This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Dynamic CT Reconstruction from Limited Views with Implicit Neural Representations and Parametric Motion Fields

Albert W. Reed[1] awreed@asu.edu
Hyojin Kim[2] hkim@llnl.gov
Rushil Anirudh[2] anirudh1@llnl.gov
K. Aditya Mohan[2] mohan3@llnl.gov
Kyle Champley[2] champley1@llnl.gov
Jingu Kang[2] kang12@llnl.gov
Suren Jayasuriya[1] sjayasur@asu.edu

[1]Arizona State University
[2]Lawrence Livermore National Laboratory

Abstract

Reconstructing dynamic, time-varying scenes with computed tomography (4D-CT) is a challenging and ill-posed problem common to industrial and medical settings. Existing 4D-CT reconstructions are designed for sparse sampling schemes that require fast CT scanners to capture multiple, rapid revolutions around the scene in order to generate high quality results. However, if the scene is moving too fast, then the sampling occurs along a limited view and is difficult to reconstruct due to spatiotemporal ambiguities. In this work, we design a reconstruction pipeline using implicit neural representations coupled with a novel parametric motion field warping to perform limited view 4D-CT reconstruction of rapidly deforming scenes. Importantly, we utilize a differentiable analysis-by-synthesis approach to compare with captured x-ray sinogram data in a self-supervised fashion. Thus, our resulting optimization method requires no training data to reconstruct the scene. We demonstrate that our proposed system robustly reconstructs scenes containing deformable and periodic motion and validate against state-of-the-art baselines. Further, we demonstrate an ability to reconstruct continuous spatiotemporal representations of our scenes and upsample them to arbitrary volumes and frame rates post-optimization. This research opens a new avenue for implicit neural representations in computed tomography reconstruction in general. Code is available at https://github.com/awreed/DynamicCTReconstruction.

1. Introduction

Computed-tomography (CT) is a mature imaging technology with vital industrial and medical applications [20, 13, 8]. CT scanners capture x-ray data or *sinograms* by scanning a sequence of angles around an object. Reconstruction algorithms then estimate the scene from these measured sinograms. CT imaging in both 2D and 3D of static objects is a well-studied inverse problem with both theoretical and practical algorithms [21, 37, 1].

However, reconstruction of dynamic scenes (*i.e.* scene features changing over time), known as *dynamic 4D-CT*, is a severely ill-posed problem because the sinogram aggregates measurements over time which yields spatio-temporal ambiguities [12, 24, 51, 52, 30]. Analogous to motion blur, a static or quasi-static scene is only captured for a small angular range of the sinogram (motion blur analogy: short exposure), and this mapping is a function of the amount of scene motion relative to the CT scanner's rotation speed. Traditional CT reconstruction algorithms have limited capability to address 4D-CT problems due to difficulty accounting for this motion. Yet solving these 4D-CT problems is critically important for a range of applications from clinical diagnosis to non-destructive evaluation for material characteristics and metrology.

4D-CT reconstruction techniques have been proposed in the literature to handle periodic motion [33, 24] as well as more general nonlinear, deformable motion [51, 52, 30]. While the latter methods achieve state-of-the-art for deformable motion, they typically assume slow motion relative to the scanner rotation speed. Specifically, these algorithms assume sparse measurements of the object that span the full angular range ($0 - 360°$) and typically require multiple revolutions around the sample to reconstruct images at each time step. This approach is called *sparse view* CT in the literature [31]. However, this setting is not always practical such as when the object motion is too fast for the scanner to make multiple revolutions. Instead, an alternative approach is to collect measurements over partial angular ranges in a *single* revolution – which is the more challenging *limited-view* reconstruction [43]. Even in the static case, the limited-view problem is challenging due to missing information often leading to significant artifacts [18].

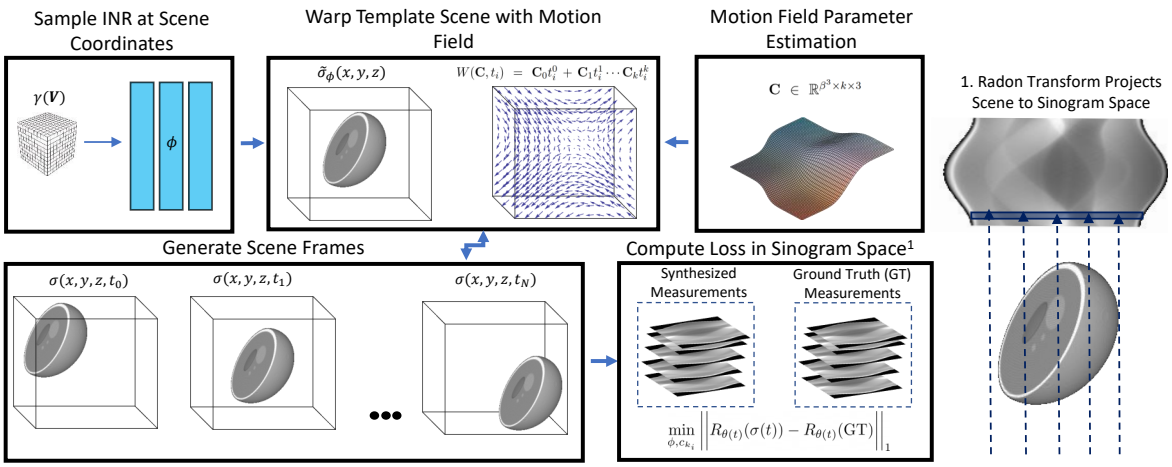

Figure 1. Given a sinogram, we jointly estimate a scene template and motion field to reconstruct the 4D scene. Here, we warp a 3D Shepp Logan template to reconstruct its linear translation in time. We simulate sinogram measurements from this 4D scene and compute a loss with the given sinogram. This loss is backpropagated to the implicit neural representation (INR) network weights and motion field parameters until convergence.

Key Contributions: In this paper, we propose a novel, training-data-free approach for 4D-CT reconstruction that works especially well in limited-view scenarios. Our method, illustrated in Figure 1, consists of an implicit neural representation (INR) [29] model that acts as the static scene prior coupled with a parametric motion field to estimate an evolving 3D object over time. The reconstruction is then synthesized into sinogram measurements using a differentiable Radon transform to simulate parallel-beam CT scanners. By minimizing the discrepancy between the synthesized and observed sinograms, we are able to optimize both the INR weights and motion parameters in a self-supervised, analysis-by-synthesis fashion to obtain accurate dynamic scene reconstructions without training data.

Validation: We are primarily interested in general, in-situ CT imaging where the scanned object cannot be frozen in time and the motion dynamics are arbitrary (e.g., non-periodic) and cannot be predicted (e.g., [17]). However, acquiring 4D-CT data is challenging and one of the primary bottlenecks for research in 4D-CT reconstruction. While this is partly due to the expense and logistics of accessing CT scanners and data, it is also because acquiring ground truth in the 4D case is exceptionally challenging. While there are examples of real CT data [51, 52], these datasets are specific to particular scanners and sampling schemes.

In order to evaluate our method on an in-situ imaging task and highlight our method's ability to resolve 4D scenes from limited angles, we introduce a synthetic dataset for parallel beam CT. This dataset is generated with an accurate physics simulator for material deformation and used to benchmark ours and competing state-of-the-art methods. Further, while not our target application, we also evaluate our method's performance on a medical imaging task of reconstructing a periodically deforming thoracic cavity using simulated x-ray measurements from real thoracic reconstructions. We observe that our method outperforms competitive baselines on both of our datasets.

2. Related Work

3D-CT Sparse and Limited View Reconstruction: Traditional CT reconstruction is a mature imaging problem with applications in security, industrial and healthcare. For sparse-view 3D-CT, common techniques include the algebraic reconstruction technique (ART) and the filtered backprojection algorithm (FBP) [2, 21]. Model-based approaches have been proposed for the limited angle case [18, 49]. More recently, deep-learning based approaches utilize training data [14, 22, 19, 53, 49, 3, 25] to estimate scenes from sparse-view or limited angles. For a more comprehensive characterization on limited-angle tomography, we refer the reader to [10]. In our paper, we are interested in the 4D-CT problem, particularly for limited view sampling.

4D-CT for Periodic Motion: When recovering periodic motion, such as the breathing phases of clinical patients, several methods [24, 33, 48] gate measurements into phase/amplitude cycles to help reconstruct 3D image volumes. This gating limits the number of angular measurements per phase and can induce motion artifacts due to phase error [41]. Similar to our approach, these methods sometimes include parametric motion models for more general non-periodic motion (*e.g.*, heart + breathing motion) [41, 38, 40]. These motion models perform image registration between estimated phase images, but typically require either enough phase gated measurements or relatively

slow motion to work effectively. In contrast, our approach does not require phase information for reconstruction.

4D-CT for Deformable Motion: State-of-the-art (SOA) 4D-CT reconstruction methods jointly estimate the scene and motion field parameters [51, 30, 52] to solve for non-periodic deforming scenes. The authors of [30] show that quickly rotating the CT scanner and sampling sparse angular views in an interlaced fashion enables the capture of high-fidelity images. In [51], the authors introduce low-discrepency sampling and jointly solve for the motion flow-field and scene volume. This work was extended in [52] by using the low-discrepency sampling and adapting the reconstruction algorithm to dynamically upsample temporal frames when the motion is rapidly changing the measured object. Importantly, all these methods require quickly rotating the CT scanner to capture sparse angular measurements that span 360 degrees, a constraint we do not require.

Implicit Neural Representations (INR): Recently, coordinate based multi-layer perceptrons, or INRs, have found success in the imaging domain. These architectures learn functions that map input coordinates (*e.g.*, (x, y, z)) to physical properties of the scene (*e.g.*, density at (x, y, z)). These networks coupled with differentiable renderers have demonstrated impressive capabilities for estimating 3D scenes [29, 28, 27, 54]. Our method is inspired by [29], who estimate a continuous 3D volume density from a limited set of 2D images. More recently, [6, 7, 9, 36, 55] exploit the NeRF architecture to solve problems like view synthesis, texture completion from impartial 3D data, non-line-of-sight imaging recognition, etc. In [34], the authors introduced a method to jointly learn a scene template and warp field for transforming the template through time for RGB video frames. Our work is similar in approach, as we also learn a scene template and then a continuous mapping through time. However our application is wildly different by leveraging 4D-CT measurements and incorporating a differentiable Radon transform into our pipeline. Concurrently at the time of writing this paper, recent work in [45] demonstrates the potential of INRs to solve ill-posed inverse problems in the tomographic imaging domain. However, this paper only considers 3D static scenes whereas we consider 4D scenes and under limited view sampling.

3. 4D-CT Forward Imaging Model

The primary task of this paper is to reconstruct 3D scenes in time from CT measurements (*i.e.* sinograms). In this section, we formulate our forward imaging model and the key assumptions our algorithm makes for reconstructing scenes from CT angular projections. Next, in section 4, we discuss our algorithmic pipeline to recover 4D scenes (*i.e.* 3D volume in time) from CT sinograms.

Mathematically, in a parallel-beam CT configuration, the three dimensional Radon transform models the CT measurement of a dynamic scene at a particular viewing angle as follows:

$$p_\theta(r, z) = \iint \sigma(x, y, z, t)\delta(x\cos(\theta(t)) + y\sin(\theta(t)) - r)dxdy,$$

where $\sigma(x, y, z, t)$ is scene's linear attenuation coefficient (LAC) at coordinates (x, y, z) and time (t), $\theta(t)$ is the view angle at time t, and $\delta(\cdot)$ is the Dirac delta function [21]. The result of this transform, $p_\theta(r, z)$, is the summation of the scene's LAC at view angle θ, detector pixel r and height z. In other words, $\{p_\theta(r, z)|(r, z) \in \mathbb{R}^2\}$ is the 2D projection of the 3D volume $\sigma(x, y, z, t)$ at the view angle θ and time t. We assume the scene's LAC remains fixed for single projection, such that $||\sigma(x, y, z, t+dt) - \sigma(x, z, y, t)||_1 = 0$ where dt is the exposure time of the CT scanner.

3.1. Modeling Assumptions:

Here we discuss the main assumptions made by our synthetic dataset and reconstruction algorithm.

Parallel-Beam Geometry: Our first key assumption is that we assume a parallel-beam geometry, the geometry usually used by synchrotron CT scanners. Synchrotron scanners have vital applications for medical and industrial applications [32, 23, 15], and we expect our algorithm to be useful in these domains. Our reconstruction geometry would need to be modified to reconstruct CT measurements from cone-beam or helical measurements, typically used in medical imaging applications. The required modification is replacing our Radon transform with a differentiable volume ray tracer, for example those shown in [29, 34]. We leave this modification of our algorithm to future work.

Data Pre-Processing: We assume CT measurements have undergone pre-processing to account for beam hardening and truncation. First, we consider the development of our algorithm for synchrotron systems which employ a monochromatic x-ray source and are therefore not affected by beam hardening. In cases where the effect is unavoidable, many methods exist for correcting beam hardened measurements and subsequently enabling high quality reconstructions [16, 4]. Truncation is not common in scientific and industrial imaging, but when present in measurements, methods exist for correcting it [50].

Limited View Sampling: Previous 4D-CT methods assume either phase-based information or a sparse set of angular samples in order to reconstruct the scene. However, objects undergoing general deformations are not amenable to phase gating and sparse sampling schemes require scanner rotation speed to be fast relative to scene motion. Our method relaxes these assumptions — we do not require a specialized sampling scheme or phase information for reconstruction. Rather, we assume only that the scene is captured sequentially between start and stop angles and that motion is static for the duration of the scanner's exposure

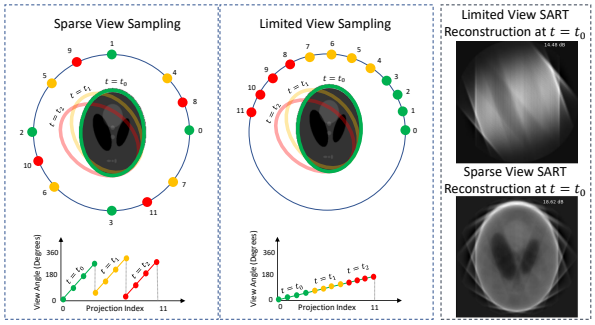

Figure 2. Left: Sparse view sampling used by baseline methods [30, 51, 52]. Right: Limited view sampling scheme we address with our method. Reconstructing objects from limited views is traditionally considered more challenging than from sparse views. We illustrate this fact on the far right of the figure where we show SART [2] reconstructions of the Shepp-Logan phantom [42] from 20 limited views versus 20 sparse views. Note the superior structural recovery of the phantom under sparse view sampling.

time. Consequently, our sampling method requires fewer revolutions of the turntable during a single scan, and we show that we outperform state-of-the-art methods under this sampling scheme. This fact implies our method will enable slower CT scanners to capture moving objects at a fidelity that was impossible with existing methods. In Figure 2, we illustrate the differences between limited view and sparse view sampling for CT imaging.

4. System Architecture

As described earlier, limited view 4D CT is an ill-posed problem both due to scene motion as well as the incomplete data captured from dense angular sampling. Our key insight is to leverage implicit neural representations to jointly learn continuous functions of the scene volume and its evolution in time. In this section, we present our algorithmic pipeline consisting of three main parts: (1) an INR to estimate a template reconstruction of the static 3D volume LACs; (2) a parametric motion field that warps the template in time; and (3) a differentiable Radon transform to synthesize an estimate of the sinogram measurements. This pipeline is optimized jointly via analysis-by-synthesis with the ground truth sinogram measurements in a self-supervised fashion.

Template Estimation: In order to estimate a template of our measured volume's LACs, we use an implicit neural representation architecture, as illustrated in the upper left portion of Figure 1. Specifically, we use a multi-layer perceptron (MLP), denoted by the function $\tilde{\sigma}_\phi$, that maps scene coordinates (x, y, z) to a template reconstruction of the scene's LACs $\tilde{\sigma}(x, y, z)$, such that $\tilde{\sigma}_\phi : (x, y, z) \mapsto \tilde{\sigma}(x, y, z)$ — we denote the tunable parameters of the MLP as ϕ. Note that $\tilde{\sigma} \neq \sigma$, *i.e.*, this template reconstruction does not equal the actual reconstruction of the scene's LACs until we use the parametric motion field to warp the template to be consistent with the measurements given in the sinogram. In implementation, we input a grid of (x, y, z) coordinates. Specifically, let $\mathbf{V} \in \mathbb{R}^{\beta^3 \times 3}$ be a voxel representation of the scene, and the scene boundaries defined as $[-1, 1]^3$. For our experiments, we set $\beta = 80$ and sample our INR with 80^3 linearly spaced coordinates at each iteration. Importantly, we perturb these coordinates randomly within their voxel to ensure the INR learns the scene's continuous representation.

Our INR architecture is inspired by [29] — we use 4 fully-connected MLP layers with ReLU activiations. We use Gaussian random Fourier feature (GRFF) [47] to randomly encode input coordinates with sinusoids of a random frequency. Formally, let $\mathbf{v} = (x, y, z)$ be a coordinate from the input grid. Its GRFF is computed as $\gamma(\mathbf{v}) = [\cos(2\pi\kappa\mathbf{B}\mathbf{v}), \sin(2\pi\kappa\mathbf{B}\mathbf{v})]$, where cos and sin are performed element-wise; $\mathbf{B}$ is a vector randomly sampled from a Gaussian distribution $\mathcal{N}(0, I)$, and κ is the bandwidth factor which controls the sharpness of the output from the INR. Similar to [47], we find that tuning the κ parameter regularizes our reconstruction. As shown in the supplemental material, setting κ too low prevents the INR from fitting high frequency content in the scene. Conversely, setting it too high causes the INR to fit spurious features in the measured sinogram resulting in poor reconstruction quality.

Motion Estimation: To map the estimated template LAC to the sinogram measurements, we introduce a parametric motion field to warp the template to different time values (*i.e.*, $\tilde{\sigma} \rightarrow \sigma(x, y, z, t_0), \sigma(x, y, z, t_1), \ldots, \sigma(x, y, z, t_N)$). Specifically, we define a tensor $\mathbf{C} \in \mathbb{R}^{\beta^3 \times k \times 3}$. This tensor contains k polynomial coefficients at each scene voxel in β^3 in the 3 spatial dimensions (x, y, z). Next, we define N time samples linearly spaced within $[0, 1]$ where N is the number of angular measurements as $t_{i=0 \cdots N-1}$. To warp a voxel to a specific time t_i, we compute the polynomial $W(\mathbf{C}, t_i) = \mathbf{C}_0 t_i^0 + \mathbf{C}_1 t_i^1 \cdots \mathbf{C}_k t_i^k$, where $W(\mathbf{C}, t_i) \in \mathbb{R}^{\beta^3 \times 3}$ is the warp field of the scene at t_i, and $\mathbf{C}_k$ is $\mathbf{C}(:, k, :)$. We generate scene frames using a differentiable grid sampling function as introduced in [35], $\textit{warp_fn}(W(\mathbf{C}, t_i), \hat{\sigma}) = \sigma(x, y, z, t_i)$. Typically, we observe that polynomials of order $k = 5$ are sufficient for describing the deformable and periodic motion present in our empirical studies.

Hierarchical motion model: We observed that attempting to estimate the warp field at the full volume of β^3 results in poor motion reconstruction. To address this issue, we introduce a hierarchical coarse-to-fine procedure for estimating motion. Specifically, our initial motion field at a base resolution α such that $\mathbf{C}_\alpha \in \mathbb{R}^{\alpha^3 \times k \times 3}$ where $\alpha < \beta$. We iteratively increase α throughout training (*e.g.* $2^3 \rightarrow 4^3 \rightarrow 8^3 \rightarrow 16^3 \rightarrow \ldots$) and use linear upsampling to progressively grow our warp field like $\mathbf{C}_{\alpha_{i+1}} = U(\mathbf{C}_{\alpha_i})$

where $U : \mathbb{R}^{\alpha_i^3} \mapsto \mathbb{R}^{\alpha_{i+1}^3}$. This strategy encourages our optimization to first recover simple motion and then iteratively recover more complex deformations.

Differentiable Radon Transform: After estimating a sequence of LAC volumes $\sigma(t_0), \ldots, \sigma(t_N)$ by applying our motion field W to the LAC volume template $\tilde{\sigma}$, we then project the LAC volumes through the CT forward imaging model to synthesize CT measurement using the 3D Radon transform. Thus we can compare our synthesized measurements to the ground truth CT measurements provided by the captured sinograms. To enforce a loss, we implement the 3D Radon transform in an autograd enabled package so that the intensity of each projected pixel is differentiable with respect to the viewing angle. We backpropagate derivatives through this operation and update the INR and motion field parameters for analysis-by-synthesis.

The weights of the INR and the coefficients of the parametric motion field are updated via gradient descent to minimize our loss function

$$\min_{\phi,\mathbf{C}} \lambda_1 \left|\left|R_{\theta(t)}(\sigma(x,y,z,t) - R_{\theta(t)}(GT)\right|\right|_1 + \lambda_2 \mathrm{TV}(\mathbf{C}), \quad t \in [0,1]. \tag{1}$$

The first term in our loss function is an L1 loss between our synthesized and given sinogram measurements. Additionally, we regularize the motion field by penalizing the spatial variation of the coefficients. The weight of λ_2 governs the allowed spatial complexity of the motion field. Higher values create smoother warp fields but may underfit complex motion, and low values allow fitting of complex motion but are more prone to noisy solutions.

Continuous Volume and Time Representation: Due to memory constraints, we sample our INR at resolution of 80^3 points during optimization. However, similar to [29], we randomly perturb these points at each iteration, encouraging our INR to learn a continuous mapping from (x, y, z) to scene LAC. This continuous mapping property is useful because it allows us to query our INR at an arbitrary resolution post-optimization. Similarly, we sample our motion field at random times t during the optimization which encourages the polynomial coefficients to fit a continuous representation of the scene motion. We use this fact to upsample our scenes to arbitrary frame rates post-optimization. Due to the parameteric representation of the motion field, we are constrained to simple trilinear interpolation for upsampling the field itself. We find that this works well in practice, but is something to be addressed in future work.

Using the upsampling functionality, we show that we can optimize our scene on a set of relatively low-resolution measurements (*e.g.* 80^3) at 10 time frames, and then upsample the measurements to 256^3 at 90 time frames ($256 \times 256 \times 256 \times 90$), making our method practically viable while also bypassing substantial GPU memory requirements. We demonstrate this fact in Section 6 and supplemental videos.

5. Implementation

Datasets: We benchmark our algorithm and competing SOA methods on a dynamic 4D-CT dataset of object deformation that we created (D4DCT Dataset). This dataset represents a time-varying object deformation to demonstrate damage evolution due to mechanical stresses over time for the study of materials science and additive manufacturing. The deformation by the damage evolution provides crucial information about the performance and safety of the material of interest, more accurate, physics-based simulation is needed. To this end, we generated dataset using the material point method (MPM) [39] to accurately represent deformation of a type of aluminum under various loading conditions. We then simulated 4D sinogram data with the provided angular ranges of 180 and 720 degrees where the number of uniformly spaced projections is 90 and the detector row size is 80. The dimension of the ground truth volume is 80^3 and the number of ground truth frames for algorithm evaluation is 10. We plan to open source this dataset to encourage reproducibility in 4D-CT research.

We also benchmark our algorithm on thoracic CT data [5]. This dataset contains volumetric reconstructions of the chest cavity at 10 breathing phases. There is motion present due to periodic functions of the diaphragm and heart. We project the 10 reconstructions from an 80^3 volume into sinogram space with 90 uniform angular projection between 0 and 180 degrees to emulate the real sinogram data.

Comparisons: To our knowledge, we are the first method to propose solving 4D CT in the limited angle regime without the use of motion phase information. However, we benchmark against two baseline methods typically used for sparse angular views on our limited angle datasets: TIMBIR [30] and Warp and Project [52]. TIMBIR uses sparse angular views with interleaved sampling to recover 4D-CT reconstructions. Warp and Project jointly solves for motion and the object reconstruction from sparse angular views. We note that these methods are expected to perform poorly in our limited angle sampling regime as both are designed for sparse sampling in time. We also benchmark the datasets with the filtered back projection (FBP) method for static 3D CT. This method is not designed to account for motion and serves to illustrates the degrading effects motion has on reconstruction performance. We note that we utilize source code for both [30] and [52] given to us by the authors for running our experiments.

Algorithm Implementation Details: Our algorithm is implemented in PyTorch to run on two Titan X GPUs for 15 minutes per recovery of an 80^3 LAC volume in time from a given sinogram. We use a Fourier Feature value of $\kappa = 1$ or $\kappa = 1.5$, as well as the ADAM optimizer [26] with learning rate .001, $\lambda_1 = 1$, and $\lambda_2 = 0.001$ for all experiments. In the supplemental material, we present a detailed list of network layers and parameters for full reproduction. We run

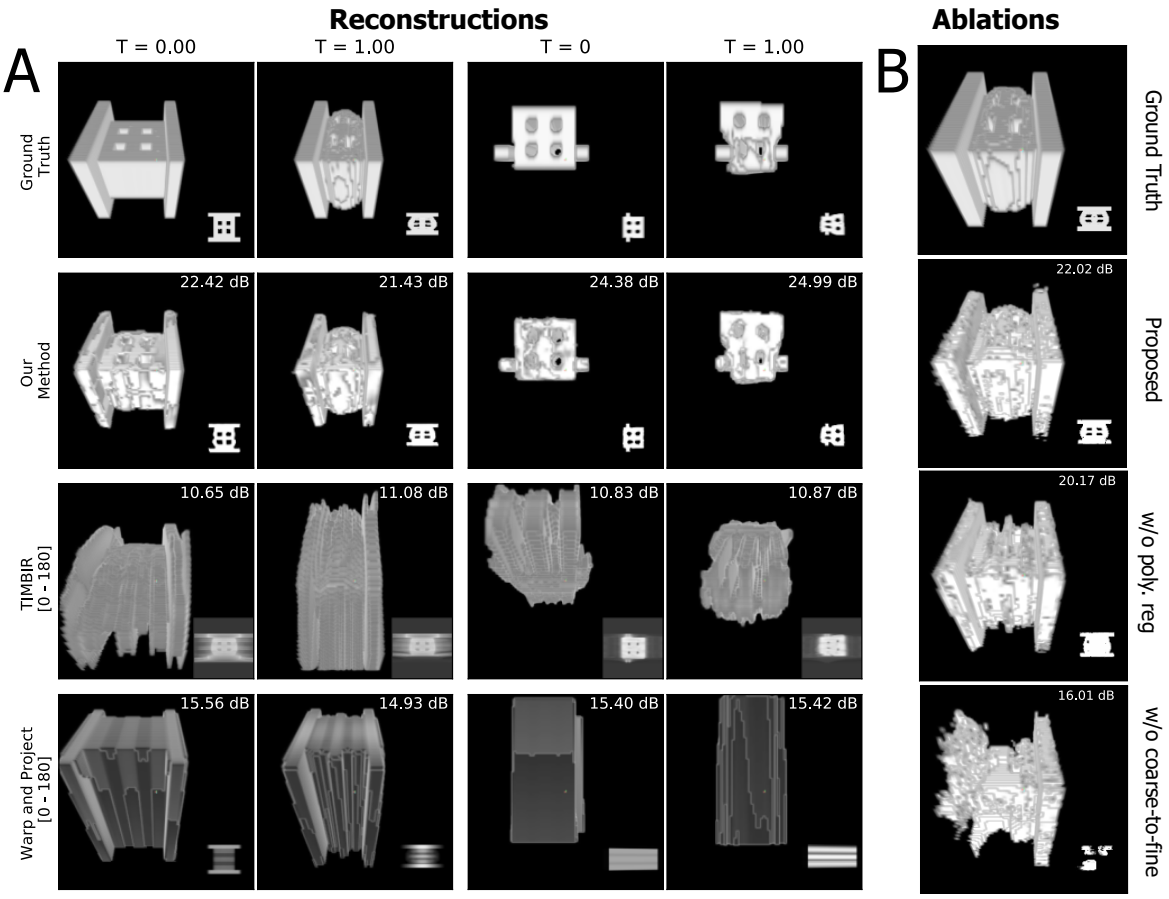

Figure 3. **(A)** Reconstruction results of our method and competing baseline methods on two objects (Left: Object Alum#1. Right: Object Alum#2.) from our aluminum deformation dataset at the beginning ($T = 0.00$) and end ($T = 1.00$) of the deformation. On the left, two plates compress the center mass of the object, and on the right, the center of the object is squeezed by two bars. In each tile, we display 3D rendering of the object to the left, and on the right, a white 2D inlet containing an XY slice through the center of the object. The PSNR of each frame is shown in white at the upper right corner of each method's tile. **(B)** Ablation of key motion field regularization components.

TIMBIR on a one P100 GPU for an average of 2 minutes per 80^3 scene. Warp and Project ran on a laptop with 16 GB of RAM and ran for approximately 6 hours to reconstruct 80^3 scenes in our dataset.

6. Experimental Results

Here we benchmark the performance of our algorithm against baselines on our D4DCT dataset and the thoracic dataset. While we show these results at a resolution of 80^3 to compare with our baselines, we also demonstrate our ability to upsample our results to a more practical CT resolution of 256^3. In addition, we show the ability to upsample our videos to arbitrary frame rates in the supplemental videos. Finally, we show ablations on our method and and its sampling schemes at the end of this section.

D4DCT Dataset: As shown in Figure 3(A) and summarized in Table 1, our proposed method drastically outperforms competing methods in peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) on the D4DCT dataset. Our method recovers the geometry and deformation of the aluminum object accurately in time. We encourage the reader to refer to the supplemental videos of these reconstructions to see the full deformations over time.

Object	Ours	TIMBIR[30]	Warp[52]	FBP
Alum#1	**22.68/0.95**	10.95/0.08	15.12/0.72	9.29/0.08
Alum#2	**24.50/0.96**	10.74/0.04	14.22/0.65	10.65/0.07
Alum#3	**26.47/0.98**	11.23/0.06	16.01/0.76	14.82/0.12
Alum#4	**26.01/0.98**	11.06/0.04	15.65/0.77	9.76/0.06
Alum#5	**26.56/0.98**	11.39/0.06	16.31/0.76	13.00/0.11
Alum#6	**24.66/0.97**	10.98/0.04	15.37/0.72	9.03/0.05
Thorax	**22.45/0.90**	14.82/0.61	8.27/0.17	15.36/0.63

Table 1. Summary of the results shown in Figures 3 and 4 benchmarking our algorithm against two SOA methods and the FBP method. We report PSNR/SSIM metrics averaged over 10 reconstructed estimated and ground truth frames. Alum #3-#6 results are shown in the supplemental material.

We observe that our method is not without flaws —

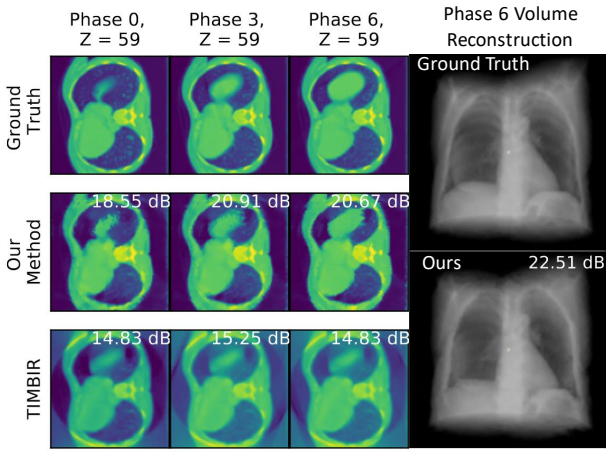

Figure 4. The top row shows ground truth XY slices of the thoracic cavity at three different breathing phases (i.e., inhalation and exhalation states). We show results of our method and TIMBIR in the middle and bottom rows respectively. On the right, we show our volumetric reconstruction of the chest cavity. Please see supplemental materials for a video of this reconstruction.

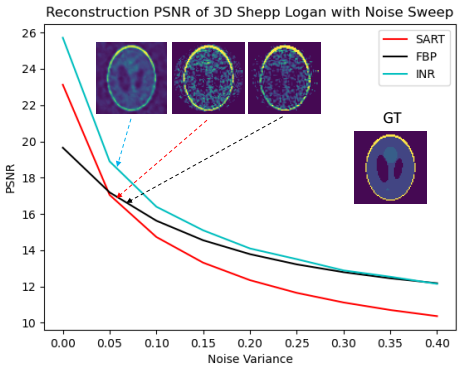

Figure 5. The INR model outperforms conventional CT reconstruction methods [21, 2] in static 3D reconstructions. We show transverse slices of the 3D Shepp-Logan for viewer reference.

our reconstructions contain artifacts on the surface of the aluminum object. However, we notice that the competing methods of Figure 3(A) contain much more severe artifacts. These artifacts exist because the limited view sampling scheme prevents these methods from constructing sufficient initial estimates of the object at each time step; these algorithms expect a sparse set of samples in order to form these initial reconstructions, as shown in 2. We believe our method performs significantly better in this sampling scheme because we optimize a single reconstruction that is warped through time, meaning our optimization leverages the full angular range to reduce artifacts.

Thoracic CT Data [5]: In Figure 4, we display our reconstruction results for 3 breathing phases of our thoracic CT data. In the upper half of each transverse slice, the top portion of the diaphragm is observed progressively rising and occupying more of the scene at each breathing phase. Our method recovers this motion and the overall geometry of the thoracic cavity. We encourage the reader to view the supplemental material videos to view the full reconstructions in time. Compared to the TIMBIR reconstruction, we observe that our method preserves sharper details and achieves a better estimate of the motion. We also benchmarked the Warp and Project method on this data, but the reconstruction quality was subpar as noted quantitatively in Table 1. This may be due to the fact that its code implementation is optimized for sparse angular sampling and not robust to limited angular sampling.

6.1. Ablation Studies

INR Reconstruction: Using an INR in our reconstruction pipeline allows us several key advantages. First, we observe that our INR outperforms conventional reconstruction methods on 3D scene reconstructions. In Figure 5, we tasked an INR and the two conventional reconstruction methods (SART and FBP), the 3D reconstruction technique used by our 4D-CT baselines [52, 30], to reconstruct a 3D Shepp-Logan phantom. The INR is implemented as a MLP and takes GRFF features of (x, y) coordinates as input to predict the LAC at each coordinate. The scene is projected to sinogram space with the Radon transform and compared to the given limited view measurements to enforce a loss via gradient updates of the INR weights. We observed that the INR gave better reconstruction PSNR. We also tested this performance in the presence of additive noise and still observe better performance under these conditions. We believe this performance gap extends to the 4D problem since our 4D reconstruction results drastically outperform baselines methods that use SART and FBP.

Secondly, our INR allows us to upsample the scene to arbitrary resolutions in the post-optimization, as shown in Figure 6. Impressively, our INR yields sharp results that resemble ground truth at the high resolution of 256^3 despite being optimized on data at the low resolution of 80^3. Further, we observe that it qualitatively and quantitatively outperforms naive trilinear upsampling methods. However, our method is not perfect and fails to capture fine details like arteries at the high resolution. This is possibly because these subtle details were too degraded at the optimization resolution 80^3 for the INR to recover this structure.

Parametric Motion Field Regularization: We regularize our motion field both by choosing an appropriate order for its polynomial equation and with the hierarchical motion model. In Figure 3(B), we illustrate the importance of these methods on the final reconstruction quality of our scenes. Of these two methods, the coarse-to-fine ablation affected the largest change on the reconstruction PSNR with a 6 dB performance improvement. This result is expected as the motion field recovery is extremely ill-posed — simplifying its initial estimation ensures the motion field does not imme-

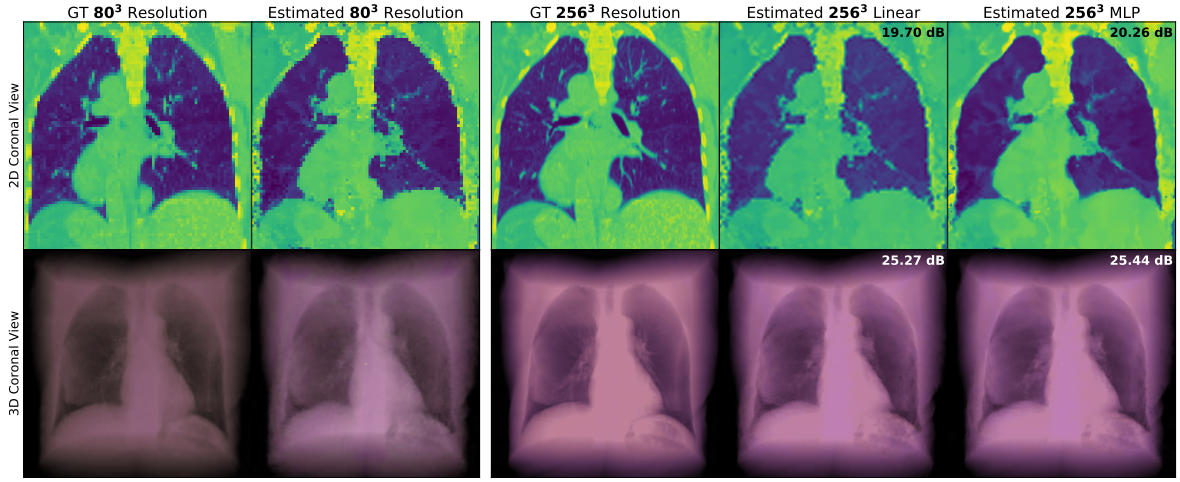

Figure 6. We demonstrate an ability to effectively upsample our scenes to arbitrary resolutions by sampling our INR's continuous representation, and these upsampled scenes outperform naive upsampling using trilinear interpolation. In this experiment, we use the thoracic data to 80^3 as ground truth (first column), and performed scene reconstruction with our method (second column). In the third column, we show the ground truth thoracic data at its native resolution 256^3. In the last two columns from left to right, we show the results of upsampling our 80^3 reconstruction with a trilinear interpolation, and upsampling by querying our INR (i.e., MLP) at dense (x, y, z) coordinates.

diately overfit to noisy solutions. For the warp field polynomial, we observed that a parametric motion field with low order polynomials underfits non-linear motion, and thus required order 3 or higher for satisfactory performance.

Effects of Angular Sampling on Performance: We observe enhanced reconstruction performance of our tested methods when we increase the angular range for a fixed number of projections (*i.e.*, make the samples more sparsely situated). As an example, we compared our method's and TIMBIR's reconstruction of an aluminum object (Alum#5) measured with 90 projections within 180 degrees — our method reconstructed the object to 25.30 dB compared with TIMBIR's 11.20 dB. When we the increased the angular range to 720 degrees, we measured 27.69 dB and 16.28 dB for our method and TIMBIR, respectively. We provide visualizations of this experiment in the supplemental material.

7. Discussion

We demonstrate that our proposed algorithm outperforms SOA methods in reconstructing limited view 4D-CT measurements of deformable motion. These results have the potential to enable CT scanners to measure rapidly moving scenes with a fidelity that was previously unattainable. Generally, this research has the potential to enable more efficient CT scans in industrial and clinical settings.

We also address two limitations of our work. First, we only consider a parallel-beam scanning geometry. While this makes our method directly applicable to synchrotron scanners, our method needs to be modified to reconstruct cone-beam data. Several other works provide implementations of differentiable ray tracers capable of modeling this geometry [46, 11], but we leave this modification to future work. Second, we show promising results of upsampling our scenes post-optimization. However, the efficacy of this upsampling needs to be further explored and compared with running the optimization at the full resolution. Even so, we believe the upsampling results we show are a promising method for achieving super-resolution in memory-hungry regimes — a few very recent works also show impressive super-resolution results with INRs [44, 45]. We hope our work sparks interest in dynamic 4D-CT reconstructions that leverage INRs in the future.

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-CONF-816780. The first author is funded by the DoD National Defense Science and Engineering Graduate Fellowship. This work was partially funded by ONR N00014-20-1-2330.

References

[1] Anders H Andersen. Algebraic reconstruction in ct from limited views. *IEEE Transactions on Medical Imaging*, 8(1):50–55, 1989. 1

[2] 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. 2, 4, 7

[3] Rushil Anirudh, Hyojin Kim, Jayaraman J Thiagarajan, K Aditya Mohan, Kyle Champley, and Timo Bremer. Lose the views: Limited angle ct reconstruction via implicit sinogram completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6343–6352, 2018. 2

[4] Rodney A Brooks and Giovanni Di Chiro. Beam hardening in x-ray reconstructive tomography. *Physics in Medicine & Biology*, 21(3):390, 1976. 3

[5] Richard Castillo, Edward Castillo, Rudy Guerra, Valen E Johnson, Travis McPhail, Amit K Garg, and Thomas Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine & Biology*, 54(7):1849, 2009. 5, 7

[6] Wenzheng Chen, Fangyin Wei, Kiriakos N Kutulakos, Szymon Rusinkiewicz, and Felix Heide. Learned feature embeddings for non-line-of-sight imaging and recognition. *ACM Transactions on Graphics (TOG)*, 2020. 3

[7] Julian Chibane and Gerard Pons-Moll. Implicit feature networks for texture completion from partial 3d data. In *European Conference on Computer Vision*. Springer, 2020. 3

[8] Leonardo De Chiffre, Simone Carmignato, J-P Kruth, Robert Schmitt, and Albert Weckenmann. Industrial applications of computed tomography. *CIRP Annals*, 63(2):655–677, 2014. 1

[9] Emilien Dupont, Miguel Bautista Martin, Alex Colburn, Aditya Sankar, Josh Susskind, and Qi Shan. Equivariant neural rendering. In *International Conference on Machine Learning*. PMLR, 2020. 3

[10] Jürgen Frikel and Eric Todd Quinto. Characterization and reduction of artifacts in limited angle tomography. *Inverse Problems*, 29(12):125007, 2013. 2

[11] Cong Gao, Xingtong Liu, Wenhao Gu, Benjamin Killeen, Mehran Armand, Russell Taylor, and Mathias Unberath. Generalizing spatial transformers to projective geometry with applications to 2d/3d registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 329–339. Springer, 2020. 8

[12] JW Gibbs, K Aditya Mohan, EB Gulsoy, AJ Shahani, X Xiao, CA Bouman, M De Graef, and PW Voorhees. The three-dimensional morphology of growing dendrites. *Scientific Reports*, 5(1):1–9, 2015. 1

[13] Gerhard W Goerres, Ehab Kamel, Thai-Nia H Heidelberg, Michael R Schwitter, Cyrill Burger, and Gustav K von Schulthess. Pet-ct image co-registration in the thorax: influence of respiration. *European Journal of Nuclear Medicine and Molecular Imaging*, 29(3):351–360, 2002. 1

[14] Yoseob Han and Jong Chul Ye. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE Transactions on Medical Imaging*, 37, 08 2017. 2

[15] Lukas Helfen, Feng Xu, Heikki Suhonen, Peter Cloetens, and Tilo Baumbach. Laminographic imaging using synchrotron radiation–challenges and opportunities. In *Journal of Physics: Conference Series*, volume 425, page 192025. IOP Publishing, 2013. 3

[16] Gabor T Herman. Correction for beam hardening in computed tomography. *Physics in Medicine & Biology*, 24(1):81, 1979. 3

[17] T Hirano, K Usami, Y Tanaka, and C Masuda. In situ x-ray ct under tensile loading using synchrotron radiation. *Journal of Materials Research*, 10(2):381–386, 1995. 2

[18] Yixing Huang, Xiaolin Huang, Oliver Taubmann, Yan Xia, Viktor Haase, Joachim Hornegger, Guenter Lauritsch, and Andreas Maier. frestoration of missing data in limited angle tomography based on helgason-ludwig consistency conditions. *Biomedical Physics & Engineering Express*, 2017. 1, 2

[19] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017. 2

[20] Marc Kachelrieß, Stefan Ulzheimer, and Willi A Kalender. Ecg-correlated image reconstruction from subsecond multi-slice spiral ct scans of the heart. *Medical Physics*, 27(8):1881–1902, 2000. 1

[21] A. C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1998. available online at http://www.slaney.org/pct/pct-toc.html. 1, 2, 3, 7

[22] Eunhee Kang, Junhong Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Medical Physics*, 44:e360–e375, 2017. 2

[23] Johann Kastner, Bernhard Harrer, Guillermo Requena, and Oliver Brunke. A comparative study of high resolution cone beam x-ray tomography and synchrotron tomography applied to fe-and al-alloys. *Ndt & E International*, 43(7):599–605, 2010. 3

[24] Paul J Keall, George Starkschall, HEE Shukla, Kenneth M Forster, Vivian Ortiz, CW Stevens, Sastry S Vedam, Rohini George, Thomas Guerrero, and Radhe Mohan. Acquiring 4d thoracic ct scans using a multislice helical method. *Physics in Medicine & Biology*, 49(10):2053, 2004. 1, 2

[25] Hyojin Kim, Rushil Anirudh, Aditya K Mohan, and Kyle M Champley. Extreme few-view ct reconstruction using deep inference. *Advances in Neural Information Processing Systems*, 33, 2019. 2

[26] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015. 5

[27] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7708–7717, 2019. 3

[28] Matthew M Loper and Michael J Black. Opendr: An approximate differentiable renderer. In *European Conference on Computer Vision*, pages 154–169. Springer, 2014. 3

[29] 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. In *ECCV*, 2020. 2, 3, 4, 5

[30] K Aditya Mohan, SV Venkatakrishnan, John W Gibbs, Emine Begum Gulsoy, Xianghui Xiao, Marc De Graef, Peter W Voorhees, and Charles A Bouman. Timbir: A method for time-space reconstruction from interlaced views. *IEEE Transactions on Computational Imaging*, 1(2):96–111, 2015. 1, 3, 4, 5, 6, 7

[31] Shanzhou Niu, Yang Gao, Zhaoying Bian, Jing Huang, Wufan Chen, Gaohang Yu, Zhengrong Liang, and Jianhua Ma. Sparse-view x-ray ct reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 59(12):2997, 2014. 1

[32] Alexandra Pacureanu, Max Langer, Elodie Boller, Paul Tafforeau, and Françoise Peyrin. Nanoscale imaging of the bone cell network with synchrotron x-ray tomography: optimization of acquisition setup. *Medical Physics*, 39(4):2229–2238, 2012. 3

[33] Tinsu Pan, Ting-Yim Lee, Eike Rietzel, and George TY Chen. 4d-ct imaging of a volume influenced by respiratory motion on multi-slice ct. *Medical Physics*, 31(2):333–340, 2004. 1, 2

[34] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. *ICCV*, 2021. 3

[35] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8026–8037, 2019. 4

[36] Gernot Riegler and Vladlen Koltun. Free view synthesis. In *European Conference on Computer Vision*, pages 623–640. Springer, 2020. 3

[37] Ludwig Ritschl, Frank Bergner, Christof Fleischmann, and Marc Kachelrieß. Improved total variation-based ct image reconstruction applied to clinical data. *Physics in Medicine & Biology*, 56(6):1545, 2011. 1

[38] Christopher Rohkohl, Gunter Lauritsch, Alois Nottling, Marcus Prummer, and Joachim Hornegger. C-arm ct: Reconstruction of dynamic high contrast objects applied to the coronary sinus. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 5113–5120. IEEE, 2008. 2

[39] A. Sadeghirad, R.M. Brannon, and J.E. Guilkey. Second-order convected particle domain interpolation (cpdi2) with enrichment for weak discontinuities at material interfaces. *International Journal for Numerical Methods in Engineering*, 95(11):928–952, 2013. 5

[40] Chris Schwemmer, Christopher Rohkohl, Günter Lauritsch, Kerstin Müller, and Joachim Hornegger. Residual motion compensation in ecg-gated interventional cardiac vasculature reconstruction. *Physics in Medicine & Biology*, 58(11):3717, 2013. 2

[41] Chris Schwemmer, Christopher Rohkohl, Günter Lauritsch, Kerstin Müller, Joachim Hornegger, and J Qi. Opening windows-increasing window size in motion-compensated ecg-gated cardiac vasculature reconstruction. In *Proc. Int. Meeting Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.*, pages 50–53, 2013. 2

[42] Lawrence A Shepp and Benjamin F Logan. The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974. 4

[43] Emil Y Sidky, Chien-Min Kao, and Xiaochuan Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct. *Journal of X-ray Science and Technology*, 14(2):119–139, 2006. 1

[44] Ivan Skorokhodov, Savva Ignatyev, and Mohamed Elhoseiny. Adversarial generation of continuous images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 8

[45] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems. *arXiv preprint arXiv:2102.05181*, 2021. 3, 8

[46] Christopher Syben, Markus Michen, Bernhard Stimpel, Stephan Seitz, Stefan Ploner, and Andreas K Maier. Pyro-nn: Python reconstruction operators in neural networks. *Medical Physics*, 46(11):5110–5115, 2019. 8

[47] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020. 4

[48] Oliver Taubmann, Mathias Unberath, Günter Lauritsch, Stephan Achenbach, and Andreas Maier. Spatio-temporally regularized 4-d cardiovascular c-arm ct reconstruction using a proximal algorithm. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 52–55. IEEE, 2017. 2

[49] Singanallur V Venkatakrishnan, Lawrence F Drummy, Michael A Jackson, Marc De Graef, Jeff Simmons, and Charles A Bouman. A model based iterative reconstruction algorithm for high angle annular dark field-scanning transmission electron microscope (haadf-stem) tomography. *IEEE Transactions on Image Processing*, 22(11):4532–4544, 2013. 2

[50] Alexander A Zamyatin and Satoru Nakanishi. Extension of the reconstruction field of view and truncation correction using sinogram decomposition. *Medical Physics*, 34(5):1593–1604, 2007. 3

[51] Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. Space-time tomography for continuously deforming objects. *ACM Trans. Graph.*, 37(4), July 2018. 1, 2, 3, 4

[52] Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. Warp-and-project tomography for rapidly deforming objects. *ACM Transactions on Graphics (TOG)*, 38(4):1–13, 2019. 1, 2, 3, 4, 5, 6, 7

[53] Hanming Zhang, Liang Li, Kai Qiao, Linyuan Wang, Bin Yan, Lei Li, and Guoen Hu. Image prediction for limited-angle tomography via deep learning with convolutional neural network. *CoRR*, abs/1607.08707, 2016. 2

[54] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 3

[55] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics (TOG)*, 2021. 3