Supplemental Material: 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

In this document, we provide additional details on our system architecture, expand on the ablation studies of the main text, and provide additional results for the D4DCT and thoracic data. Additionally, we present attached videos showing dynamic reconstructions of our method compared with baselines. Finally, we show videos that demonstrate our ability to upsample scenes in space and time post-optimization.

1. INR Architecture: Additional Details

Network Architecture: As discussed in the main text, we use a neural network as our implicit neural representation (INR) of the scene's volume template. We illustrate the implementation of our network in Figure 1. The architecture is comprised of four fully-connected layers that map the Fourier features of input coordinates to a template volume. We use a Swish [2] activation between intermediate network layers and a sigmoid activation at the final layer's output to normalize the scene LACs.

GRFF Details: As discussed in the main text, we use Gaussian random Fourier features (GRFF) [3] at the input of our INR. 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 network. We compute 512 Fourier features for each voxel. As such, each layer of the network has 512 input channels. At each iteration of the optimization, we randomly sample coordinates within 80^3 voxels, compute 512 features per voxel, and query the network for the scene's linear attenuation coefficient (LAC) at each input coordinate. Post-optimization, we can upsample the scene by inputting a finer gird of coordinates (*e.g.*, 256^3) to the network.

Parameter Ablations: The bandwidth of the Fourier features κ regularizes our INR output. As shown in the right plot of Figure 2, choosing an appropriate value for κ helps to ensure a quality reconstruction. If κ is too low, the reconstruction is unable to fit high frequency content such as edges, decreasing the PSNR performance. Alternatively, if κ is too high, the reconstruction begins to distort with high frequency artifacts, also decreasing the PSNR performance. This tradeoff is illustrated by the upside down V shape of the right plot of Figure 2 — there is a clear optimum for the κ value against PSNR performance.

2. Motion Field: Additional Details

Coarse-to-fine: We use a motion field to warp the INR's predicted scene template into a sequence of frames. 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 $\mathbf{C}(:, k, :)$. In the left plot of Figure 2, we show that increasing the polynomial order k past $2-5$ has limited effect on reconstruction quality. For all of our reconstruction experiments, we use a polynomial of order 5.

We observe that estimating our warp field at the full resolution results in supbar motion estimates, and introduce a coarse-to-fine optimization procedure for learning warp field parameters. We start our motion field at a 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 and use linear upsampling to progressively grow our warp field like $\mathbf{C}_{\alpha_{i+1}} = U(\mathbf{C}_{\alpha_i})$

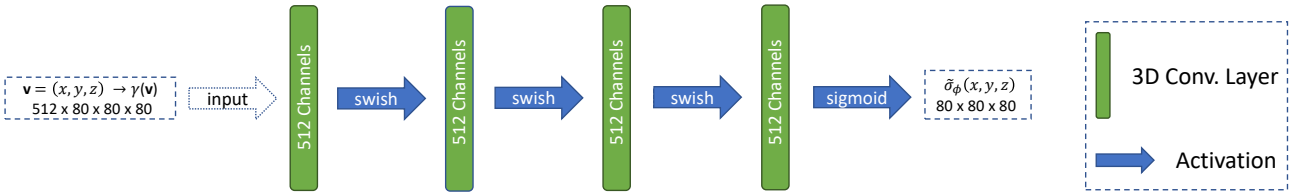


Figure 1: Illustration of our network architecture. We utilize fully connected layers that map input coordinates to a volume template that is warped by the motion field.

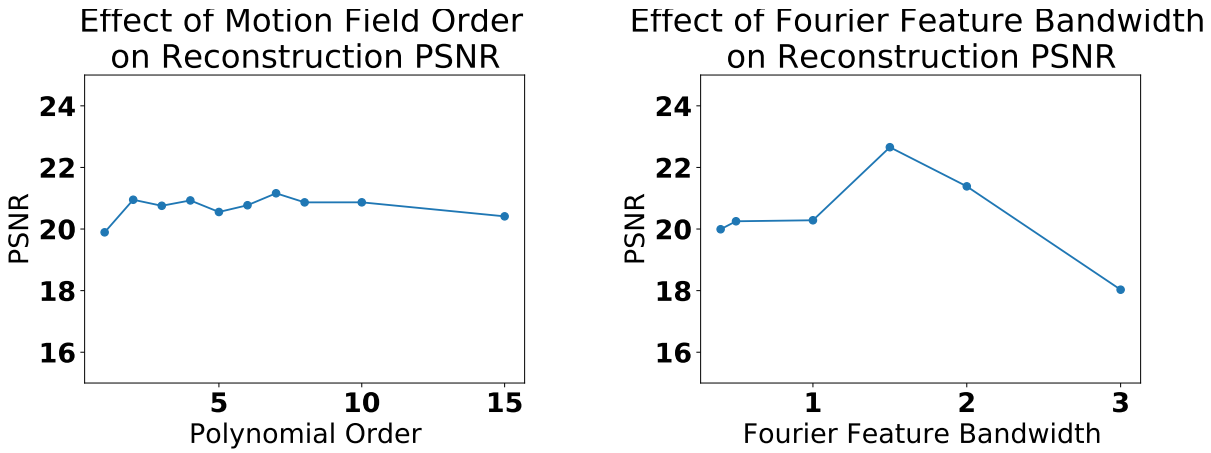


Figure 2: Left: The effect of motion field polynomial order on the reconstruction PSNR. Right: The effect of the Fourier feature bandwidth κ on reconstruction PSNR.

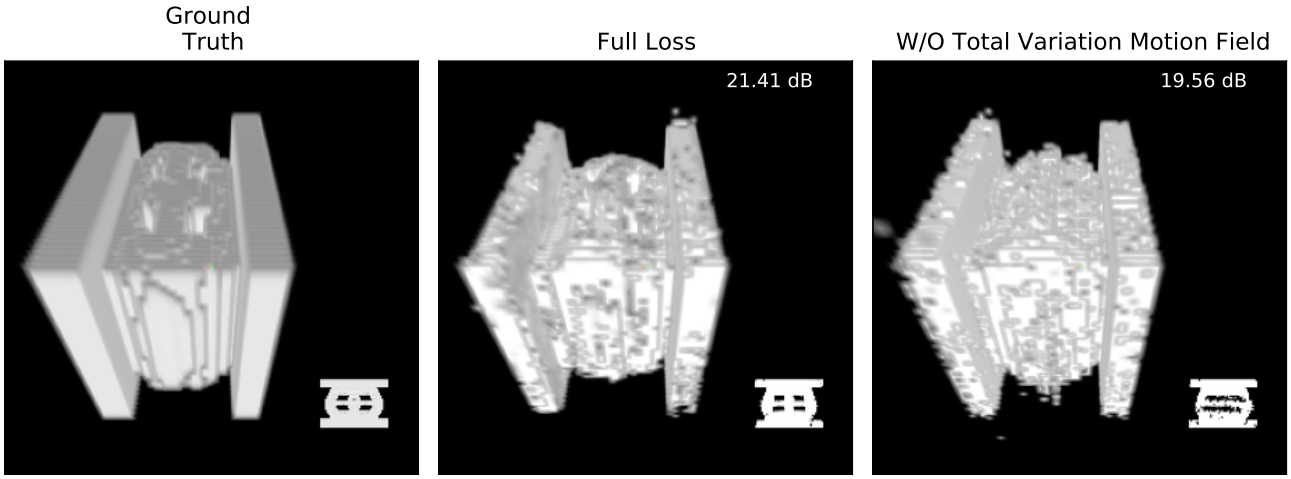


Figure 3: Effect of the total variation loss on the motion field. Notice that with the loss we resolve a smoother surface on the reconstructed object. We also notice cleaner reconstruction of the object edges as shown by 2D object slice in the bottom right corner of the tiles.

where $U : \mathbb{R}^{\alpha_i^3} \mapsto \mathbb{R}^{\alpha_{i+1}^3}$. For our experiments, we initialize the field as a 2^3 volume at the start of optimization and upsample by a powers of 2 (*e.g.*, $2^3, 8^3, \cdots, 64^3$) every 100 iterations. We note that the motion field is upsampled with a trilinear interpolation to the full volume resolution (80^3) in order to warp each voxel of the template volume. In Figure 4 we provide an example of a typical progression in reconstruction quality during the optimization. For the D4DCT and thoracic datasets, we run our optimization for 700 and 1200 iterations, respectively.

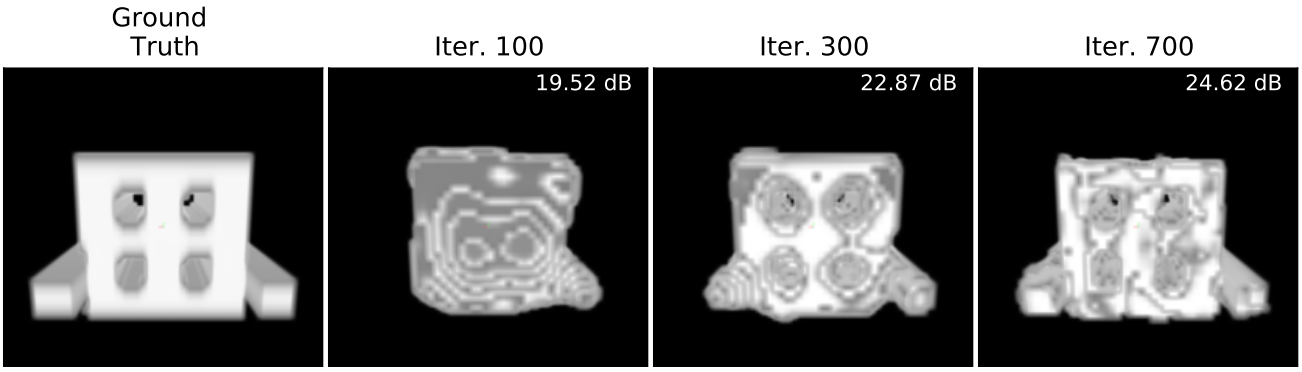


Figure 4: Progression of object reconstruction over 700 iterations. In each tile, the object (alum. #2) is shown at the first time step ($T = 0.00$).

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

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

In Figure 3 we show the impact of total variation term on the reconstruction performance. We notice that adding the regularization term boosts PSNR and reconstructs a smoother surface on the reconstructed object at each time step. In implementation, we normalize the total variation loss by the number of parameters in our motion field at each growing stage.

Visualizing the Motion Field: In Figure 5 we illustrate slices of the optimized motion field for object alum. #1 in time. Motion in the (x, y, z) direction is encoded as RGB colors. For this aluminum object, shown in Figure 9d, a laterally moving plate compresses a center cube against another plate. As shown in the bottom row of Figure 5, the green channel increases in intensity along the moving plate, indicating motion of the plate in the y direction. The middle row demonstrates this same motion in the y direction within the xz plane. The top row shows the center of the aluminum object deforming, as we see blue and red colors indicating deformation in the z and x directions.

3. Additional Results and Reconstruction Videos

D4DCT Dataset: We provide additional 4D reconstructions from our D4DCT dataset in Figure 9. We observe that our method reconstructs the geometry of the object at each of the 10 ground truth frames. However, our algorithm is not perfect and sometimes adds high frequency artifacts on the objects. This said, our method drastically outperforms baseline methods in average frame PSNR (Table 1 in main text) and qualitatively. In Figure 8, we provide visualizations of our method versus baseline reconstructions on objects alum. #3-6. Additionally, we attach two videos, *alum1.mp4* and *alum2.mp4*, showing comparisons of our method to baselines. In these videos, we observe that our method captures the geometry and deformation of the aluminum object. Baseline methods [1] and [4] reconstruct only rough estimations of the aluminum object at each time frame because of the limited angle sampling scheme.

Number of Projections	90	45	30	18	10
PSNR (dB)	22.73	21.54	19.91	19.46	17.24

Table 1: Average Reconstruction PSNR of 10 frames for Object (Alum. 1) against the number of uniform projections taken within 180 degrees. Our algorithm's performance remains robust to a limited number of projections such that our reconstruction using only 10 projections outperforms baseline methods using 90 projections.

In Table 1, we compute the reconstruction performance of our algorithm against a range of equiangular measurements within 180 degrees. Our algorithm retains reasonable frame PSNR with as few as 10 projections within a limited view. Specifically, our algorithm reconstructs object (alum. # 1.) at a higher PSNR from 10 projections than both baseline methods which used 90 projections (see Table 1 of main text for baseline PSNR scores).

Additional Results on Thoracic CT Data: In Figure 7 we show our reconstruction of the 10 breathing phases for the thoracic data. We achieve sharp reconstructions of the ground truth volume at each breathing phase. Since motion can be difficult to observe in these static images, we encour-

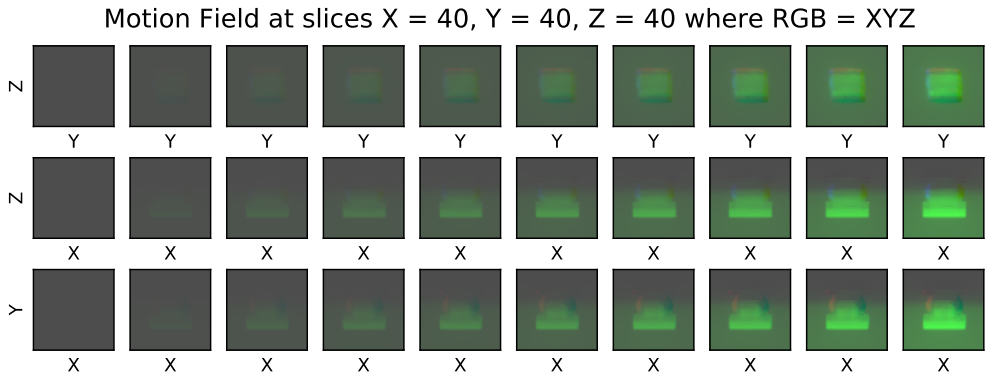


Figure 5: Motion field visualization in time for object alum #1. We encode the intensity of motion in the (x, y, z) direction as (RGB). The bottom two rows show a plate moving in the y direction over time, indicated by the green color. The top row shows the object deforming into the x and z directions indicated by the blue and red colors.

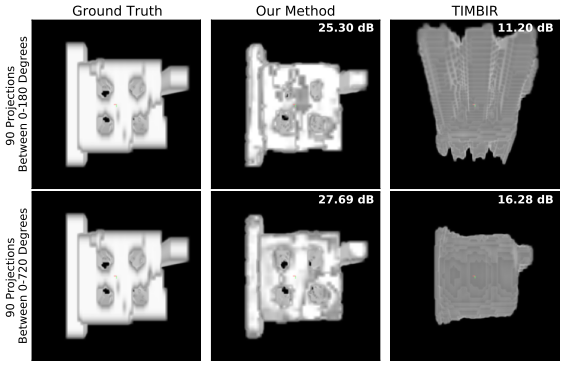


Figure 6: More sparsely sampled projections results in better reconstruction performance for ours and SOA method TIMBIR [1]. Here we show the PSNR and visual difference of reconstructing with 90 uniform projections between 180 degrees (top row) and between 720 degrees (bottom row).

age the reader to view our video reconstruction of this data (see *thoracic.mp4*). In this video, there are 10 ground truth frames, 10 TIMBIR frames, 28 warp and project frames, and 60 frames from our method. In the attached video, we demonstrate an ability to recover the vertical movement of the diaphragm at the bottom of the chest cavity, but notice our method incorrectly adds motion to the middle section of the chest. We hope to address this issue in future work by adaptively defining regions of interest for our motion field to ensure rigid portions of the object remain static over time. We notice that TIMBIR fails to capture the majority of breathing movement, and the warp and project method fails to reconstruct an initial volume of the object, likely because of the limited view nature of the measurements.

Continuous Volume and Time: We provide videos demonstrating our ability to upsample our reconstructions in space and time to arbitrary resolutions, post-optimization. In *heart_upsampled.mp4* we show an animated version of the main text's Figure 5. In this figure, our INR was optimized to reconstruct the thoracic data at a resolution of 80^3. After the optimization, we upsample the INR with finer grid (256^3) of coordinates. This strategy yields sharper reconstructions compared with naively upsampling the 80^3 scene with a linear interpolation. We observe that our method recovers breathing motion from the diaphragm and global motion around the heart. While our method misses finer motion around the heart, it far outperforms baseline methods in per frame PSNR and qualitatively as shown in *thoracic.mp4*.

In addition to upsampling scene resolution, we provide an example of sampling our scene at different frame rates. We accomplish this by sampling our warp field at an arbitrarily fine sequence of times. In the provided video, *dynamic_frame_rate.mp4*, two bars compress an aluminum cylinder. We show a time-varying reconstruction of this process at frame rates (5, 10, 30, 60).

Effects of Angular Sampling on Performance: We observe enhanced reconstruction performance of our tested methods when we increase the angular range of our projections (*i.e.*, make the samples more sparsely situated). In Figure 6 we show the reconstruction results from our method and TIMBIR [1] on data captured with 90 projections within 180 and 720 degrees. Our method resolves the object geometry and deformation in both cases, whereas TIMBIR only begins to capture the underlying geometry in the latter case.

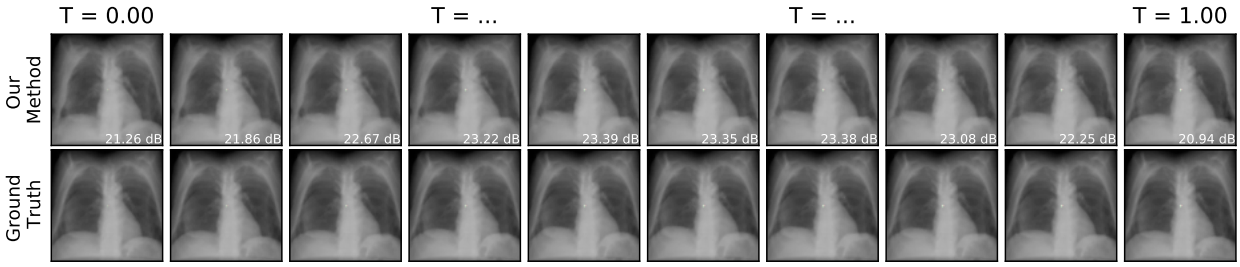
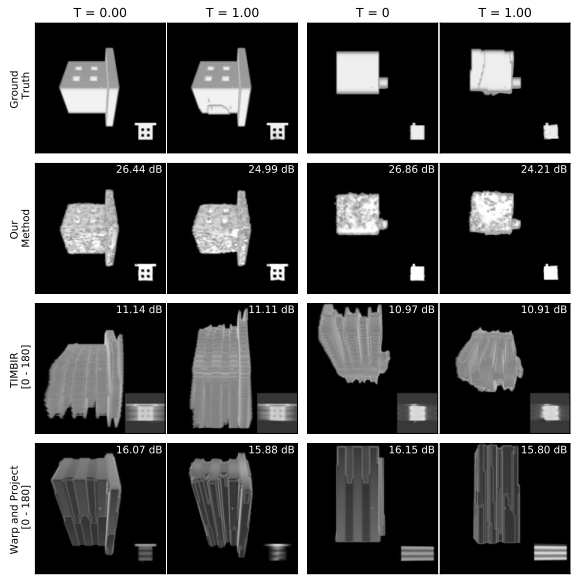
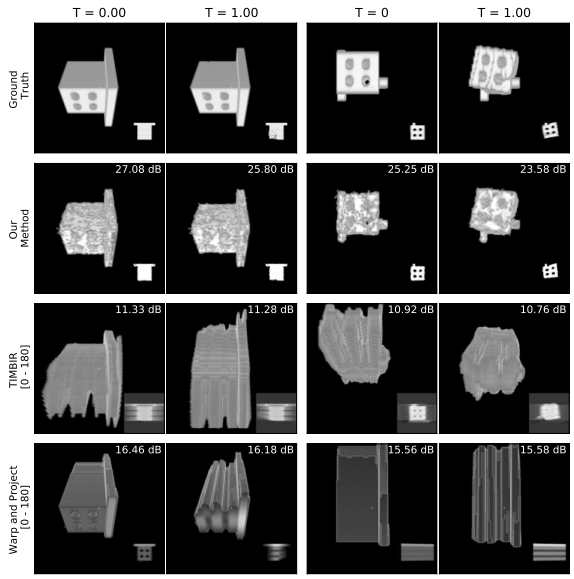


Figure 7: Our reconstruction of 10 breathing phases from the thoracic data from 90 projections between $0-180$ degrees.

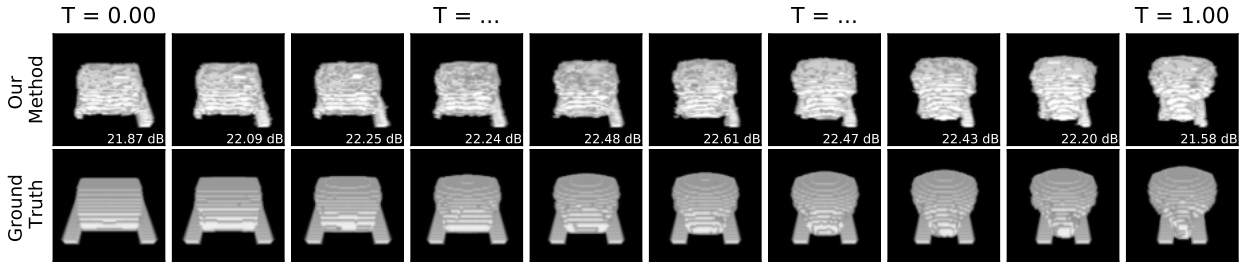


(a) Left columns: Object Alum#3. Right columns: Object Alum#4.

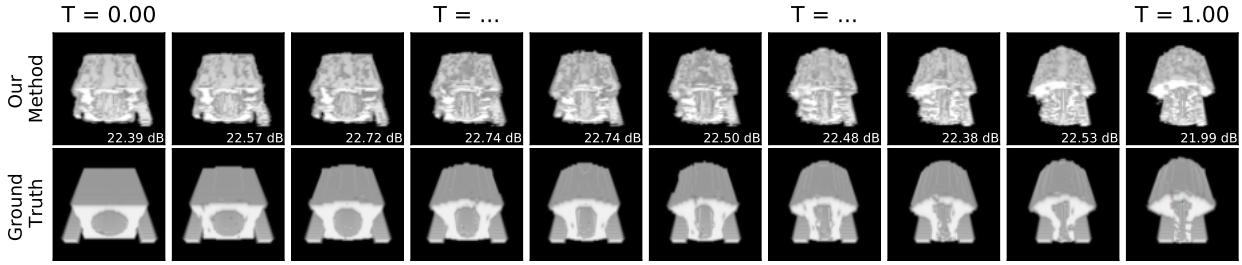


(b) Left columns: Object Alum#5. Right columns: Object Alum#6.

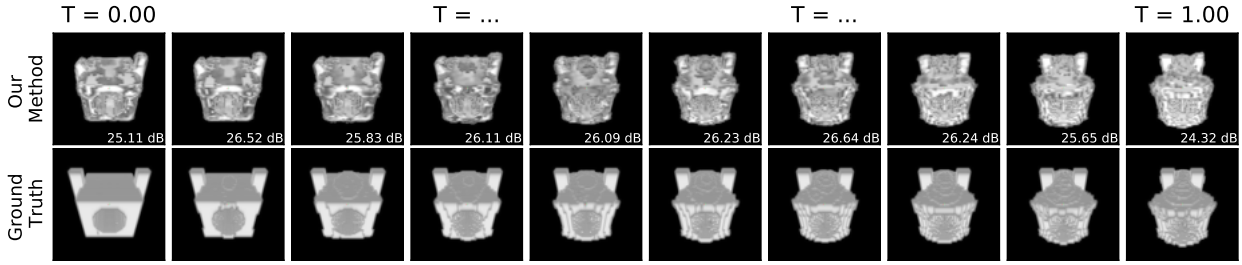
Figure 8: Reconstruction results of our method and competing baseline methods for two objects at first (T = 0.00) and last (T = 1.00) time steps. PSNR values for these reconstructions are found in Table 1 of the main text.



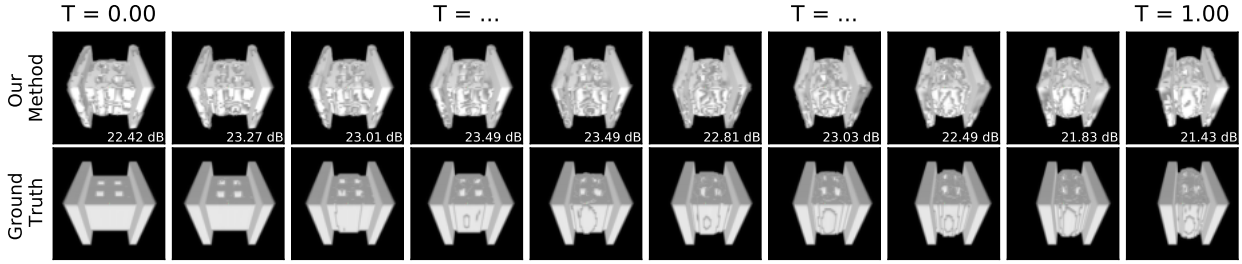
(a) Two horizontal bars deform a cylindrical center mass.



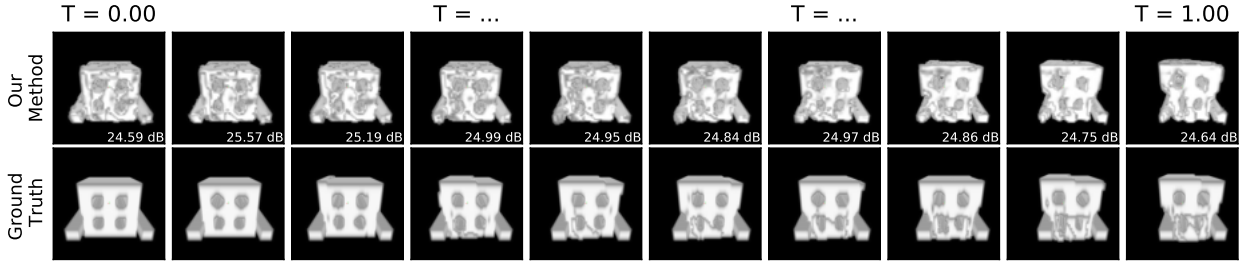
(b) Two horizontal bars deform a cube center mass containing a through-hole.



(c) Two vertical bars deform a cube center mass containing a through-hole.



(d) Object alum. # 1 where two plates deform a center mass containing four through-holes.



(e) Object alum. # 2 where two horizontal bars deform a center mass containing four through-holes

Figure 9: Limited view reconstructions of 5 objects from our D4DCT dataset displayed above ground truth for 10 time frames. The top three object rows are reconstructed from 90 uniformly spaced projections between $0 - 720$ degrees. The bottom two object rows are reconstructed from 90 projections uniformly spaced between $0 - 180$ degrees.

References

[1] 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. 3, 4

[2] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017. 1

[3] 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. *arXiv preprint arXiv:2006.10739*, 2020. 1

[4] 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. 3