Author Manuscript

Volumetric MRI with Sparse Sampling for MR-Guided 3D Motion Tracking via Sparse Prior-Augmented Implicit Neural Representation Learning

Lianli Liu[1], Liyue Shen[2], Adam Johansson[3,4,5], James M Balter[3], Yue Cao[3], Lucas Vitzthum[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: lliliu@stanford.edu, lei@stanford.edu

Running title: Sparse Volumetric MRI via Implicit Neural Representation Learning

Abstract.

Background: Volumetric reconstruction of MRI from sparse samples is desirable for 3D motion tracking and promises to improve MR-guided radiation treatment precision. Data-driven sparse MRI reconstruction, however, requires large-scale training datasets for prior learning, which is time consuming and challenging to acquire in clinical settings.

Purpose: To investigate volumetric reconstruction of MRI from sparse samples of 2 orthogonal slices aided by sparse priors of 2 static 3D MRI through implicit neural representation (NeRP) learning, in support of 3D motion tracking during MR-guided radiotherapy.

Methods: A multi-layer perceptron network was trained to parameterize the NeRP model of a patient-specific MRI dataset, where the network takes 4D data coordinates of voxel locations and motion states as inputs and outputs corresponding voxel intensities. By first training the network to learn the NeRP of 2 static 3D MRI with different breathing motion states, prior information of patient breathing motion was embedded into network weights through optimization. The prior information was then augmented from 2 motion states to 31 motion states by querying the optimized network at interpolated and extrapolated motion state coordinates. Starting from the prior-augmented NeRP model as an initialization point, we further trained the network to fit sparse samples of 2 orthogonal MRI slices and the final volumetric reconstruction was obtained by querying the trained network at 3D spatial locations. We evaluated the proposed method using 5-minute volumetric MRI time series with 340 ms temporal resolution for 7 abdominal patients with

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.16845.

This article is protected by copyright. All rights reserved.

Author Manuscript

hepatocellular carcinoma, acquired using golden-angle radial MRI sequence and reconstructed through retrospective sorting. Two volumetric MRI with inhale and exhale states respectively were selected from the first 30 sec of the time series for prior embedding and augmentation. The remaining 4.5-min time series was used for volumetric reconstruction evaluation, where we retrospectively subsampled each MRI to 2 orthogonal slices and compared model-reconstructed images to ground truth images in terms of image quality and the capability of supporting 3D target motion tracking.

Results: Across the 7 patients evaluated, the peak signal-to-noise-ratio between model-reconstructed and ground truth MR images was 38.02 ± 2.60 dB and the structure similarity index measure was 0.98 ± 0.01. Throughout the 4.5-min time period, gross tumor volume (GTV) motion estimated by deforming a reference state MRI to model-reconstructed and ground truth MRI showed good consistency. The 95-percentile Hausdorff distance between GTV contours was 2.41 ± 0.77 mm, which is less than the voxel dimension. The mean GTV centroid position difference between ground truth and model estimation was less than 1 mm in all 3 orthogonal directions.

Conclusion: A prior-augmented NeRP model has been developed to reconstruct volumetric MRI from sparse samples of orthogonal cine slices. Only one exhale and one inhale 3D MRI were needed to train the model to learn prior information of patient breathing motion for sparse image reconstruction. The proposed model has the potential of supporting 3D motion tracking during MR-guided radiotherapy for improved treatment precision and promises a major simplification of the workflow by eliminating the need of large-scale training datasets.

1. Introduction

Precise radiation treatment to abdominal targets has been challenged by complex motions (1, 2). If not carefully accounted for during treatment planning and delivery, displacements and deformations of abdominal organs may result in underdosage to the target and/or over dosage to nearby normal tissues (3-5). MR-guided radiotherapy permits simultaneous radiation delivery and motion monitoring, and promises to improve treatment precision by gating the dose delivery based on motion information from MR imaging (6). The long imaging time of MRI, however, hinders the use of volumetric MRI for 3D motion tracking. Currently, MR-guided treatment gating is based on 2D motion information extracted from one or a few slices (7-9), which limits our ability in capturing out-of-plane motions and compromises the quality of patient care.

To support 3D motion tracking during MR-guided radiotherapy, various work has been done to reconstruct volumetric MRI from sparse samples with sub-second acquisition time. However, to regularize the ill-posed problem MRI reconstruction from subsampled data, most work requires population priors learnt from MR images of different patients (10-13), or patient-specific priors learnt from 4D MRI acquired during each imaging session of motion tracking (14-17). The image reconstruction algorithm recently proposed by Liu et al.,(18) has demonstrated the robustness to longitudinal patient anatomy changes through geometry-informed deep learning and showed the potential of real time volumetric MRI over multiple imaging sessions with a single 4D MRI training dataset acquired long before, yet 4D MRI is still not widely available on commercial MRI scanners and requires additional data processing for sufficient image quality (19-21). As a result, curation of

This article is protected by copyright. All rights reserved.

Author Manuscript

training data remains a bottleneck for clinical implementation of data-driven MRI-based 3D motion tracking.

In this study, we investigated volumetric MR imaging with both sparse samples and sparse priors including only an exhale state 3D MRI and an inhale state 3D MRI, which can be easily obtained using clinical MRI sequences. To achieve this goal, we adopted a novel machine learning technique termed implicit neural representation (NeRP) learning (22-24) to model and augment the sparse priors. The prior-augmented NeRP model was then used to regularize the subsequent task of image reconstruction from sparse samples. Comparing to conventional machine learning methods, NeRP offers two unique advantages. First, instead of learning a mapping between data values (e.g., observed versus unobserved image voxel values), NeRP learns a mapping from data coordinates to data values through neural network optimization. Since the inputs to the network are data coordinates, the optimized network can effectively encode the underlying data structure without large-scale training datasets and the final reconstructed image can be obtained by a simple query of the optimized network at desired voxel locations. Secondly, NeRP learns a continuous data representation instead of grid-based discrete data representation, which permits the modeling of data value distributions outside input data coordinates. Therefore, NeRP offers an efficient tool for data augmentation by querying the network at interpolated/extrapolated coordinates. NeRP has shown advantages over conventional techniques in modeling various types of data, such as images (25, 26), videos (27), 3D shapes (28), treatment plans (29), linear accelerator beam (30) and arrangements of objects (31). Here we advanced real time volumetric MRI by leveraging the NeRP strategy to effectively model prior information of patient dynamics from only 2 static 3D MRI. The NeRP-modeled prior information was then used to facilitate the subsequent volumetric MR imaging task with sub-second sampling time. The proposed method was evaluated using free-breathing volumetric MRI time series for 7 hepatocellular carcinoma patients. The potential of the NeRP model in supporting 3D motion tracking was demonstrated by tumor motion analysis using model-reconstructed images. Through NeRP modeling, an efficient prior embedding strategy was developed that significantly reduced the amount of training data required without compromising the reconstructed image quality. NeRP therefore provides a clinically viable solution for data-driven volumetric MRI with ultra-sparse sampling and has the potential of supporting 3D motion tracking during MR-guided radiotherapy for improved patient care.

2. Methods and Materials

2.A Problem Formulation

For a set of volumetric MR images with different motion states, we can implicitly define a continuous mapping function over the image dataset as $\Phi: \boldsymbol{v} \rightarrow d$, where $\boldsymbol{v} = [x, y, z, k] \in R^4$ is the 4D coordinate of the dataset with x, y, z being the spatial location of an image voxel and k being the motion state of the image. Given the function input $\boldsymbol{v}$, the function output d is the corresponding image voxel intensity at the specific location and motion state. Implicit neural representation (NeRP) learning aims at training a neural network M with network weights θ to parameterize the function Φ

$$M_\theta: \boldsymbol{v} \rightarrow d \tag{1}$$

Through this NeRP formulation, the problem of image modeling is transformed into the searching of the optimal network with weights θ that accurately map image coordinates to image intensities.

This article is protected by copyright. All rights reserved.

Author Manuscript

After the network optimization, we can use the trained NeRP model to reconstruct an image by querying the network at the corresponding coordinates and assigning voxel intensities based on the network outputs.

Training a NeRP model for volumetric MRI reconstruction from sparse samples without 4D MRI priors, however, is faced with two challenges: the ill-posedness in fitting a network to subsampled data and the lack of prior information to regularize the ill-posed problem. To address these challenges, we proposed a prior-augmented NeRP learning strategy. Firstly, we trained a NeRP model with 2 motion states only, by fitting the network to 2 fully sampled volumetric MRI at inhale and exhale states. Leveraging the capability of continuous data representation of NeRP, we augmented the 2 MRI to a series of MRI by querying the trained network at interpolated/extrapolated motion state coordinates to obtain MRI with various motion states. The augmented priors were used to facilitate the NeRP learning from sparse MRI samples, for example, orthogonal cine MRI, where we started from the NeRP model of the augmented MRI series and further trained the network to fit the sparse samples. Figure 1 illustrates the framework of prior-augmented NeRP learning for MRI reconstruction with both sparse samples and sparse priors.

2.B Prior-Augmented NeRP Learning for MRI reconstruction

2.B.1 Network Structure

Following previous work on NeRP modeling of images and isodose plans(26, 29), we implemented a multi-layer perceptron (MLP) network to parameterize the NeRP model. The network consists of 8 fully connected layers. Each layer has 256 neural nodes followed by a periodic activation function (24). To effectively model high frequency components of the MRI dataset, a Fourier feature mapping (32) F was used to encode the coordinates $\boldsymbol{v}$ as

$$F(\boldsymbol{v}) = [\cos(2\pi \boldsymbol{B}\boldsymbol{v}), \sin(2\pi \boldsymbol{B}\boldsymbol{v})] \quad (2)$$

before the coordinates were input to the network M_θ, where entries of the matrix **B** were sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. In this study, we experimentally set the size of the Fourier feature encoding (number of rows for **B**) as 256 and the scale σ as 4.

2.B.2 Prior Embedding and Augmentation

We first trained the NeRP model using an image dataset of two fully sampled volumetric MRI to learn prior information needed for the subsequent sparse imaging task. The dataset consists of an inhale state MRI and an exhale state MRI, both with image size $X \times Y \times Z$. The MLP network was optimized to fit the image dataset, where we searched for the optimal network weights Θ to minimize the L2 loss between network outputs and MRI voxel values

$$\Theta^* = \arg\min \sum_{i=1}^{N} \left\| M_\Theta\big(F(\boldsymbol{v}_i)\big) - d_i^{prior} \right\|_2^2 \quad (2)$$

where $\boldsymbol{v}_i = [x_i, y_i, z_i, k_i]$, with $x_i \in \{0,1,\cdots,X-1\}, y_i \in \{0,1,\cdots,Y-1\}, z_i \in \{0,1,\cdots,Z-1\}$ being the voxel locations and $k_i \in \{0,1\}$ indicating the two motion states (0 for inhale and 1 for exhale). d_i^{prior} is the corresponding MRI voxel value in the 2 prior images.

This article is protected by copyright. All rights reserved.

Author Manuscript

By learning the NeRP model of the two fully sampled MRI, we embedded prior knowledge of image properties into the optimized network weights, which may be used to facilitate image reconstruction from sparse samples, if the target image to be reconstructed shares similar properties with the two prior images. However, as the prior images contain only two motion states, the target image may be very different from the prior images due to varying motion states such as mid-breath or deep inhale. Utilizing the continuous and nonlinear data presentation offered by NeRP modeling, we augmented the prior images to include model-generated motion states, by interpolating and extrapolating the motion state coordinates and querying the NeRP model at those coordinate values. In this way, volumetric MRI representing different respiratory-induced anatomical deformations can be obtained. With the model-augmented priors, for any image to be reconstructed, we may find a prior image with motion state similar enough so that we can use the prior-embedded network to facilitate image reconstruction from sparse samples. Specifically, in this study, we augmented the prior images from 2 volumetric MRI with motion coordinates $k_i \in \{0,1\}$ to 31 volumetric MRI with motion coordinates $\hat{k}_i \in \{-1, -0.9, \cdots, 1.9, 2\}$. Details of the extrapolation range selection and example model-generated motion states are presented in the result section.

2.B.3 NeRP Learning from Sparse Samples for MRI Reconstruction

With the prior-embedded network M_{Θ^*} and the augmented prior images, for a set of sparse samples d_i^{sub} acquired at subsampled 3D voxel locations that correspond to the sparse acquisition scheme $\left[x_i^{sub}, y_i^{sub}, z_i^{sub}\right],\ i = 1, \cdots, M,\ M \ll N$, we first determined the motion state coordinate of d_i^{sub} as

$$k^{sub} = \arg\min \sum_{i=1}^{M} \left\| d_i^{sub} - M_{\Theta^*}\left(F(\boldsymbol{v}_i)\right) \right\|_2^2 \tag{3}$$

, where $\boldsymbol{v}_i = \left[x_i^{sub}, y_i^{sub}, z_i^{sub}, \hat{k}_i\right]$ with $\hat{k}_i \in \{-1, -0.9, \cdots, 1.9, 2\}$. Basically, we searched for the prior image that best matches the sparse samples at sampling locations and used the motion state coordinate of that prior image as the motion state coordinate of the sparse samples.

After determining the motion state coordinate, we further trained the prior-embedded network M_{Θ^*} to learn the NeRP model of the target image to be reconstructed, by optimizing the network weights to fit the sparse samples

$$\theta^* = \arg\min \sum_{i=1}^{M} \left\| M_\theta\left(F\left(\boldsymbol{v}_i^{sub}\right)\right) - d_i^{sub} \right\|_2^2 \tag{4}$$

, where $\boldsymbol{v}_i = \left[x_i^{sub}, y_i^{sub}, z_i^{sub}, k_{sub}\right]$. With the prior-embedded network M_{Θ^*} as an initialization point for network training, the image reconstruction process is implicitly regularized by the network parameterization, including both the image property priors embedded in the network weights Θ^* and the low-level statistics(33) priors captured by the network structure itself (34). After the training is done, we can reconstruct the volumetric image by querying the network as $M_{\theta^*}(\boldsymbol{v}_i)$ with $\boldsymbol{v}_i = [x_i, y_i, z_i, k_{sub}], i = 1, \cdots, N$.

2.C Model Training and Evaluation

2.C.1. Datasets

This article is protected by copyright. All rights reserved.

Author Manuscript

To evaluate the proposed NeRP model for MRI reconstruction, MRI data from 7 patients with hepatocellular carcinoma were collected under an institution review board approved protocol (The University of Michigan Institutional Review Board, HUM00068061). Each MRI examination consisted of a 5-min free-breathing scan acquired using a golden-angle stack-of-stars spoiled gradient echo sequence with fast suppression. The imaging parameters ranged from 1.14 to 1.21 ms for echo time, 2.71–4.51 ms for repetition time, 10°-14°for flip angle, 2-2.45 mm for in-plane voxel size and 3-4 mm for slice thickness. A stack of radial k-space samples (termed 'spoke' for convenience) was collected every 170 ms. All MR images were acquired using an in-house 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 size of the imaging matrix was 192 × 192, the number of slices was 64 and the bandwidth was 900 Hz per pixel.

Following previous studies on high temporal and spatial resolution breathing motion characterization for abdominal patients (33, 35, 36), a 5-min volumetric time series with a temporal resolution of 340 ms was reconstructed for each patient. Briefly, High temporal resolution but low spatial resolution image series was first reconstructed using a Gaussian temporal view sharing filter that centered around each k-space radial spoke with $\sigma = 1$ at the center and $\sigma = 3$ at the periphery. The high temporal resolution image series was used to assess and label the breathing motion state of each spoke. Using the breathing motion state labels, retrospective k-space sorting and re-binning was performed for high spatial resolution breathing motion state reconstruction. Combing the motion state label of each spoke and the high spatial resolution breathing motion states, 3D breathing motion information can be obtained for each spoke. The accuracy of the above breathing motion characterization scheme has been validated in previous studies that use the characterized breathing motion information to correct artifacts of DCE-MRI (35) and analyze non-breathing induced involuntary anatomical changes (33, 36).

In this study, we constructed the volumetric image time series using motion states associated with every other spoke, which resulted in a temporal resolution of 340 ms. In addition to the volumetric MRI time series, a 4D MRI series with 21 phases was also obtained for each patient through retrospective sorting based on motion state labels. The exhale phase of the 4D MRI was selected as the reference volume with gross tumor volume (GTV) defined. B-spline deformable registration was performed between the 5-min time series and the reference MRI to transfer the GTV contour to the time series volumes. The volumetric MR images and GTV motion information obtained through the above reconstruction and registration process served as the ground truth for model training and evaluation. Details of the reconstruction/registration methods and accuracy validation of the methods can be found in previous publications (33, 35, 36).

2.C.2. NeRP Model-Based Motion State Generation

To evaluate the NeRP model in generating motion states from interpolated and extrapolated coordinates for prior augmentation, we trained NeRP models by selecting two phases from the 21-phase 4D MRI for each patient and evaluated the trained model in interpolating/extrapolating to other motion phases. The motion magnitudes between the two training phases were varied from 90% (the 2nd and the 20th phases) to 14% (10th and 12th phases) of the full motion magnitude (the

This article is protected by copyright. All rights reserved.

Author Manuscript

1st and the 21st phases). During model training, data normalization was performed to scale image intensities and coordinates to be within [0,1] and an Adam optimizer with a learning rate of 10^{-4} was used to optimize the MLP network for 1000 iterations. After the model was trained, we queried the model using a series of motion state coordinate values and compared model-generated motion phases to the ground truth in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM).

2.C.3 NeRP Model-Based Volumetric Image Reconstruction.

For each patient, one inhale state and one exhale state MRI from the first 30 sec of the 5-min volumetric image time series were used for prior embedding and augmentation. The breathing motion state was determined using the GTV centroid position in the superior-inferior direction, where images with the lowest/highest GTV centroid position during the first 30 sec were chosen as the inhale state/exhale state MRI respectively. MR images from the remaining 4.5 min were used for model evaluation. As orthogonal cine MRI has been previously investigated and evaluated to have an acquisition time less than 1 sec for large field-of-view imaging such as abdominal MRI (37), in this study, we retrospectively subsampled each 3D image of the 4.5-min testing time series to two orthogonal slices (one coronal slice and one sagittal slice) for NeRP model-based image reconstruction. The slice locations were determined using the GTV centroid position from the reference MRI. With a 340 ms temporal resolution, a total of 795 pairs of orthogonal slices were obtained. Data normalization was performed to scale image intensities and coordinates to be within [0,1]. During the prior embedding stage, the MLP network was optimized using an Adam optimizer with a learning rate of 10^{-4} for 1000 iterations. The prior-embedded network was further optimized to fit sparse samples for image reconstruction, where we used the same Adam optimizer with a reduced learning rate of 10^{-6} for 30 iterations. We evaluated the image reconstruction quality by comparing NeRP model-reconstructed images to ground truth images in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). A baseline method that directly uses deformation vector fields (DVF) between the 2 prior MRI to estimate 3D motion states from sparse samples was also implemented. Following existing work on 3D DVF estimation from orthogonal cine MRI (38), a pattern search algorithm was used to determine the optimal coefficient that scales the DVF between the 2 prior MRI, so that the mean square errors between the deformed exhale state prior and the orthogonal cine MRI were minimized at sampled voxel locations. The exhale state prior deformed by the optimally scaled DVF was compared to the ground truth volumetric MRI in terms of PSNR and SSIM. The capability of NeRP model-based reconstruction in supporting 3D motion tracking was evaluated by deforming the reference state MRI with GTV contours to model-reconstructed images. GTV contours estimated using model-reconstructed images were compared to those estimated using ground truth images, where we calculated GTV centroid position differences in anterior-posterior, left-right and superior-inferior directions as well as 95-percentile Hausdorff distances between GTV contours.

3. Results

3.A NeRP Model-Based Motion State Generation for Prior Augmentation

The highest PSNR/SSIM between NeRP model-generated 4D MRI phases and the ground truth was 34.10 dB/0.95. The highest PSNR/SSIM was observed when the testing phase was in-between the 2

This article is protected by copyright. All rights reserved.

Author Manuscript

training phases. The lowest PSNR/SSIM between model synthesis and the ground truth was observed when the testing phase was outside the 2 training phases. The lowest PSNR/SSIM decreased from 31.97 dB/0.93 to 30.27 dB/0.90, as we decreased the motion magnitude between the two training phases from 90% to 14% of the full motion magnitude. Figure 2 shows example NeRP model-generated and ground truth MRI phases. These results demonstrated the capability of NeRP model in generating motion states that represent respiratory-induced anatomical deformations, while the accuracy depends on the motion magnitudes of both training and testing phases. Across different patients/NeRP models, the averaged motion state coordinate values that produce the best-matching model-generated phases with the ground truth ranged from -1 to 1.6. Extrapolating the motion state coordinates to extended ranges introduced image artifacts and did not improve the PSNR/SSIM between model-generated and ground truth MRI phases, as shown in Figure 2. Therefore, in the subsequent prior augmentation step for sparse volumetric MRI reconstruction, a coordinate range of [-1,2] was used when querying the prior-embedded NeRP model.

3.B Sparse Volumetric MRI using Prior-Augmented NeRP Model

Across the 7 patients evaluated, the PSNR/SSIM between model-based reconstruction from orthogonal cine MRI and the ground truth volumetric MRI were 38.02±2.60 dB/0.98±0.01. Increasing the number finetuning iterations from 30 to 100 did not change image quality metrics (PSNR/SSIM were 38.04±2.63 dB/0.98±0.01 after 100 iterations). Figure 3 compares sample axial, coronal and sagittal views of model-reconstructed MRI to ground truth MRI. Both visual comparison and quantitative evaluation suggested good consistency between volumetric reconstruction through prior-augmented NeRP learning and the ground truth. Figure 4 shows ground truth and model-reconstructed MRI with motion states different from the two prior images. The proposed NeRP model is able to reconstruct MRI with motion states both in between and outside the two priors, as indicated by the difference in liver dome positions.

Figure 5 compares images reconstructed by finetuning the prior-augmented NeRP model with sparse cine acquisition using Eq (4), and images obtained by directly querying the prior-augmented NeRP model with the motion state coordinate that best matches the sparse samples using Eq (3). The finetuning step visibly improves the consistency between model reconstruction and the ground truth, as demonstrated by the difference maps. Without the finetuning step, the PSNR/SSIM decreased from 38.02±2.60 dB/0.98±0.01 to 32.88±2.81 dB/0.94±0.02. Figure 5 also compares the baseline reconstruction through DVF scaling to NeRP model-based reconstruction. The images reconstructed by scaling the DVF between the 2 prior MRI showed visible discrepancy from the ground truth MRI at multiple organ locations, as highlighted by the difference maps, while the discrepancy was visibly reduced through NeRP model-based reconstruction. The baseline reconstructions also showed a lower PSNR/SSIM of 21.23±3.77 dB/0.76±0.16, as comparing to NeRP model-based reconstructions. These results suggested that a single DVF between two MRIs may not be sufficient in modeling the complex breathing motion. DVF series extracted from 4D MRI and principal component analysis will be required to build an accurate breathing motion model, as demonstrated in previous studies (38). NeRP learning on the other hand, has the potential of modeling breathing motion accurately from limited data samples.

3.C 3D Motion Tracking using NeRP Model-Based Reconstruction

This article is protected by copyright. All rights reserved.

Author Manuscript

Figure 6 shows example GTV contours generated by deforming the reference state MRI to model-reconstructed and ground truth MRI respectively. Both contours were also overlaid with the reference state MRI with reference GTV contour defined. The GTV contours determined using the model-based reconstruction and the ground truth agreed well with each other, while the reference GTV contour showed visible discrepancies from the two due to breathing motion. The 95-percentile Hausdorff distance between GTV contours estimated using model-reconstructed images and ground truth images was 2.41 ± 0.77 mm across the 7 patients, which is of sub-voxel order. Figure 7 plots example motion traces of GTV centroids in the superior-inferior direction over 4.5-min. Despite of varying motion patterns and breathing irregularities, the motion trace estimated using the model-based reconstruction showed good consistency with the ground truth. The GTV centroid position differences across patients were 0.26 ± 0.35 mm, 0.24 ± 0.23 mm and 0.22 ± 0.08 mm in anterior-posterior, left-right and superior-inferior directions respectively. Paired t-test showed no statistically significant difference ($p<0.05$) between the DVFs estimated using the ground truth and model-reconstructed MRI, with p-values equal to 0.56, 0.42 and 0.16 in AP, LR and SI directions respectively. Linear correlation between model-estimated target centroid position and the ground truth was also evaluated with R^2 values greater than 0.99 in all 3 directions, which are higher than previously reported (14).

4. Discussion

In this study, we investigated volumetric MRI reconstruction from sparse samples of 2 orthogonal slices, aided by a patient-specific prior model learnt from 2 static 3D MRI only, which can be easily acquired using clinical MRI sequences. The reconstructed volumetric MRI showed sufficient accuracy in supporting 3D target motion tracking, as demonstrated for 7 free-breathing hepatocellular carcinoma patients. The GTV contour and centroid position estimated using the sparse reconstruction showed good consistency with the ground truth, with 95-percentile Hausdorff distances between GTV contours less that voxel dimensions and GTV centroid position differences less than 1 mm. The proposed method therefore may offer a clinically viable solution to real time 3D motion tracking during MR-guided radiotherapy.

Through NeRP learning, we eliminated the need of population dataset or 4D MRI dataset for model training before volumetric reconstruction from sparse samples. Instead, prior information of motion-induced patient anatomy changes was learnt from 2 static 3D MR images and augmented to include various motion states in between and outside the 2 MR images. This was achieved by training the NeRP model to learn a continuous mapping from voxel coordinates of spatial locations and motion states to voxel intensities. As each pair of voxel coordinate/voxel intensity serves as a training sample, even with 2 volumetric MRI there are a lot of samples available for model training. The continuous data representation also supports flexible image augmentation with different motion states. Therefore, we can find a prior image with motion state sufficiently close to the target image to be reconstructed and use the prior-embedded network to facilitate the image reconstruction process. Comparing to existing work, the removal of large-scale training dataset represents a major simplification of the workflow for real time 3D motion tracking and makes the method valuable for clinical implementation.

Several limitations will be addressed in future work. Firstly, due to GPU memory limits, the input data coordinates had to be divided into batches for network optimization. Currently, volumetric MRI

This article is protected by copyright. All rights reserved.

Author Manuscript

reconstruction with 30 finetuning iterations took about 8 seconds. Better GPU resources and parallel computing on multiple GPUs will allow optimization in a single batch to reduce the image reconstruction time. Given the spatial smoothness of respiratory-induced deformation fields, it is also possible to perform the optimization on a coarser grid to reduce the memory usage and accelerate the optimization process, which will be investigated in future studies. Alternatively, a NeRP model-based 3D motion prediction technique may be developed to compensate for motion tracking latency. Secondly, in this study, we used inhale and exhale state MRI from a dynamic image time series acquired on a 3T MR scanner for prior embedding and augmentation. A prospective study using imaging sequences from clinical MR-Linac systems, for example the bSSFP sequence, will be performed for a more thorough evaluation of the proposed method before clinical implementation. The difference in image intensities between 3D acquisition and cine acquisition will be accounted for by proper data normalization and objective function selection, as well as imaging protocol optimization, which will be investigated in future studies. Future work will also evaluate the proposed method using a larger patient dataset with more tumor motion variations. Thirdly, this study investigated orthogonal cine acquisition as an example sparse sampling scheme, while the proposed method is quiet general and applicable to other acquisition methods, such as radial MRI. Future work will apply the NeRP model to reconstruct volumetric images from sparse samples acquired with different sequences. For example, golden-angle radial acquisition using MR-RIDDLE sequence (39) and Cartesian acquisition using orthogonal cine bSSFP sequence (40) have been implemented on clinical MR-Linac systems with sub-second acquisition time for each radial spoke/orthogonal slice pairs. Evaluation of the proposed model using these sequences will further validate the clinical applicability of the proposed method. Future work will apply the NeRP model to reconstruct volumetric images from sparse samples acquired with different sequences. Finally, while the proposed model has shown the capability of reconstructing volumetric MRI with motion states in between and outside the prior images, dramatic changes of breathing motion patterns, e.g., very deep inhale may degrade the image reconstruction quality due to the limited model extrapolation accuracy. A confidence level for image reconstruction will be of interest for robust clinical implementation. Another alternative would be an outlier detection algorithm that detects significant variations of breathing motion and warns potential problem of image reconstruction. Furthermore, as non-breathing induced motions, such as slow anatomical rearrangements (36) and organ peristalsis (41, 42) have been shown to impact radiation delivery significantly, the extension of the proposed model in tracking those motions should also be investigated in the future.

5. Conclusion

A prior-augmented implicit neural representation (NeRP) learning method has been proposed for volumetric MR imaging with sparse samples of orthogonal slices and sparse priors of 2 static 3D MRI. We demonstrated that the NeRP model can reconstruct volumetric MRI with sufficient accuracy to support 3D motion tracking of abdominal targets. By eliminating the need for large scale training datasets such as population data or 4D MRI, the method promises to enable the clinical workflow of 3D motion tracking for precision radiation therapy.

This article is protected by copyright. All rights reserved.

Author Manuscript

Acknowledgements

This research was supported by research grants from ViewRay Inc., Stanford Institute for Human-Centered Artificial Intelligence (HAI) and the National Institutes of Health (NIH) 5R01CA256890, 5R01CA260426, 1R01CA223667, 1R01CA227713, NIH R01 EB016079. The radial scanning sequence was provided under a research agreement from Siemens Healthineers.

References

1. Li W, Purdie TG, Taremi M, Fung S, Brade A, Cho BJ, et al. 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;81(5):1568-75.
2. Wysocka B, Kassam Z, Lockwood G, Brierley J, Dawson LA, Buckley CA, et al. Interfraction and respiratory organ motion during conformal radiotherapy in gastric cancer. International Journal of Radiation Oncology* Biology* Physics. 2010;77(1):53-9.
3. Balter JM, Ten Haken RK, Lawrence TS, Lam KL, Robertson JM. Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing. International Journal of Radiation Oncology* Biology* Physics. 1996;36(1):167-74.
4. Aruga T, Itami J, Aruga M, Nakajima K, Shibata K, Nojo T, et al. Target volume definition for upper abdominal irradiation using CT scans obtained during inhale and exhale phases. International Journal of Radiation Oncology* Biology* Physics. 2000;48(2):465-9.
5. Jayachandran P, Minn AY, Van Dam J, Norton JA, Koong AC, Chang DT. Interfractional uncertainty in the treatment of pancreatic cancer with radiation. International Journal of Radiation Oncology* Biology* Physics. 2010;76(2):603-7.
6. Wojcieszynski AP, Rosenberg SA, Brower JV, Hullett CR, Geurts MW, Labby ZE, et al. Gadoxetate for direct tumor therapy and tracking with real-time MRI-guided stereotactic body radiation therapy of the liver. Radiotherapy and Oncology. 2016;118(2):416-8.
7. Keiper TD, Tai A, Chen X, Paulson E, Lathuilière F, Bériault S, et al. Feasibility of real - time motion tracking using cine MRI during MR - guided radiation therapy for abdominal targets. Medical physics. 2020;47(8):3554-66.
8. de Muinck Keizer D, Pathmanathan A, Andreychenko A, Kerkmeijer L, Van Zyp JVDV, Tree A, et al. Fiducial marker based intra-fraction motion assessment on cine-MR for MR-linac treatment of prostate cancer. Physics in Medicine & Biology. 2019;64(7):07NT2.
9. Ginn JS, Ruan D, Low DA, Lamb JM. An image regression motion prediction technique for MRI - guided radiotherapy evaluated in single - plane cine imaging. Medical Physics. 2020;47(2):404-13.
10. Huttinga NR, Bruijnen T, Van Den Berg CA, Sbrizzi A. Real-time non-rigid 3D respiratory motion estimation for MR-guided radiotherapy using MR-MOTUS. IEEE Transactions on Medical Imaging. 2021;41(2):332-46.
11. Huttinga NR, Bruijnen T, van den Berg CA, Sbrizzi A. Nonrigid 3D motion estimation at high temporal resolution from prospectively undersampled k - space data using low - rank MR - MOTUS. Magnetic resonance in medicine. 2021;85(4):2309-26.
12. Romaguera LV, Mezheritsky T, Mansour R, Carrier J-F, Kadoury S. Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy. Medical Image Analysis. 2021;74:102250.

This article is protected by copyright. All rights reserved.

Author Manuscript

13. Shao H-C, Li T, Dohopolski MJ, Wang J, Cai J, Tan J, et al. Real-time MRI motion estimation through an unsupervised k-space-driven deformable registration network (KS-RegNet). Physics in Medicine & Biology. 2022;67(13):135012.
14. Feng L, Tyagi N, Otazo R. MRSIGMA: magnetic Resonance SIGnature MAtching for real - time volumetric imaging. Magnetic resonance in medicine. 2020;84(3):1280-92.
15. Mickevicius NJ, Paulson ES. Simultaneous acquisition of orthogonal plane cine imaging and isotropic 4D-MRI using super-resolution. Radiotherapy and Oncology. 2019;136:121-9.
16. Han P, Chen J, Xiao J, Han F, Hu Z, Yang W, et al. Single projection driven real-time multi-contrast (SPIDERM) MR imaging using pre-learned spatial subspace and linear transformation. Physics in Medicine & Biology. 2022;67(13):135008.
17. Harris W, Yin FF, Cai J, Ren L. Volumetric cine magnetic resonance imaging (VC-MRI) using motion modeling, free-form deformation and multi-slice undersampled 2D cine MRI reconstructed with spatio-temporal low-rank decomposition. Quant Imaging Med Surg. 2020;10(2):432-50.
18. Liu L, Shen L, Johansson A, Balter JM, Cao Y, Chang D, et al. Real time volumetric MRI for 3D motion tracking via geometry - informed deep learning. Medical Physics. 2022;49(9):6110-9.
19. Xiao H, Ni R, Zhi S, Li W, Liu C, Ren G, et al. A dual-supervised deformation estimation model (DDEM) for constructing ultra-quality 4D-MRI based on a commercial low-quality 4D-MRI for liver cancer radiation therapy. Med Phys. 2022;49(5):3159-70.
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. Magn Reson Med. 2016;75(2):775-88.
21. Keijnemans K, Borman PTS, van Lier A, Verhoeff JJC, Raaymakers BW, Fast MF. Simultaneous multi-slice accelerated 4D-MRI for radiotherapy guidance. Phys Med Biol. 2021;66(9).
22. Eslami SA, Jimenez Rezende D, Besse F, Viola F, Morcos AS, Garnelo M, et al. Neural scene representation and rendering. Science. 2018;360(6394):1204-10.
23. Sitzmann V, Zollhöfer M, Wetzstein G. Scene representation networks: Continuous 3d-structure-aware neural scene representations. Advances in Neural Information Processing Systems. 2019;32.
24. Sitzmann V, Martel J, Bergman A, Lindell D, Wetzstein G. Implicit neural representations with periodic activation functions. Advances in Neural Information Processing Systems. 2020;33:7462-73.
25. Chen Y, Liu S, Wang X, editors. Learning continuous image representation with local implicit image function. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021.
26. Shen L, Pauly J, Xing L. NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. IEEE Transactions on Neural Networks and Learning Systems. 2022.
27. Dupont E, Goliński A, Alizadeh M, Teh YW, Doucet A. Coin: Compression with implicit neural representations. arXiv preprint arXiv:210303123. 2021.
28. Shen T, Gao J, Yin K, Liu M-Y, Fidler S. Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis. Advances in Neural Information Processing Systems. 2021;34:6087-101.
29. Vasudevan V, Shen L, Huang C, Chuang C, Islam MT, Ren H, et al. Implicit neural representation for radiation therapy dose distribution. Physics in Medicine & Biology. 2022;67(12):125014.
30. Liu L, Shen L, Yang Y, Schüler E, Zhao W, Wetzstein G, et al. Modeling linear accelerator (Linac) beam data by implicit neural representation learning for commissioning and quality assurance applications. Med Phys. 2023.
31. Qureshi AH, Mousavian A, Paxton C, Yip MC, Fox D. Nerp: Neural rearrangement planning for unknown objects. arXiv preprint arXiv:210601352. 2021.

This article is protected by copyright. All rights reserved.

Author Manuscript

32. Tancik M, Srinivasan P, Mildenhall B, Fridovich-Keil S, Raghavan N, Singhal U, et al. Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems. 2020;33:7537-47.
33. 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;66(17):175028.
34. Ulyanov D, Vedaldi A, Lempitsky V, editors. Deep image prior. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
35. Johansson A, Balter JM, Cao Y. Abdominal DCE - MRI reconstruction with deformable motion correction for liver perfusion quantification. Medical physics. 2018;45(10):4529-40.
36. 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;66(8):085002.
37. Seregni M, Paganelli C, Lee D, Greer P, Baroni G, Keall P, et al. Motion prediction in MRI-guided radiotherapy based on interleaved orthogonal cine-MRI. Physics in Medicine & Biology. 2016;61(2):872.
38. Stemkens B, Tijssen RH, De Senneville BD, Lagendijk JJ, Van Den Berg CA. Image-driven, model-based 3D abdominal motion estimation for MR-guided radiotherapy. Physics in Medicine & Biology. 2016;61(14):5335.
39. Bruijnen T, Stemkens B, Lagendijk JJ, Van Den Berg CA, Tijssen RH. Multiresolution radial MRI to reduce IDLE time in pre-beam imaging on an MR-Linac (MR-RIDDLE). Physics in Medicine & Biology. 2019;64(5):055011.
40. Bjerre T, Crijns S, af Rosenschöld PM, Aznar M, Specht L, Larsen R, et al. Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes. Physics in Medicine & Biology. 2013;58(14):4943.
41. Mostafaei F, Tai A, Omari E, Song Y, Christian J, Paulson E, et al. Variations of MRI-assessed peristaltic motions during radiation therapy. PloS one. 2018;13(10):e0205917.
42. Johansson A, Balter JM, Cao Y. Gastrointestinal 4D MRI with respiratory motion correction. Medical Physics. 2021;48(5):2521-7.

This article is protected by copyright. All rights reserved.

Author Manuscript

List of Figures

Figure 1. Framework of prior-augmented NeRP learning for MRI reconstruction. a) During prior embedding and augmentation stage, the network was trained using sparse priors of 2 fully sampled volumetric MRI. The network took 3D image voxel locations and motion state coordinates as input and output the corresponding voxel intensities. b) During sparse image reconstruction stage, the network took voxel locations that correspond to the sparse sampling schemes, for example, locations sampled by orthogonal cine acquisition as inputs and was optimized to output the corresponding voxel intensities. Volumetric MRI was obtained by querying the optimized network at desired 3D voxel locations.

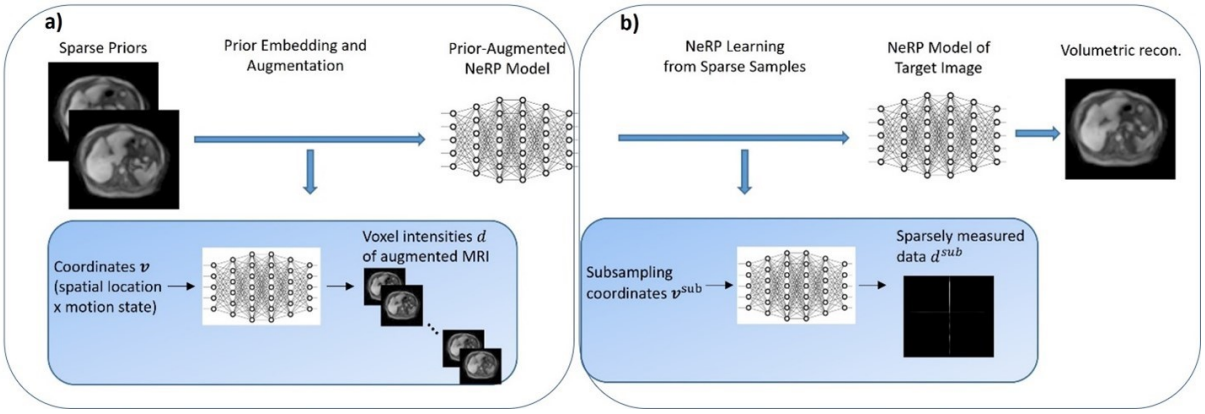
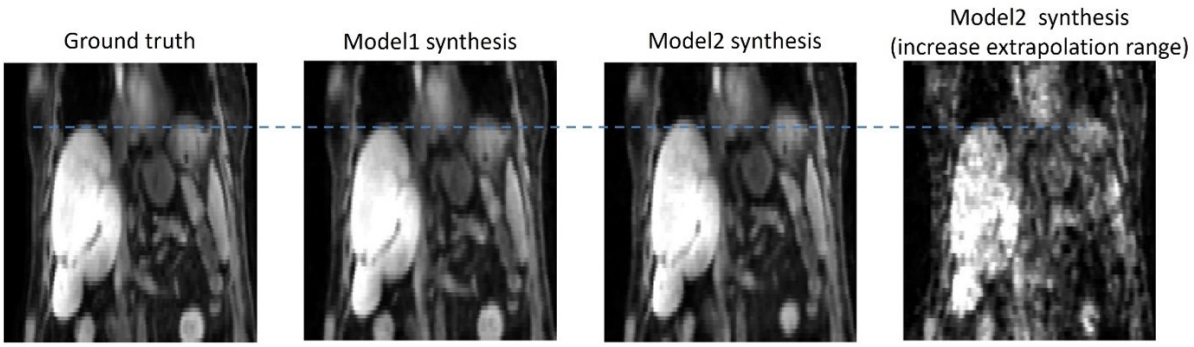


Figure 2. Comparison of NeRP models in synthesizing 4D MRI phases. The synthesis quality decreases as the motion magnitude between two training phases decreases from 90% (model 1) to 14% (model 2) of the full motion magnitudes. Increasing the extrapolation range of model 2 to extreme results in image artifacts and does not improve the overall matching.



This article is protected by copyright. All rights reserved.

Author Manuscript

Figure 3. Sample views of MRI reconstructed from 2 orthogonal slices using the NeRP model.

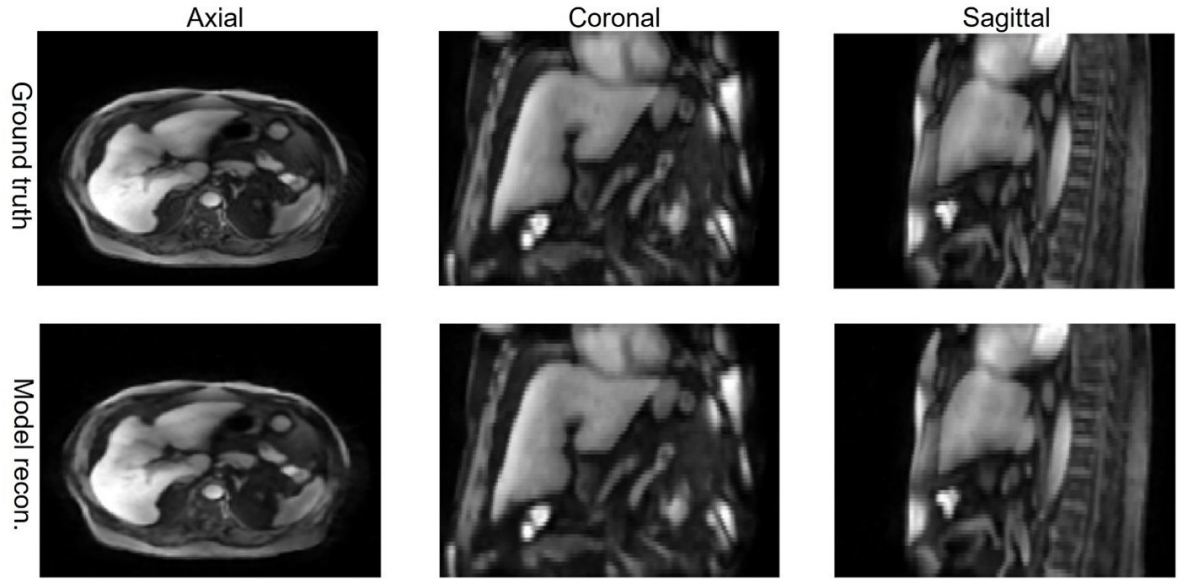


Figure 4. Example prior images, ground truth images and model reconstructed images. Blue, red and green lines pass through the liver dome of the exhale state prior image, the inhale state prior image and the ground truth image respectively. The model is able to reconstruct MRI with motion states both in between (patient 1) and outside (patient 2, deeper inhale state) the two priors.

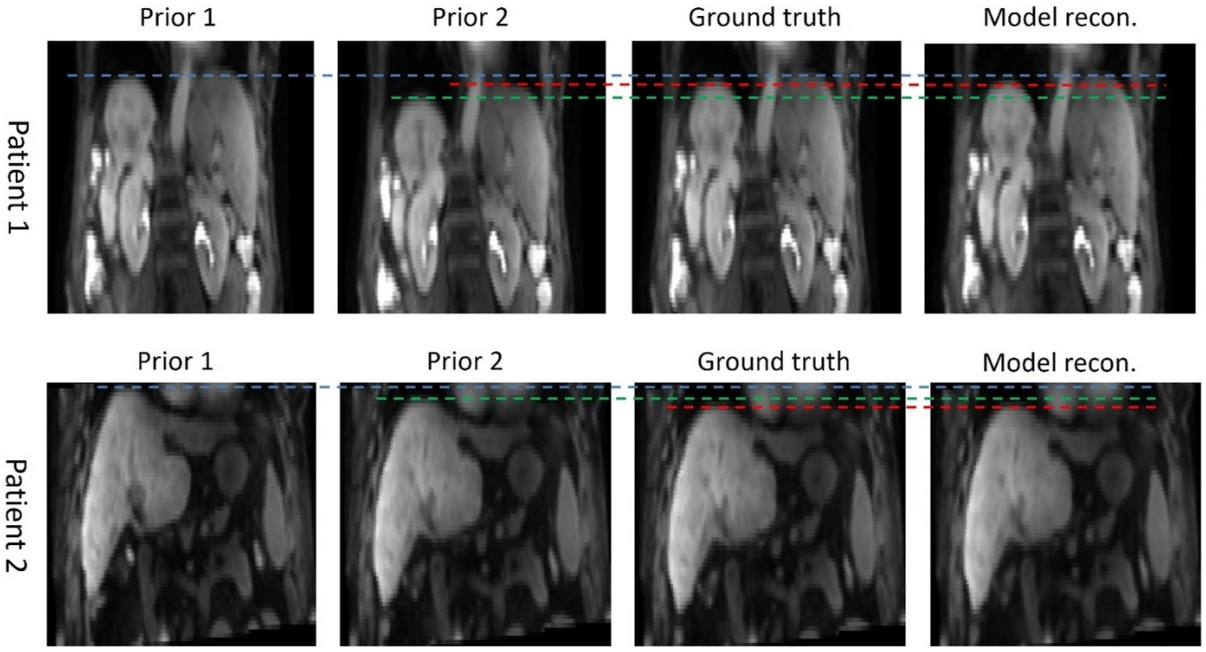


Figure 5. Comparison of images reconstructed using the proposed method through NeRP modeling with/without finetuning and using the baseline method through DVF scaling. The difference map was shown at the bottom. Images reconstructed by DVF scaling showed discrepancies at multiple organ locations (liver, bowel and spleen), while the discrepancy was reduced through NeRP

This article is protected by copyright. All rights reserved.

Author Manuscript

modeling. The finetune step further improved the consistency between the reconstructed image and the ground truth.

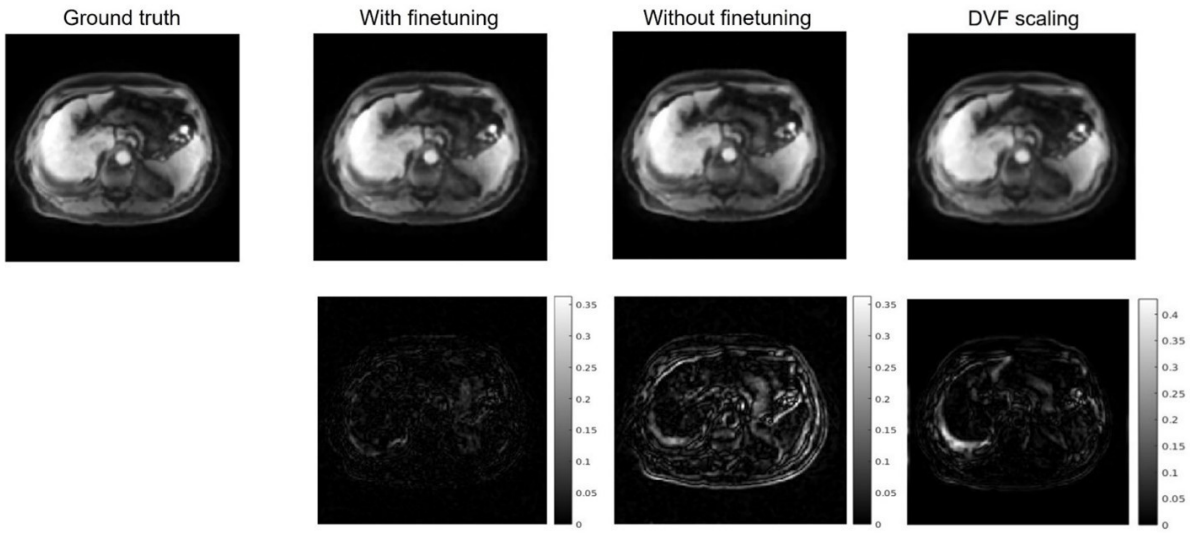
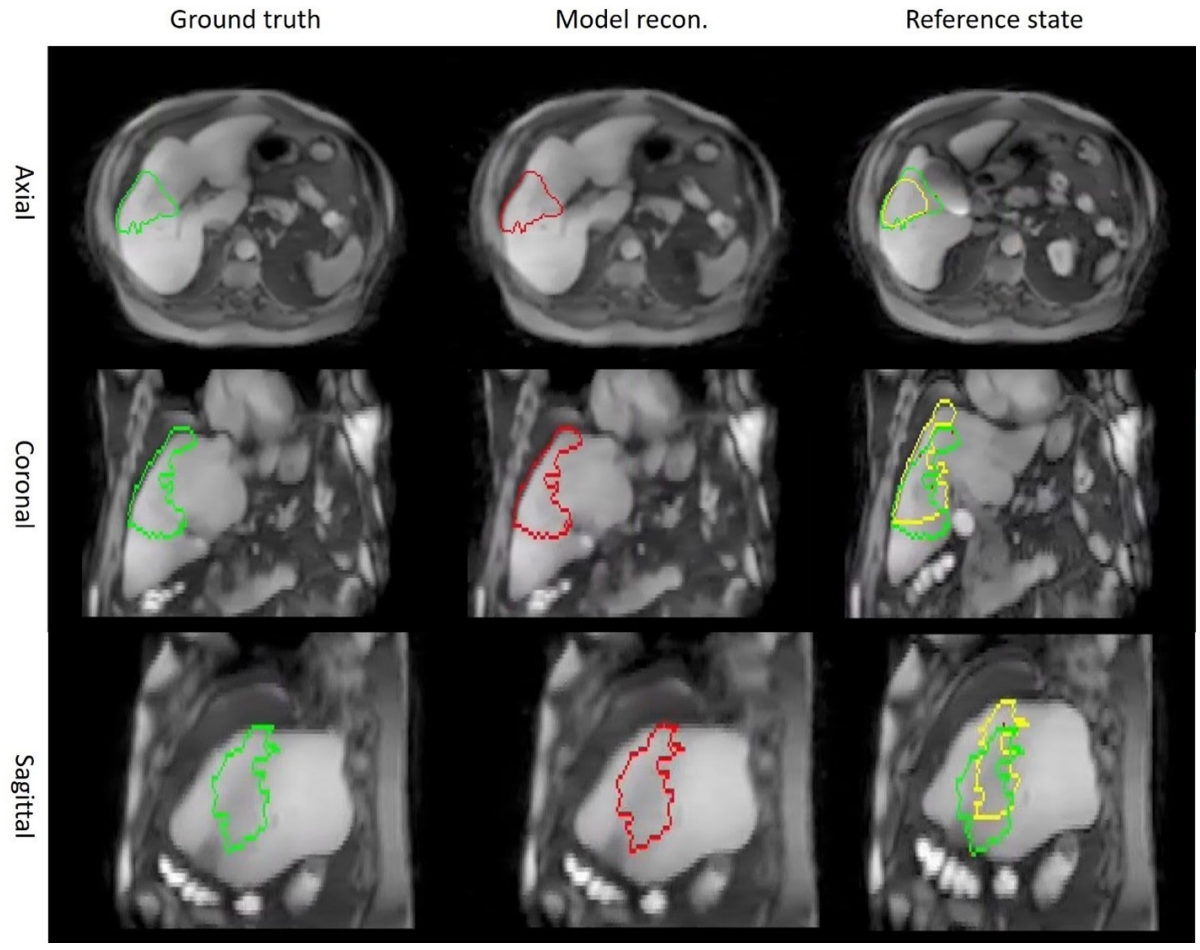


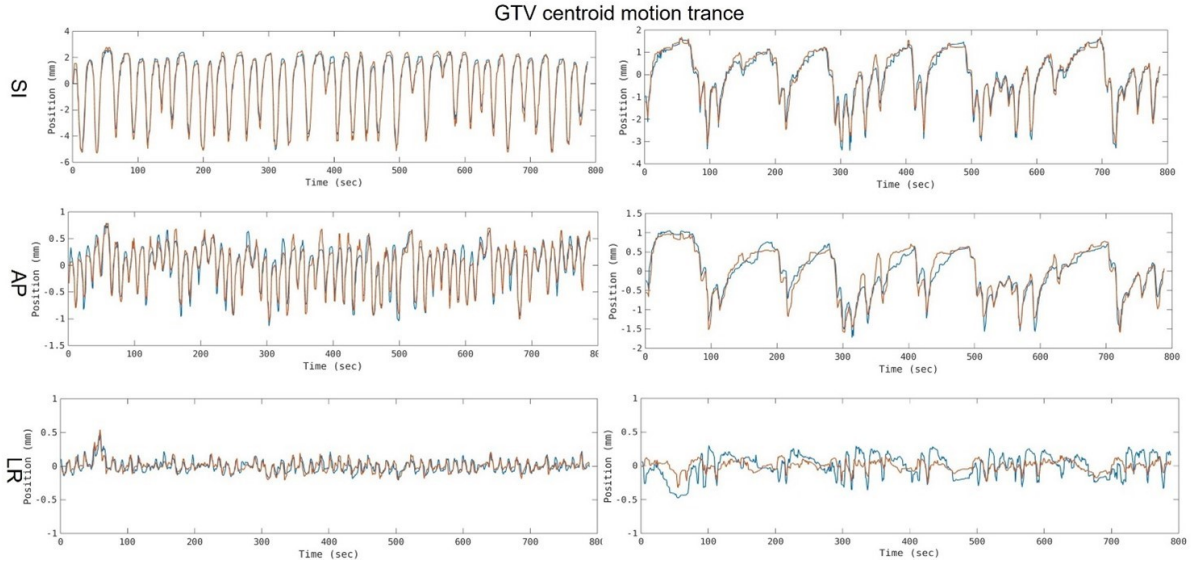
Figure 6. Example GTV contours generated by deforming the reference state MRI to ground truth (green) and model reconstructed (red) MRI. The GTV contours were also overlaid with the reference state MRI with reference GTV contour defined (yellow).



This article is protected by copyright. All rights reserved.

Author Manuscript

Figure 7. Example motion trace of GTV centroid in the superior-inferior (SI), anterior-posterior (AP) and left-right (LR) directions over 4.5 min.



This article is protected by copyright. All rights reserved.