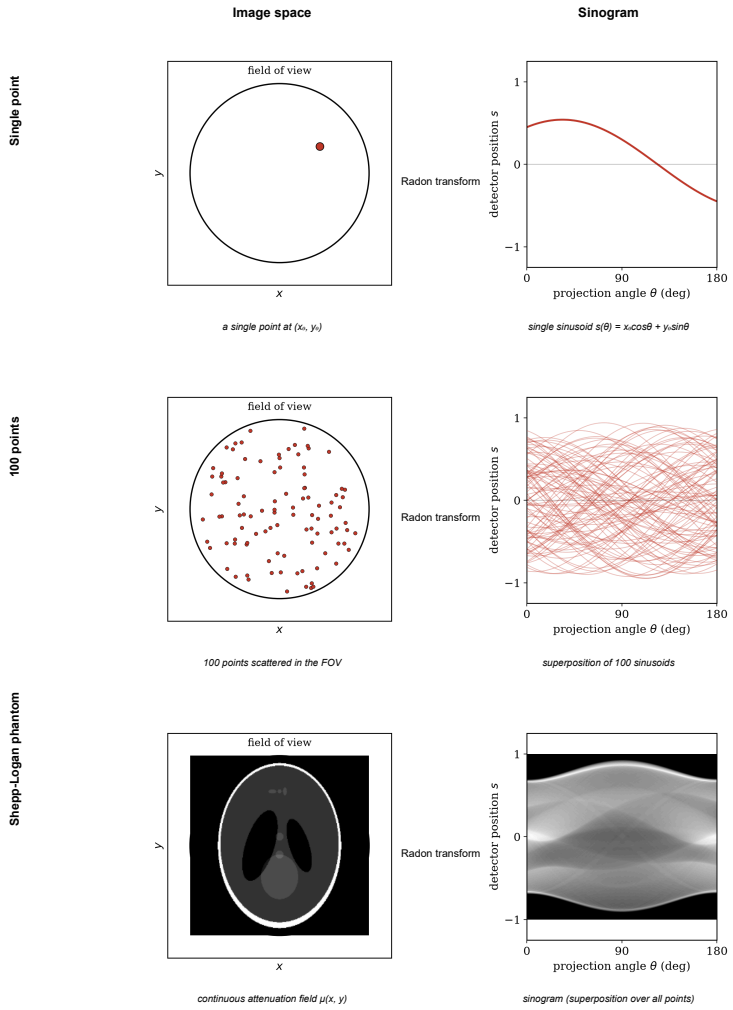


Figure 2.2: Sinogram formation as a superposition of point sinusoids. Top: a single point at (x_0, y_0) in image space (left) maps to a single sinusoidal curve $s(\theta) = x_0 \cos\theta + y_0 \sin\theta$ in the sinogram (right). Middle: 100 points scattered in the FOV (left) each contribute their own sinusoid, and the resulting sinogram (right) is the superposition of all 100 curves. Bottom: the Shepp–Logan phantom, a continuous attenuation field $\mu(x, y)$ (left), and its sinogram (right), the continuum-limit superposition of every point's sinusoid.

As introduced in Chapter 1, reconstruction algorithms that recover x from Equation (2.2) fall into two families: analytic methods such as FBP and FDK, which invert the Radon transform directly via a filtered backprojection, and iterative methods such as SART, CGLS, and ASD-POCS, which pose reconstruction as an optimization problem solved over successive passes through the data. Figure 2.3 compares all five on the

Shepp–Logan phantom under a sparse-view acquisition (50 views): FBP and FDK differ only in ray geometry (parallel- vs. fan-beam) and both show the characteristic streaking of analytic methods under sparse sampling, while the iterative methods progressively suppress these artifacts, with the Total Variation (TV)-regularized ASD-POCS producing the cleanest result at the cost of additional computation.

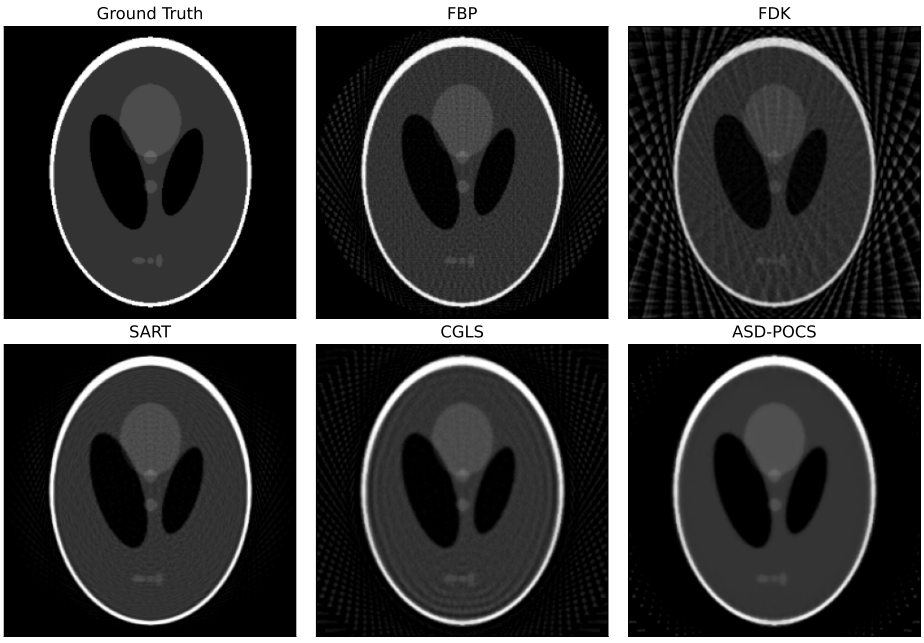
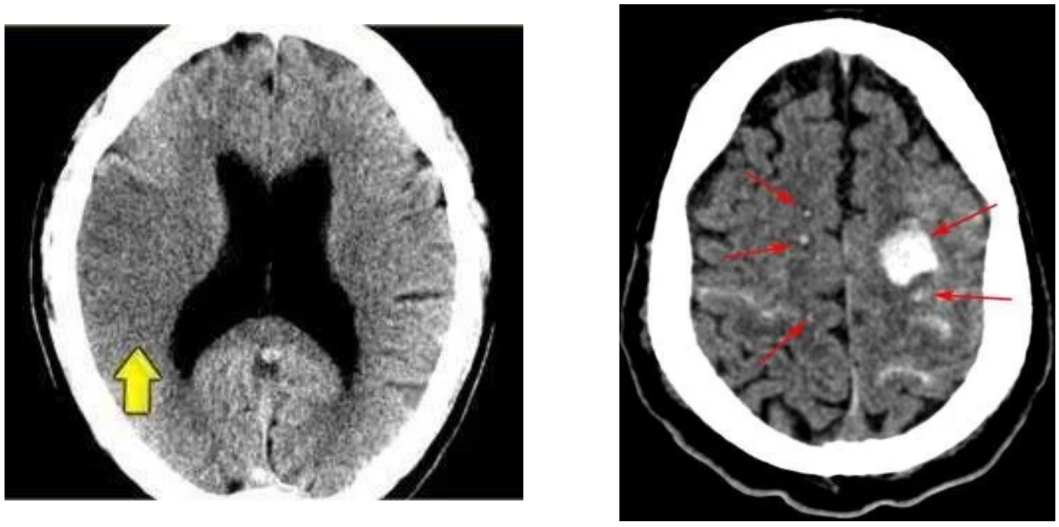


Figure 2.3: Reconstruction of the Shepp–Logan phantom from 50 sparse views using analytic (FBP, FDK) and iterative (SART, CGLS, ASD-POCS) algorithms, alongside the ground truth. FDK and SART/CGLS/ASD-POCS use a parallel-beam sinogram over 180°; FBP uses a fan-beam sinogram over 360°, matching each algorithm's native acquisition geometry.

Reconstructed images are typically stored in DICOM or NIFTI format. Both of them capture the voxel grid information and a collection of metadata related to the scan. DICOM is the most common standard used as the scanner output in hospitals and institutions. It uses either a slice per file or a voxel grid per file structure and has rich metadata with scanning parameters and patient information for accurate protocol control and patient monitoring. NIFTI format is primarily used in research applications

and supports a smaller header file with less metadata. In order to keep scans used for research de-identified, NIFTI metadata focuses on the geometry rather than patient- or scanner-specific information.

Reconstruction quality can be measured both quantitatively and qualitatively. The most common metrics used include the referential Peak Signal-to-Noise Ratio (PSNR), which is a direct pixel-to-pixel comparison, and Structural Similarity Index Measure (SSIM) [WBSS04], which uses a sliding window to compare statistical and structural discrepancies between the ground truth and the reconstruction. In addition, with DL perception models, Learned Perceptual Image Patch Similarity (LPIPS) [ZIE+18] has been introduced, which uses a pre-trained DL model such as Visual Geometry Group (VGG) Network to assign a quantitative human perception score to an image. Perceptually, the best approach in qualitative evaluation is to share the ground truth and the reconstructed image with a radiologist. For example, when diagnosing ischemic stroke or petechial hemorrhages, a reconstructed image may look sharp and similar to the ground truth, but the critical information lies in small biomarkers that can be hard to identify for anyone without clinical training. Figure 2.4 shows two such examples: ischemic stroke (Figure 2.4a) and petechial hemorrhage (Figure 2.4b).



(a) Ischemic stroke on CT: the hypodense region (dark area) in the left middle cerebral artery territory indicates infarcted tissue.

(b) Petechial hemorrhage on CT: small hyperdense foci (bright punctate spots) indicating micro-bleeds that require expert radiological interpretation.

Figure 2.4: Examples of clinically critical CT findings that highlight the limits of quantitative metrics alone. A reconstruction may score well on PSNR or SSIM yet obscure the subtle biomarkers needed for accurate diagnosis.

The notation related to CBCT and image reconstruction that will be used throughout the dissertation can be seen in Table 2.2.

2.3 Digitally Reconstructed Radiographs

A **digitally reconstructed radiograph (DRR)** is a simulation of the CT signal acquisition process given a 3D reconstructed image as the ground truth x and a system matrix A defined digitally. This process is useful for research in image reconstruction in CBCT, as there are very few available public radiograph datasets.

This dissertation uses exclusively DRRs for the training and the evaluation of all projects. We use two software tools for this purpose. The first one is Plastimatch [SLW+10], and the second is TIGRE [BDHS16]. We define the simulation

Table 2.2: Notation for detector array and radiograph dataset definitions.

Symbol	Description
N, M	Detector array dimensions (rows × columns)
d	Detector panel: discrete grid $\{(i, j) \mid i = 1, \ldots, N,\ j = 1, \ldots, M\}$
$d_{i,j}$	Physical detector cell at index (i, j)
w, h	Physical width/height of a single detector element
(u_i, v_j)	Continuous position of detector cell (i, j)
s_u, s_v	Total physical detector size $(N{\cdot}w \times M{\cdot}h)$
$\mu(\mathbf{x})$	Attenuation coefficient field of the imaged volume
I_0	Incident X-ray source intensity
$L_{i,j}$	Ray from source S to detector cell (i, j)
$I_{i,j}(\theta)$	Measured signal at cell (i, j), gantry angle θ
$P_\theta \in \mathbb{R}^{N \times M}$	Single radiograph at angle θ
K	Total number of angular acquisitions
$\Delta\theta$	Angular increment between acquisitions
$\mathcal{P} \in \mathbb{R}^{K \times N \times M}$	Full radiograph dataset (projection tensor)

parameters for both software tools using scripts in Python, while the software itself is written in C++.

To define a geometry for the scan, we begin with the definition of the detector panel dimensions. This requires the height and width of each detector in the array, and the total number of detectors in the panel vertically and horizontally. We do not change the default detector material, the energy levels of the X-ray emission beam, or the sensitivity of the system. Table 2.3 maps the corresponding geometry parameters between the two software tools.

The input to the DRR tools is a volume grid 3D image in DICOM format. The outputs are .png format images and a text file that captures the camera intrinsic and extrinsic parameters. These parameters are necessary for transformation between world-to-camera and camera-to-world coordinates.

Table 2.3: Geometry parameter mapping between Plastimatch and TIGRE.

Concept	Plastimatch	TIGRE
Source-to-axis / source-to-origin distance	`-sad`	`self.DSO`
Source-to-detector / source-to-image distance	`-sid`	`self.DSD`
Detector resolution (pixels)	`-r` (`-dim`)	`nDetector`
Detector physical size	`-z` (`-detector-size`)	`dDetector` × `nDetector` = `sDetector`
Object/volume offset	`-o` (isocenter)	`offOrigin`
Detector centering	`-c` (`-image-center`)	`offDetector`
Angular sampling	`-N` (gantry-angle spacing) + `-a` (num angles)	`angles` array passed to projector
Forward projection algorithm	`-algorithm {exact, uniform}`	`accuracy` (samples/voxel along ray)

2.4 Machine Learning Fundamentals

Because the contributions in this dissertation repurpose standard deep learning machinery for a non-standard task, we clarify a few key terms as we use them here, which is not always how they are used in their general form. In conventional deep learning, a model is trained on a population of examples so that it generalizes to unseen inputs; the methodologies in this work invert that setting, since each model is deliberately overfit to a *single* scan, and its trained weights serve as the reconstruction itself.

- **Dataset**: The set of measurements to which a single-scene representation is fit. Depending on the methodology, these are the acquired (or simulated) radiographs together with their projection geometry, as in GaSpCT and ReNAF, or points sampled in the reconstruction volume, as in RibPull. Because each model reconstructs one scan, the conventional split into training, validation, and testing sets does not apply in the usual sense. This means that there is essentially no

held-out population to generalize to. Where we report the held-out projection views, these are unseen *poses* of the same scan, used to verify that the representation has recovered the underlying volume rather than merely memorizing the input projections.

- **Loss Function**: The measure of disagreement between the model's output and the acquisition it is fit to. Throughout this dissertation the primary term is the discrepancy between a rendered projection and the corresponding acquired radiograph at the same pose. Each contribution augments this data term with additional objectives suited to the sparse-view setting, such as the total-variation and Beta priors in GaSpCT and the classical reconstruction prior in ReNAF.

- **Regularization**: The imposition of additional constraints on an otherwise ill-posed problem. Sparse-view reconstruction admits many volumes consistent with the available projections, most of which are anatomically implausible. Regularization narrows this solution space toward physically and anatomically reasonable reconstructions.

All contributions in this dissertation follow ML principles, with the main objective being a function that produces CT imaging information as output given 3D coordinates and camera poses as input. The model's weights are formed during training, using the sparse radiograph dataset as ground truth against which specific coordinates are tested. We assume the reader is familiar with the fundamentals of deep learning, such as backpropagation, gradient-based optimization, and iterative training, and instead focus on the terms whose role in this dissertation differs from their conventional use.

2.5 Deep Learning for Off-Grid Representation

ML has had a significant impact on image reconstruction in healthcare for CBCT [ABVGT+21]. The ability to draw on large datasets to translate a set of radiographs $\mathcal{P}$ directly into a 3D image has improved quantitative metrics such as PSNR and SSIM [ZZL22, CWY+24]. While there are numerous research works that demonstrate the ability of DL models to accurately reconstruct voxel grids from sparse-view and low-dose sinograms, these methodologies have also been marred by issues with dataset availability of a large and diverse population and long training times. Moreover, optimizing for a voxel grid can be computationally intense, especially when a medical 3D scene could use a much sparser representation with lower computation and storage costs.

Because of those constraints, there has been a rise in research interest regarding off-grid representations of 3D medical scans, in particular learned representations using machine learning approaches. These approaches can be *implicit*, such as Neural Radiance Fields (NeRF) [MST+21], NAF [ZZL22], and SDFs [LWX+24], or *explicit*, such as 3D Gaussian Splatting (3DGS) [KKLD23]. In all cases, a single scene is overfit in the weights of the model. In NeRF, an MLP is used to convert a 5-dimensional input representing the camera coordinates (x, y, z, θ, ϕ) directly to a novel view from the given camera pose. The approach is similar with 3DGS, where instead of the weights of an MLP, the 3D scene is encoded in the parameters of a collection of Gaussians $\mathcal{G} = \{g\}$, with each Gaussian $g = \{\text{mean}, \text{covariance}, \text{alpha}, \text{color}\}$. The same 5-dimensional camera pose will then create a new view. With SDF and NAF, the input is a 3-dimensional (x, y, z) coordinate. Both these representations can be used directly for volume rendering rather than novel view synthesis. In SDF, the output is the distance of the coordinate from the surface of an object Ω, with the sign indicating whether the point is located

within or outside the object. Similarly, NAF is an alternative approach to NeRF where the MLP is used to directly synthesize the attenuation value for a 3D coordinate rather than going through novel view synthesis first. NAF is the baseline off-grid learned approach for CT. A related coordinate-based method, NeRP [SPX22], instead pretrains an MLP with periodic activations to overfit a patient-specific prior image, then fine-tunes that network against the sparse measurements through the imaging forward model. This substitutes the prior for a large training dataset and generalizes across 2D/3D CT and MRI. COIL [SLX+21a] takes a different approach and operates in the measurement domain rather than the image domain. A self-supervised MLP maps sensor coordinates to their measured responses and interpolates a dense measurement field from sparse-view CT data, which is then passed to a standard reconstruction method rather than being an attenuation field itself. This dissertation explores the NAF, SDF, and 3DGS approaches as representations of a CBCT scan.

A 3D scene or object can be represented in several ways, illustrated in Figure 2.5 for a simple circular object: as a **depth map** or a **voxel grid**, both of which discretize space onto a regular grid; as a **point cloud** or a **triangular mesh**, which represent the object's surface using an irregular set of discrete samples; or as an **implicit surface**, a continuous coordinate-based function of space whose zero level-set defines the object's boundary. The chapters that follow make use of several of these representations: GaSpCT (Chapter 3) represents the scene explicitly as a set of 3D Gaussians, RibPull (Chapter 4) learns an implicit occupancy field and converts it to a point cloud and mesh for geometric analysis, and ReNAF (Chapter 5) learns an implicit attenuation field directly from radiographs.

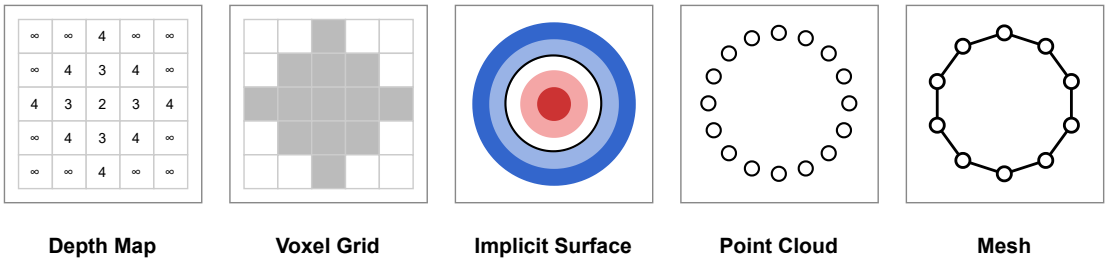


Figure 2.5: The same circular object represented as a depth map, a voxel grid, an implicit surface, a point cloud, and a triangular mesh. The depth map and voxel grid discretize the object onto a regular grid; the implicit surface encodes it as a continuous field (red: inside, blue: outside, with the black ring marking the zero level-set); the point cloud and mesh represent its boundary with an irregular set of samples.

GaSpCT and ReNAF both learn their scene representation through differentiable rendering. The network weights θ that parameterize the scene (a set of 3D Gaussians in GaSpCT, or the attenuation field $\mu(\mathbf{x})$ in ReNAF) are used to render a prediction, which is compared against a reference image to compute a loss. The gradient $\partial\text{loss}/\partial\theta$ is then backpropagated through the renderer to update θ, and the process repeats until convergence. Figure 2.6 illustrates this loop using the Shepp–Logan phantom [SL74] as an example: early in training, the rendered image is a blurred, Poisson-noise-corrupted version of the reference, which is progressively resolved as θ converges.

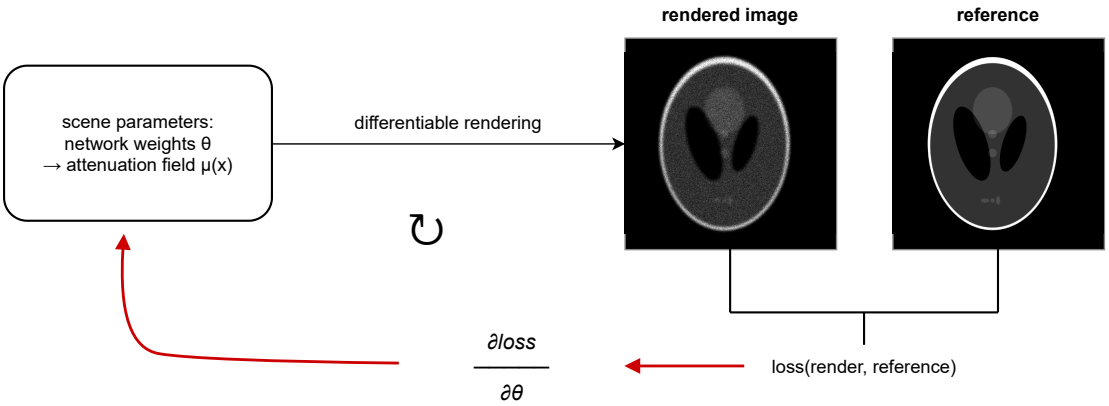


Figure 2.6: The differentiable rendering training loop shared by GaSpCT and ReNAF. Scene parameters θ are rendered into a prediction, compared against a reference (here, the Shepp–Logan phantom) to compute a loss, and the gradient $\partial \text{loss}/\partial\theta$ is backpropagated to update θ. The rendered image shows a mid-training state: a blurred, Poisson-noise-corrupted version of the reference that has not yet converged.

The main advantage of these methodologies is that, due to overfitting to a single scene, a validated method can be adapted for out-of-distribution data of the same scan type. A lot of the protocol or geometry differences between scanners are directly encoded into the DICOM or NIFTI metadata. Therefore, all necessary information for the camera pose parameters can be directly inferred from the header files. A comparison of the storage size of DICOM and the sparser off-grid representations shown in RibPull in Chapter 4 shows that the models are about 57% lighter. Also, the reconstruction can take only 10 minutes, achieving quantitative metrics comparable with traditional reconstruction methodologies. Finally, since the learned approaches are continuous representations of the 3D scene, the representation preserves geometrical and topological features such as the curvature, which we leverage during the dissertation for certain operations.

Chapter 3

GaSpCT: Gaussian Splatting for Novel Brain CBCT Projection View Synthesis

---

This chapter is based on:

Emmanouil Nikolakakis, Utkarsh Gupta, Jonathan Vengosh, Justin Bui & Razvan Marinescu. GaSpCT: Gaussian Splatting for Novel Brain CBCT Projection View Synthesis. *SPIE Medical Imaging 2024* [NGV+24].

We present GaSpCT, a novel view synthesis and 3D scene representation method used to generate novel projection views for Cone Beam Computed Tomography (CBCT) medical scans. We adapt the Gaussian Splatting framework to enable novel view synthesis in CBCT using limited sets of 2D image projections as inputs and without the need for Structure from Motion methodologies. Therefore, we reduce the total duration of the scan and the amount of radiation dose administered to the patient during the scan. The low-frequency components of radiographs and the smooth and gradual variation of radiodensity across soft tissues make Gaussian Splatting an ideal method of scene

representation in CBCT. The loss function is adapted to our use case by encouraging a stronger background and foreground distinction using sparsity and smoothness-promoting regularizers. Finally, we initialize the Gaussian locations across the 3D space using an ellipsoid-shaped uniform prior distribution of where the brain's positioning would be expected to be within the field of view. We evaluate the performance of our model using de-identified brain CBCT scans from the PPMI dataset and demonstrate that the rendered novel views closely match the original projection views of the simulated scan and have better performance than other implicit 3D scene representation techniques. Furthermore, we empirically observe reduced training time and memory requirements compared to analogous models applied to CBCT reconstruction tasks.

3.1 Introduction

Implicit and explicit scene representation models have had a significant impact on Computer Vision and Computer Graphics applications. One recent 3D scene representation methodology is 3D Gaussian Splatting (3DGS) [KKLD23], which uses 3D Gaussians to encode view-dependent color values. These models are better suited for preserving low-detail visuals, effectively reducing both the graininess and motion-induced artifacts in the image. They require considerably less training time than other state-of-the-art models such as MipNeRF360 [BMV+22]. However, their memory footprint is much larger.

Novel view synthesis (NVS) and 3D scene representation would be highly desirable for Cone Beam Computed Tomography (CBCT) since raw CBCT signal is acquired as projection sinograms, which can then be converted into 2D angular radiographs [GG22]. To minimize the overall dose [PMT+16b] administered to the patient during the scanning procedure, a typical approach is to reduce the number of projection views or lower the X-

ray intensity. As a result, ML techniques have been used for inverse problems in low-dose CT reconstruction. These aimed to remove statistical noise such as graininess caused by a decreased amount of photons detected [CZK+17a, YYZ+18a, GS22a], while other techniques dealt with sparse-view reconstruction in the spatial [ZLD+18, HY18, SE22] or the sinogram domain [DVC19, LLK+18, GYZ+23]. However, since the input data consists of 2D projections, 3D scene representation models can potentially be used for CBCT reconstruction that renders novel projection views in the spatial domain, and which would naturally compensate for the lack of sufficient views that typically can result in structural artifacts such as streaks.

There are several reasons why 3DGS is the preferred 3D scene representation methodology for CBCT. First, given that radiodensities in CBCT imaging vary anisotropically [MYK+15], Gaussians form an ideal representation basis. Second, 3DGS is better suited for smooth images consisting of low-frequency components, such as X-ray radiographs, and is less prone to floaters and graininess. Finally, training a 3DGS scene representation is faster than NeRF-based models [MST+21], which is critical in medical imaging, while the higher memory footprint is not a constraint in the field.

In this work, we present GaSpCT, a Gaussian Splatting-based model enhanced specifically for CBCT brain imaging applications. The result is a model that accurately encodes the 3D information of a CBCT scan given only half or less of the total projection views as a training dataset. The reconstructed 3D scenes preserve the signal with high accuracy for any camera pose within the FOV. In our experiments, the training process takes 5–10 minutes while the memory footprint of the 3-dimensional data in polygon file format (27–42 MB) is smaller than that of other voxel-based or mesh-based approaches. We summarize our contributions as follows:

- We introduce GaSpCT, a 3D scene representation and novel view synthesis model

that allows for rendering novel CBCT brain projections from a limited projection dataset. The model leaves a small memory footprint, and rendering new views is computationally inexpensive.

- In CT imaging, it is expected that pixels not occupied by the patient will have empty or background intensity values. To promote smoothness and sparsity in the synthesized views, we augment the baseline loss function used in Gaussian Splatting by adding a TV loss and a Beta distribution negative log-likelihood loss.

- We introduce a script to extract CT camera parameters from the DICOM metadata of the reconstructed images by approximating them to those of a pinhole camera. Thus, we remove the necessity of **Structure from Motion (SfM)** [SF16], which performs poorly on CT radiographs due to the lack of distinct edges [AKB+08]. Additionally, our approach also initializes an ellipsoid 3D point cloud representing the expected patient brain volume.

- We verify our methodology on de-identified brain CBCT projection images and provide all used datasets to the medical imaging community.

3.2 Method

3.2.1 Baseline 3D Gaussian Splatting

Our model is based on the original 3DGS. Thus, we now proceed to describe in more detail some of the main mechanisms introduced in the original paper.

Differentiable Rasterization

Rendering must be fast, since training requires many iterations, and differentiable, since gradients must reach the Gaussian parameters. 3DGS [KKLD23] projects each Gaussian onto the image plane, then splits the screen into small tiles processed independently and in parallel on the GPU; within a tile, only the Gaussians that project onto it are considered, ordered by depth so that closer Gaussians blend on top of farther ones. Each pixel accumulates color front-to-back until it becomes effectively opaque, at which point the remaining Gaussians are skipped. Training reuses this same depth order in reverse; only the final accumulated opacity per pixel is stored, and this alone recovers the gradients needed for every Gaussian that contributed, keeping memory and computation bounded regardless of how many overlap.

A 3D Gaussian is chosen as the primitive because gradient descent optimizes the whole scene directly. Unlike a point, a mesh triangle, or a voxel, a Gaussian has no hard boundary; its influence fades smoothly from its center, so small updates to its position, shape, or color produce correspondingly small, smooth changes in the rendered image. A Gaussian's shape is set by its covariance matrix; a spherical covariance forces every Gaussian into a round blob, requiring many of them to cover a flat or thin structure. An anisotropic, non-spherical covariance instead lets a single Gaussian stretch and orient itself along the surface it represents. Optimizing the covariance matrix directly is unsafe, since arbitrary gradient updates can produce a matrix that no longer represents a valid Gaussian; the covariance is instead built from a separate scale and rotation, which keeps every update valid while still allowing full anisotropy.

Adaptive Density Control of Gaussians

During optimization, the initial set of Gaussians is periodically densified and pruned so that the representation can allocate more Gaussians to under-reconstructed regions and fewer to over-reconstructed ones. Under-reconstruction occurs where the geometry of a region of the scene is missing, typically covered by too few or too small Gaussians; this is addressed by *cloning* the Gaussians within the region to add capacity in that area. Over-reconstruction occurs where a single large Gaussian covers a region with more variation than the region's geometry; this is addressed by *splitting* the Gaussian into smaller ones that jointly cover the same region with finer detail. Figure 3.1 illustrates both cases. Gaussians are pruned when they are too small or too large given specific thresholds or when their opacity levels are too low to contribute meaningful information to the scene.

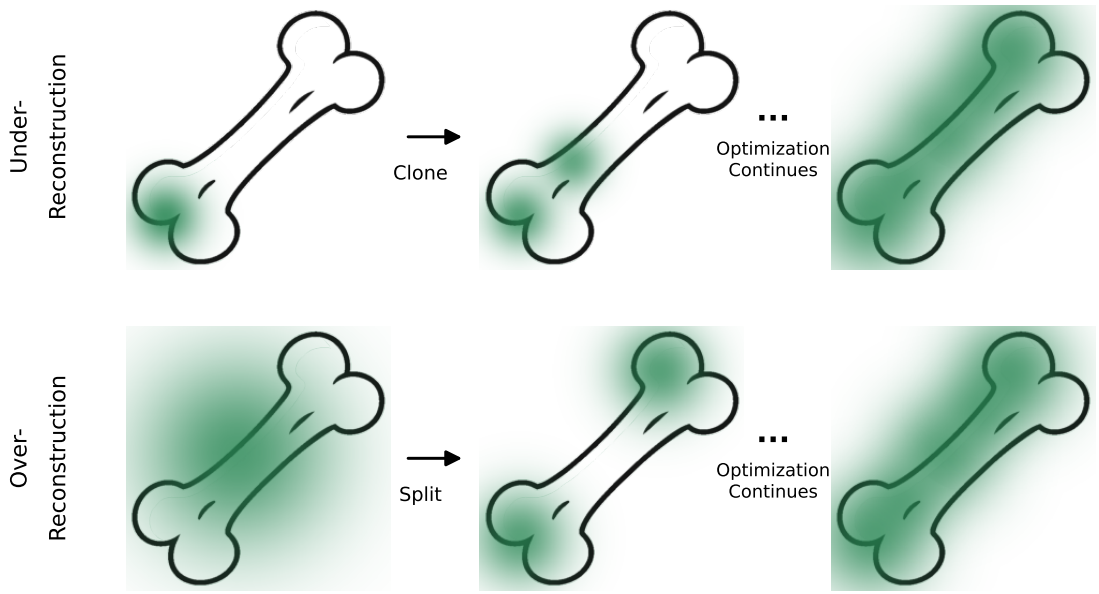


Figure 3.1: Adaptive density control of Gaussians. Top: under-reconstruction, where a region is covered by too few Gaussians; cloning duplicates the Gaussian, and the pair drifts apart during subsequent optimization to cover more of the missing geometry. Bottom: over-reconstruction, where a single large Gaussian covers more of the scene than it can represent accurately; splitting divides it into smaller Gaussians that jointly refine the same region. Both processes converge toward an accurate coverage of the target geometry (illustrated here on a bone shape in place of a generic scene silhouette).

Spherical Harmonics

Spherical harmonics (SH) are a set of orthonormal basis functions defined on the surface of a sphere. In 3DGS, each Gaussian is parameterized by a set of SH coefficients per color channel, and its view-dependent color is obtained by evaluating that expansion in the direction toward the camera. This choice is motivated by the fact that the color observed at a point in a scene depends on the viewing direction rather than being fixed.

While attenuation does not require spherical harmonics and the output images are grayscale rather than RGB, we did not remove this encoding from GaSpCT. Another adaptation for radiology published in parallel with our work, R2-Gaussian [ZLC+24], eliminates spherical harmonics, thus making the optimization faster and the representation

more efficient.

3.2.2 GaSpCT

We adapted the model specifically for CBCT brain scans, adding two sparsity-promoting regularizers in the loss function and initializing the point cloud to an ellipsoid, similar to the expected brain structure seen in the training images. This is equivalent to what is typically referred to as geometric initialization. An overview of our model can be seen in Figure 3.2.

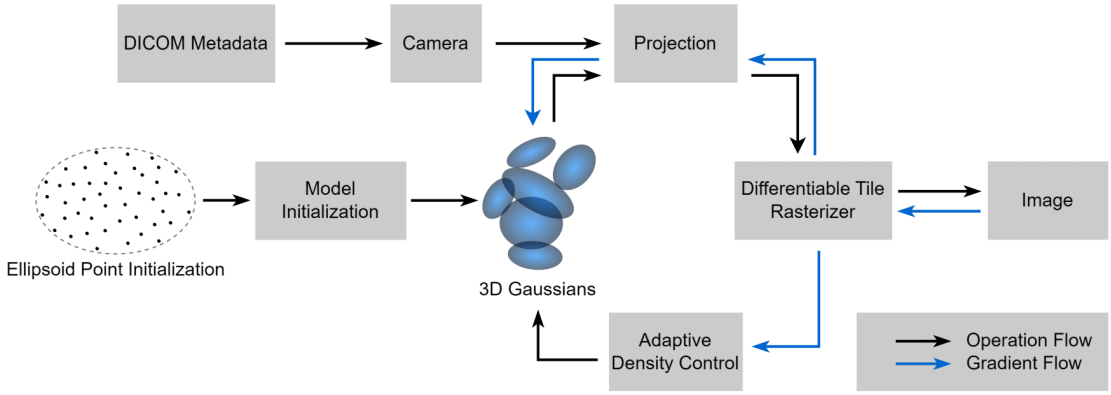


Figure 3.2: Optimization of 3D Gaussians representing the CBCT scan. Using the camera poses extracted from the DICOM metadata, we can perform forward passes and backpropagation through the differentiable Gaussian rasterizer.

3.2.3 Loss Function Adaptation

Each Gaussian in the model consists of parameters that encode its position, covariance, color, and opacity. During optimization, a 2D image and camera pose are sampled from the distribution of the training dataset. With the use of a differential Gaussian rasterizer, the equivalent image is rendered from the point cloud for the given pose. The loss between the rendered image and the ground truth is calculated, and the backpropagation step is performed using the Adam optimizer on the following loss function's gradients:

$$\mathcal{L}_{Original} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{D\text{-}SSIM} \tag{3.1}$$

which is a combination of an L1 loss promoting sparsity and a Dynamical Structure Similarity Index (D-SSIM) term, which encourages the similarity between the rendered image and the ground truth for a given camera pose. This baseline loss does not by itself encourage the pixels outside the patient to render as uniformly empty, which motivates the two regularizers below: one enforcing local smoothness, and one directly pushing background pixels toward zero opacity. Adding a TV regularizer penalizes large variations between neighboring pixels and enhances smoothness while also reducing the impact of Poisson noise artifacts. We implement the TV loss as done by Rudin et al. [RO94], where p represents the pixel value at coordinates i, j, and N and M are the image's height and width, respectively, matching the detector array dimensions of Table 2.2:

$$\mathcal{L}_{TV} = \lambda_{TV} \sum_{i,j}^{N,M} \left|p_{i+1,j} - p_{i,j}\right| + \left|p_{i,j+1} - p_{i,j}\right| \tag{3.2}$$

We also adapt the negative log-likelihood of a Beta distribution (Beta(0.5, 0.5)) as used by Lombardi et al. [LSS+19], where $N_p = N \cdot M$ is the total number of pixels p in the image and I_α is the image opacity. This loss promotes sparsity by pushing background values to zero and enhancing the pixel intensities of the foreground.

$$\mathcal{L}_{Beta} = \frac{1}{N_p} \sum_{p} \left[\log(I_\alpha(p)) + \log(1 - I_\alpha(p))\right] \tag{3.3}$$

We combine the TV regularizer and the Beta distribution regularizer to give the total loss function:

$$\mathcal{L}_{Final} = \lambda_1 \mathcal{L}_1 + \lambda_{D\text{-}SSIM} \mathcal{L}_{D\text{-}SSIM} + \lambda_{Beta} \mathcal{L}_{Beta} + \lambda_{TV} \mathcal{L}_{TV} \tag{3.4}$$

3.2.4 Retrieving Camera Poses

3DGS requires the output of SfM [SF16] software as an input to the training script. This includes the camera intrinsics and extrinsics, along with the point cloud representing the identified features in the 3D scene. Nevertheless, applying SfM to CT images is particularly challenging since there is a lack of the refined edges needed for accurate and robust feature extraction. As a result, we generate the camera intrinsics and extrinsics mathematically, leveraging information from the DICOM metadata, such as the FOV of the imaged space, the dimensions of the detector array, and the distance of the source to the detectors and the patient. These variables are used to calculate the Cartesian coordinates and the polar angle of each camera pose. The azimuth angle remains fixed at 0, given that we set the origin of the world coordinates in the center of the scanner's FOV. Finally, we initialize an ellipsoid as a point cloud instead of using the output from SfM. We derive the camera intrinsic and extrinsic matrices directly from these quantities as follows:

$$C_{int} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \qquad C_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

The focal lengths f_x and f_y are replaced with the distance between the source and the patient denoted as SP. The principal points c_x and c_y are set to ($\text{FOV}_x/2$, 0) by appropriately setting the point of origin in the center of the scanner's field of view ($\text{FOV}_y = 0$). Using the scanner's parameters, the intrinsic matrix can be approximated

as follows:

$$C_{int} = \begin{bmatrix} \mathrm{SP} & 0 & \mathrm{FOV}_x/2 \\ 0 & \mathrm{SP} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Calculating the extrinsic matrix involves computing the rotation matrix R and the translation vector T. Knowing that the X-ray detector plane rotates only along the y-axis for our chosen reference frame, we can set the roll and the yaw of the rotation to 0 while the pitch increments in steps of the angular resolution. The pitch angle is therefore equivalent to the polar angle in spherical coordinates.

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\mathrm{yaw}) & -\sin(\mathrm{yaw}) \\ 0 & \sin(\mathrm{yaw}) & \cos(\mathrm{yaw}) \end{bmatrix}, \quad R_y = \begin{bmatrix} \cos(\mathrm{pitch}) & 0 & \sin(\mathrm{pitch}) \\ 0 & 1 & 0 \\ -\sin(\mathrm{pitch}) & 0 & \cos(\mathrm{pitch}) \end{bmatrix}$$

$$R_z = \begin{bmatrix} \cos(\mathrm{roll}) & -\sin(\mathrm{roll}) & 0 \\ \sin(\mathrm{roll}) & \cos(\mathrm{roll}) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R = R_x \cdot R_y \cdot R_z = \begin{bmatrix} \cos(\mathrm{pitch}) & 0 & \sin(\mathrm{pitch}) \\ 0 & 1 & 0 \\ -\sin(\mathrm{pitch}) & 0 & \cos(\mathrm{pitch}) \end{bmatrix}$$

The translation vector is then calculated as follows:

$$T = -R \cdot C \tag{3.5}$$

where $C = [x, y, z]$ is the coordinate vector of the detector plane.

3.3 Results

3.3.1 Dataset

We use de-identified CBCT brain scans acquired from the PPMI study. The data, originally 3D images in DICOM format, are used as input to Plastimatch [SLW+10], a synthetic radiograph generation software, which generates Digitally Reconstructed Radiographs (DRR). We set the angular increment $\Delta\theta$ (Table 2.2) of the DRRs to 1° and the image resolution to 128×128. The input 3D volume and a few of the DRRs can be seen in Figure 3.3.

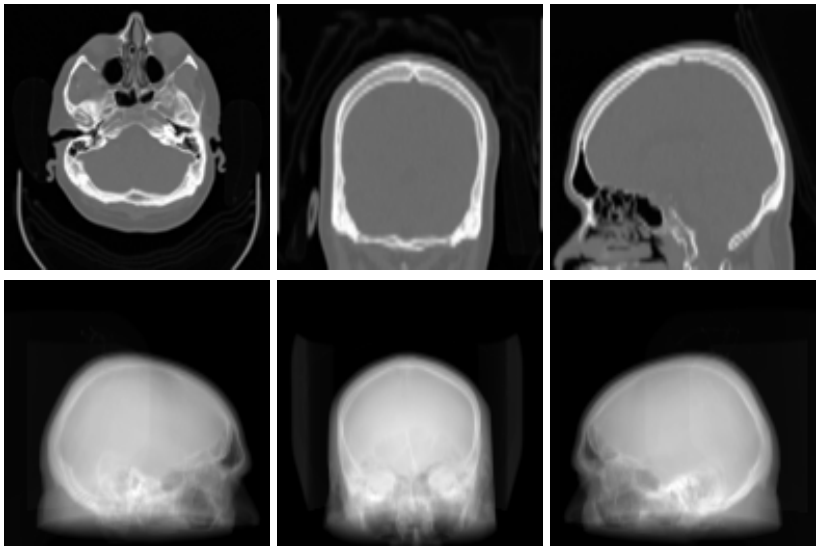


Figure 3.3: Top row: the original 3D voxel-based DICOM image retrieved from the PPMI dataset, used as input to the DRR algorithm. Bottom row: resulting projection views at 0°, 90°, and 180°.

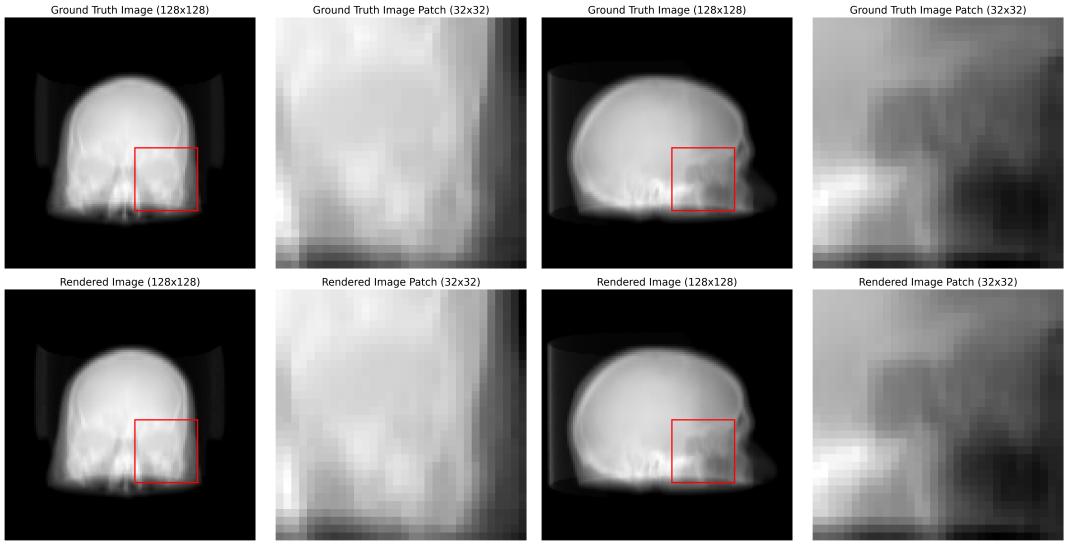


Figure 3.4: Two angular views (0° left, 90° right) for 50% of the radiograph training dataset. The ground truth was hidden during training. Zoomed-in patches show that the perceptual differences between the ground truth and rendered images are negligible.

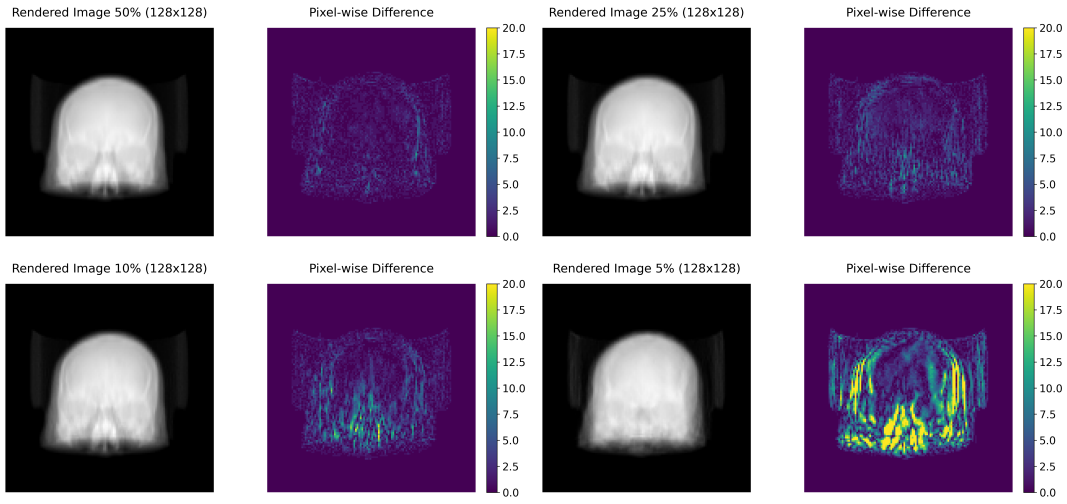


Figure 3.5: Pixel-wise absolute difference between rendered images and ground truth for various proportions of training projections. Even at 5% (18 out of 360 radiographs), rendered images remain indistinguishable from the ground truth to human perception.

3.3.2 Quantitative Results

All training and rendering were performed on a single NVIDIA RTX A4000 (16 GB GDDR6 SDRAM). All GaSpCT tests were run for 20K iterations over 20 projection view scans of 20 different subjects, using 180 out of 360 images for training and the remainder for testing. Loss regularizer weights: $\lambda_{L1} = 0.79$, $\lambda_{D\text{-}SSIM} = 0.2$, $\lambda_{Beta} = 0.005$, $\lambda_{TV} = 0.005$. Evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), SSIM [WBSS04], and Learned Perceptual Image Patch Similarity (LPIPS) [ZIE+18] using a VGG network. Results are shown in Table 3.1. Training took 5–10 minutes per scan; the total number of optimized parameters was 4.6×10^5–6.2×10^5; output files in polygon file format occupied 27–42 MB.

Table 3.1: Performance of GaSpCT compared to other models using 50% of views from the training dataset.

Metric	MedNeRF	MipNeRF360	3DGS	GaSpCT
PSNR ↑	30.36 ± 0.33	27.67 ± 0.23	40.79 ± 1.06	**43.93 ± 1.12**
SSIM ↑	0.558 ± 0.055	0.14 ± 0.03	0.99 ± 0.002	**0.997 ± 0.0007**
LPIPS ↓	0.341 ± 0.031	0.9 ± 0.05	0.017 ± 0.0037	**0.0059 ± 0.0015**

Table 3.2: Ablation studies for GaSpCT.

Metric	Baseline	Ellipsoid + TV	Ellipsoid + Beta	TV + Beta	GaSpCT
PSNR ↑	40.79 ± 1.06	43.56 ± 1.095	43.72 ± 0.95	43.55 ± 1.08	**43.93 ± 1.12**
SSIM ↑	0.9905 ± 0.0021	0.997 ± 0.0008	0.997 ± 0.0008	0.997 ± 0.0008	**0.997 ± 0.0007**
LPIPS ↓	0.017 ± 0.0037	0.0054 ± 0.0014	**0.0052 ± 0.0014**	0.0054 ± 0.0014	0.0059 ± 0.0015

Table 3.3: Performance of GaSpCT for various proportions of training projections.

Metric	50% of views	25% of views	10% of views	5% of views
PSNR ↑	43.93 ± 1.12	42.03 ± 0.95	38.5 ± 1.21	34.01 ± 1.59
SSIM ↑	0.997 ± 0.0007	0.994 ± 0.0015	0.976 ± 0.0015	0.936 ± 0.017
LPIPS ↓	0.0059 ± 0.0015	0.01 ± 0.0028	0.037 ± 0.0089	0.08 ± 0.016

In Table 3.1, we compare our metric scores to the baseline 3DGS model. The quantitative evaluation metrics indicate that optimizing the hyperparameters, initializing the splats within an ellipsoid, and tailoring the loss function for CBCT data collectively contribute to synthesizing 3D views that exhibit closer adherence to the ground truth observations than the baseline model. Additionally, we compare our results to MedNeRF [CFFBT+22a], which synthesizes 2D CBCT projection images using Implicit Scene Representation (ISR) from a single view, and MipNeRF360, which allows us to benchmark against one of the most powerful ISR methodologies. Our metric scores are superior to MedNeRF's, which is expected since GaSpCT overfits to a single CBCT scene rather than a generalized latent distribution. We found that the CBCT brain dataset failed to optimize using MipNeRF360 following 80K iterations with a batch size of 4096 rays. A comparison between the ground truth and rendered image for front and side views can be seen in Figure 3.4. Ablation studies are shown in Table 3.2 and the effect of hiding increasing proportions of training data in Table 3.3. Pixel-wise differences are visualized in Figure 3.5.

3.4 Discussion

Failure of MipNeRF360. Table 3.1 shows that MipNeRF360 performs worst of all evaluated methods, reaching only 27.67 dB PSNR with an SSIM of 0.14 and an LPIPS of 0.9 – a reconstruction that is perceptually unusable, not just imprecise. MipNeRF360 is built for unbounded, real-world photographic scenes with view-dependent radiance, parallax, and high-frequency texture, and it relies on a scene-contraction warp and a proposal-network sampling scheme tuned for that setting. None of this matches DRR data: DRRs are grayscale, low-frequency, and largely smooth, they have no view-dependent appearance to exploit, and they are formed by a transmission model (a line integral of

attenuation along each ray, Equation (2.1)) instead of the surface-oriented emission–absorption model that NeRF-style volume rendering assumes. The scene-contraction step exists to allocate capacity to distant unbounded backgrounds, which is the opposite of what helps a single compact object centered in the scanner field of view. Without the radiance cues and edge structure its sampling network relies on, and with fewer views to work with, MipNeRF360 ends up converging to a degenerate solution. MedNeRF, despite also being NeRF-based, does not run into this: it was designed for CT in the first place and never assumed unbounded scenes or rich view-dependent radiance.

3DGS and GaSpCT, on the other hand, operate on explicit primitives initialized near the expected anatomy, so they never have to discover the scene geometry from scratch out of edgeless, low-texture, sparse projections. They start close to a plausible solution and refine it from there, which is most likely the reason why both remain stable where MipNeRF360 does not.

Ablation and metric saturation. The ablation in Table 3.2 points to redundancy among the three enhancements rather than any single one dominating. Adding any two of them to the baseline recovers almost the entire improvement: the hyperparameter-tuned 3DGS baseline reaches 40.79 dB, every two-component variant lands within a narrow 0.17 dB band (43.55–43.72 dB), and the full model only adds a further 0.2 dB on top of that. The variant without the geometric initialization (TV + Beta, 43.55 dB) is essentially tied with the two variants that include it, so no single component is individually necessary at this view density – any one of them can be dropped at negligible cost as long as the other two remain. The perceptual metrics hit a ceiling the same way: SSIM is pinned at 0.997 across every augmented configuration, and LPIPS only moves within its confidence interval (the full model's 0.0059 is marginally higher than the pairs' 0.0052–0.0054, well inside one standard deviation). In short, once two of the three enhancements are in

place the reconstruction is already saturated, and it becomes hard to say how much any individual component is still contributing.

We stress, however, that this ablation was conducted at 50% of views, which Table 3.3 shows to be a comparatively easy regime: at this density the projections alone nearly determine the reconstruction (43.93 dB), whereas quality degrades substantially by 5% of views (34.01 dB). We therefore expect the relative importance of the components to change under greater sparsity. When projections are few, the geometric initialization should matter more, since it supplies anatomical shape that the views can no longer constrain, and the Beta background term should matter more, since it suppresses the floating artifacts that proliferate in the large unconstrained empty regions of an underdetermined scene. Verifying this would require repeating the ablation at 5% and 10% of views; we identify it as a direct extension, and note that it would convert the present saturation observation into a characterization of *when* each component is essential rather than optional.

3.5 Conclusions

We have presented GaSpCT, a 3D scene representation methodology adapted specifically for brain CBCT imaging. We adapted the baseline 3DGS model to use suitable sparsity regularizers and adjusted the point initialization for brain CBCT radiographs, removing the necessity of SfM methodologies in the process. Testing on 20 sets of DRR generated using PPMI brain images demonstrates that our model outperforms current state-of-the-art methodologies and can render highly accurate views from as little as 5% of the original projection dataset.

3.6 Limitations and Future Work

We have demonstrated that our methodology can generate highly accurate novel radiographs for a CBCT scan, which can help eliminate geometric artifacts. Nevertheless, there are critical steps before this methodology can be considered for clinical applications. We need to compare the accuracy of GaSpCT against analytic methodologies such as Feldkamp–Davis–Kress (FDK) reconstruction, iterative methodologies such as the Simultaneous Algebraic Reconstruction Technique (SART), and deep-learning-based methodologies such as IntraTomo [ZIL+21]. Moreover, the current data are DRRs generated from 3D CT volumes, and the methodology must still be validated against original radiograph datasets directly from CBCT scanners.

Future work involves shape analysis in the point cloud domain. Using point clouds instead of voxel-based 3D volumes preserves an accurate geometrical representation of the ground truth [ZZR+21]. We intend to tackle fracture detection using geometry information extraction on point cloud representations of rib cages following mesh skeletonization, building on the foundation laid by RibSeg [JGW+23]. We also aim to investigate the geometrical information preserved under a compressed sensing acquisition strategy by studying the coherence between the medial axis and the sinogram domain, establishing the minimum signal required to detect and classify fractures with near-optimal confidence, using the coherence formulation of Candès et al. [CT06].

Chapter 4

RibPull: Implicit Occupancy Fields and Medial Axis Extraction for CT Ribcage Scans

---

This chapter is based on:

Emmanouil Nikolakakis, Amine Ouasfi, Julie Digne & Razvan Marinescu. RibPull: Implicit Occupancy Fields and Medial Axis Extraction for CT Ribcage Scans. *SPIE Medical Imaging 2025* [NODM25].

We present RibPull, a methodology that utilizes implicit occupancy fields to bridge computational geometry and medical imaging. Implicit 3D representations use continuous functions that handle sparse and noisy data more effectively than discrete methods. While voxel grids are standard for medical imaging, they suffer from resolution limitations, topological information loss, and inefficient handling of sparsity. Coordinate functions preserve complex geometrical information and offer a better solution for

sparse data representation, while allowing for further morphological operations. In addition, implicit scene representations enable neural networks to encode entire 3D scenes within their weights. The result is a continuous function that can implicitly compensate for sparse signals and infer further information about the 3D scene by passing any combination of 3D coordinates as input to the model. In this work, we use neural occupancy fields that predict whether a 3D point lies inside or outside an object to represent CT-scanned ribcages. We also apply a Laplacian-based contraction to extract the medial axis of the ribcage, thus demonstrating a geometrical operation that benefits greatly from continuous coordinate-based 3D scene representations versus voxel-based representations. We evaluate our methodology on 20 medical scans from the RibSeg dataset, which is itself an extension of the RibFrac dataset.

4.1 Introduction

Medical imaging is a critical tool in diagnosing various conditions in patients such as anatomical injuries, malignant tumors, or neurological disorders. It commonly relies on discrete voxel representations from a variety of scanners, such as computed tomography (CT) or positron emission tomography (PET), to provide anatomical and functional information. Although this standardized format enables straightforward operations such as segmentation and classification tasks, voxel grids impose fundamental limitations on geometric analysis due to their discrete nature and fixed resolution. These limitations prevent the extraction of clinically valuable morphological properties, such as precise surface curvature, **medial axes**, and smooth deformation fields, which could enhance diagnostic capabilities and provide a real-time, accurate, and interpretable visualization of the ground truth.

Morphological operations like skeletonization require smooth interpolation between

data points to accurately capture the underlying geometry, which is challenging with discrete voxel grids that introduce artifacts at grid boundaries. While there are a few methodologies that leverage skeletonization [SBB+00b, JS13], the Laplace–Beltrami operator [QBM06, SLK+08], and other shape analysis techniques, the medical field would strongly benefit from continuous coordinate-based representations such as SDFs that enable seamless interpolation at arbitrary resolutions [ZWX+23].

Recent advances in neural 3D scene representation have demonstrated the potential of encoding volumetric data as continuous functions. Examples include implicit representations such as Neural Radiance Fields (NeRF) [MST+21] and explicit representations such as 3D Gaussian Splatting (3DGS) [KKLD23]. When using neural fields or differentiable rendering, the goal is to overfit MLPs to encode entire 3D scenes within their weights. These continuous representations naturally support interpolation by providing smooth function values at any 3D coordinate, making them ideal for geometric operations that require precise spatial derivatives and gradual transitions. These approaches offer reduced memory requirements and enable novel querying capabilities such as novel view synthesis and point interpolation. However, existing applications of neural 3D representations in medical imaging [NGV+24, CFFBT+22b] focus primarily on angular acquisitions for sparse-view reconstruction rather than geometric analysis. Alternative approaches, such as NAF [ZZL22] or Coordinate-Based Internal Learning [SLX+21b], offer continuous representations of measurements and capture the scene in the weights of a simple MLP. Nevertheless, their emphasis is primarily on accurate reconstruction rather than thorough geometrical analysis of a single object of interest, such as a bone structure. Therefore, there is significant interest in the intersection between computational geometry, implicit scene representations, and medical imaging.

We introduce RibPull, the first implicit surface reconstruction methodology for medical imaging that leverages **occupancy fields** to reconstruct complex anatomical

structures from CT scans. While **SDFs** represent the distance from the main object based on the **eikonal equation**, Occupancy Fields represent a simplified form that only predicts if the query point is within or outside the object, which makes them considerably simpler to learn and query than a full signed distance. Figure 4.1 illustrates this distinction: Occupancy and SDF both encode the sign of the query point relative to the surface (inside vs. outside), whereas Occupancy collapses the continuous distance to a single binary label and discards the metric information that SDF retains. A related alternative, the **Unsigned Distance Field (UDF)**, keeps the continuous distance magnitude but drops the sign altogether, which allows it to represent open, non-watertight surfaces at the cost of a well-defined inside/outside region.

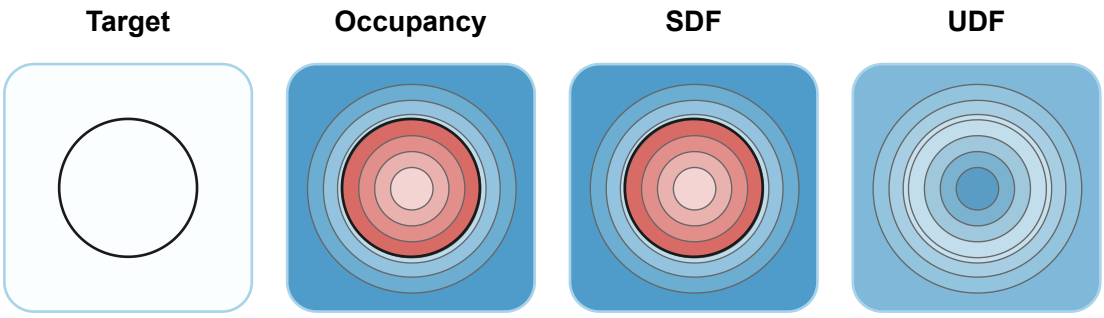


Figure 4.1: Given a target shape (left), Occupancy and SDF both label the interior (red) and exterior (blue) relative to the boundary, but Occupancy only stores a binary inside/outside label while SDF retains the continuous distance to the surface. UDF (right) retains the continuous distance magnitude but is unsigned, so it treats the interior and exterior symmetrically.

We summarize our contributions as follows:

- We employ a neural pull-based optimization framework with minimax entropy regularization to handle the challenging thin structures characteristic of skeletal anatomy. The methodology is specifically designed to accommodate the noise and sparsity inherent in medical imaging data.

- Our results demonstrate that Occupancy Field-based representations advance morphological analysis capabilities in medical imaging while maintaining computational efficiency. We are also able to create a representation that reduces the memory storage of the 3D scene by ~57%, from an average of 4.2 MB of the input point cloud training data to 1.8 MB used by the model's weights.

- We establish a foundation for future extensions to compressed sensing scenarios and broader CT anatomical structure reconstruction. Moreover, this work bridges computational geometry and medical imaging, opening new possibilities for fracture analysis, surgical planning, and biomechanical modeling of anatomical structures.

4.2 Method

In this section, we discuss RibPull. Our methodology generates implicit occupancy field representations of segmented ribcages from CT scans, then extracts an explicit surface directly from the occupancy field's uncertainty values for downstream geometric analysis. CT was chosen for its ability to provide detailed anatomical and structural information of bone tissue through measurement of mass attenuation coefficients. Furthermore, ribcages represent a challenging 3D scene from a geometrical perspective, with fracture detection or rib labeling being tasks that could be tackled with a coordinate-based representation.

Our approach first learns binary occupancy probabilities from sparse point clouds by querying a given number of samples. The surface can then be extracted with a traditional surface reconstruction algorithm, such as Marching Cubes (MC), using the k-th percentile of the occupancy field's uncertainty levels around the surface. Unlike previous

approaches, our method specifically targets geometric analysis applications, enabling resolution-independent queries, direct medial axis extraction, and smooth morphological operations previously impossible with discrete voxel data. The continuous nature of our representation allows for precise interpolation and smooth connected component analysis during skeletonization, avoiding the staircase artifacts and topological inconsistencies common in voxel-based approaches. We validate our approach using 20 segmented ribcages from the RibFrac challenge dataset [JYK+20]. An overview of our methodology can be seen in Figure 4.2.

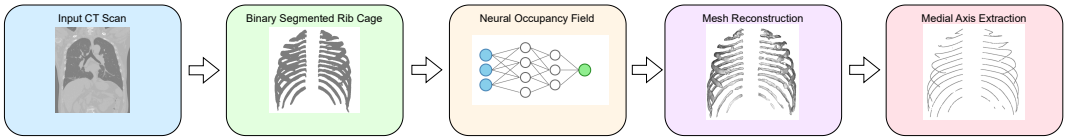


Figure 4.2: From left to right: CT scan input, binary ribcage segmentation using RibSeg, neural occupancy field training using SparseOcc, isosurface extraction using the MC algorithm, and medial axis extraction for morphological analysis using Laplacian-based contraction.

4.2.1 RibSeg

RibSeg [YGW+21] is a ribcage segmentation model that takes a volumetric CT scan as input and converts it to a point cloud using models such as PointNet++ [QYSG17] and DGCNN [WSL+19]. More specifically, the segmentation preserves the floating ribs while removing background structures such as the clavicle, scapula, or sternum. To train the model, RibSeg leverages ground truth segmented ribcages manually annotated by radiologists. We use these manual annotations for all testing.

4.2.2 SparseOcc

Given a point cloud $\mathcal{X}$ obtained from a volumetric CT scan using RibSeg, we propose to learn a neural implicit representation of the corresponding shape that can be used to perform geometrical operations on medical scans. In this context, the SDF is the most popular representation. For an object Ω, the SDF represents the shortest distance from any point to the object's surface, with negative values inside the object, zero on the boundary, and positive values outside. The key property of SDFs is that $|\nabla f_\Omega(\mathbf{x})| = 1$ almost everywhere, known as the eikonal equation.

Still, without dense ground-truth supervision, learning SDFs from point clouds remains a challenging problem. To stabilize training, different smoothness priors leveraging, for example, the gradient [GYH+20] or the Laplacian [BSHKG23] of the field have been introduced. In contrast, Ouasfi et al. [OB24] suggested the use of binary occupancy fields, which are simpler to learn as a binary signal compared to the continuous SDF field that needs to satisfy additional constraints such as the eikonal equation [GYH+20]. The proposed method defines the margin uncertainty $U_\theta(\mathbf{x})$ of a neural occupancy function as the difference between the probability of the query point $\mathbf{x} \in \mathbb{R}^3$ being inside and its probability of being outside the shape. It shows that sampling from the zeros of this function corresponds to sampling from the surface of the shape. Subsequently, the neural occupancy function can be supervised by minimizing the distance between the sampled points and the input point cloud $\mathcal{X}$. Samples from the surface of the occupancy function are obtained using a Newton–Raphson iteration on the margin-uncertainty function $U_\theta(\mathbf{x})$. The corresponding loss is:

$$\mathcal{L}_{samp}(\theta, Q) = \mathop{\mathbb{E}}_{p,q\sim Q} \left\| q - U_\theta(q) \cdot \frac{\nabla U_\theta(q)}{\|\nabla U_\theta(q)\|_2} - p \right\|_2^2, \tag{4.1}$$

where Q consists of pairs of query points q sampled around the input point cloud

and their closest input point in $\mathcal{X}$.

To further stabilize training, an entropy-based regularization loss $\mathcal{L}_{entr}$ is introduced. It guides the occupancy field toward minimal entropy almost everywhere in space and maximizes its entropy at the input point cloud:

$$\mathcal{L}_{\text{entr}}(\theta, \Omega, \mathcal{X}) = \mathop{\mathbb{E}}_{\mathbf{x}\sim\Omega} \mathbb{H}(y|\mathbf{x}, \theta) - \mathop{\mathbb{E}}_{\mathbf{p}\sim\mathcal{X}} \mathbb{H}(y|\mathbf{p}, \theta), \tag{4.2}$$

where Ω represents the 3D spatial domain, $\mathcal{X}$ is the input point cloud, and $\mathbb{H}(y|\mathbf{x}, \theta)$ denotes the conditional entropy of the binary occupancy label y given point location and network parameters θ.

The final loss, with the entropy loss regularized by λ, is:

$$\min_{\theta} \; \mathcal{L}_{samp}(\theta, Q) + \lambda \, \mathcal{L}_{entr}(\theta, \Omega, \mathcal{X}) \tag{4.3}$$

Definition of surface normals. We learn our representation from an unoriented point cloud. The input rib points don't have surface normals pre-calculated, and SparseOcc does not require them. Moreover, we do not learn surface normals during training. Instead, training learns the occupancy field U_θ, which implicitly defines a normal direction at any point q on its decision boundary as the normalized gradient of the field,

$$\mathbf{n}(q) = \frac{\nabla U_\theta(q)}{\|\nabla U_\theta(q)\|}, \tag{4.4}$$

obtainable through automatic differentiation. Since U_θ is an occupancy field, its gradient does not satisfy the eikonal constraint $\|\nabla U_\theta(q)\| = 1$, so only its direction is meaningful as a normal and not its magnitude. Because the network predicts a continuous occupancy probability rather than a binary output, this gradient is well-defined near the surface, with the entropy regularization sharpening the transition around the boundary, and its

orientation follows from the inside and outside convention of the field. In practice, once the sampling and entropy losses have shaped the field, we extract an explicit mesh with Marching Cubes and take the per-vertex normals directly from the mesh geometry.

4.2.3 Laplacian-Based Contraction

A direct benefit of continuous coordinate-based representations is that they allow for geometrical operations on medical scans by querying necessary samples and connecting relevant components. We demonstrate this capability by performing point cloud skeletonization of the reconstructed ribcages.

Among a variety of applicable skeletonization methodologies [SBB+00b, HWCO+13b, CD23], we choose a Laplacian-based contraction [CTO+10] due to its robustness, simplicity, and interpretability. We also observe that it performs particularly well with sparse and noisy data. This algorithm contracts the shape while preserving geometry as dictated by the linear system in Equation (4.5):

$$\begin{bmatrix} W_L L \\ W_H \end{bmatrix} \mathbf{P}' = \begin{bmatrix} \mathbf{0} \\ W_H \mathbf{P} \end{bmatrix} \tag{4.5}$$

$$E(\mathbf{P}') = \|W_L L \mathbf{P}'\|_F^2 + \sum_{i=1}^{n} W_{H,i}^2 \|\mathbf{p}_i' - \mathbf{p}_i\|_2^2 \tag{4.6}$$

where $L \in \mathbb{R}^{n \times n}$ is the Laplacian matrix with cotangent weights, $\mathbf{P} \in \mathbb{R}^{n \times 3}$ is the point cloud sampled from the reconstructed surface after Marching Cubes (distinct from the raw input point cloud $\mathcal{X}$ used to train the occupancy field), $\mathbf{P}' \in \mathbb{R}^{n \times 3}$ represents the contracted point cloud, $W_L, W_H \in \mathbb{R}^{n \times n}$ are diagonal weight matrices balancing contraction and attraction forces, and $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius and Euclidean norms, respectively. The Laplacian provides average neighborhood information for

each point, pulling it toward the local average and resulting in contraction. The process is structured as the minimization of energy E (Equation (4.6)), where the first term represents the contraction force and the second encourages geometry preservation by minimizing the distance between updated and original point clouds across iterations.

4.3 Results

4.3.1 Dataset

The datasets we use are directly provided by RibSeg. RibSeg leverages CT scans sourced from the RibFrac challenge [JYK+20], which consists of a large collection of 3D DICOM CT chest scans. To validate segmentation accuracy against ground truth images, RibSeg used point clouds manually annotated by radiologists. These annotated point clouds are used in our experiments due to their accuracy and structural consistency.

4.3.2 Quantitative and Qualitative Results

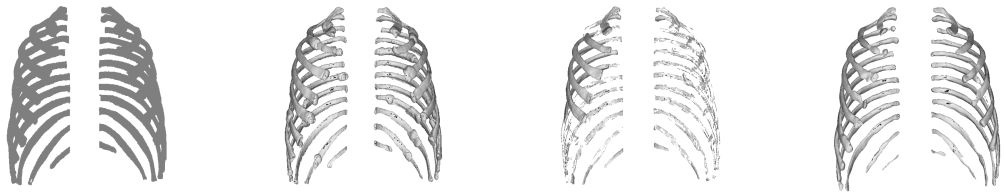
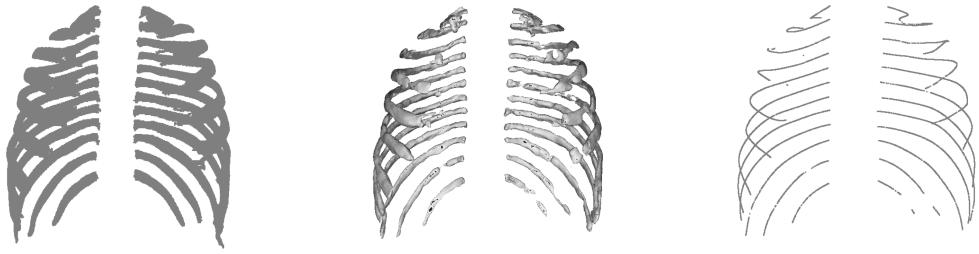
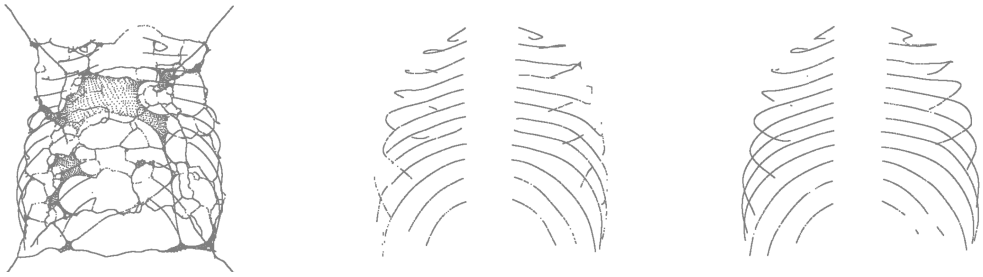


Figure 4.3: Ribcage reconstruction comparison using various implicit surface reconstruction methodologies. Ground truth segmented ribcage from the RibSeg dataset (left), followed by reconstructions using DiGS, Neural-Pull, and SparseOcc methods, respectively.



Ground Truth Reconstruction Skeletonization

Figure 4.4: Left: ground truth segmented ribcage using RibSeg. Middle: surface reconstruction from the implicit occupancy field using the MC algorithm, with the threshold set to its median uncertainty value. Right: the final skeleton using Laplacian-based contraction for medial axis extraction.



DiGS Neural-Pull SparseOcc

Figure 4.5: Skeleton generated by applying Laplacian-based contraction to the outputs of DiGS, Neural-Pull, and SparseOcc.

We compare our methodology against DiGS [BSHKG23] and Neural-Pull [MHLZ20]. In evaluating all models, 100,000 points were sampled randomly across the ribcage. Surface reconstruction is performed using the MC algorithm with the median (50th percentile) uncertainty as a surface level. Our reconstruction results are shown in Figure 4.3 and the skeletonization output in Figure 4.4. We evaluate reconstruction quality using three metrics: Chamfer-L1, Chamfer-L2, and Hausdorff distance. The results in Table 4.1 show that SparseOcc performs best in Chamfer-L1 and Chamfer-L2,

while Neural-Pull is superior in Hausdorff distance. Figure 4.5 provides a qualitative evaluation of skeletonization from the output of each model; we observe that DiGS outputs often create large connected artifacts outside the object of interest, resulting in a greatly deformed skeleton.

Table 4.1: Quantitative performance of uncertainty-based Occupancy Field extraction. Distances between the ground truth and reconstructed point clouds after sampling 100,000 points with normalization applied.

Metric	DiGS	Neural-Pull	SparseOcc
Chamfer-L1 ↓	8.04 ± 10.1	5.81 ± 0.38	**1.55** ± **0.24**
Chamfer-L2 ↓	5.36 ± 6.71	3.90 ± 0.26	**1.04** ± **0.16**
Hausdorff ↓	36.8 ± 23.0	**8.25** ± **3.22**	12.6 ± 4.39

4.4 Discussion

Chamfer-L1 and Chamfer-L2 average the nearest-neighbor distance over all sampled points, while the Hausdorff distance reports only the single worst-case deviation, and that difference explains the split in Table 4.1: SparseOcc reconstructs the ribcage more accurately almost everywhere, but it incurs a small number of large, localized errors that Neural-Pull avoids. The occupancy field encodes only a binary inside/outside label and discards the metric distance that an SDF retains (Figure 4.1), so it gives no distance gradient toward the surface in regions where the boundary is ambiguous. Ribs are thin, high-curvature, and often closely adjacent, and in these regions the Marching Cubes extraction can produce isolated stray components far from the true surface. These artifacts inflate the worst-case Hausdorff distance without doing much to the averaged Chamfer distances, so SparseOcc wins on Chamfer-L1/L2 but loses to Neural-Pull on Hausdorff.

Methods designed for dense inputs struggle on this task. DiGS performs substantially

worse than both other methods across every metric, with a Hausdorff distance several times larger and a high variance across scans (Table 4.1), and Figure 4.5 demonstrates the issues: DiGS frequently produces large connected components outside the object of interest, which severely deform the extracted skeleton. Its divergence-guided formulation relies on a smooth, well-sampled surface to constrain the field, and the sparse, unevenly sampled ribcage point clouds do not follow that assumption. Neural-Pull pulls query points onto the surface via a learned signed distance, so it is more robust than DiGS, but it tends to over-smooth. This is exactly the reason why it beats SparseOcc on worst-case Hausdorff distance but loses to it on average Chamfer accuracy. SparseOcc's occupancy formulation was built for sparse, noisy inputs in the first place, and that is why it strikes the best balance for this anatomy.

4.5 Conclusions

We have presented RibPull, a methodology that combines RibSeg for segmenting ribcages, SparseOcc for encoding the CT scene in a neural network's weights, and Laplacian contraction for medial axis extraction. Our main goal was anatomy preservation, which is critical for analyzing CT scans. Following the training of the neural occupancy fields, we observe that SparseOcc outperforms other methodologies on Chamfer-L1 and Chamfer-L2 metrics by accurately preserving the topological and geometrical structure of the object. This is expected as SparseOcc is specifically designed for sparse and noisy point clouds, while dense point cloud models typically collapse and are less robust under such conditions.

Our methodology provides a simple, minimal, and interpretable visualization of ribcage anatomy that enables fracture detection and scoliosis assessment, while serving as a potential tool for surgical planning. The Laplacian-based contraction performs well

for medial axis extraction, demonstrating the advantages of continuous representations for geometric operations that would be challenging with traditional voxel-based approaches. One of the benefits of this architecture is that it is easily transferable to other datasets. While we have demonstrated the effectiveness of this methodology with ribcages, it would be straightforward to adapt and test for other bone structures and shapes, such as pelvic, dental, knee, elbow, or foot scans.

4.6 Limitations and Future Work

While our work represents a novel approach in extracting, analyzing, and visualizing anatomical structures, several limitations remain to be addressed. The reconstruction accuracy does not yet reach desired levels, and some information loss occurs during training. Additionally, our evaluation was performed on datasets that have undergone manual curation by radiologists, which may not reflect real-world clinical scenarios with inherent noise and artifacts.

Future work will focus on improving 3D scene encoding and skeletonization performance and testing our methodology on sparse, noisy CT image scans. We also intend to investigate the preservation of geometric properties such as surface normals and curvature under extreme compressed sensing constraints, which could further expand the clinical applicability of our approach. Finally, we acknowledge the difficulty with sparse and noisy data from real scanners and will explore optimal denoising methodologies for the input point cloud format, such as DBSCAN [EKSX96], while making the overall methodology robust to a wide variety of input ribcages from diverse demographics.

Chapter 5

ReNAF: Regularized Neural Attenuation Fields with 3D Reconstruction Priors for Sparse-View CT

This chapter is based on:

Emmanouil Nikolakakis, Daksh K. Shah, Julia Wong, Julie Digne & Razvan Marinescu. ReNAF: Regularized Neural Attenuation Fields with 3D Reconstruction Priors for Sparse-View CT. *Ready to be submitted* [NSW+26].

Traditional Computed Tomography (CT) medical image reconstruction relied on analytical algorithms such as FDK or SART. While machine learning models achieve impressive reconstruction quality, they tend to overfit to the training distribution and fail to generalize to unseen anatomies or acquisition protocols. Rather than viewing

traditional algorithms as obsolete, we argue they serve as enduring physics-informed priors that can complement and stabilize learning-based reconstruction models. Implicit neural representations, specifically Neural Radiance Fields adapted for CT as NAF, offer a compelling alternative by overfitting an MLP to a single CT acquisition set. NAF overcomes the generalization problem by being trained on a single patient rather than a large database, but remains vulnerable under extremely sparse input, where the underconstrained optimization leads to poor convergence. To alleviate this, we propose Regularized Neural Attenuation Fields (ReNAF), which incorporates a traditional algorithmic reconstruction as a physics-informed 3D prior to guide and regularize the MLP optimization. We show that this prior allows us to handle compressed sensing where only a few radiographs are provided to the model. On PSNR and SSIM scores, ReNAF achieves competitive performance and remains robust down to 10 input radiographs, conditions where existing methods degrade significantly.

5.1 Introduction

Medical imaging has long been associated with grid representations, such as pixel matrices in 2D images or voxel matrices in 3D images. Grid structures offer multiple benefits, in particular with visualization and rendering on monitors or performing linear algebra operations, which can be useful in signal processing or physics-informed reconstruction. More recently, DL models are designed to work seamlessly with matrices as inputs. Nevertheless, while they offer structure and simplicity, grids are memory-intensive since empty or non-informative voxels still encode unnecessary information, an example being the air attenuation values in CT scans. Moreover, a reconstruction method in healthcare should be fast for a quicker diagnosis and reliable for various data distributions. DL methods often fail in that regard [CKD+25].

Recently, the field of novel view synthesis (NVS) has seen a major breakthrough with the emergence of ISR, with the introduction of Neural Radiance Field (NeRF) [MST+21] and 3D Gaussian Splatting (3DGS) [KKLD23]. ISR models provide a continuous representation of the scene that can be rendered from any viewpoint. These methods were later adapted for medical image reconstruction through NeRF-based approaches [CFFBT+22a, CWY+24, ZH23, RKA+21], implicit SDF representations [NODM25], and 3DGS approaches [LFL+25, NGV+24, CLW+24, ZLC+24] for representing CT scans and generating additional radiographs using NVS. Even so, the main question remains: why use ISR over explicit voxel grids in clinical applications, and can it provide direct benefits for radiologists?

ISR immediately delivers two direct benefits. The first is the ability to infer the values of missing data points and create an optimized representation, which is ideal for sparse data, often the case in medical scans. This is due to a large number of homogeneous regions in the scanned subject's body and a large portion of the field of view (FOV) typically being air, which is information that does not need to be stored explicitly. As a result, a neural network whose weights overfit to a single 3D medical scan requires less storage space than an equivalent explicit voxel grid representation, such as DICOM or NIFTI [SSdS+21]. The second is that ISR only uses data from the scanner acquisition to create a full 3D scene representation. Consequently, it offers a fast and accurate reconstruction without the artifacts of analytical and iterative methodologies and without requiring the large amount of training data and training time associated with DL methodologies [ABVGT+21]. This also means that the method is generalizable and can be applied to various scanner protocols and population demographics.

ISR methodologies such as NAF [ZZL22] provide volumetric querying that directly retrieves the linear attenuation coefficient (LAC) for the given coordinate. This is optimal for radiologists, who study LAC in Hounsfield Units (HU) along with qualitatively evalu-

ating the reconstructed CT scans before making a diagnosis. Nevertheless, there are still improvements that can be made, with a large emphasis on fast and accurate reconstruction with minimum overall patient exposure to radiation in a CT scan [PMT+16a].

Priors have already been shown to help implicit representations cope with sparse measurements. NeRP [SPX22] pretrains its MLP on a patient-specific prior image before fine-tuning against the sparse acquisition, substituting the prior for a large training dataset. Still, that prior is an external reference image rather than information derived from the sparse acquisition itself, and it does not exploit the projection geometry used in analytical CT reconstruction. We instead investigate using a classical algorithmic reconstruction of the same sparse-view acquisition as a **physics-informed 3D prior**, directly regularizing the NAF optimization without requiring a separate reference scan.

To address these challenges, we introduce ReNAF, a model that improves on the original NAF with a focus on fast, improved reconstruction quality while maintaining robustness under ultra-sparse constraints (as few as 10 views for training). We make the following contributions:

- We present an ISR-based method that leverages a 3D CT reconstruction as a physics-informed prior. We investigate FDK, SART, CGLS, and ASD-POCS algorithms to choose the most appropriate for our training process.

- We perform thorough investigations of **compressed sensing** constraints. Our results show that even when providing as few as 10 radiographs, our reconstruction still produces usable results, while we identify 5 radiographs as the breaking point.

- We compare our model's reconstruction metrics against other NAF-based models. Our design is consistently robust under extreme sparsity on four datasets for different body parts. Moreover, we show that our model converges in a much shorter training time.

5.2 Related Work

Classical Methods. Classical CT reconstruction algorithms have involved analytic methodologies such as FDK [FDK84], or iterative algorithms such as CGLS [HS+52], SART [AK84], Ordered Subsets – Modified Simultaneous Iterative Reconstruction Technique (OS-MSIRT) [BJ19], and ASD-POCS [SP08], among others. While these methodologies are reliable and generalize well due to their mathematical and statistical foundation, they also have disadvantages. FDK, while fast, produces significant streaking artifacts and statistical noise under limited sparse views. Iterative algorithms can reduce the total noise in the reconstruction, but can be computationally expensive.

Deep Learning approaches. DL models had an impressive impact on the field by producing high-quality reconstructions, in particular for low-dose constraints [CZK+17b, YYZ+18b, GS22b], but required large datasets, long training times, and often failed to generalize well to out-of-distribution data. Thus, their transition to clinical settings was often a challenging process [DBS+22].

Implicit approaches. ISR has been an interesting approach to 3D scene representation and reconstruction over the last few years. NAF learns from the subjects' 2D projections by parameterizing LAC as a continuous function of 3D spatial coordinates using an MLP. To capture high-frequency details efficiently, NAF utilizes a learning-based hash encoding strategy [MESK22a] for points along each ray, exploiting the inherent sparsity and smoothness of human organs. For volume rendering, NAF employs a stratified sampling approach and optimizes the network by minimizing the error between real and synthesized projections. The research closest to our work was published in 2025, when Zhou et al. [ZFM+25] introduced ρ-NeRF. This model extends the 3D spatial coordinate input to NAF and SAX-NeRF [CWY+24], with a fourth coordinate corresponding to an

estimate of the LAC at that location. The LACs, initialized using a variety of classical CT reconstruction algorithms, are encoded with a simple linear transformation and concatenated with the encoded 3D coordinates. In 2025, Zhu et al. [ZIK+25] introduced Neural Attenuation Surfaces (NeAS), an SDF-based framework for volumetric CT reconstruction that utilizes two decoupled MLPs to represent surface geometry using SDFs and spatial attenuation. The SDF MLP allows for 3D surface representations, which create a feature vector as input for the NAF-like attenuation MLP. Furthermore, NeAS proposed a multi-material architecture for multi-object surface extraction, where the LAC of distinct materials is constrained by disjoint, pre-defined attenuation bounds.

5.3 Methodology

Baseline: Neural Attenuation Fields

Since ReNAF extends NAF [ZZL22], we first describe the baseline model in the notation used throughout this chapter. NAF represents the linear attenuation coefficient (LAC) as a continuous function $\mu(\mathbf{x})$ of a 3D spatial coordinate $\mathbf{x}$, parameterized by an MLP Φ together with a learnable position encoder Θ. Training is self-supervised and needs no external CT volumes – the only supervision is the set of 2D projections of the subject being reconstructed. The pipeline has four stages.

Ray sampling. For each detector pixel, NAF casts a ray through the volume according to the scanner geometry and samples N points along the segment where the ray intersects the reconstruction cube, using the stratified scheme of [MST+21]: the ray is divided into N evenly spaced bins and one point is drawn uniformly within each. We choose N larger than the target volume resolution so that every voxel a ray passes through gets at least one sample.

Position encoding. A plain coordinate MLP is biased toward low-frequency functions and struggles to recover fine anatomical detail, so NAF encodes each sampled coordinate with the multiresolution hash encoding of Instant-NGP [MESK22b] instead of the sinusoidal frequency encoding used in the original NeRF. This suits CT well: human organs are largely homogeneous media whose attenuation only varies near tissue boundaries, so a sparse, learnable encoder can spend its capacity on the few regions that actually need it. The hash encoder's feature vector is also much smaller than a frequency encoder's, so the MLP itself stays compact and reconstruction takes on the order of tens of minutes.

Attenuation prediction. The encoded coordinate is passed to the MLP Φ, which outputs the scalar attenuation coefficient $\mu(\mathbf{x})$ at that location. The network is deliberately small, since the hash encoder already does most of the work.

Projection synthesis. Given the predicted coefficients along a ray, NAF synthesizes the corresponding detector reading through the discrete Beer–Lambert model introduced in Chapter 2 (Equation (2.1)):

$$\hat{I} = I_0 \exp\left(- \sum_{i=1}^{N} \mu(\mathbf{x}_i)\, \delta_i \right), \tag{5.1}$$

where I_0 is the incident intensity, $\hat{I}$ the synthesized pixel intensity, and $\delta_i = \|\mathbf{x}_{i+1} - \mathbf{x}_i\|$ the spacing between adjacent samples. The model is optimized by minimizing the squared error between synthesized and measured projections over a batch of rays $\mathcal{B}$:

$$\mathcal{L}_{\text{NAF}}(\Theta, \Phi) = \sum_{r \in \mathcal{B}} \left\| \hat{I}(r) - I(r) \right\|_2^2, \tag{5.2}$$

updating both the hash encoder Θ and the attenuation network Φ. Once trained, the final volume is recovered by querying Φ at every voxel coordinate of the desired grid.

This projection loss (Equation (5.2)) is precisely the rendering term that ReNAF retains as the first component of its objective (Equation (5.3)); our contribution is the addition of a prior loss that constrains the reconstruction directly in the volume domain, described next.

ReNAF

Our architecture expands on NAF models by integrating a prior loss. An overview of our ReNAF model can be seen in Figure 5.1.

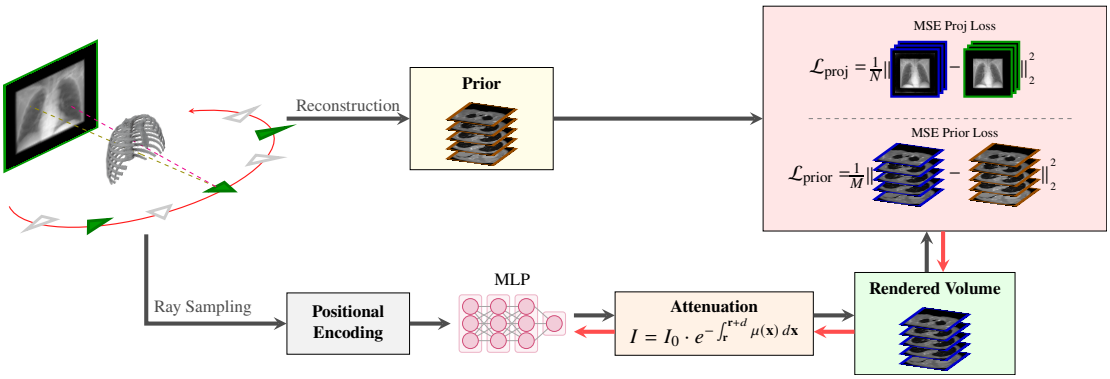


Figure 5.1: ReNAF architecture: The red arrows show the gradient flow from the loss function to the weights of the MLP, while the black arrows show the output and input between blocks.

As established in Chapter 2 (Equation (2.1)), the synthesis of LAC follows the Beer–Lambert law, where I_0 is the incident X-ray intensity, I is the intensity measured at the detector, and $\mu(\mathbf{x})$ is the LAC along the ray path.

Our methodology relies on using a reconstruction prior to create a physics-informed model. We employ the same digitally reconstructed radiographs (DRRs) used as training data to create a 3D reconstruction using traditional analytic (FDK) or iterative (SART, CGLS, ASD-POCS) algorithms via the TIGRE toolbox [BDHS16].

In addition to the mean squared error (MSE) loss between the pixel intensity value on a rendered image and the ground truth (as in NAF), we add a prior loss (Equation (5.3)). The prior loss is an MSE between N_v randomly sampled voxels of the reconstruction prior and the rendered volume, weighted by a factor that is cosine-annealed during training to allow for a smooth geometrical initialization. This decay is useful since the loss is computationally intensive and has the most impact in the first stages of the training process. Our final loss function is:

$$\mathcal{L}_{Total} = w_{rendering} \cdot \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\hat{I}_i - I_i\right)^2 + w_{prior} \cdot \frac{1}{N_v} \sum_{j=1}^{N_v} \left(\hat{\mu}_j - \mu_j\right)^2 \tag{5.3}$$

where $N_p = N \cdot M$ is the total number of pixels in each 2D projection I used for the rendering loss, with N, M the detector array dimensions defined in Table 2.2, and N_v the number of voxels sampled for the prior loss from each volume μ. The prior weight decays as:

$$w_{prior}(t) = \frac{w_{initial}}{2} \left(1 + \cos\left(\frac{\pi t}{T_{decay}}\right)\right), \quad t < T_{decay} \tag{5.4}$$

We set $w_{initial} = 0.5$ and the total decay horizon T_{decay} to 20% of the training epochs. Since calculating the prior loss for all voxels in a 3D image would be computationally expensive, we randomly sample 8,192 voxels from the current network to approximate the total loss.

5.4 Results

5.4.1 Setup

All training and rendering were performed on a single NVIDIA RTX A4000. We ran all ReNAF tests for 150K training steps. We chose to go with the prior generated by ASD-POCS, which gave the best results as shown in Table 5.1. The ASD-POCS prior took approximately 36 seconds for the chest dataset and 4 minutes for the jaw and foot datasets to generate. The abdomen dataset was an exception due to its significantly larger volumetric size; thus, we used the CGLS prior across all model training. Before settling on these choices, we tried several settings for each classical method; the exact settings we tested, along with how bright each resulting reconstruction was (in Hounsfield Units) and how long it took to compute, are listed at the end of this chapter in Table 5.4 and Table 5.5.

Since the authors did not respond to our attempts to contact them, our results on ρ-NeRF and NeAS are based on our own implementation. We follow the information shared in the publications, since no code was provided for these papers. For NeAS, we report metrics for both the 1-material (1M) and 2-material (2M) design in Table 5.1; in Table 5.3, evaluations are performed for the 2-material (2M) design. We will provide a few details for their implementation below.

5.4.2 ρ-NeRF

We reimplemented ρ-NeRF following the architecture described by Zhou et al. [ZFM+25]. The novelty of the paper was in concatenating the encoded initialized attenuation value before producing the output refined attenuation. The implementation included making a simple 8-layer learnable linear transformation or encoder to transform

the attenuation value to a feature vector f that gets concatenated with the encoded 3D coordinates, finally producing a 32-dimensional vector. Since the anatomic prior for ρ-NeRF is generated using traditional reconstruction algorithms and is therefore discrete, the required attenuation value for any continuous 3D coordinate is initialized using trilinear interpolation. We chose this initialization scheme based on the paper's results, which show that it performs better than the alternatives of nearest neighbor and mean value. It is important to note that ρ-NeRF underperformed given the published results from the original authors on the same dataset, so further investigation into the implementation or an explanation for the discrepancy is required.

5.4.3 NeAS

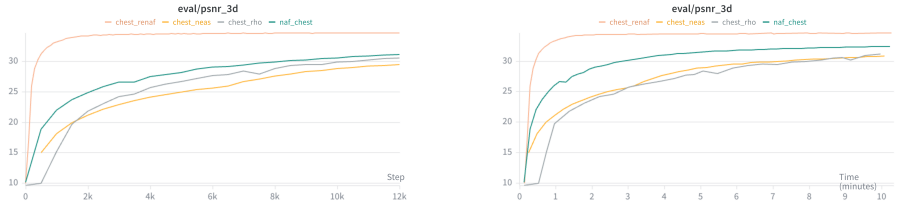
We reimplemented NeAS following the architecture described by Zhu et al. [ZIK+25]. The model represents the scene with two coupled MLPs, described in turn below: an SDF network Θ_{sdf}, which maps an encoded 3D coordinate to a signed distance d_{sdf} and a feature vector f, and an attenuation network Θ_{att}, which maps f to a raw attenuation value $\bar{\mu}$. We use the surface boundary function $\Omega(d_{sdf}, s) = \text{sigmoid}(-s\, d_{sdf})$, with the steepness parameter s initialized to 20 and learned jointly with the rest of the network, to mask $\bar{\mu}$ so that attenuation is only retained inside the predicted surface. The final layer of Θ_{att} uses the bounded activation $\alpha \cdot \text{sigmoid}(x) + \beta$, which restricts the predicted attenuation for each material to a manually chosen range $[\beta, \alpha + \beta]$, estimated per dataset from the attenuation histogram as described in the original paper. We use hash encoding for both MLPs, matching the paper's 14-level, multiresolution configuration, and optimize the network with the intensity MSE loss plus an Eikonal regularization term on the SDF gradients. For datasets with two relevant materials (e.g., bone and soft tissue), we use two independent attenuation networks Θ_{att1} and Θ_{att2}, and select between

their outputs with the hard selector from [ZIK+25]: the coefficient predicted by the inner material is used wherever its signed distance is negative, and the outer material's coefficient is used everywhere else.

We later extended this implementation to support more than two materials, replacing the independent per-material networks with a shared attenuation backbone and a soft, differentiable selector, and automating the manual α, β tuning described above by fitting a Gaussian Mixture Model to the attenuation distribution. This extension is not used for the comparisons in this chapter; we discuss it further as future work in Chapter 6.

5.4.4 Evaluation

We demonstrate our architecture's ability to generate higher metric values (Table 5.1) while preserving anatomical details under extreme sparsity with 5 radiographs as input (Table 5.2). Table 5.3 demonstrates how ReNAF outperforms NAF for reduced input radiographs, while ρ-NeRF fails to optimize for 5 or 10 radiographs. Figure 5.2 shows that our model converges much faster than the other methods. We also notice that our scores for the abdomen dataset are lower, which we attribute to our inability to generate an accurate prior in a time-efficient way using ASD-POCS for that dataset, thus using the CGLS prior instead. Figure 5.3 illustrates the resulting 3D volume renderings compared to alternative methods. In contrast, Figure 5.4 and Figure 5.5 show reconstructed slices and pixel-wise error maps, respectively, across varying percentages of input radiographs.



(a) PSNR vs. Training Steps

(b) PSNR vs. Training Time

Figure 5.2: Comparison of PSNR convergence across methods by training steps and training time for the chest dataset.

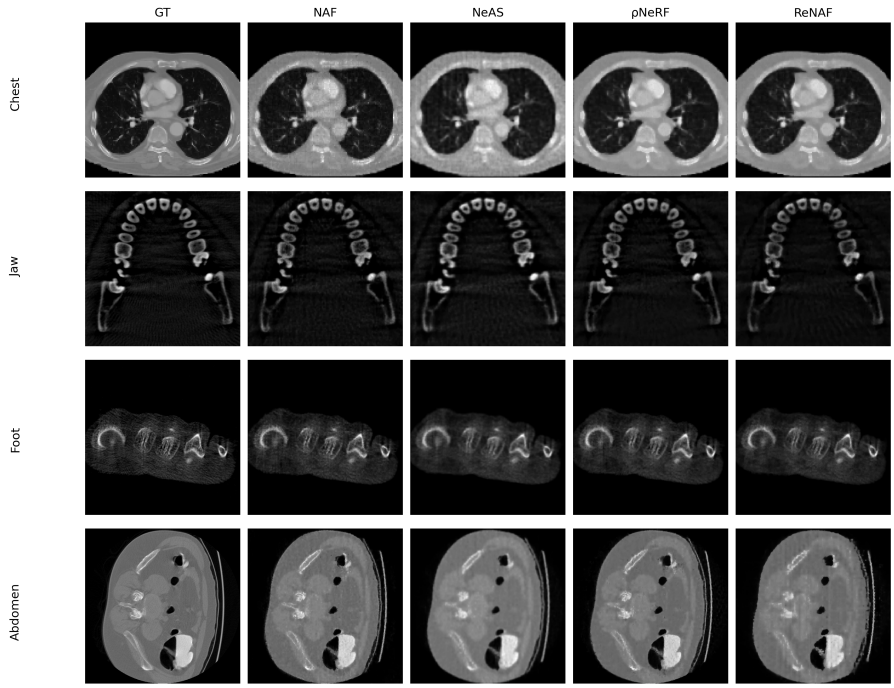


Figure 5.3: 3D volume rendering results compared to alternative methodologies.

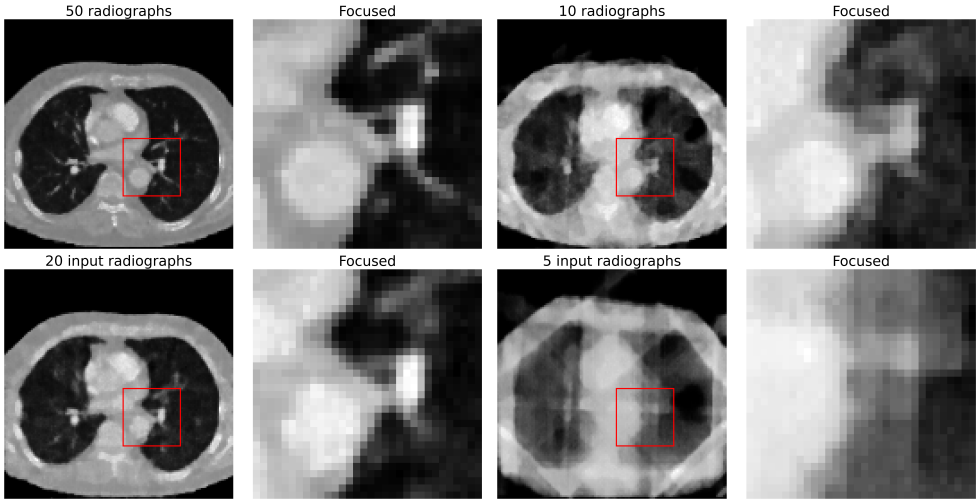


Figure 5.4: Four slices obtained from ReNAF trained using various percentages of the radiographs from the original chest scan. Rendered images are 128×128, patches are 32×32.

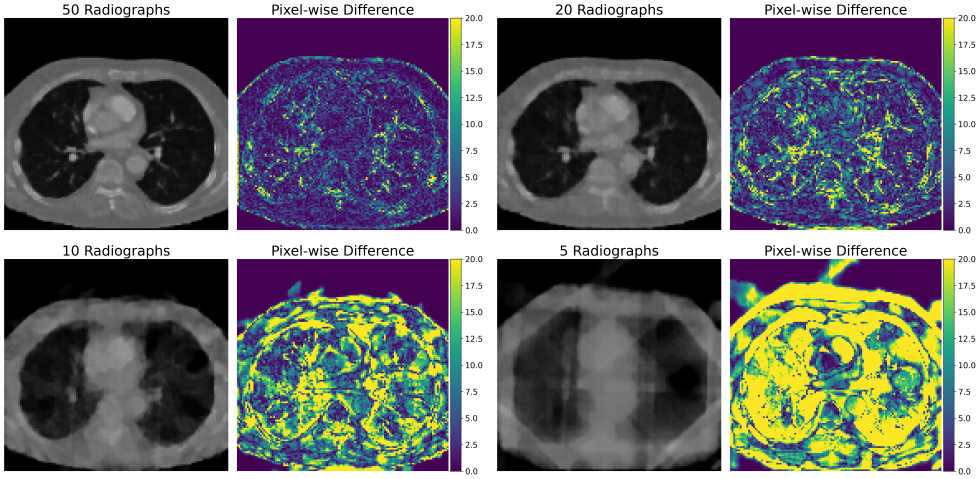


Figure 5.5: Pixel-wise difference between rendered images and ground truth for 50%, 20%, 10%, and 5% of projections used for training. Image resolution is 128×128.

Table 5.1: 3D PSNR / 3D SSIM performance. Top: Performance of ReNAF for the chest dataset using various reconstruction priors. Bottom: Comparison with other models (50% of views).

ReNAF performance using various priors (Chest)				
FDK	SART	CGLS	ASD-POCS	
33.29/.96	34.49/.97	34.28/.97	34.85/.97	
Model Comparison				
	Chest	Jaw	Foot	Abdomen
IntraTomo3D [ZIL+21]	31.94/.95	31.95/.91	31.43/.91	30.43/.90
NAF	33.05/.96	34.14/.94	31.63/.94	**34.45/.95**
1M-NeAS Hash	30.38/.89	33.77/.88	31.62/.90	32.58/.84
2M-NeAS Hash	32.76/.93	33.57/.83	31.32/.89	32.76/.84
ρ-NeRF	33.9/.97	35.38/.95	**31.94/.94**	32.43/.93
ReNAF (Ours)	**34.85/.97**	**35.49/.95**	31.77/.93	31.15/.93

Table 5.2: 3D PSNR / 3D SSIM performance across various percentages of used images.

% of views	Chest	Abdomen	Jaw	Foot
50%	34.85/.97	31.15/.93	35.49/.95	31.77/.93
20%	30.85/.93	28.11/.86	29.63/.85	28.58/.9
10%	25.77/.83	25.54/.82	26.88/.7	26.22/.86
5%	21.55/.69	17.44/.64	23.1/.62	22.47/.79

Table 5.3: 3D PSNR / 3D SSIM performance comparison between models across various percentages of input images for the chest CT scan.

% of views	ReNAF	NAF	ρ-NeRF	NeAS
50%	**34.85/.97**	33.05/.96	33.9/.97	32.76/.93
20%	**30.85/.93**	29.83/.92	27.94/.88	26.27/.76
10%	**25.77/.83**	24.28/.77	11.07/.27	23.32/.63
5%	**21.55/.69**	18.27/.55	9.96/.23	17.07/.31

5.5 Discussion

ρ-NeRF and ReNAF both incorporate an anatomic prior derived from a classical reconstruction, yet ReNAF remains consistently more robust as the number of input views decreases (Table 5.1), and the difference seems to come down to *how* each method

supplies the prior, not the prior itself. ρ-NeRF appends the estimated LAC as an additional input coordinate, so the prior is only advisory: the network has to learn to exploit the extra channel, and nothing constrains the predicted attenuation to stay close to it. ReNAF instead supplies the prior as an explicit regularization loss that directly penalizes deviation of the predicted volume from the reconstruction. At 50% of views, where supervision is plentiful, this barely matters and the two models are comparable, but under extreme sparsity it does. When training with only 5–10 radiographs, the projection loss constrains the volume too weakly for ρ-NeRF to learn a reliable mapping from its input prior to the output, and its reconstruction collapses (Table 5.3). ReNAF's loss-based prior keeps constraining the predicted attenuation wherever it is defined, regardless of how many projections remain, so it degrades gracefully instead.

The prior in ReNAF does add a per-step cost, since the regularization term is evaluated against a full 3D volume. However, this term is annealed out after the first 20% of training, once it has guided the optimization toward an anatomically plausible solution. The result is markedly faster early convergence without a lasting computational penalty. Indeed, as shown in Figure 5.2, ReNAF attains higher metrics than the baselines within the first ten minutes of training.

NeAS couples an SDF with the attenuation field, borrowing surface-based ideas from computational geometry instead of representing the volume directly. Its material model, however, assumes a small number of discrete materials with *non-overlapping* attenuation ranges, separated by a hard, non-differentiable selector and delimited by attenuation bounds (α, β) that must be tuned by hand for each scene [ZIK+25]. This design suits the settings NeAS was demonstrated on, such as knee and skull surfaces or leg and head phantoms with at most two materials (bone and skin/muscle), but it does not transfer cleanly to the CBCT anatomies in our benchmark, where many tissues contribute overlapping, continuously varying attenuation and no small set of hand-chosen bounds

captures the full distribution. As a result, NeAS yields lower volumetric fidelity on these scenes despite its geometric sophistication. Notably, the authors of NeAS identify the manual setting of α and β, together with the restriction to at most two materials, as the method's principal limitations, and explicitly propose extending it to three or more materials as future work.

Geometric information, then, can meaningfully improve reconstruction quality, but only with a material model expressive enough for real anatomy: differentiable material assignment and automatically determined attenuation bounds, instead of a hard selector and hand-tuned ranges. This is what motivates our extension of NeAS to an arbitrary number of materials, K-NeAS, discussed in Section 6.5.

While our model generates high 3D metric values with much faster training, we experience sub-optimal performance when testing with large volumetric data such as the abdomen dataset, where generating an accurate ASD-POCS prior in a time-efficient manner is challenging.

5.6 Conclusion

We have presented ReNAF, a physics-informed NAF-based model that leverages a 3D prior to provide fast and accurate 3D reconstructions. Our results demonstrate the model's ability to deliver robust reconstructions preserving the global anatomy even under extremely sparse-view dataset constraints. It also converges faster than other NAF-based models.

Our future work includes two directions. The first is to incorporate anatomy-aware sampling of points across the ray by leveraging the image gradient of the prior to identify anatomical structures of high importance, replacing hierarchical and stratified volume sampling in CT scans. Second, we will study cases where previous scans of a patient

are available to provide the reconstruction priors. An earlier higher-dose reconstruction could complement a more recent sparse-view scan to supply a high-quality reconstruction that captures the differences between the two. We plan to validate this hypothesis with time-series patient experimental data.

5.7 Additional Reconstruction Details

Table 5.4 lists the exact hyperparameter settings tried for each classical reconstruction prior (SART, CGLS, ASD-POCS), and Table 5.5 reports the mean HU value and reconstruction time achieved by each setting across our four datasets.

Table 5.4: Reconstruction hyperparameters for all priors. FDK has no tunable hyperparameters.

Method	ID	n_{iter}	λ	Block	α	ϵ	n_g	δ
FDK	–	–	–	–	–	–	–	–
SART	1	30	1.0	1	–	–	–	–
SART	2	100	0.8	20	–	–	–	–
SART	3	200	0.5	10	–	–	–	–
SART	4	50	1.2	5	–	–	–	–
CGLS	1	20	–	–	–	–	–	–
CGLS	2	50	–	–	–	–	–	–
CGLS	3	100	–	–	–	–	–	–
CGLS	4	200	–	–	–	–	–	–
ASD-POCS	1	30	–	20	0.001	10^{-4}	20	-0.005
ASD-POCS	2	50	–	10	0.01	10^{-5}	50	0.0
ASD-POCS	3	80	–	5	0.05	10^{-6}	100	-0.01
ASD-POCS	4	100	–	1	0.002	10^{-3}	30	0.005

Table 5.5: Reconstruction statistics after HU conversion and reconstruction time across our four datasets. Dashes (—) indicate reconstructions not yet performed for that anatomy. FDK has no tunable hyperparameters.

		Chest		Foot		Jaw		Abdomen	
Method	ID	Mean HU	Time (s)	Mean HU	Time (s)	Mean HU	Time (s)	Mean HU	Time (s)
FDK	–	-474.84	0.07	-575.23	0.31	-434.07	0.24	-540.02	1.45
SART	1	-477.28	36.25	-642.70	222.44	-345.53	215.41	—	—
SART	2	-478.86	119.48	-643.81	714.57	-346.92	705.74	—	—
SART	3	-479.93	224.99	-643.99	1416.50	-347.48	1419.13	—	—
SART	4	-478.01	70.05	-643.35	370.12	-346.41	352.62	—	—
CGLS	1	-471.97	3.73	-609.52	14.82	-341.06	15.19	-472.88	64.93
CGLS	2	-472.42	4.88	-608.87	35.60	-344.45	34.75	-449.66	157.12
CGLS	3	-472.42	5.54	-608.70	38.92	-344.57	36.30	-442.10	266.49
CGLS	4	-472.42	5.33	-608.70	39.55	-344.57	36.19	-437.26	397.52
ASD-POCS	1	-472.12	35.97	-641.54	239.55	-311.36	238.67	—	—
ASD-POCS	2	-471.01	15.25	-638.77	153.27	-288.95	106.65	—	—
ASD-POCS	3	-468.97	19.26	-636.50	295.65	-276.47	251.45	—	—
ASD-POCS	4	-471.87	27.54	-641.30	285.40	-301.58	201.46	—	—

Chapter 6

Conclusions and Future Work

6.1 Summary of Contributions

This dissertation has explored learned gridless representations as an alternative methodology for reconstructing 3D CT images from radiographs. Moreover, the encoded 3D scene in off-grid continuous representations serves as a memory-efficient methodology to store CT scans due to their sparsity. Additionally, we have suggested and demonstrated that continuous representations are useful for preserving the topology of the scene and performing geometric operations with clinical value. Finally, we observe a methodology that leverages the benefits of machine learning without requiring large medical datasets, which are sparse in the field, primarily due to privacy concerns for existing scans and the complexity of acquiring new scans directly from clinical institutions.

6.2 Synthesis Across Chapters

The main motivation for this dissertation was identifying reconstruction methodologies that remain robust under sparse acquisition constraints. Such a methodology

should encourage a CBCT protocol with less radiation administered to the patient, with enough reliability to preserve the clinical information for detection, recognition, and discrimination.

GaSpCT (Chapter 3) was the first evidence that learned gridless representations are viable for CBCT at all. An explicit set of 3D Gaussians reconstructed a CBCT scene from half or less of the usual projection views while staying more memory-efficient than a voxel grid, and it still held up down to 18 of 360 views. RibPull (Chapter 4) then asked a different question. Instead of investigating the minimum number of views a reconstruction needs, we explored what a gridless representation makes possible once you have one. Learning an implicit occupancy field instead of a voxel grid preserved the ribcage's topology well enough to extract a medial axis directly, something voxel-based methods handle poorly under sparsity. ReNAF (Chapter 5) went back to the reconstruction problem and pushed the sparsity further. As opposed to a fraction of a large view set, it was tested down to as few as 10 absolute radiographs across four different anatomies, a regime where the baseline NAF collapses, and regularizing an implicit attenuation field with a classical reconstruction as a physics-informed prior kept it robust well past that point.

So the three chapters move from an explicit representation (GaSpCT) to implicit ones (RibPull, ReNAF), and the question being asked shifts along with it: first whether learned off-grid reconstruction works at all, then what it buys you geometrically, then how far it can be pushed under sparsity. Memory efficiency, topological fidelity, and robustness all matter for a reconstruction method meant for clinical use, and each chapter happens to supply the evidence for one of them, each with a different representation from the taxonomy introduced in Chapter 2.

6.3 Limitations

The work in this dissertation also reveals several limitations that apply to the current state of the art in learned gridless representations, two of which are shared across all three contributions. First, these methodologies overfit to a single scene. Each scan must be reconstructed by training a model from scratch, which means the reconstruction cost is paid per patient rather than once, and the learned representation does not transfer to any other scan. Moreover, since CT acquisitions vary substantially in anatomy and attenuation distribution, a configuration tuned for one region does not carry over to another, and none of the off-grid methodologies presented here applies uniformly across anatomies without adjustment. Second, and more fundamentally, every methodology in this dissertation was evaluated on digitally reconstructed radiographs or on curated volumetric data rather than on projections acquired directly from a scanner. DRRs do not contain the electronic noise, scatter, beam hardening, and calibration errors that are present in clinical acquisitions. As a result, the reported metrics should be interpreted as an upper bound on real-world performance, and validation on clinical radiographs remains an open task for all three methodologies.

Each methodology also carries a limitation specific to its own design. GaSpCT's ellipsoidal geometric initialization was tailored to the brain, and it encodes a prior on the compact, convex shape of the imaged volume; an extremity or a ribcage enclosing a large hollow interior would violate that assumption, and there is no principled way yet to pick the right initialization for a new anatomy. RibPull preserves topology well on ribcages, but ribcages are among the simplest geometries in the human body, and the occupancy field's real difficulty is with thin, adjacent, high-curvature structures, where the binary inside/outside representation loses the metric information needed to place a surface accurately. With this in mind, extending it to denser anatomy such as the pelvis

or the jaw would likely hurt reconstruction quality before it improves on the ribcage results. ReNAF, finally, is only as accurate as the classical reconstruction it uses as a prior: it inherits that prior's low-frequency structure and cannot fully correct its errors. This shows up most on the abdomen, the largest of our test volumes, which is also where a high-quality ASD-POCS prior was computationally out of reach, forcing us to fall back on a weaker CGLS prior instead. Therefore, the weakest prior ends up paired with the hardest reconstruction, and the final metrics show it. This is a real constraint on the method, not an incidental one: a good prior is exactly what is least affordable where it is needed most.

Taken together, these limitations define the boundary of what off-grid representations have so far been shown to achieve, which is the reconstruction of single, region-specific scans from simulated projections. The geometric operations that motivate this direction, and that we discuss in Section 6.5, have yet to be demonstrated under the sparse and clinically realistic conditions that this dissertation targets.

6.4 Deploying the Proposed Methods in Real Clinical Applications

A common discussion topic when presenting at conferences was that many of the published research projects in the ML-based image reconstruction field need further validation before they can be applied clinically. The same issue is present in the works introduced in this thesis. For example, one of the limitations of our DRR dataset is that it is not raw data directly provided by the scanners themselves, but instead the output of simulation software. Consequently, for their successful deployment, we identify the following three tasks as necessary steps for future researchers in this field:

1) Thorough testing and evaluation with a variety of CBCT-specific datasets that are the result of a real scan reflecting detector miscalibration and quantum noise, among other common CBCT artifacts. 2) Approval by certified clinicians and radiologists. As discussed in the introduction, radiologists need as much signal as possible accurately reconstructed to diagnose severe diseases such as petechial hemorrhages and ischemic strokes. Thus, only a trained individual can approve whether the final reconstruction or representation adequately preserves the clinically valuable information. 3) Evaluation in further downstream tasks such as classification or anomaly detection. The motivation of our work is to preserve clinical information under low-dose constraints. The preserved information should consistently result in comparable diagnostic scores to the alternative, higher-dose approaches. Only then will our work be useful, regardless of what reconstruction metric scores it achieves.

Finally, regarding the technical capabilities of our solution, each scan has to be represented by a different model, so training is repeated for every new scan, adding per-patient compute time. Whether that translates well to clinical environments, both financially and logistically, is subject to further deployment tests in clinical centers.

6.5 Future Work

The current literature shows that research in this field is still in its early stages, and the results in this dissertation point to more directions than they close. The most compelling of these follow from the following observation. Once staircase artifacts and partial volume effects are removed, a continuous representation makes geometric operations on anatomy possible that a voxel grid handles poorly. Point-cloud representations have been used to guide accurate puncture surgery [MNX+25], and topology-preserving representations have improved segmentation accuracy [OUG+26]. What learned gridless representations

add is the ability to perform such operations *under sparse-acquisition constraints*: the missing signal is recovered by interpolation, yielding a smooth, differentiable anatomical reconstruction on which these operations can be carried out. Realizing that potential, however, requires progress on three fronts, each of which this dissertation has touched but none of which it resolves.

Learned gridless methods are currently validated almost entirely on simulated radiographs, and the field lacks diverse, clinically sourced **data** to establish that they generalize. Before a reconstruction can inform a diagnosis or a downstream task, its output also needs to be interpretable enough to **trust**, so a clinician can see which anatomical features drove a result instead of accepting it as a black box. And real anatomy comprises many tissues with overlapping attenuation, while current surface-based methods only offer a small, hand-configured set of materials, which is a limit on **representational expressiveness**. The following subsections take these up in turn: 1) appropriate datasets for off-grid learning, 2) an explainability framework for medical image restoration, and 3) scalable multi-material reconstruction. I will be describing for each the motivation, a suggested solution, and an estimate of the task's complexity and timeframe.

Gridless Learning Appropriate Datasets

It is a well-known issue in the medical imaging AI field that public radiograph datasets are not easy to find. This issue resulted in the widespread use of TIGRE and Plastimatch to generate such data for training off-grid learned representations. In particular, the chest, jaw, foot, and abdomen dataset introduced by NAF has been widely used as the de facto radiograph dataset in various works, as it provides a metric for four types of CBCT that vary significantly between them anatomically.

However, this still leaves out other widely used scanning areas such as breast, brain,

neck, and shoulder. Moreover, a dataset generated by imaging tools from a 3D voxel grid is still a simulated dataset. Finally, to establish confidence in these methodologies, it is critical to evaluate a model architecture against multiple instances of the same type of scan, similar to how GaSpCT was evaluated for 20 different brain radiograph sets.

With these constraints in mind, I encourage clinical institutions and research centers to publicly release de-identified radiographs from the CBCT scans they already perform on a daily basis. Rather than a single new benchmark, the field would benefit most from broad coverage of all common CBCT scan types, with multiple instances of each to capture outliers, anatomical variation, and abnormalities. Such diversity is what ultimately establishes the robustness and reliability of gridless reconstruction methodologies. DRRs remain valuable as a proof-of-concept dataset for developing entirely new models, but access to clinical radiographs acquired directly from the scanner is critical for evaluating any proposed method under realistic acquisition constraints.

A second, complementary direction is the construction of datasets with exactly known ground-truth geometry. By designing objects in 3D modeling software such as Blender, 3D-printing them as physical phantoms of prescribed volume and material composition [VBR20], and then scanning them, we obtain a reference against which geometric operations can be measured directly rather than approximated. This enables quantitative comparison between gridless and voxel-grid representations on tasks that voxel grids handle poorly: for example, edge and outline definition of anatomical structures for surgical and robotic guidance, or total-volume computation of an organ or skeletal part, where the printed phantom supplies the true volume as ground truth.

Explainability framework for medical image restoration

With the emergence of ML as a methodology to perform image restoration and augmentation, the need for appropriate evaluation of the output images for clinical

applicability has been highlighted. In most DL models introduced, the focus is on preserving pixel or voxel fidelity, but there is no mechanism to ensure the preservation of clinical importance. Nevertheless, it is always critical that the perceptual value of the restored image is also evaluated, which can be done by a certified clinician or radiologist.

As mentioned in the introduction, low-level details in the image can be critical for diagnosing diseases such as ischemic stroke, which can be easily lost under an image transformation or sparse sampling. These are the details that a clinician investigates to diagnose a given disease. Still, a similar process is followed when a classification model is trained to classify certain diseases. Therefore, a question can be asked: can we evaluate an image restoration model by how well it preserves features that are relevant for disease classification?

As a proof of concept, we plan to evaluate a CT dataset used for COVID-19 detection [MAP+20]. We pass this dataset through a classification model and observe an explainable classification. We then compare this explanation for the classification with another one for the same image that has undergone corruption and restoration. We can easily corrupt the image either by using a translation model or by applying Poisson and Gaussian noise. The image is then denoised using both a DL model and a domain transfer model. We prioritize pretrained models that function off the shelf, falling back to fine-tuning on the Low-Dose CT Grand Challenge dataset [MCHI+21] if necessary, and investigate both methodologies to ensure that the pipeline functions well as a plug-and-play methodology. We plan to use Diffusion Probabilistic Priors [LXL+25] as the DL and Optimal Transport Cycle-GAN [SOK+20] as the domain transfer example models. Once the image is restored, we can use it as an input to the COVID-19 classification model [ZKHS21] and evaluate the following: 1) Is the previous classification preserved? 2) What features in the image drove this classification? By quantifying the shift in the region of interest, we can evaluate the perceptual value of

the restored image, similar to how it would be performed by a clinician. Figure 6.1 illustrates this proposed pipeline. With further evaluation for other imaging modalities, datasets, classification models, and degradation or restoration techniques, a generalized explainability framework could be introduced to evaluate image restoration models across multiple modalities in medical imaging and could potentially be applied to fields outside of healthcare as well.

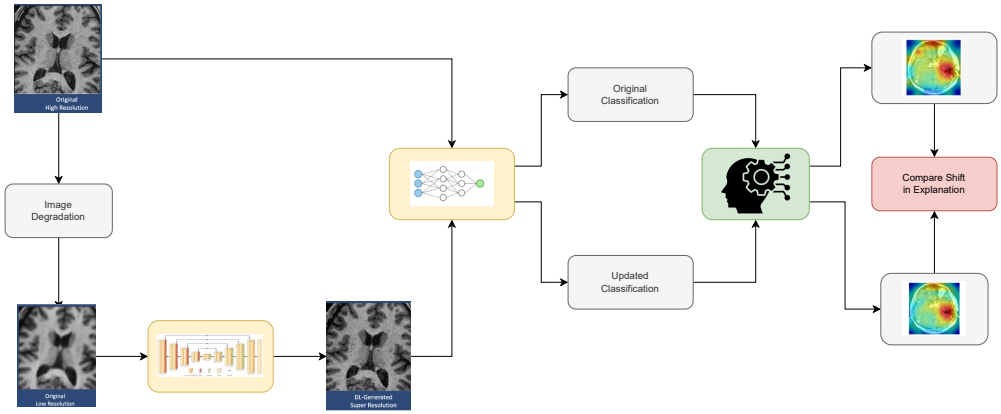


Figure 6.1: Proposed explainability framework for evaluating medical image restoration models. The original high-resolution image is degraded and passed through an image restoration model to obtain a super-resolution reconstruction. Both the original and reconstructed images are passed through a disease classification model, and an explainability method (e.g., GradCAM) is applied to each classification to generate a saliency map. Comparing the shift between the two explanations quantifies whether the restoration preserves the clinically relevant features driving the classification, rather than only pixel-level fidelity.

For the explainability framework, we use the following methods, compared below in a compact and descriptive list outlining their theoretical basis, applicability, output resolution, and computational cost:

- **Grad-CAM** [SCD+17]: Grounded in gradient flow (calculus); applicable only to gradient-based (differentiable) models. Produces high-resolution, smooth heatmaps and is computationally fast (one forward and one backward pass), enabling real-time use. Explains where the model looks (smooth localization).

- **LIME** [RSG16]: Based on local linear approximation; model-agnostic (universal). Yields coarse explanations at the level of superpixels/regions of interest. Computationally slow since cost scales with the number of perturbation samples and superpixels. Explains where (discrete regions).

- **Shapley Values** [LL17]: Founded on cooperative game theory; model-agnostic (universal). Offers a principled trade-off between computation and resolution via numerical feature attribution. Computationally expensive, exponential in the number of features. Explains how much each feature contributes.

- **Sobol (Seg-Sobol)** [SGG23]: Founded on variance-based sensitivity analysis (Sobol indices); gradient-free and model-agnostic, requiring no access to internal architecture. Perturbs the input with multiple noisy masks and measures how each region drives the variance of the output. Resolution is grid-dependent (set by the masking grid size), producing region-importance saliency maps. Slower than Grad-CAM (needs many perturbed forward passes) but more efficient than exhaustive Shapley estimation. Uniquely captures higher-order feature interactions, explaining where and how much regions matter.

With the completion of this research work, the downstream tasks on restored medical images will be more transparent and explainable, augmenting their clinical value and bridging the gap between state-of-the-art restoration models and clinical applicability.

Scalable Multi-Material Reconstruction

NeAS, one of the baselines compared against in Chapter 5, is restricted to two materials: it separates tissues with a hard, non-differentiable selector and requires manually tuned attenuation bounds for each new scene. Work has started on K-NeAS [SNM26], which extends NeAS to an arbitrary number of materials and is

currently in progress. Instead of independent material networks, K-NeAS shares a single latent backbone across all materials, with a lightweight prediction head per material branching from it; a differentiable sequential soft selector then resolves material membership at each point without the discontinuities of a hard selector. Attenuation bounds are estimated automatically by fitting a Gaussian Mixture Model to a single-material reconstruction instead of being tuned by hand, and a scheduled floater loss suppresses spurious geometry in empty space during early training. Figure 6.2 illustrates the resulting architecture. Preliminary results on four DRR CBCT datasets are shown qualitatively in Figure 6.3 and quantitatively in Table 6.1: 3D fidelity improves at three materials. High-contrast, low-variation anatomy such as the jaw, however, remains a limiting case for the automatic bound estimation. This work was accepted for an oral presentation at "Off-Grid: 1st Workshop on Continuous Representations and Grid-Free Methods in Medical Imaging", a new workshop hosted at MICCAI 2026.

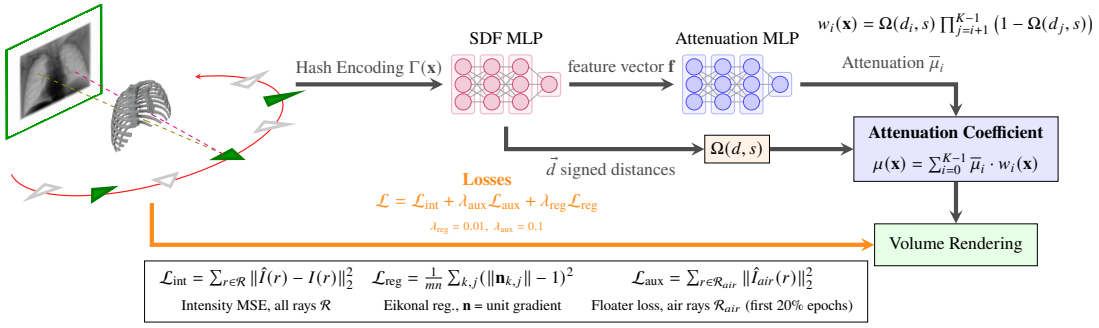


Figure 6.2: K-NeAS architecture: a shared latent backbone with K lightweight attenuation heads replaces NeAS's independent per-material networks, with a differentiable soft selector combining their outputs into a single attenuation coefficient $\mu(\mathbf{x})$.

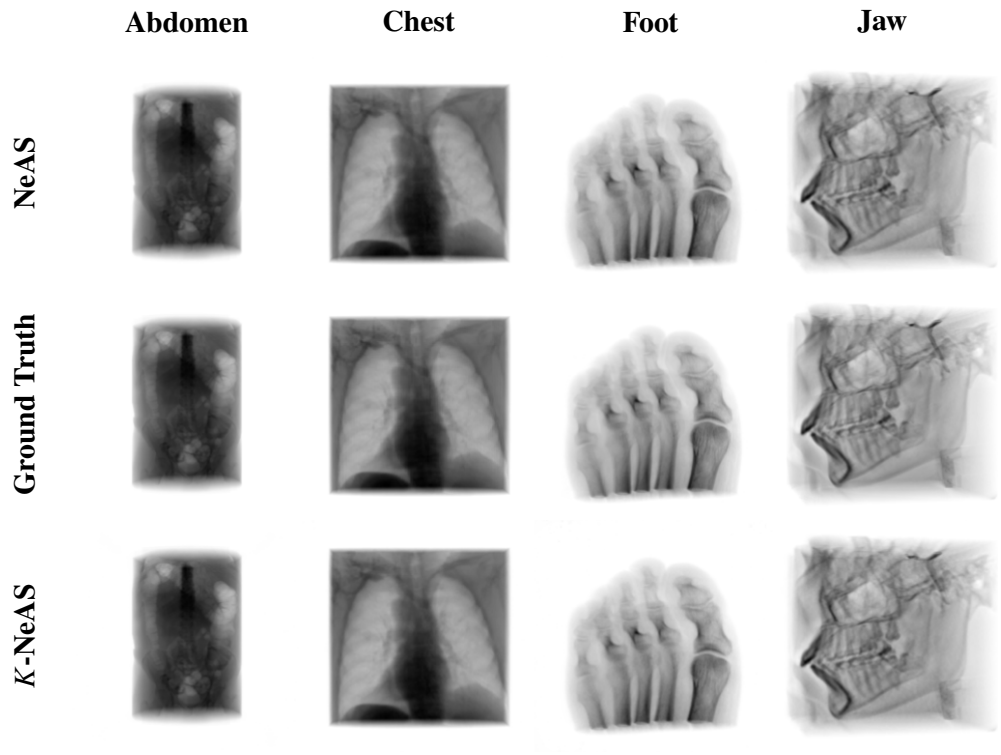


Figure 6.3: Comparisons between K-NeAS, ground truth, and NeAS. Abdomen, Chest, Foot, and Jaw are evaluated on 2 materials, using held-out validation projection indices 24, 1, 1, and 30, respectively. These are evaluated on a single run.

		2D PSNR ↑		2D SSIM ↑		3D PSNR ↑		3D SSIM ↑	
Scene	Config	NeAS	*K*-NeAS	NeAS	*K*-NeAS	NeAS	*K*-NeAS	NeAS	*K*-NeAS
Abdomen	1M	46.744	46.850	0.992	0.993	31.399	32.648	0.820	0.845
	2M	47.098	47.204	0.993	0.993	31.391	32.722	0.823	0.846
	3M	—	**47.687**	—	**0.994**	—	**33.276**	—	**0.858**
	4M	—	47.346	—	**0.994**	—	32.916	—	0.850
Chest	1M	45.949	45.230	0.991	0.991	31.508	31.396	0.913	0.908
	2M	**46.609**	45.672	0.992	0.992	32.171	31.784	0.918	0.915
	3M	—	46.036	—	**0.993**	—	**32.181**	—	**0.920**
	4M	—	45.427	—	0.992	—	31.592	—	0.911
Foot	1M	42.230	42.717	0.981	**0.983**	31.007	31.579	**0.900**	0.893
	2M	42.224	42.570	0.981	0.982	31.092	31.388	0.891	0.889
	3M	—	**42.800**	—	**0.983**	—	**31.603**	—	0.889
	4M	—	42.717	—	**0.983**	—	31.462	—	0.887
Jaw	1M	**39.451**	34.321	**0.966**	0.953	**34.099**	31.672	**0.882**	0.797
	2M	37.472	34.366	0.963	0.956	33.514	31.828	0.832	0.808
	3M	—	34.332	—	0.955	—	31.847	—	0.806
	4M	—	34.541	—	0.956	—	31.879	—	0.810

Table 6.1: Preliminary quantitative comparison between NeAS and *K*-NeAS across scenes and material configurations (1M–4M). NeAS only supports up to 2 materials, so its 3M/4M entries are marked with a dash. Best score per metric and scene is in bold. Results are averaged over 3 training runs.

Bibliography

[ABVGT+21] Emmanuel Ahishakiye, Martin Bastiaan Van Gijzen, Julius Tumwiine, Ruth Wario, and Johnes Obungoloch. A survey on deep learning in medical image reconstruction. *Intelligent Medicine*, 1(03):118–127, 2021. 27, 65

[AK84] Anders H Andersen and Avinash C Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging*, 6(1):81–94, 1984. 6, 18, 67

[AKB+08] Stéphane Allaire, John J Kim, Stephen L Breen, David A Jaffray, and Vladimir Pekar. Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis. In *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 1–8. IEEE, 2008. 34

[Ame24] American Association for Cancer Research. Cancer in 2024, 2024. Accessed: July 1, 2026. 2

[BAA+25] Biykem Bozkurt, Tariq Ahmad, Kevin Alexander, William L. Baker, Kelly Bosak, Khadijah Breathett, Spencer Carter, Mark H. Drazner, Shannon M. Dunlay, Gregg C. Fonarow, Stephen J. Greene, Paul Heidenreich, Jennifer E. Ho, Eileen Hsich, Nasrien E. Ibrahim, Lenette M. Jones, Sadiya S. Khan, Prateeti Khazanie, Todd Koelling, Christopher S. Lee, Alanna A. Morris, Robert L. Page, Ambarish Pandey, Mariann R. Piano, Alexander T. Sandhu, Josef Stehlik, Lynne W. Stevenson, John Teerlink, Amanda R. Vest, Clyde Yancy, and Boback Ziaeian. HF Stats 2024: Heart Failure Epidemiology and Outcomes Statistics — An Updated 2024 Report from the Heart Failure Society of America. *Journal of Cardiac Failure*, 31(1):66–116, 2025. 2

[BDHS16] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction. *Biomedical Physics & Engineering Express*, 2(5):055010, 2016. 9, 23, 70

[BJ19] DM Bappy and Insu Jeon. High-quality x-ray computed tomography reconstruction using projected and interpolated images. *IET Image Processing*, 13(7):1074–1080, 2019. 67

[BMV+22] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 32

[BSHKG23] Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, and Stephen Gould. DiGS: Divergence guided shape implicit neural representation for unoriented point clouds, 2023. 55, 59

[CD23] Mattéo Clémot and Julie Digne. Neural skeleton: Implicit neural representation away from the surface. *Computers & Graphics*, 114:368–378, 2023. 11, 57

[CFFBT+22a] Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3843–3848. IEEE, 2022. 45, 65

[CFFBT+22b] Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3843–3848. IEEE, 2022. 51

[CKD+25] Devina Chatterjee, Adway Kanhere, Florence X Doo, Jerry Zhao, Andrew Chan, Alexander Welsh, Pranav Kulkarni, Annie Trang, Vishwa S Parekh, and Paul H Yi. Children are not small adults: addressing limited generalizability of an adult deep learning ct organ segmentation model to the pediatric population. *Journal of Imaging Informatics in Medicine*, 38(3):1628–1641, 2025. 64

[CLW+24] Yuanhao Cai, Yixun Liang, Jiahao Wang, Angtian Wang, Yulun Zhang, Xiaokang Yang, Zongwei Zhou, and Alan Yuille. Radiative gaussian splatting for efficient x-ray novel view synthesis, 2024. 65

[CT06] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006. 48

[CTO+10] Junjie Cao, Andrea Tagliasacchi, Matt Olson, Hao Zhang, and Zhinxun Su. Point cloud skeletons via laplacian based contraction. In *2010 Shape Modeling International Conference*, pages 187–197. IEEE, 2010. 9, 57

[CWY+24] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11174–11183, 2024. 27, 65, 67

[CZK+17a] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017. 33

[CZK+17b] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017. 67

[DBS+22] Nicola K Dinsdale, Emma Bluemke, Vaanathi Sundaresan, Mark Jenkinson, Stephen M Smith, and Ana IL Namburete. Challenges for machine learning in clinical translation of big data imaging studies. *Neuron*, 110(23):3866–3881, 2022. 67

[DVC19] Xu Dong, Swapnil Vekhande, and Guohua Cao. Sinogram interpolation for sparse-view micro-ct with deep learning neural network. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, pages 692–698. SPIE, 2019. 33

[DWL+13] Xinhui Duan, Jia Wang, Shuai Leng, Bernhard Schmidt, Thomas Allmendinger, Katharine Grant, Thomas Flohr, and Cynthia H McCollough. Electronic noise in ct detectors: impact on image noise and artifacts. *American Journal of Roentgenology*, 201(4):W626–W632, 2013. 17

[EHR07] Andrew J Einstein, Milena J Henzlova, and Sanjay Rajagopalan. Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography. *Jama*, 298(3):317–323, 2007. 5

[EKSX96] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 226–231, 1996. 62

[FAAF+21] Eduardo Kaiser Ururahy Nunes Fonseca, Antonildes Nascimento Assunção, Jose de Arimateia Batista Araujo-Filho, Lorena Carneiro Ferreira, Bruna Melo Coelho Loureiro, Daniel Giunchetti Strabelli, Lucas de Pádua Gomes de Farias, Rodrigo Caruso Chate, Giovanni Guido Cerri, Marcio Valente Yamada Sawamura, et al. Lung lesion burden found on chest ct as a prognostic marker in hospitalized patients with high clinical suspicion of covid-19 pneumonia: A brazilian experience. *Clinics*, 76:e3503, 2021. 3

[FDK84] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984. 6, 18, 67

[GG22] Vivek Gopalakrishnan and Polina Golland. Fast auto-differentiable digitally reconstructed radiographs for solving inverse problems in intra-operative imaging. In *Workshop on Clinical Image-Based Procedures*, pages 1–11. Springer, 2022. 32

[GS22a] Qi Gao and Hongming Shan. Cocodiff: a contextual conditional diffusion model for low-dose ct image denoising. In *Developments in X-Ray Tomography XIV*, volume 12242. SPIE, 2022. 33

[GS22b] Qi Gao and Hongming Shan. CoCoDiff: a contextual conditional diffusion model for low-dose ct image denoising. In *Developments in X-Ray Tomography XIV*, volume 12242, pages 92–98. SPIE, 2022. 67

[GYH+20] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes, 2020. 8, 55

[GYZ+23] Bing Guan, Cailian Yang, Liu Zhang, Shanzhou Niu, Minghui Zhang, Yuhao Wang, Weiwen Wu, and Qiegen Liu. Generative modeling in sinogram domain for sparse-view ct reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2023. 33

[HML+14] WY Huang, CH Muo, CY Lin, YM Jen, MH Yang, JC Lin, FC Sung, and CH Kao. Paediatric head ct scan and subsequent risk of malignancy and benign brain tumour: a nation-wide population-based cohort study. *British journal of cancer*, 110(9):2354–2360, 2014. 5

[HS+52] Magnus R Hestenes, Eduard Stiefel, et al. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49(6):409–436, 1952. 6, 67

[HWCO+13a] Hui Huang, Shihao Wu, Daniel Cohen-Or, Minglun Gong, Hao Zhang, Guiqing Li, and Baoquan Chen. L1-medial skeleton of point cloud. *ACM Trans. Graph.*, 32(4):65–1, 2013. 11

[HWCO+13b] Hui Huang, Shihao Wu, Daniel Cohen-Or, Minglun Gong, Hao Zhang, Guiqing Li, and Baoquan Chen. L1-medial skeleton of point cloud. *ACM Transactions on Graphics*, 32(4):65–1, 2013. 57

[HY18] Yoseob Han and Jong Chul Ye. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE transactions on medical imaging*, 37(6):1418–1429, 2018. 33

[HY23] Shady Hermena and Michael Young. Ct-scan image production procedures. *StatPearls*, 2023. 15

[JGW+23] Liang Jin, Shixuan Gu, Donglai Wei, Jason Ken Adhinarta, Kaiming Kuang, Yongjie Jessica Zhang, Hanspeter Pfister, Bingbing Ni, Jiancheng Yang, and Ming Li. Ribseg v2: A large-scale benchmark for rib labeling and anatomical centerline extraction. *IEEE Transactions on Medical Imaging*, 2023. 11, 48

[JOT+19] Jayapramila Jayamani, Noor Diyana Osman, Abdul Aziz Tajuddin, Zaker Salehi, Mohd Hanafi Ali, and Mohd Zahri Abdul Aziz. Determination of computed tomography number of high-density materials in 12-bit, 12-bit extended and 16-bit depth for dosimetric calculation in treatment planning system. *Journal of Radiotherapy in Practice*, 18(3):285–294, 2019. 19

[JS13] Dakai Jin and Punam K Saha. A new fuzzy skeletonization algorithm and its applications to medical imaging. In *International Conference on Image Analysis and Processing*, pages 662–671. Springer, 2013. 51

[JSLR24] Giavanna Jadick, Geneva Schlafly, and Patrick J La Rivière. Dual-energy computed tomography imaging with megavoltage and kilovoltage x-ray spectra. *Journal of Medical Imaging*, 11(2):023501–023501, 2024. 17

[JYK+20] Liang Jin, Jiancheng Yang, Kaiming Kuang, Bingbing Ni, Yiyi Gao, Yingli Sun, Pan Gao, Weiling Ma, Mingyu Tan, Hui Kang, et al. Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of fracnet. *EBioMedicine*, 62, 2020. 54, 58

[KKLD23] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 8, 9, 27, 32, 35, 51, 65

[KS01] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001. 6, 14

[LFL+25] Yingtai Li, Xueming Fu, Han Li, Shang Zhao, Ruiyang Jin, and S Kevin Zhou. 3dgr-ct: Sparse-view ct reconstruction with a 3d gaussian representation. *Medical Image Analysis*, 103:103585, 2025. 65

[LL17] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017. 91

[LLK+18] Hoyeon Lee, Jongha Lee, Hyeongseok Kim, Byungchul Cho, and Seungryong Cho. Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2):109–119, 2018. 33

[LSS+19] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*, 2019. 39

[LWX+24] Lizhe Liu, Bohua Wang, Hongwei Xie, Daqi Liu, Li Liu, Zhiqiang Tian, Kuiyuan Yang, and Bing Wang. Surroundsdf: Implicit 3d scene understanding based on signed distance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21614–21623, 2024. 27

[LXL+25] Xuan Liu, Yaoqin Xie, Chenbin Liu, Jun Cheng, Songhui Diao, Shan Tan, and Xiaokun Liang. Diffusion probabilistic priors for zero-shot low-dose ct image denoising. *Medical Physics*, 52(1):329–345, 2025. 89

[MAP+20] Sergey P Morozov, Anna E Andreychenko, Nikolay A Pavlov, AV Vladzymyrskyy, Natalya V Ledikhova, Victor A Gombolevskiy, Ivan A Blokhin, Pavel B Gelezhe, AV Gonchar, and V Yu Chernina. Mosmeddata: Chest ct scans with covid-19 related findings dataset. *arXiv preprint arXiv:2005.06465*, 2020. 89

[MCHI+21] Taylor R Moen, Baiyu Chen, David R Holmes III, Xinhui Duan, Zhicong Yu, Lifeng Yu, Shuai Leng, Joel G Fletcher, and Cynthia H McCollough. Low-dose ct image and projection dataset. *Medical physics*, 48(2):902–911, 2021. 89

[MESK22a] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022. 67

[MESK22b] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 69

[MHLZ20] Baorui Ma, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. *arXiv preprint arXiv:2011.13495*, 2020. 59

[MNX+25] Satoshi Miura, Masayuki Nakayama, Kexin Xu, Zhang Bo, Ryoko Kuromatsu, Masahito Nakano, Yu Noda, and Takumi Kawaguchi. Point cloud registration algorithm using liver vascular skeleton feature with computed tomography and ultrasonography image fusion. *International Journal of Computer Assisted Radiology and Surgery*, 20(12):2469–2478, 2025. 86

[MST+21] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 8, 27, 33, 51, 65, 68

[MYK+15] Cynthia H McCollough, Lifeng Yu, James M Kofler, Shuai Leng, Yi Zhang, Zhoubo Li, and Rickey E Carter. Degradation of ct low-contrast spatial resolution due to the use of iterative reconstruction and reduced dose levels. *Radiology*, 276(2):499–506, 2015. 33

[NGV+24] Emmanouil Nikolakakis, Utkarsh Gupta, Jonathan Vengosh, Justin Bui, and Razvan Marinescu. GaSpCT: Gaussian splatting for novel brain CBCT projection view synthesis. In *Medical Imaging 2024: Image Processing*. SPIE, 2024. 10, 31, 51, 65

[NODM25] Emmanouil Nikolakakis, Amine Ouasfi, Julie Digne, and Razvan Marinescu. RibPull: Implicit occupancy fields and medial axis extraction for CT ribcage scans. In *Medical Imaging 2025: Image Processing*. SPIE, 2025. 12, 49, 65

[NSW+26] Emmanouil Nikolakakis, Daksh K. Shah, Julia Wong, Julie Digne, and Razvan Marinescu. ReNAF: Regularized neural attenuation fields with 3d reconstruction priors for sparse-view CT, 2026. Ready to be submitted. 13, 63

[OB24] Amine Ouasfi and Adnane Boukhayma. Unsupervised occupancy learning from sparse point cloud. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21729–21739, 2024. 9, 55

[OUG+26] Seher Ozcelik, Sinan Unver, Ilke Ali Gurses, Rustu Turkay, and Cigdem Gunduz-Demir. Topology-aware loss for segmentation in computed tomography images. *Biomedical Signal Processing and Control*, 117:109512, 2026. 86

[PGZ+26] Shyam Prabhakaran, Nestor R Gonzalez, Kori S Zachrison, Opeolu Adeoye, Anne W Alexandrov, Sameer A Ansari, Sherita Chapman, Alexandra L Czap, Oana M Dumitrascu, Koto Ishida, et al. 2026 guideline for the early management of patients with acute ischemic stroke: a guideline from the american heart association/american stroke association. *Stroke*, 2026. 2

[PMT+16a] Stephen P Power, Fiachra Moloney, Maria Twomey, Karl James, Owen J O'Connor, and Michael M Maher. Computed tomography and patient risk: Facts, perceptions and uncertainties. *World journal of radiology*, 8(12):902, 2016. 66

[PMT+16b] Stephen P Power, Fiachra Moloney, Maria Twomey, Karl James, Owen J O'Connor, and Michael M Maher. Computed tomography and patient risk: Facts, perceptions and uncertainties. *World journal of radiology*, 8(12):902, 2016. 4, 5, 32

[PNBSB20] Sangeetha Prabhu, Divya Kumari Naveen, Sandhya Bangera, and B Subrahmanya Bhat. Production of x-rays using x-ray tube. In *Journal of Physics: Conference Series*, volume 1712, page 012036. IOP Publishing, 2020. 16

[QBM06] Anqi Qiu, Dmitri Bitouk, and Michael I Miller. Smooth functional and structural maps on the neocortex via orthonormal bases of the laplace-beltrami operator. *IEEE Transactions on Medical Imaging*, 25(10):1296–1306, 2006. 51

[QYSG17] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 54

[RKA+21] Albert W. Reed, Hyojin Kim, Rushil Anirudh, K. Aditya Mohan, Kyle Champley, Jingu Kang, and Suren Jayasuriya. Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2258–2268, 2021. 65

[RO94] Leonid I Rudin and Stanley Osher. Total variation based image restoration with free local constraints. In *Proceedings of 1st international conference on image processing*, volume 1, pages 31–35. IEEE, 1994. 39

[RSG16] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016. 91

[Sav06] Jeffrey L Saver. Time is brain—quantified. *Stroke*, 37(1):263–266, 2006. 2

[SBB+00a] Mie Sato, Ingmar Bitter, Michael A Bender, Arie E Kaufman, and Masayuki Nakajima. Teasar: tree-structure extraction algorithm for accurate and robust skeletons. In *Proceedings the Eighth Pacific Conference on Computer Graphics and Applications*, pages 281–449. IEEE, 2000. 11

[SBB+00b] Mie Sato, Ingmar Bitter, Michael A Bender, Arie E Kaufman, and Masayuki Nakajima. TEASAR: tree-structure extraction algorithm for accurate and robust skeletons. In *Proceedings the Eighth Pacific Conference on Computer Graphics and Applications*, pages 281–449. IEEE, 2000. 51, 57

[SBG+26] Ty Skyles, Samantha M Bouchal, Anna Giarratana, Jacob Wengler, Ian Hart, Erin Greig, Harmanjeet Singh, Steve S Huang, Felipe Martinez, Ba Nguyen, et al. Pet imaging in alzheimer disease in the era of antiamyloid therapy in the united states: clinical utility, quantification, and policy landscape. *Journal of nuclear medicine technology*, 54(1):10–17, 2026. 3

[SCD+17] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 90

[SE22] Ziyu Shu and Alireza Entezari. Sparse-view and limited-angle ct reconstruction with untrained networks and deep image prior. *Computer Methods and Programs in Biomedicine*, 226:107167, 2022. 33

[SF16] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 34, 40

[SGCB+21] Teemu Siiskonen, Aoife Gallagher, Olivera Ciraj Bjelac, Leos Novak, Marta Sans Merce, Jad Farah, Jérémie Dabin, Françoise Malchair, Željka Knežević, and Mika Kortesniemi. A european perspective on dental cone beam computed tomography systems with a focus on optimisation utilising diagnostic reference levels. *Journal of radiological protection*, 41(2):442–451, 2021. 5

[SGG23] Hossein Shreim, Abdul Karim Gizzini, and Ali J Ghandour. Trainable noise model as an explainable artificial intelligence evaluation

method: Application on sobol for remote sensing image segmentation. *Environmental Sciences Proceedings*, 29(1):49, 2023. 91

[SL74] Lawrence A Shepp and Benjamin F Logan. The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974. 29

[SLK+08] Yonggang Shi, Rongjie Lai, Sheila Krishna, Nancy Sicotte, Ivo Dinov, and Arthur W Toga. Anisotropic laplace-beltrami eigenmaps: bridging reeb graphs and skeletons. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–7. IEEE, 2008. 51

[SLW+10] Gregory C Sharp, Rui Li, John Wolfgang, G Chen, Marta Peroni, Maria Francesca Spadea, Shinichro Mori, Junan Zhang, James Shackleford, and Nagarajan Kandasamy. Plastimatch: an open source software suite for radiotherapy image processing. In *Proceedings of the XVI'th International Conference on the use of Computers in Radiotherapy (ICCR), Amsterdam, Netherlands*, volume 3, 2010. 9, 23, 42

[SLX+21a] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for tomographic imaging. *IEEE Transactions on Computational Imaging*, 7:1400–1412, 2021. 28

[SLX+21b] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for tomographic imaging. *IEEE Transactions on Computational Imaging*, 7:1400–1412, 2021. 51

[SNM26] Daksh K Shah, Emmanouil Nikolakakis, and Razvan V Marinescu. *k*-neas: Scalable multi-material ct reconstruction using neural sdfs. *arXiv preprint arXiv:2607.14415*, 2026. 91

[SOK+20] Byeongsu Sim, Gyutaek Oh, Jeongsol Kim, Chanyong Jung, and Jong Chul Ye. Optimal transport driven cyclegan for unsupervised learning in inverse problems. *SIAM Journal on Imaging Sciences*, 13(4):2281–2306, 2020. 89

[SP08] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777, 2008. 6, 18, 67

[SPX22] Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image

reconstruction. *IEEE transactions on neural networks and learning systems*, 35(1):770–782, 2022. 28, 66

[SSdS+21] Luiz Schirmer, Guilherme Schardong, Vinícius da Silva, Hélio Lopes, Tiago Novello, Daniel Yukimura, Thales Magalhaes, Hallison Paz, and Luiz Velho. Neural networks for implicit representations of 3d scenes. In *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 17–24. IEEE, 2021. 65

[SSP+22] Hamayak Sisakian, Syuzanna Shahnazaryan, Sergey Pepoyan, Armine Minasyan, Gor Martirosyan, Mariam Hovhannisyan, Ashkhen Maghaqelyan, Sona Melik-Stepanyan, Armine Chopikyan, and Yury Lopatin. Reduction of hospitalization and mortality by echocardiography-guided treatment in advanced heart failure. *Journal of Cardiovascular Development and Disease*, 9(3):74, 2022. 2

[TBA11] Pang-yu Teng, Ahmet Murat Bagci, and Noam Alperin. Automated prescription of an optimal imaging plane for measurement of cerebral blood flow by phase contrast magnetic resonance imaging. *IEEE transactions on biomedical engineering*, 58(9):2566–2573, 2011. 11

[TSS16] Steven Tilley, Jeffrey H Siewerdsen, and J Webster Stayman. Model-based iterative reconstruction for flat-panel cone-beam ct with focal spot blur, detector blur, and correlated noise. *Physics in Medicine & Biology*, 61(1):296–319, 2016. 18

[Tub96] Maurice Tubiana. Wilhelm conrad röntgen and the discovery of x-rays. *Bulletin de l'Academie nationale de medecine*, 180(1):97–108, 1996. 2

[VBR20] Alejandra Valladares, Thomas Beyer, and Ivo Rausch. Physical imaging phantoms for simulation of tumor heterogeneity in pet, ct, and mri: an overview of existing designs. *Medical physics*, 47(4):2023–2037, 2020. 88

[vdBvSB+23] Judith van der Bie, Marcel van Straten, Ronald Booij, Daniel Bos, Marcel L Dijkshoorn, Alexander Hirsch, Simran P Sharma, Edwin HG Oei, and Ricardo PJ Budde. Photon-counting ct: review of initial clinical results. *European journal of radiology*, 163:110829, 2023. 17

[WBSS04] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 22, 44

[WPP+18] Martin J Willemink, Mats Persson, Amir Pourmorteza, Norbert J Pelc, and Dominik Fleischmann. Photon-counting ct: technical principles and clinical prospects. *Radiology*, 289(2):293–312, 2018. 17

[WSL+19] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5):1–12, 2019. 54

[YGW+21] Jiancheng Yang, Shixuan Gu, Donglai Wei, Hanspeter Pfister, and Bingbing Ni. Ribseg dataset and strong point cloud baselines for rib segmentation from ct scans. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021*, pages 611–621. Springer, 2021. 11, 54

[YYZ+18a] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018. 33

[YYZ+18b] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018. 67

[ZFM+25] Li Zhou, Changsheng Fang, Bahareh Morovati, Yongtong Liu, Shuo Han, Yongshun Xu, and Hengyong Yu. ρ-NeRF: leveraging attenuation priors in neural radiance field for 3d computed tomography reconstruction. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 1636–1641. IEEE, 2025. 67, 72

[ZGS14] Z Zhao, GJ Gang, and JH Siewerdsen. Noise, sampling, and the number of projections in cone-beam ct with a flat-panel detector. *Medical physics*, 41(6Part1):061909, 2014. 5, 18

[ZH23] Yanjie Zheng and Kelsey B Hatzell. Ultrasparse view x-ray computed tomography for 4d imaging. *ACS Applied Materials & Interfaces*, 15(29):35024–35033, 2023. 65

[ZIE+18] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 22, 44

[ZIK+25] Chengrui Zhu, Ryoichi Ishikawa, Masataka Kagesawa, Tomohisa Yuzawa, Toru Watsuji, and Takeshi Oishi. NeAS: 3d reconstruction from x-ray images using neural attenuation surface. *arXiv preprint arXiv:2503.07491*, 2025. 68, 73, 74, 78

[ZIL+21] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021. 48, 77

[ZKHS21] Abolfazl Zargari Khuzani, Morteza Heidari, and S Ali Shariati. Covid-classifier: An automated machine learning model to assist in the diagnosis of covid-19 infection in chest x-ray images. *Scientific Reports*, 11(1):9887, 2021. 89

[ZLC+24] Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R^2-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction, 2024. 37, 65

[ZLD+18] Zhicheng Zhang, Xiaokun Liang, Xu Dong, Yaoqin Xie, and Guohua Cao. A sparse-view ct reconstruction method based on combination of densenet and deconvolution. *IEEE transactions on medical imaging*, 37(6):1407–1417, 2018. 33

[ZWX+23] Zhengwu Zhang, Yuexuan Wu, Di Xiong, Joseph G Ibrahim, Anuj Srivastava, and Hongtu Zhu. LESA: Longitudinal elastic shape analysis of brain subcortical structures. *Journal of the American Statistical Association*, 118(541):3–17, 2023. 51

[ZZL22] Ruyi Zha, Yanhao Zhang, and Hongdong Li. NAF: neural attenuation fields for sparse-view CBCT reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–452. Springer, 2022. 8, 9, 27, 51, 65, 68

[ZZR+21] Xiaojun Zhu, Zheng Zhang, Jian Ruan, Houde Liu, and Hanxu Sun. Ressanet: Learning geometric information for point cloud processing. *Sensors*, 21(9):3227, 2021. 48