Real Time Volumetric MRI for 3D Motion Tracking via Geometry-Informed Deep Learning

Lianli Liu[1], Liyue Shen[2], Adam Johansson[3,4,5], James M Balter[3], Yue Cao[3], Daniel Chang[1] and Lei Xing[1,2]

1 Department of Radiation Oncology, Stanford University, Palo Alto, CA 94304
2 Department of Electrical Engineering, Stanford University, Palo Alto, CA 94304
3 Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109
4 Department of Immunology Genetics and pathology, Uppsala University, Sweden 75105
5 Department of Surgical Sciences, Uppsala University, Sweden 75015
Email: llliu@stanford.edu, liyues@stanford.edu, adam.johansson@igp.uu.se, jbalter@med.umich.edu, yuecao@med.umich.edu, lei@stanford.edu

Mailing address: 875 Blake Wilbur Dr, Palo Alto, CA, 94304

Abstract

Purpose: To develop a geometry-informed deep learning framework for volumetric MRI with sub-second acquisition time in support of 3D motion tracking, which is highly desirable for improved radiotherapy precision but hindered by the long image acquisition time.

Methods: A 2D-3D deep learning network with an explicitly defined geometry module that embeds geometric priors of the k-space encoding pattern was investigated, where a 2D generation network first augmented the sparsely sampled image dataset by generating new 2D representations of the underlying 3D subject. A geometry module then unfolded the 2D representations to the volumetric space. Finally, a 3D refinement network took the unfolded 3D data and outputted high-resolution volumetric images. Patient-specific models were trained for 7 abdominal patients to reconstruct volumetric MRI from both orthogonal cine slices and sparse radial samples. To evaluate the robustness of the proposed method to longitudinal patient anatomy and position changes, we tested the trained model on separate datasets acquired more than one month later and evaluated 3D target motion tracking accuracy using the model-reconstructed images by deforming a reference MRI with gross tumor volume (GTV) contours to a 5-min time series of both ground truth and model-reconstructed volumetric images with a temporal resolution of 340 ms.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mp.15822.

Results: Across the 7 patients evaluated, the median distances between model-predicted and ground truth GTV centroids in the superior-inferior direction were 0.4±0.3 mm and 0.5±0.4 mm for cine and radial acquisitions respectively. The 95-percentile Hausdorff distances between model-predicted and ground truth GTV contours were 4.7±1.1 mm and 3.2±1.5 mm for cine and radial acquisitions, which are of the same scale as cross-plane image resolution.

Conclusion: Incorporating geometric priors into deep learning model enables volumetric imaging with high spatial and temporal resolution, which is particularly valuable for 3D motion tracking and has the potential of greatly improving MRI-guided radiotherapy precision.

Key words MRI-guided radiotherapy, Motion tracking, Deep learning, Image reconstruction

1. Introduction

Intrafractional patient motion challenges the precision of radiotherapy [1,2]. The advent of MRI-guided radiation therapy enables frequent monitoring of patient during treatment delivery and promises to improve treatment precision by gating treatment delivery based on tumor and/or nearby tissue position changes [3]. The long acquisition time of 3D MRI however, limits its capability of capturing patient motion in volumetric space with sufficient temporal resolution. Currently motion tracking during MRI-guided radiotherapy is mostly based on 2D imaging with a small number of slices such as cine MRI [4-6], which is limited in tracking out of plane motion. Estimating 3D image or motion information from lower dimensional surrogates including 2D images has been investigated in previous works[7-10]. However, the procedure is either iterative and not suitable for real time motion tracking or is based on template matching that requires a library of 2D samples and 3D volumes in the same coordinate space. As radiotherapy is typically delivered in fractions over multiple days, such template matching process may not be sufficiently robust to longitudinal changes of patient position and anatomy, with limited capability in extrapolating to unseen patient motion states.

Substantial effort has been made to accelerate MRI acquisition by reconstructing MRI images from sparsely sampled k-space data, including compressed sensing-based methods and low rank model-based methods [11-13], where prior knowledge of imaging subjects, such as sparsity in a transform domain or low rankness of image matrices were exploited to regularize the ill-posed problem of reconstructing MRI from subsampled k-space data. More recently, deep neural networks that are capable of learning complex data-driven priors from a training dataset have shown improved performance over conventional methods that use pre-defined priors [14-16]. The acceleration factor is, however, still insufficient in supporting volumetric motion tracking during MRI-guided radiotherapy, which requires sub-second temporal resolution to capture patient dynamics. A major deficiency of most deep learning-based methods is that the reconstruction process is purely data-driven without fully utilizing geometric priors of the image system, which play a vital role in associating intensity information with correct spatial locations. Indeed, while the prior knowledge of k-space sampling

pattern is critically important for reliable image reconstruction, it is used at most as a consistency constraint for algorithm estimations at sampled k-space locations [17,18]. As demonstrated in tomographic CT reconstruction[19], the reconstruction quality would be greatly enhanced if the geometric prior of the imaging system could be incorporated into the data-driven image reconstruction process.

In this study, we demonstrate the feasibility of real-time volumetric MRI for 3D motion tracking by combining ultra-sparse sampling with a geometry-informed deep learning model. A 2D-3D deep learning network with an explicitly defined geometry module that embeds both k-space sampling patterns and the known transform between k-space and image domain was investigated. Instead of creating volumetric images directly from k-space samples, which often suffers from severe artifacts due to extreme subsampling, we started our reconstruction by first enhancing 2D representations of the underlying 3D subject using a 2D generation network. The geometry module was then used to unfold the 2D representations to a volumetric space. Finally, a 3D refinement network took the unfolded 3D data and outputted high-resolution volumetric images. By simply changing the geometry module based on k-space sampling patterns, the same network structure was trained to reconstruct volumetric images from both cine MRI slices and sparse radial samples with sub-second acquisition time. To evaluate the robustness of the proposed method to longitudinal patient changes, we trained and tested the model on separate datasets acquired more than one month apart. The capability of the model-reconstructed images in support of 3D motion tracking was evaluated for 7 abdominal patients over a 5-min time period.

2. Materials and Methods

2.A. Problem Formulation

We investigated sparse sampling schemes for both cine and radial MRI. For cine MRI, interleaved acquisition was considered, which samples two orthogonal MRI slices of coronal and sagittal views respectively. For radial MRI, a stack-of-star acquisition pattern [20] was used, where radial readout

lines were acquired in the axial plane and Cartesian phase encoding was performed in the superior-inferior direction. The sequence sampled all radial lines with the same angle throughout the superior-inferior direction before moving to the next radial angle. A collection of radial lines with the same angle forms a radial spoke. In our study, we sampled two radial spokes with radial angles of 0° and 111.25° (the golden angle) [21] respectively. Inverse Fourier transforming each of the 2 radial spokes gives 2 projection images of the patient. Both acquisition schemes take less than 1 second (about 500 ms for orthogonal cine acquisition [22] and 340 ms for radial acquisition [23]) for large field-of-view imaging such as abdominal MRI.

With the ultra-sparse sampling scheme, filling out missing data samples in the volumetric space directly is challenging. Instead, we formulated a 2D data completion problem first before reconstructing 3D volumetric images. Denote the underlying volumetric image with an image size of $M \times M \times K$ as $I \in \mathbb{R}^{M\times M\times K}$, for cine acquisition, we constructed 2D representations of the 3D subject by sampling rotating planar images from the volumetric image. The sampling coordinates were defined in the cylindrical coordinate system with origin at the volume center and longitudinal axis parallel to the superior-inferior direction of the volume. Under this coordinate system, sampling locations for the acquired coronal cine slice $p_0 \in \mathbb{R}^{M\times K}$ were $\left\{\rho,\theta,z \middle|\ \rho = 1,2,\cdots,\frac{M}{2};\theta = 0,\pi;\ z = -\frac{K}{2},\cdots,\frac{K}{2}\right\}$ and the sampling locations for the sagittal slice $p_1 \in \mathbb{R}^{M\times K}$ were similarly defined with $\theta = \{\pi/2\,, 3\pi/2\}$. With the two acquired slices, the goal is to complete n rotating slices $p_{2,}\cdots,p_{n+1}$ sampled with $\theta = \{\pi\frac{i-1}{n+1}, \pi\left(\frac{i-1}{n+1}+1\right)\}$ for slice $p_i, i = 2,\cdots,n+1$. For radial acquisition, 2D representations were similarly constructed by sampling the k-space with varying radial angles and performing inverse Fourier transform of radial spokes to generate projection images. Denote the projection images generated with acquired radial spokes as $p_0 \in \mathbb{R}^{M\times K}$ and $p_1 \in \mathbb{R}^{M\times K}$, we aim to complete n more projection images $p_{2,}\cdots,p_{n+1}$that correspond to radial spokes with radial angle equals $i \times 111.25^{\circ}$ for projection p_i .

After defining the target 2D representations, a 2D generation network Φ_1 with network weights W_1 was trained to complete missing data in 2D space by synthesizing target 2D representations from acquired 2D samples $\Phi_1(p_0, p_1; W_1) = (p_0, p_1, \hat{p}_2, \cdots, \hat{p}_{n+1})$. A geometry module Φ_2 with fixed weights W_2 then unfolded both acquired and network-generated 2D representations to the volumetric space $\Phi_2(p_0, p_1, \hat{p}_2, \cdots, \hat{p}_{n+1}; W_2) = \hat{I} \in \mathbb{R}^{M\times M\times K}$. Finally, a 3D refinement network Φ_3 with learnt weights W_3 took the unfolded volumetric data from the geometric module as input and outputted final reconstructed images $\Phi_3(\hat{I}; W_3) = I_{recon} \in \mathbb{R}^{M\times M\times K}$. Based on previous experimental results for ultra-sparse imaging [19], in this study we set $n = 10$. Figure 1 presents the framework of the geometry-informed deep learning model, and the following sections describe each model component in detail.

2.B. Model Architecture

2.B.1. 2D generation network

Similar with previous work for sparse-view computed tomography reconstruction [19], a multi-domain image translation model [24] synthesizes new 2D representations from acquired 2D representations. The model views each 2D representation as a separate domain and completes missing domain information by exploiting both shared content between different domains (same patient anatomy) and unique information associated with each domain (different sampling locations). The network consists of 1) a shared content encoder E^c that encodes information shared by different representations into a content code c; 2) domain-specific encoders E_i^a that encode unique information of different representations into domain-specific codes $a_i,\ i = 0,1,\cdots n+1$; 3) generators G_i that generate new representations by combining shared content code c and domain specific codes $a_i,\ i = 0,1,\cdots n+1$ and 4) discriminators $D_i,\ i = 0,1,\cdots n+1$ that distinguish generated representations and real representations. The network was trained by optimizing the total loss including the reconstruction loss, the adversarial loss and the cycle consistent loss

$$\min_{E^c,E_i^a,G_i}\max_{D_i}\sum_{i=0}^{1}L_{cyc}^{p_i}+L_{cyc}^{c}+\sum_{i=0}^{n+1}(L_{cyc}^{a_i}+L_{adv}^{p_i}+L_{rec}^{pi}) \quad (1)$$

where L_{cyc} is the L1-norm cycle-consistent loss including representation consistency loss that enforces synthesized 2D representations to be consistent with input representations (planar images or projections), which is optimized over the 2 acquired representations.

$$L_{cyc}^{p_i}=\left\|G_i\left(E^c(p_0,p_1),E_i^a(p_i)\right)-p_i\right\|_1, i=0,1 \quad (2)$$

and coding consistency loss that enforces consistency of both shared content code and domain-specific codes before decoding and after encoding.

$$L_{cyc}^{c}=\|E^c(G_0(c,a_0),\cdots,G_{n+1}(c,a_{n+1}))-c\|_1 \quad (3)$$

$$L_{cyc}^{a_i}=\left\|E_i^a\left(G_i(c,a_i)\right)-a_i\right\|_1, i=0,1,\cdots n+1 \quad (4)$$

Following previous work[24], we assume prior distribution for the domain-specific codes a_i is a standard Gaussian distribution $\mathcal{N}(0,I)$. During image generation, the generator samples domain-specific codes from the prior distribution and combines the domain-specific codes with the shared content code to synthesize 2D representations. L_{adv} and L_{rec} are the adversarial loss and L1-norm reconstruction loss respectively.

$$L_{adv}^{p_i}=\log\left(1-D_i\left(G_i(c,a_i)\right)\right)+\log\left(D_i(p_i)\right), i=0,1,\cdots n+1 \quad (5)$$

$$L_{rec}^{p_i}=\|G_i(c,a_i)-p_i\|_1, i=0,1,\cdots n+1 \quad (6)$$

2.B.2. Geometry Module

A geometry module was defined for cine and radial MRI respectively. The geometry module is deterministic for given a sampling pattern and requires no additional training. For cine MRI, the

geometry module populates voxel information of the 3D volume based on volumetric sampling coordinates of both acquired orthogonal cine slices and network-synthesized rotating planar images. Specifically, the geometry module estimates the voxel intensity at $(m,n,k), -\frac{M}{2} < m,n \leq \frac{M}{2}, -\frac{K}{2} < k \leq \frac{K}{2}$, as

$$\hat{I}(m,n,k) = \hat{I}(r,\theta,k) = w_{1,1}p_i(r_1,k) + w_{2,1}p_j(r_1,k) + w_{1,2}p_i(r_2,k) + w_{2,2}p_j(r_2,k) \quad (7)$$

where $r = \sqrt{m^2+n^2}$ and $\theta = \cos^{-1}\left(\frac{m}{r}\right)$ are the cylindrical coordinates of the target voxel. p_i and p_j are the two rotating planar images with sampling coordinates $\theta_i \leq \theta \leq \theta_j$. The geometry module performs a bilinear interpolation using pixel information from p_i and p_j at location (r_1,k) and (r_2,k) with $r_1 \leq r \leq r_2$ and $w_{11},\cdots w_{22}$ are the interpolation weights, determined based on the distance between the interpolation locations and the sampled locations.

For radial MRI, the geometry module first performs a 2D uniform Fourier transform on each of the projection images to get the corresponding radial spokes $\mathcal{F}(p_i) = P_i, i = 0,1,\cdots,n+1$ and fills out the volumetric k-space based on both acquired and network-synthesized radial samples $\mathcal{F}(p_i)$. A non-uniform 3D Fourier transform (3DNUFFT) is then performed on the volumetric k-space to reconstruct 3D images. As the radial sampling pattern leads to a denser sampling of the k-space center than the periphery, a ρ filter [23] was used to compensate for the difference in sampling density in k-space before 3D NUFFT. Specifically, the filter multiplies k-space samples at sampling location $\boldsymbol{v} = (k_x, k_y)$ with density compensation weights $w(\boldsymbol{v}) = w_0(\boldsymbol{v})e^{-2\pi\sigma^2\|\boldsymbol{v}\|^2}$, where $w_0(\boldsymbol{v})$ equals the k-space volume of a semicylindrical shell with central radius $|\boldsymbol{v}|$. The shell height and width equal to the sample spacing in the longitudinal and radial directions, respectively. To reduce ringing, a Gaussian window [23] with $\sigma = 1$ voxel was also included in the density compensation function.

2.B.3. Volumetric Refinement Network

The 3D image created by the geometry module is generally not perfect and may contain imaging artifacts due to sparse sampling. A 3D network was used to refine the quality of volumetric images. A U-Net [25] based encoder-decoder network was built, where the encoder consists of four-stage down-sampling blocks and the decoder consists of four-stage up-sampling blocks. Similar with previous work on CT reconstruction [19], each down-sampling and up-sampling block consists of double 3D convolution layers followed by rectified linear unit activation layers and group normalization layers. The down sampling operation was implemented by a max pooling layer with a step of (2,2,2) and up sampling was implemented by interpolation. Hierarchical skip connections were built by concatenating feature maps between encoder and decoder of the same feature level. A final 3D convolution layer with kernel size of 1×1×1 and hyperbolic tangent activation function was used to output the reconstructed 3D image. The network was trained by minimizing the L1-norm difference between network outputs and the ground truth.

2.C. Model Training and Evaluation

Under an institution review board approved protocol (The University of Michigan Institutional Review Board, HUM00068061), 7 patients with intrahepatic tumors were involved in this study. A patient-specific deep learning model was trained separately for each of the 7 patients using a 4D MRI dataset of 21 breathing motion phases. The 4D MRI was acquired with a golden-angle stack-of-stars spoiled gradient echo sequence and reconstructed through retrospective k-space sorting and re-binning using a previously published technique [26]. All MRI data were acquired using a 3 Tesla scanner (Skyra, Siemens Medical Systems, Erlangen, Germany) with an 18-channel flexible surface coil (BodyMatrix) placed anteriorly and 1 or more posterior 4-channel coils embedded in the patient couch (Spine Matrix). The field of view covered the liver, stomach and a large portion of the intestines. The imaging parameters ranged from 1.14 to 1.21 ms for echo time, 2.71 to 4.51 ms for repetition time, 10° to 14° for flip angle, 2 to 2.45 mm for in-plane voxel size and 3 to 4 mm for slice thickness. The size of the imaging matrix was 192 × 192 and the number of slices was 64.

The training dataset was augmented by applying 3D deformations to volumetric MR images. To generate new deformation fields, principal component analysis (PCA) was performed on deformation fields associated with each of the 21 MR images that align the exhale phase MR image to other breathing motion phases, which were extracted in previous studies using B-spline deformable registration [26]. New deformation fields were generated using the mean PC mode $\boldsymbol{m}_0$ and the leading 3 PC modes $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ as $\boldsymbol{d} = \boldsymbol{m}_0 + \alpha_1 \boldsymbol{m}_1 + \alpha_2 \boldsymbol{m}_2 + \alpha_3 \boldsymbol{m}_3$ where α was varied between ±3 standard deviation around the mean PC coefficients. Through this data augmentation process, we obtained a total number of 2500 volumetric images, where 2400 images were randomly selected for training purpose and the remaining for validation purpose.

All MR images were then normalized to an intensity range of [0,1]. To train the 2D generation network, the volumetric MR images were retrospectively subsampled to 2 orthogonal cine slices or 2 radial projection images, which served as the model input. After the 2D training was completed, the geometry module was used to generate volumetric images from the network-predicted 12 rotating planar images or 12 projection images. The geometry module-generated volumetric images were then paired with ground truth images to train the 3D refinement network. Both 2D and 3D networks were trained using an Adam optimizer and a batch size of 1. The learning rate and number of iterations were 0.001/0.005 and 50000/30000 for 2D and 3D training respectively. Random 3D translation and rotation were also applied to MR images before each training epoch. To evaluate the impact of incorporating geometric priors, we compared the proposed model to a baseline deep learning model that is purely data-driven, following the work by Shen *et. al.*[29] The baseline model consists of a 2D encoder and a 3D generator with a feature transformation module that connects 2D and 3D feature space. We trained the baseline model using the same dataset of paired 2D slices/projections and 3D images as the proposed model.

To evaluate the robustness of the trained model to longitudinal patient anatomy and position changes during a radiotherapy course, we applied the trained model to a testing dataset that was

acquired more than one month after the training dataset. The same imaging sequence and parameters were used. A 5-min time series of dynamic MRI was reconstructed from the acquired k-space samples with a temporal sampling rate of 340 ms. The reconstruction was based on a previously published technique for high temporal and spatial resolution breathing motion characterization via retrospective motion state labeling and served as the ground truth for model evaluation [27]. We subsampled the ground truth volumetric images to 2 orthogonal slices or 2 projections images and input the sparse samples to the trained model. The model-reconstructed volumetric images were compared to the ground truth images in support of 3D abdominal target tracking. Specifically, the first image of the ground truth image time series was chosen as the reference image with clinically defined gross tumor volume (GTV) contours transferred to it. The reference image was then deformed to match both ground truth and model-reconstructed volumetric images using B-spline deformable registration implemented in NiftyReg [28], a registration method which has been validated in previous studies for aligning different breathing motion states [26]. Target motion during the 5-min time period was characterized by deforming the reference GTV volume with calculated deformation fields. Tracking accuracy using model-reconstructed volumetric images was evaluated by calculating the difference between centroid positions of deformed GTV volumes and 95-percentile Haursdorff distance between deformed GTV contours. The linear correlation of motion estimation using model-reconstructed images and ground truth images was also assessed similarly with previous work[8], where linear fitting was performed between the ground truth and model-estimated GTV centroid displacements in the superior-inferior direction and the R-square value was calculated.

3. Results

3.A Model Validation

After the model training, generating one volumetric MRI from sparse samples took 8.8 ms using a Nvidia Tesla K40C GPU. Figure 2 shows example model validation results for the 2D generation network. The trained network is able to synthesize new 2D representations that closely match the

ground truth from both cine and radial acquisitions. Figure 3 shows volumetric images produced by the geometry module, which serve as the input to the 3D refinement network and the final network-reconstructed volumetric images. Table 1 summarizes quantitative evaluation results including structural similarity index (SSIM), peak signal-to-noise ratio (PSNR) and mean square error between the network reconstruction and the ground truth. Figure 4 compares sample slices of 3D images reconstructed by the proposed model and the baseline model. The proposed model reduces blurry artifacts and outperforms the baseline model in terms of various image quality metrics, including structural similarity index (SSIM), peak signal-to-noise ratio (PSNR) and mean square error between model reconstruction and ground truth, as summarized in Table 1.

3.B Volumetric Target Tracking

Figure 5 shows sample views of training and testing volumetric MRI. Positional and anatomical changes can be observed in all planes due to long acquisition time intervals between the training and testing datasets, while network-reconstructed images show good agreement with testing images. Figure 6 plots example motion traces of gross tumor volume (GTV) centroid in the superior-inferior direction during the 5-min examination. Despite of varying motion patterns and motion irregularities across patients, the motion traces estimated using the model-reconstructed volumetric images are consistent with the ground truth. Across the 7 patients evaluated, the motion was most significant in the superior-inferior (SI) direction. The median distances between GTV centroids during the 5-min period and the reference GTV centroid ranged between 0.4 mm to 2.6 mm in the SI direction. The median distances in the left-right (LR) and anterior-posterior (AP) direction were less than 1 mm for most patients except for one patient where the distance was 1.4 mm in the AP direction. By updating the reference GTV centroid position using model-reconstructed volumetric images, the median distances in the SI direction were reduced and were less than 1 mm for all patients evaluated. Table 2 summarizes the differences in GTV centroid position estimation using ground truth and model-reconstructed images. The GTV contour agreement, evaluated as 95-percentile Hausdorff distance between predicted and ground truth GTV contours were 4.7±1.1 mm

and 3.2±1.5 mm for cine and radial acquisitions respectively, which is of the same scale as cross-plane imaging resolution (3.5~4 mm). The model-estimated GTV centroid displacements showed excellent linear correlation with the ground truth, with an R-square of 0.99, as comparing to the previously reported R-square of 0.95[8]. It is also worth noting the model was tested on datasets that exhibit longitudinal anatomical changes from the training dataset, as opposed to training and testing using images acquired in the same imaging session. The robustness of the proposed method to longitudinal changes may remove the need of acquiring 4D MRI before each treatment session and has the potential of supporting a simplified clinical workflow.

4. Discussion

A geometry-informed deep learning model for volumetric MRI reconstruction with ultra-sparse k-space sampling is proposed. The technique makes it possible to obtain volumetric MRI images with sub-second acquisition time, which is highly desirable for real time 3D motion tracking during MRI-guided interventions. To regularize the ill-posed problem of image reconstruction from sparse samples, both patient-specific priors learnt by the deep neural network and geometric priors that are inherent to the imaging system were exploited, which is different from previous deep learning-based image reconstruction strategies that are purely data-driven [14-16]. The proposed deep learning framework consisted of a 2D generation network that completes subsampled image dataset in 2D space, a geometry module that bridges the gap between 2D and 3D space, and a 3D refinement network that reconstructs final volumetric images. By simply changing the geometry module based on pre-defined acquisition schemes, the same network structure can be trained to reconstruct volumetric MRI from both cine and radial samples. The trained model was evaluated for 7 abdominal patients in support of 3D target tracking during a 5-min time period. The median distances between the GTV centroid positions predicted by the model and derived from the ground truth in the superior-inferior direction were less than 1 mm on average and around 1 mm in the other two directions, for both cine and radial acquisitions.

While consistency of model estimation with acquired data at sampled k-space locations has been utilized to constrain the image reconstruction process, the geometry of encoding an image subject into Fourier samples at different k-space locations has not been fully exploited. In this study, we introduced a network that generates new 2D representations based on known spatial encoding patterns of different representations (e.g. different planar or radial angles) and a geometry module that bridges the 2D representations and 3D images based on the sampling geometry and known transform between k-space and image space. Incorporating such geometric prior that is deterministic with the imaging system leverages the learning task of deep neural network and permits image reconstruction from ultra-sparse k-space samplings with sub second acquisition time. Furthermore, both network-learnt and geometric priors are not bound to a specific imaging session or acquisition position, which makes the method desirable for real time imaging guidance over an entire radiotherapy course that is delivered over multiple days, as demonstrated by testing the model on separate datasets acquired more than one month later than the training dataset.

Future work will address several limitations of the current method. Firstly, the proposed deep learning model is patient-specific, meaning that separate models need to be trained for different patients. Also, the training process requires a 4D MRI dataset, which may not be readily available. Future work will investigate both patient-specific model training using a single static MRI with synthetic motion phases, and a population-based model with a larger number of patients. Other network structures for image synthesis and refinement may also be interesting and should be investigated in the future. Secondly, target and organ motions are triggered by many factors and uncertainties introduced by non-breathing motion have been shown to be significant for radiotherapy [30,31,32]. While the current work focuses on breathing motion tracking, the capability of the proposed method in tracking non-breathing motions such as organ peristalsis and slow anatomical rearrangements will be evaluated in future studies. We also noted that the size of the patient dataset used in the current study is limited with limited motion variations. One interesting study in future work could be a larger patient dataset with more significant motion variations.

Thirdly, the proposed sparse sampling scheme samples the superior-inferior direction more densely than other directions. This may contribute to the larger tracking error in anterior-posterior and left-right directions. Future work will investigate other sampling schemes, such as 3D radial sampling for potentially improved tracking performance. Finally, the proposed model was trained and evaluated using image samples acquired on a 3T MRI scanner. Impacts of the low magnetic field strength of MR-Linac systems on the proposed method (e.g., the impact of low image signal-to-noise ratio on the model's capability of extracting features) will be investigated in future studies to ensure the clinical applicability of the proposed method.

5. Conclusion

A geometry-informed deep learning model that reconstructs volumetric MRI from ultra-sparse k-space samples has been developed in support of real time 3D motion tracking during MRI-guided radiotherapy. By exploiting geometric priors that are inherent to the imaging system, the learning task of the neural network is simplified and can be focused on learning patient-specific priors. Model-reconstructed volumetric MRI from both cine and radial samples with sub-second acquisition time shows sufficient accuracy in tracking 3D abdominal target motion. Furthermore, we demonstrated the robustness of the trained model to patient position and anatomy changes over time by testing the model using a longitudinal dataset, which makes the proposed method desirable for providing imaging guidance during a radiotherapy course that is fractionated over multiple days.

6. Acknowledgements

This work was supported by NIH 1R01CA256890, 1R01CA227713, and 1R01 EB016079. The radial scanning sequence was provided under a research agreement from Siemens Healthineers.

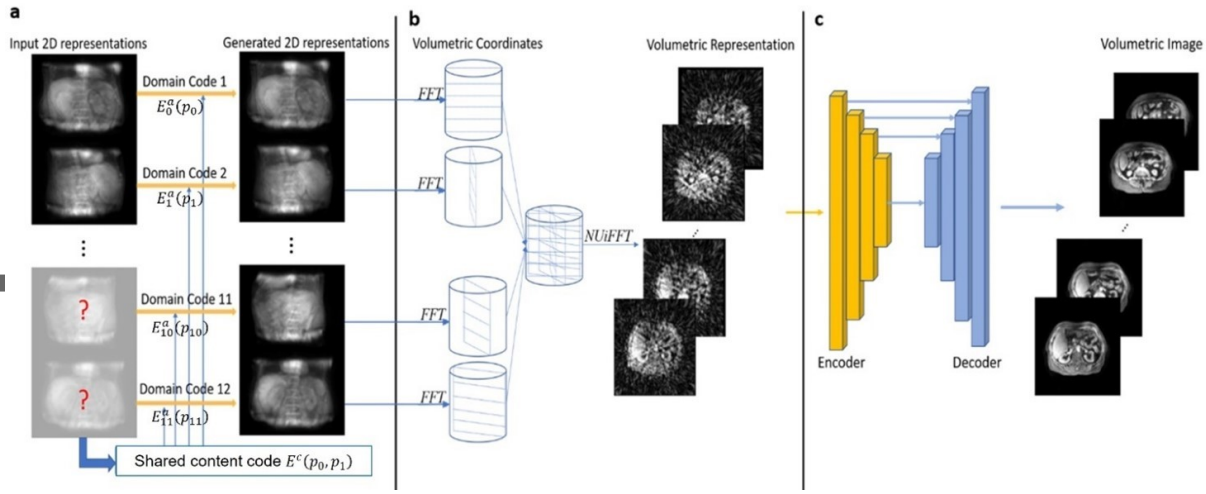
Reference

1. Li W, Purdie TG, Taremi M, Fung S, Brade A, Cho BJ, Hope A, Sun A, Jaffray DA, Bezjak A, Bissonnette JP. Effect of immobilization and performance status on intrafraction motion for stereotactic lung radiotherapy: analysis of 133 patients. International Journal of Radiation Oncology* Biology* Physics. 2011 Dec 1;81(5):1568-75.
2. Wysocka B, Kassam Z, Lockwood G, Brierley J, Dawson LA, Buckley CA, Jaffray D, Cummings B, Kim J, Wong R, Ringash J. Interfraction and respiratory organ motion during conformal radiotherapy in gastric cancer. International Journal of Radiation Oncology* Biology* Physics. 2010 May 1;77(1):53-9.
3. Wojcieszynski AP, Rosenberg SA, Brower JV, Hullett CR, Geurts MW, Labby ZE, Hill PM, Bayliss RA, Paliwal B, Bayouth JE, Harari PM. Gadoxetate for direct tumor therapy and tracking with real-time MRI-guided stereotactic body radiation therapy of the liver. Radiotherapy and Oncology. 2016 Feb 1;118(2):416-8.
4. Keiper TD, Tai A, Chen X, Paulson E, Lathuilière F, Bériault S, Hébert F, Cooper DT, Lachaine M, Li XA. Feasibility of real - time motion tracking using cine MRI during MR - guided radiation therapy for abdominal targets. Medical physics. 2020 Aug;47(8):3554-66.
5. de Muinck Keizer DM, Pathmanathan AU, Andreychenko A, Kerkmeijer LG, Van Zyp JV, Tree AC, Van Den Berg CA, De Boer JC. Fiducial marker based intra-fraction motion assessment on cine-MR for MR-linac treatment of prostate cancer. Physics in Medicine & Biology. 2019 Apr 4;64(7):07NT02.
6. Ginn, J.S., Ruan, D., Low, D.A. and Lamb, J.M., 2020. An image regression motion prediction technique for MRI - guided radiotherapy evaluated in single - plane cine imaging. Medical Physics, 47(2), pp.404-413.
7. Mickevicius, Nikolai J., and Eric S. Paulson. "Simultaneous acquisition of orthogonal plane cine imaging and isotropic 4D-MRI using super-resolution." Radiotherapy and Oncology 136 (2019): 121-129.

8. Feng, Li, Neelam Tyagi, and Ricardo Otazo. "MRSIGMA: Magnetic Resonance SIGnature MAtching for real-time volumetric imaging." Magnetic resonance in medicine 84.3 (2020): 1280-1292.
9. Bjerre, T., Crijns, S., af Rosenschöld, P. M., Aznar, M., Specht, L., Larsen, R., & Keall, P. (2013). Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes. Physics in Medicine & Biology, 58(14), 4943.
10. Stemkens, B., Tijssen, R.H., De Senneville, B.D., Lagendijk, J.J. and Van Den Berg, C.A., 2016. Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy. Physics in Medicine & Biology, 61(14), p.5335.
11. Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. IEEE signal processing magazine. 2008 Mar 21;25(2):72-82.
12. Zhao B, Haldar JP, Brinegar C, Liang ZP. Low rank matrix recovery for real-time cardiac MRI. In2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2010 Apr 14 (pp. 996-999). IEEE.
13. Otazo R, Candes E, Sodickson DK. Low - rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. Magnetic resonance in medicine. 2015 Mar;73(3):1125-36.
14. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 2018 Mar;555(7697):487-92.
15. Mardani M, Gong E, Cheng JY, Vasanawala SS, Zaharchuk G, Xing L, Pauly JM. Deep generative adversarial neural networks for compressive sensing MRI. IEEE transactions on medical imaging. 2018 Jul 23;38(1):167-79.
16. Wu Y, Ma Y, Capaldi DP, Liu J, Zhao W, Du J, Xing L. Incorporating prior knowledge via volumetric deep residual network to optimize the reconstruction of sparsely sampled MRI. Magnetic resonance imaging. 2020 Feb 1;66:93-103.

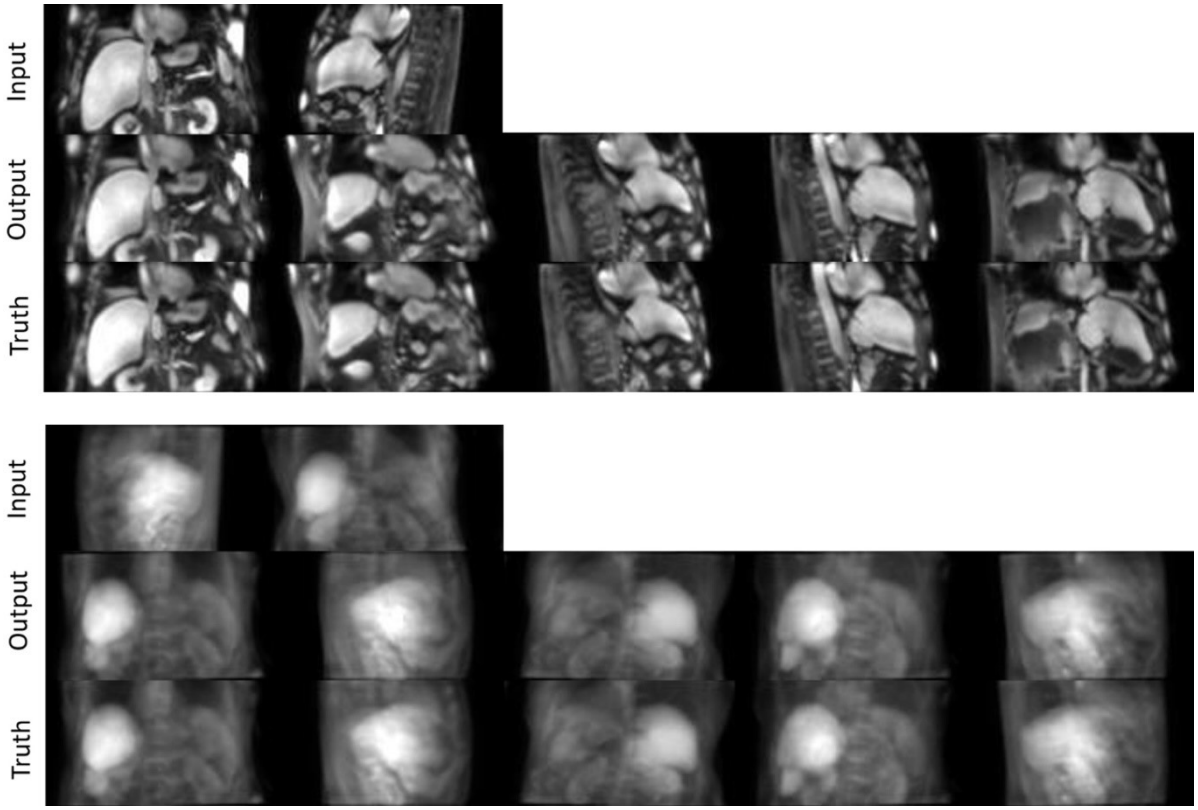
17. Hyun CM, Kim HP, Lee SM, Lee S, Seo JK. Deep learning for undersampled MRI reconstruction. Physics in Medicine & Biology. 2018 Jun 25;63(13):135007.
18. Aggarwal HK, Mani MP, Jacob M. Model based image reconstruction using deep learned priors (MODL). In2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) 2018 Apr 4 (pp. 671-674). IEEE.
19. Shen L, Zhao W, Capaldi D, Pauly J, Xing L. A Geometry-Informed Deep Learning Framework for Ultra-Sparse 3D Tomographic Image Reconstruction. arXiv preprint arXiv:2105.11692. 2021 May 25.
20. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD - GRASP: golden - angle radial MRI with reconstruction of extra motion - state dimensions using compressed sensing. Magnetic resonance in medicine. 2016 Feb;75(2):775-88.
21. Chandarana H, Block KT, Winfeld MJ, Lala SV, Mazori D, Giuffrida E, Babb JS, Milla SS. Free-breathing contrast-enhanced T1-weighted gradient-echo imaging with radial k-space sampling for paediatric abdominopelvic MRI. European radiology. 2014 Feb;24(2):320-6.
22. Seregni M, Paganelli C, Lee D, Greer PB, Baroni G, Keall PJ, Riboldi M. Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI. Physics in Medicine & Biology. 2016 Jan 7;61(2):872.
23. Johansson, A., Balter, J. and Cao, Y., 2018. Rigid - body motion correction of the liver in image reconstruction for golden - angle stack - of - stars DCE MRI. Magnetic resonance in medicine, 79(3), pp.1345-1353.
24. Shen L, Zhu W, Wang X, Xing L, Pauly JM, Turkbey B, Harmon SA, Sanford TH, Mehralivand S, Choyke PL, Wood BJ. Multi-domain image completion for random missing input data. IEEE Transactions on Medical Imaging. 2020 Dec 22;40(4):1113-22.
25. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. InInternational Conference on Medical image computing and computer-assisted intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.

26. Johansson, Adam, James M. Balter, and Yue Cao. "Abdominal DCE - MRI reconstruction with deformable motion correction for liver perfusion quantification." Medical physics 45.10 (2018): 4529-4540.
27. Liu L, Johansson A, Cao Y, Lawrence TS, Balter JM. Volumetric prediction of breathing and slow drifting motion in the abdomen using radial MRI and multi-temporal resolution modeling. Physics in Medicine & Biology. 2021 Aug 19.
28. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S. Fast free-form deformation using graphics processing units. Comput Meth Prog Bio. 2010;98(3):278–84.
29. Shen, Liyue, Wei Zhao, and Lei Xing. "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning." Nature biomedical engineering 3.11 (2019): 880-888.
30. Mostafaei F, Tai AN, Omari E, Song Y, Christian J, Paulson E, Hall W, Erickson B, Li XA. Variations of MRI-assessed peristaltic motions during radiation therapy. PloS one. 2018 Oct 25;13(10):e0205917.
31. Liu L, Johansson A, Cao Y, Kashani R, Lawrence TS, Balter JM. Modeling intra-fractional abdominal configuration changes using breathing motion-corrected radial MRI. Physics in Medicine & Biology. 2021 Apr 12;66(8):085002.
32. Johansson A, Balter JM, Cao Y. Gastrointestinal 4D MRI with respiratory motion correction. Med Phys. 2021 May;48(5):2521-2527.



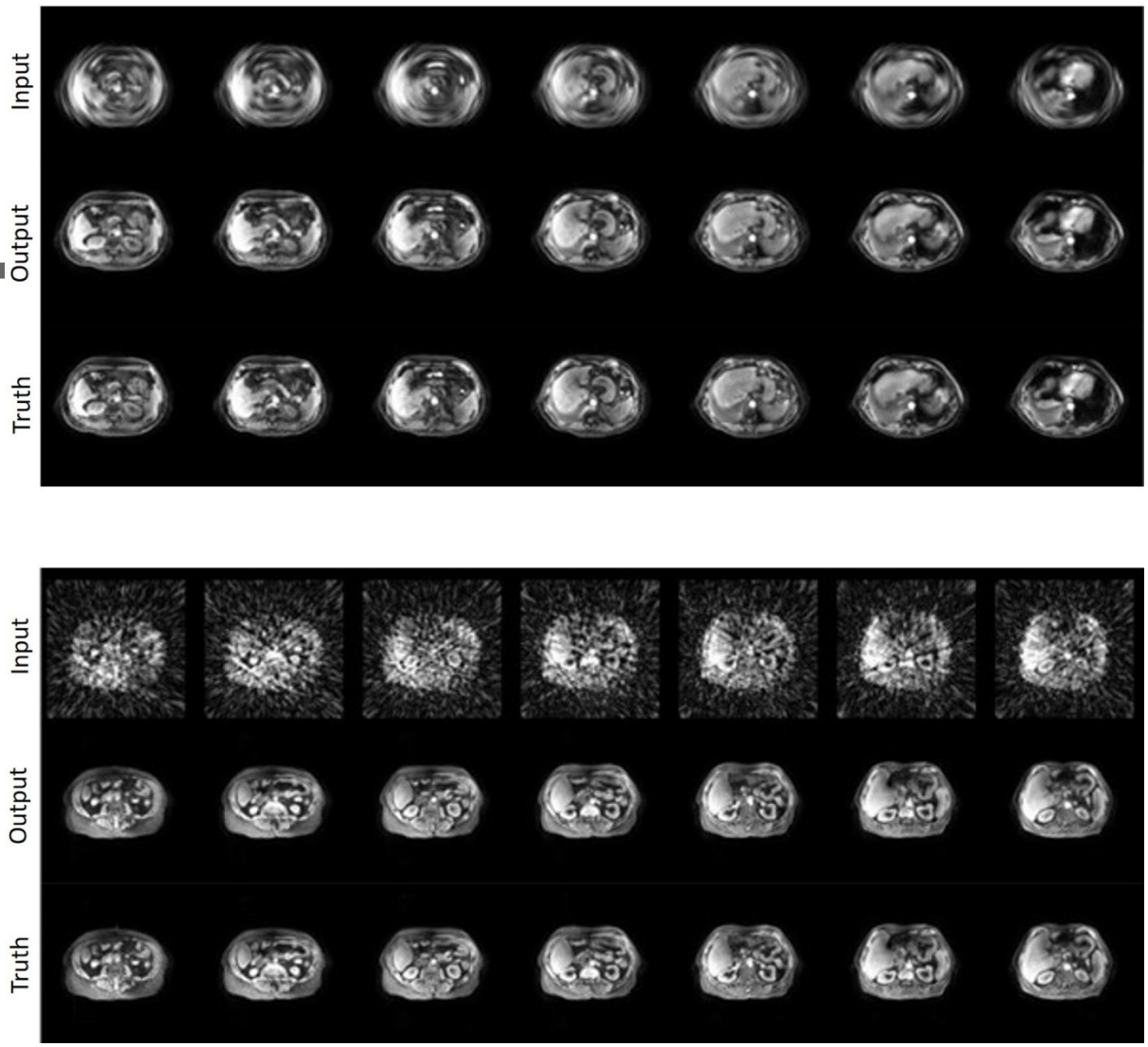
List of Figures

- Figure 1. Three modules of the proposed deep learning framework (with radial MRI as an example). a. 2D generation network that synthesizes new 2D representations associated with unacquired k-space samples. b. Geometry module that utilizes both sampling pattern and known k-space-to-image domain transforms to generate volumetric representations. c. 3D refinement network that refines image quality for final volumetric image reconstruction.

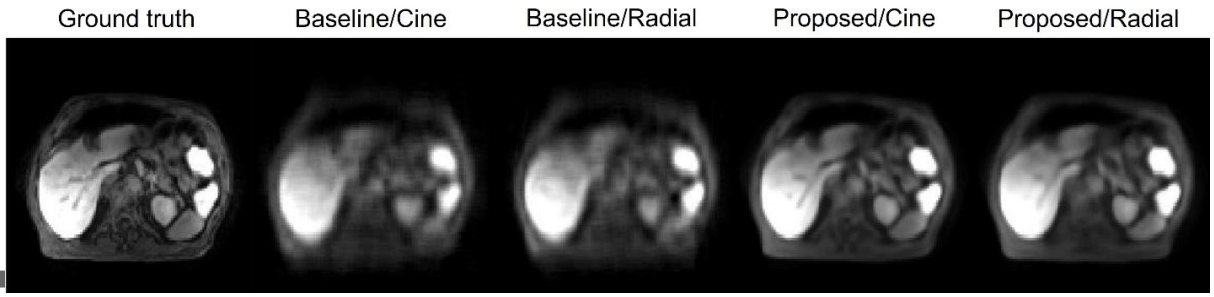


- Figure 2a. Example network-predicted rotating planar images from 2 orthogonal cine slices.

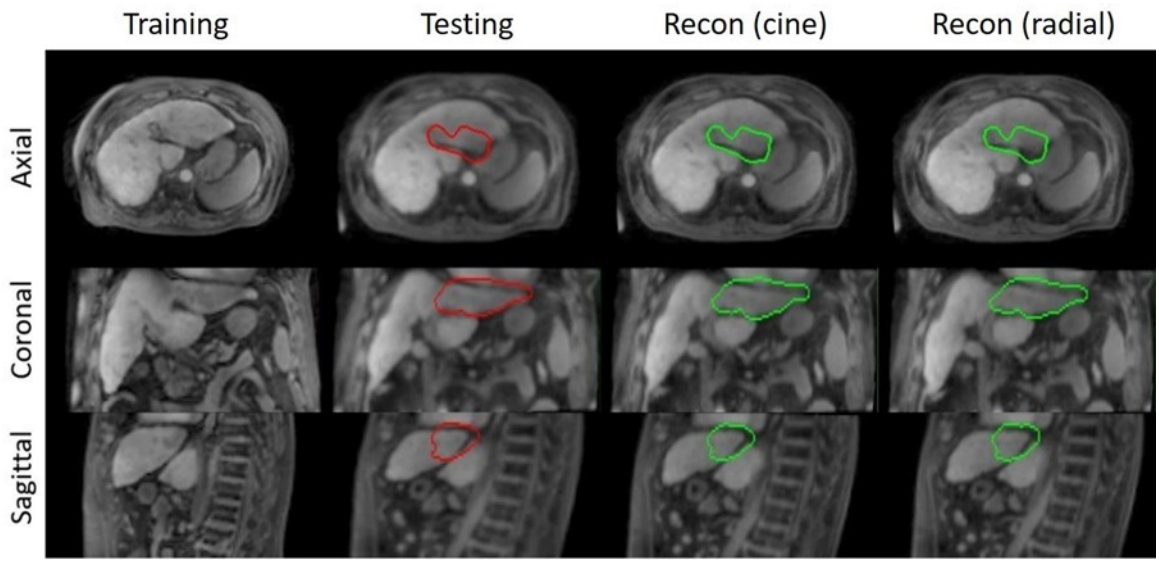
  Figure 2b. Example network-predicted radial projections from 2 radial projections.



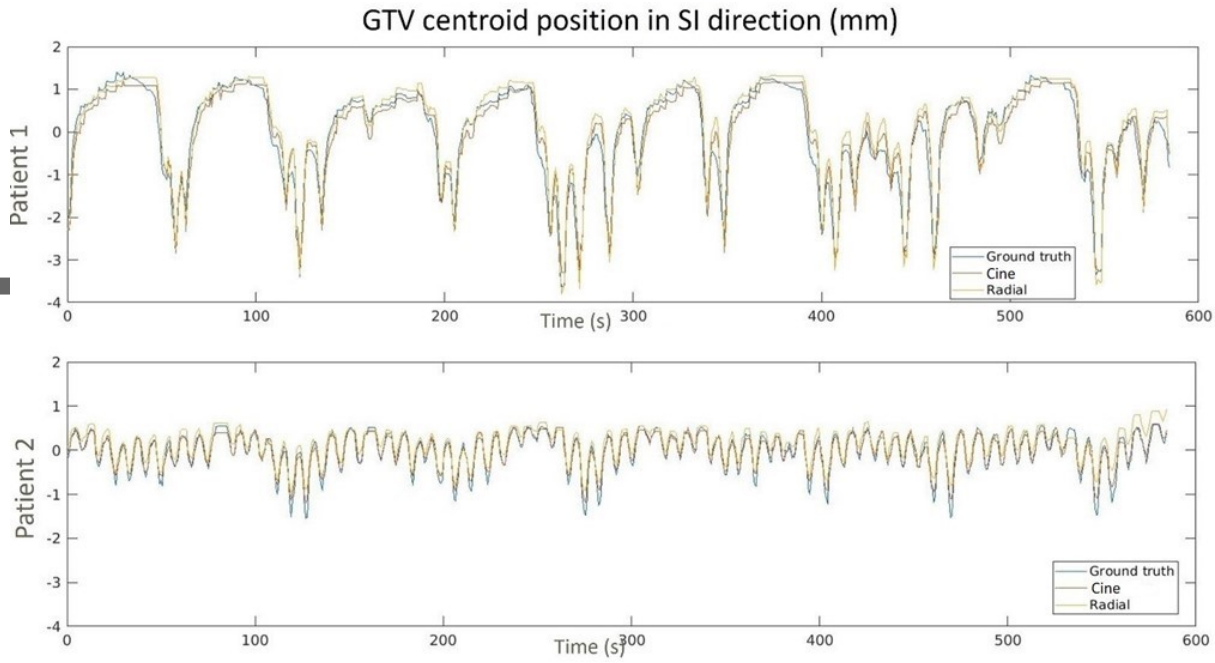
- Figure 3a. Sample slices of volumetric input generated from the geometry module and output generated from the 3D refinement network for cine acquisition. Figure 3b. Sample slices of volumetric input generated from the geometry module and output generated from the 3D refinement network for radial acquisition.



- Figure 4. Sample slices of volumetric images reconstructed using the baseline model and the proposed model respectively, with cine slices and radial projections as inputs.



- Figure 5. Sample views of training, testing and model-reconstructed MRI images. Red and green contours show GTV volume defined by deforming a reference image to testing and model-reconstructed images respectively.



- Figure 6. Example motion traces of GTV centroid in superior-inferior direction.

Table 1. Quantitative evaluation of reconstructed image quality across 7 patients

Acquisition/Reconstruction Scheme	SSIM	PSNR	MSE
Cine/Proposed	0.85±0.05	25.1±2.1	0.004±0.001
Cine/Baseline	0.75±0.04	23.5±1.8	0.005±0.002
Radial/Proposed	0.85±0.05	25.1±2.4	0.004±0.002
Radial/Baseline	0.75±0.06	23.5±2.0	0.005±0.003

Table 2. Median distances between predicted and ground truth GTV centroid positions across patients

Acquisition scheme	AP	LR	SI
Cine	1.1±0.5 mm	1.0±0.5 mm	0.4±0.3mm
Radial	1.2±0.6 mm	0.5±0.5 mm	0.5±0.4 mm