Received: 24 April 2023 | Revised: 7 July 2023 | Accepted: 31 July 2023

DOI: 10.1002/mp.16702

RESEARCH ARTICLE

MEDICAL PHYSICS

Intra-frame motion deterioration effects and deep-learning-based compensation in MR-guided radiotherapy

Zhuojie Sui[1] | **Prasannakumar Palaniappan**[1] | **Jakob Brenner**[1] | **Chiara Paganelli**[2] | **Christopher Kurz**[3] | **Guillaume Landry**[3] | **Marco Riboldi**[1]

[1]Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Garching, Germany

[2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

[3]Department of Radiation Oncology, University Hospital, LMU Munich, Munich, Germany

Correspondence
Marco Riboldi, Department of Medical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Am Coulombwall 1, 85748, Garching b. München,Germany.
Email:Marco.Riboldi@physik.uni-muenchen.de

Funding information
China Scholarship Council

Abstract

Background: Current commercially available hybrid magnetic resonance linear accelerators (MR-Linac) use 2D+t cine MR imaging to provide intra-fractional motion monitoring. However, given the limited temporal resolution of cine MR imaging, target intra-frame motion deterioration effects, resulting in effective time latency and motion artifacts in the image domain, can be appreciable, especially in the case of fast breathing.

Purpose: The aim of this work is to investigate intra-frame motion deterioration effects in MR-guided radiotherapy (MRgRT) by simulating the motion-corrupted image acquisition, and to explore the feasibility of deep-learning-based compensation approaches, relying on the intra-frame motion information which is spatially and temporally encoded in the raw data (k-space).

Methods: An intra-frame motion model was defined to simulate motion-corrupted MR images, with 4D anthropomorphic digital phantoms being exploited to provide ground truth 2D+t cine MR sequences. A total number of 10 digital phantoms were generated for lung cancer patients, with randomly selected eight patients for training or validation and the remaining two for testing. The simulation code served as the data generator, and a dedicated motion pattern perturbation scheme was proposed to build the intra-frame motion database, where three degrees of freedom were designed to guarantee the diversity of intra-frame motion trajectories, enabling a thorough exploration in the domain of the potential anatomical structure positions. U-Nets with three types of loss functions: L1 or L2 loss defined in image or Fourier domain, referred to as $NN_{ImgLoss\text{-}L1}$, $NN_{Floss\text{-}L1}$ and $NN_{L2\text{-}Loss}$ were trained to extract information from the motion-corrupted image and used to estimate the ground truth final-position image, corresponding to the end of the acquisition. Images before and after compensation were evaluated in terms of (i) image mean-squared error (MSE) and mean absolute error (MAE), and (ii) accuracy of gross tumor volume (GTV) contouring, based on optical-flow image registration.

Results: Image degradation caused by intra-frame motion was observed: for a linearly and fully acquired Cartesian readout k-space trajectory, intra-frame motion resulted in an imaging latency of approximately 50% of the acquisition time; in comparison, the motion artifacts exhibited only a negligible contribution to the overall geometric errors. All three compensation models led to a decrease

in image MSE/MAE and GTV position offset compared to the motion-corrupted image. In the investigated testing dataset for GTV contouring, the average dice similarity coefficients (DSC) improved from 88% to 96%, and the 95th percentile Hausdorff distance (HD_{95}) dropped from 4.8 mm to 2.1 mm. Different models showed slight performance variations across different intra-frame motion amplitude categories: $NN_{ImgLoss\text{-}L1}$ excelled for small/medium amplitudes, whereas $NN_{Floss\text{-}L1}$ demonstrated higher DSC median values at larger amplitudes. The saliency maps of the motion-corrupted image highlighted the major contribution of the later acquired k-space data, as well as the edges of the moving anatomical structures at their final positions, during the model inference stage.
Conclusions: Our results demonstrate the deep-learning-based approaches have the potential to compensate for intra-frame motion by utilizing the later acquired data to drive the convergence of the earlier acquired k-space components.

KEYWORDS
deep learning, imaging latency, intra-frame motion, motion compensation, MR-Linac

1 | INTRODUCTION

Inter- and intra-fractional motion management is an important issue for precision radiation therapy, which has driven the development of Image-Guided Radio-Therapy (IGRT).[1] MR-guided radiotherapy (MRgRT)[2–5] integrating Magnetic Resonance (MR) imaging in external beam delivery provides continuous high soft tissue contrast MR sequences during the treatment without ionizing radiation, and offers considerable potential in real-time treatment adaptation.[6]

According to the online adaptive MRgRT workflow,[4] any introduced error or uncertainty in one step is propagated to the final dose deposition. As MR image acquisition is the first step of the whole feedback chain, image quality represents a dominant factor towards the accuracy of the following image processing (e.g., contouring), and beam control via respiratory gating[7,8] or MultiLeaf Collimator (MLC) tracking.[9,10] For the current commercially available magnetic resonance linear accelerators (MR-Linac)s,[2,5] intra-fractional motion monitoring is facilitated by 2D+t cine MR imaging.[5] In the past few years, evaluation of cine-MR has been focused on the spatial/temporal resolution,[11] geometric distortion[12,13] and artifacts.[14]

In contrast to static imaging, cine-MR depicts moving structures, and this involves a motion-dependent sampling of the raw image data during acquisition. Object motion deterioration effects in radiography have been thoroughly studied and described mathematically as impulse responses, which are then represented by the Modulation Transfer Functions (MTF) formalism to combine both spatial and temporal deterioration of the imaging system.[15–17] For instance, uniform motion at a given velocity smears a point into a line, resulting in box-function impulse response, in which case the system's spatial MTF needs to be multiplied with a sinc function. However, MR imaging is essentially different as the raw signal is acquired in Fourier domain. Therefore, additional approaches need to be explored to study these effects. Since the acquisition of a single cine-MR frame and the physiological motion are of the same time scale, the final acquired k-space comprises signals of the target at different positions. This is reflected in the image domain, manifesting as motion artifacts and imaging latency. The latter has been thoroughly investigated by Borman et al.,[18] characterizing the lag effects between the actual object position at the moment the acquisition is completed and the apparent object position derived from the MR image. As an additional example, Liu et al.[19] from the Australian MR-Linac project have further confirmed that the imaging latency (194 ± 43 ms) is the largest contribution to the total end-to-end latency (328 ± 44 ms) in MRgRT treatment delivery; Glitzner et al.[20] have conducted a technical study on the Elekta Unity MR-Linac and have found that the latency caused by MR imaging exceed the MLC-related delays by several factors, thus confirming the importance of further research to reduce the MR imaging latency. Hereinafter, we will refer to the physical motion of objects occurring within one cine-MR frame acquisition as the intra-frame motion, and the corresponding combined image degradation effects of both motion artifacts and imaging latency as intra-frame motion deterioration effects. In recent years, organ motion due to respiration has been measured extensively,[21] and from the published results, fast anatomical variations within one breathing cycle could be expected, especially in a deep breathing mode. Observations on lung tumor motion showed that the speed could reach up to 72.6 mm/s,[22] and considering the limited temporal resolution of cine-MR imaging (four or eight frames per second), the intra-frame motion can be appreciable.

In Borman and colleagues' work,[18] a reduction of imaging latency was achieved by changing the phase encoded ordering scheme in Cartesian readouts, or by implementing a spatial-temporal (k-t) filter for the golden angle sequence in radial readouts. However, the effective high-low ordering scheme suggested in the literature can introduce more severe eddy current artifacts with additional compensation techniques required. Moreover, in high-low ordering, larger differences in the higher frequency components between the intra-frame motion-corrupted image and the final-position image (i.e., the ideal motion-deterioration-free image) could be expected, because they are earlier acquired. This also challenges image quality, as the higher frequency components offer useful semantic information which can be used for contouring.[23,24] Retrospectively weighting the k-space data with the k-t filter halves the imaging latency but there is still potential for optimization. Over the past few years, artificial intelligence (AI) algorithms have been playing an increasingly important role in MRI or MRgRT regarding multiple application fields, such as motion correction,[25,26] image segmentation,[27,28] synthetic CT generation[29] and online treatment planning.[30] A large number of results have highlighted the promising contributions of deep learning.[31]

The aim of this work is to investigate the intra-frame motion deterioration effects in MRgRT and the feasibility of deep-learning-based compensation approaches, by carrying out a breathing MR digital phantom study. Compared to the clinically acquired cine-MR sequences which are likely to have been contaminated already by the intra-frame motion of the target, digital phantoms offer several advantages. Firstly, they overcome the spatial resolution limitation typically found in clinical cine-MR images, which may not be sufficient for investigating intra-frame motion. Secondly, obtaining the corresponding ground truth image for motion-corrupted images remains challenging in clinical settings, whereas digital phantoms provide precise final-position images as well as target segmentation for reference and evaluation. Additionally, they allow for image reconstruction with various dedicated k-space readout trajectories and noise models, facilitating MR imaging optimization studies. Moreover, digital phantoms hold great potential in addressing the issue of limited data for deep-learning-based model training.

In the following paragraphs, we first explain the generation of 4D MRI digital anthropomorphic phantoms, followed by the implemented motion-corrupted image simulation with the predefined intra-frame motion model. Unlike previous literature on motion correction, which involves mainly image restoration to filter out motion artifacts,[25,32] a deep-learning-based intra-frame motion compensation method is proposed in this work. Specifically, based on the original 2D+t cine MR sequences selected from the 4D phantoms, a motion pattern perturbation scheme is proposed for intra-frame motion database creation and augmentation; the previous simulation code serves as the data generator while a Convolutional Neural Network (CNN)[33] is applied to (i) remove noise by learning the sequence-specific noise distribution and (ii) to estimate the exact final-position image from the motion-corrupted image, by correlating the intra-frame motion information spatially and temporally encoded in the raw k-space data to the final position image.

To the best of our knowledge, this study represents the first attempt to validate the feasibility of deep-learning-based intra-frame motion compensation for MRgRT. The contribution of this work can be summarized as follows:

- By constructing the intra-frame motion model and simulating the motion-corrupted MR images, we emphasize the intra-frame motion deterioration effects on image quality or the geometric accuracy, highlighting the practical value of implementing motion compensation, especially for patients with fast breathing mode.
- We propose a dedicated motion pattern perturbation scheme to fully explore the domain of the potential anatomical structure positions and ensure the diversity of intra-frame motion trajectories, paving the road towards the motion database generation and augmentation with real clinical data for deep learning applications.
- We demonstrate the superior performance of the compensation models through extensive evaluation, indicating the feasibility of deep learning based intra-frame motion compensation in MRgRT.
- In terms of the network interpretability, we present the saliency maps that reveal the ability of U-Net to detect and learn information from the later acquired data in the k-space, which thereby serves as the image filter for processing the earlier acquired components. This behavior is interesting and particularly important in addressing concerns regarding the potential and reliability of deep learning approaches to be implemented in clinical applications.

2 | METHOD AND MATERIALS

2.1 | 4D MRI digital anthropomorphic phantom generation

The 4D MRI digital anthropomorphic phantom is the MRI version of the extended 4D XCAT phantom,[34,35] which generates a moving virtual patient based on a parametric motion model of each organ. As shown in Figure 1, the workflow required a static representation of the virtual patient and the spherical tumor at the initial position of the breathing cycle. The spherical tumor here was used for locating and propagating the centroid of a realistic tumor selected among non-small cell

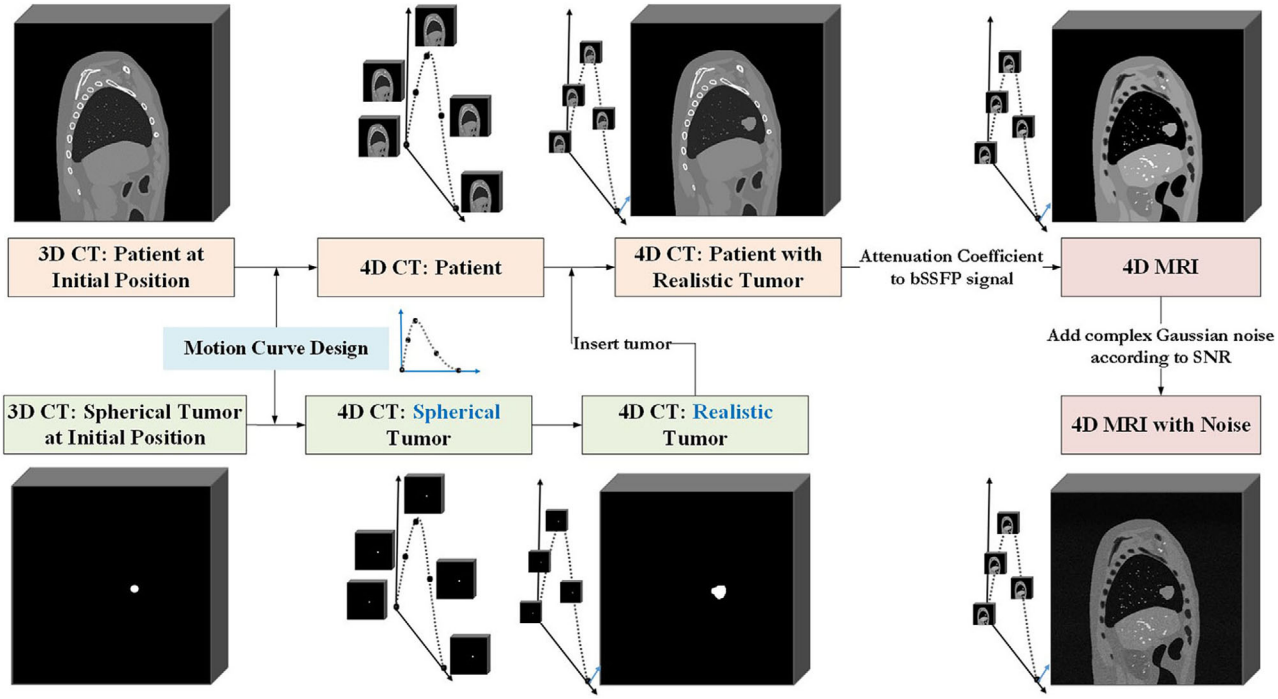

FIGURE 1 Workflow of 4D MRI digital anthropomorphic phantom generation. The phantom is schematically binned in five phases for each breathing cycle, but in the actual application process, more breathing phases were used.

lung cancer patients.[36,37] The realistic tumor shape was segmented on the treatment planning 4D CT scan, relying on the exhale phase. A respiratory motion curve was designed by defining the time-resolved amplitude of superior-inferior (SI) diaphragm motion and anterior-posterior (AP) chest-wall expansion, to generate a 4D CT of the patient and tumor, respectively. By mapping the centroid positions extracted from the 4D CT of the spherical tumor to the static 3D realistic tumor, 4D CT images of a realistic moving tumor could be derived and finally combined to the anatomical image.

In order to translate 4D CT images to 4D MRI, the bSSFP (balanced steady state free precession) signal intensity S was simulated by Equation 1,[35] with the assumption that TR/T1 and TR/T2 tend to 0, as TR is always much shorter than T1 and T2 in this sequence:

$$S = \rho \sin\alpha \frac{1}{1+\cos\alpha+(1-\cos\alpha)(T1/T2)} e^{-TE/T2} \tag{1}$$

where α is the flip angle; T1, T2 and ρ are tissue-specific values for the longitudinal, transverse relaxation and proton density, respectively; further details on these values are given in Table S-1 of the supporting material. In this work, we considered $\alpha = 60°$ and TE = 1.27 ms, to match the acquisition parameters that are used in the Viewray MRIdian[5] at the LMU University Hospital. After converting the attenuation coefficient values in the 4D CT to the bSSFP signals of each tissue based on the corresponding T1, T2 and ρ maps,[35] ideal 4D MR images $\mathbf{I}$ were hereby generated. Noisy MRI images $\mathbf{I}^*$ were simulated by adding i.i.d. (independent and identically distributed) complex Gaussian noise to the k-space data $F(k_x, k_y)$ of each slice, which leads to an additive Rician distributed noise in the magnitude of the image domain:

$$\mathrm{I}^* = \mathrm{M}^{-1}(F(k_x, k_y) + \delta_{\mathrm{Re}} + j\delta_{\mathrm{Im}}); \quad \delta_{\mathrm{Re}}, \delta_{\mathrm{Im}} \sim \mathcal{N}(0, \sigma^2) \tag{2}$$

where $\mathbf{M}^{-1}$ represents the inverse Fourier transformation matrix; σ is the standard deviation of the Gaussian distribution, which can be derived from the pre-defined signal-to-noise ratio (SNR) with:

$$\sigma = \frac{\|F(k_x, k_y)\|_2}{\sqrt{N}} \times \frac{10^{-\frac{SNR}{20}}}{\sqrt{2}} \tag{3}$$

where N is the matrix size of k-space data. In the following sections, 2D+t cine MR sequences (referred to as original sequences) were obtained by selecting specific slices from the simulated 4D MRI scans.

2.2 | Motion corrupted image simulation and evaluation

Intra-frame motion deterioration effects should be investigated as a function of the k-space signal acquisition process. Figure 2 schematically summarizes the proce-

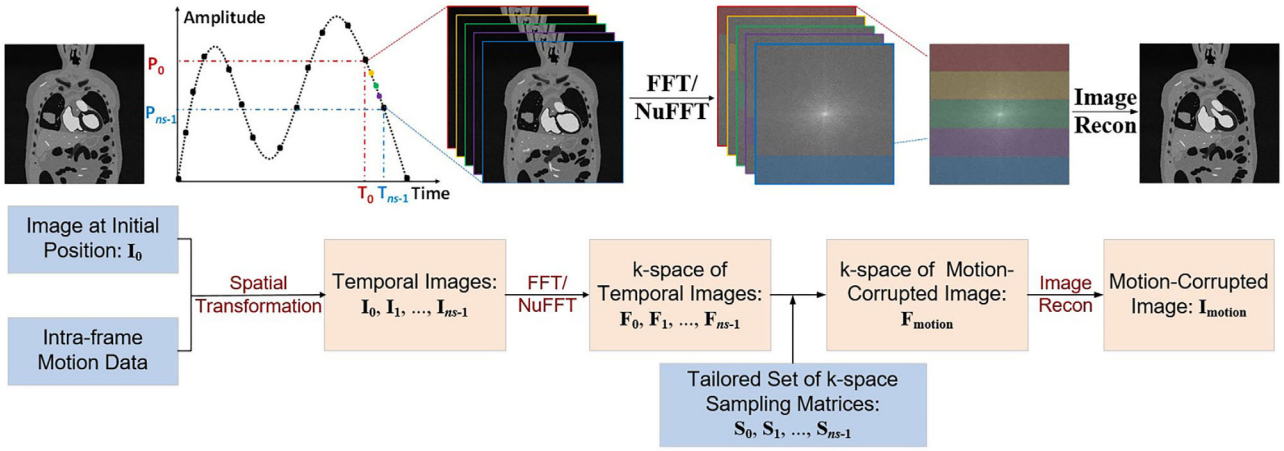

FIGURE 2 Procedure of intra-frame motion corrupted image simulation. Here *ns* is the number of time steps within the intra-frame motion. This schematic diagram only shows the case of $ns = 5$, but in the actual application process, a much larger number of time steps was used.

dure of simulating intra-frame motion-corrupted images. In the Amplitude/Time curve, the black dots indicate the start/end time of acquiring a specific cine-MR frame and the corresponding anatomical position as determined by the original sequences (see Section 2.1); the dotted line between such black dots shows the intra-frame motion trajectories; *ns* here represents the number of time steps within the intra-frame motion, which can also be interpreted as the number of shots, where a shot is defined as a subsample of k-space whose corresponding Fourier components are acquired simultaneously,[38] for example, along the frequency encoding direction.

Assume the acquisition of one frame began at time T_0, while the target was at its initial position P_0 and the corresponding image is denoted as $\mathbf{I_0}$. During the acquisition time of this frame, the target had moved to its final position $P_{ns\text{-}1}$, and the acquisition is completed at $T_{ns\text{-}1}$ which is also assumed to be the start time of the next frame. With the image at initial position $\mathbf{I_0}$ and the intra-frame motion model, the real-valued temporal images ($\mathbf{I_0}$, $\mathbf{I_1}$, …, $\mathbf{I_{ns\text{-}1}}$) of the target throughout the acquisition time could be derived by image spatial transformations like affine transformation or free form deformation. Depending on the k-space readout patterns, Fast Fourier Transformation (FFT) for Cartesian or Non-uniform Fast Fourier Transformation (NuFFT)[39,40] for radial or spiral will be performed to obtain the complex-valued k-space data ($\mathbf{F_0}$, $\mathbf{F_1}$, …, $\mathbf{F_{ns\text{-}1}}$) of each temporal image. Their corresponding components were then sequentially incorporated into the k-space arrays of the simulated motion-corrupted image over time. This process can be achieved by designing a tailored set of sampling matrices ($\mathbf{S_0}$, $\mathbf{S_1}$, …, $\mathbf{S_{ns\text{-}1}}$) based on the k-space readout trajectories with respect to the shot number, where MRI acceleration techniques like partial Fourier or parallel imaging methods[41,42] can be considered. These matrices have only two values, 0 and 1, with 1 indicating that the corresponding part of the matrix will be sampled. In this case, the complex-valued k-space of the motion-corrupted image $\mathbf{F_{motion}}$ is formulated as:

$$F_{motion} = \sum_{j=0}^{ns-1} F_j \circ S_j \tag{4}$$

where j denotes the shot number, and $\circ$ is the Hadamard product (element-wise product) operator. Finally, the motion-corrupted image $\mathbf{I_{motion}}$ was created in the form of complex numbers by the related image reconstruction method like inverse FFT/NuFFT, GRAPPA,[42] etc. For the results part of this study, we considered a linearly and fully acquired Cartesian k-space, where FFT was applied. Differences between $\mathbf{I_{motion}}$ and the ground truth final-position image $\mathbf{I_{ns\text{-}1}}$ reflect the intra-frame motion deterioration effects. From the simulation process, it can also be inferred that the finally acquired k-space consists of signals from the target at different positions. Consequently, the intra-frame motion data is spatially and temporally encoded in the k-space of the motion-corrupted image.

In the scope of this work, the intra-frame motion model was built with a linear approximation between consecutive control points (piece-wise linear approximation). That is, the overall intra-frame motion was divided into several consecutive time step intervals, with their end points being referred to as the control points. Motion between the control points is represented by the displacement vector field (DVF) between the corresponding images, which was then discretized with respect to the shot number. An optical flow[43] deformable image registration (DIR) algorithm was implemented to estimate the DVF. In order to minimize the errors introduced by optical flow and depict the most accurate possible final-position image, for a specific time step interval $[i, i+m]$ between $[0, ns\text{-}1]$, the temporal images $\mathbf{I_j}$

($j = i, i+1, \ldots, i+m$) were calculated as:

$$I_j = I_{i+m} \oplus \left(\frac{i+m-j}{m} \times DVF_m \right);\ j = i, i+1, \ldots, i+m;$$
$$DVF_m = \arg\min MSE\left(I_{i+m} \oplus dvf, I_i\right),$$
$$dvf \in \{DVF_{i+m \to i}, -DVF_{i \to i+m}\} \quad (5)$$

where the symbol $\oplus$ represents the image deformation based on the given DVF; *MSE*() denotes computing the mea-squared error (MSE) between two images; $\mathbf{I}_{i+m}$ is the image of control point $i+m$, and $\mathbf{I}_i$ is that of control point i; $DVF_{i+m \to i}$ represents the DVF from $\mathbf{I}_{i+m}$ to $\mathbf{I}_i$, while $DVF_{i \to i+m}$ represents the DVF from $\mathbf{I}_i$ to $\mathbf{I}_{i+m}$. Theoretically, $DVF_{i+m \to i}$ and $-DVF_{i \to i+m}$ should be identical, but considering the accuracy limit of the optical flow algorithm, we aimed at selecting the one leading to better residual MSE (i.e. a better $\mathbf{I}_i$ restoration) after registration (see Equation 5).

For the purpose of evaluation, a gross tumor volume (GTV) contouring error analysis was designed to match the clinical MR-Linac implementation of online structure tracking. In clinical practice, before the treatment, a preview cine MRI scan is acquired for selecting a tracking key frame $\mathbf{I}_{key}$; during the treatment, the live cine MRI frames are aligned to $\mathbf{I}_{key}$ with DIR and subsequently the GTV contour defined in the key frame is propagated.[5] Similarly, in this work, $\mathbf{I}_{key}$ as well as its corresponding binary GTV segmentation were defined: $\mathbf{I}_{key}$ was directly selected from the original sequences, while its GTV contour was obtained by choosing the same slice and frame from the 4D CT realistic tumor files. Then the GTV of the key frame was warped based on (i) the reference DVF from $\mathbf{I}_{key}$ to $\mathbf{I}_{ns\text{-}1}$ (DVF_{ref}) and (ii) the measured DVF from $\mathbf{I}_{key}$ to $\mathbf{I}_{motion}$ (DVF_{meas}), to get GTV_{ref} and GTV_{meas}, respectively. Finally, the contours were compared quantitatively by means of dice similarity coefficients (DSC), HD_{95} and centroid shift (COM shift). This allowed us to quantify the intra-frame motion deterioration effects in tumor tracking based on cine-MR images.

2.3 | Intra-frame motion compensation with deep learning

2.3.1 | Database

Original sequences

A total number of 10 4D MRI digital phantoms of lung cancer patients (five females and five males) with different anatomy were generated according to the workflow outlined in Section 2.1. Table 1 lists the basic information as well as the motion data assignment for each simulated patient. In this study, we predominantly focused on deep respiratory motion, thus tumors were mostly inserted in the middle or lower lobe of the lung, where the intra-frame motion was expected to be larger. We utilized the motion waveform multiplied by an amplitude amplification coefficient (AAC) to characterize the original patient-specific respiratory motion pattern. Specifically, 6 types of diaphragm motion waveforms were designed using the amplitude w.r.t. frame number curves, as shown in Figure 3, to mimic regular and irregular breathing patterns; 20 frames were acquired for each breathing cycle, in light of a 5s respiration period imaged with four frames per second (FPS) cine-MR; the amplitude for the corresponding type of chest-wall expansion waveforms was obtained following the same curve but with a different AAC.

From the published observations on the respiratory motion of the investigated patients, the diaphragm can move up to 100 mm in a deep breathing mode[21]; the peak-to-peak lung tumor motion amplitude ranges $0 \sim 50$ mm in the SI direction and $0 \sim 24$ mm in the AP direction,[21] while the maximum tumor speed is 72.6 ± 22.5 mm/s.[22] Our motion parameter settings for original sequences were designed based on these reported data, taking into account both normal and rapid movements. Table 2 shows the generated tumor motion data, with the largest intra-frame average motion speed calculated by dividing the largest intra-frame motion displacement (combining both the SI and AP directions) by the frame acquisition time of 250 ms. Among all patients, Patient 10 exhibited the greatest tumor motion in terms of both the motion range (55.5 mm in the SI direction) and the largest speed (73.3 mm/s). It should be noted that these motion data only represent the position information of the anatomical structure at the exact beginning/finishing moment of each frame acquisition in the original sequence. The more important intra-frame motion trajectory will be discussed in the following sections.

During the simulation of noisy MR images, the SNR was set to 10 dB; for each patient, four original 2D+t cine-MR sequences were chosen from their 4D MRI digital phantoms, comprising two sagittal and two coronal slices, with one sagittal and one coronal slice containing the tumor centroid being selected; to enhance the diversity of the selected individual slices, the other two slices were randomly chosen from the ones without tumor. These slices were specifically selected to differ significantly regarding the anatomic structure from the slices containing the tumor centroid, which had already been chosen. The image matrices were produced as 512×512 pixels, with spatial resolution of 1 mm $\times$ 1 mm; additionally, all of them are represented in the form of complex numbers, which were finally normalized through division by their maximum magnitude values.

Intra-frame motion pattern perturbation scheme

In order to build the intra-frame motion database, making the model able to generalize to various

TABLE 1 Basic information and motion data assignment for the simulated patients.

Patient ID	01	02	03	04	05	06	07	08	09	10
Gender	F	F	F	F	F	M	M	M	M	M
Age	63	65	57	65	56	63	70	52	67	50
Weight(kg)	81.3	78.6	105.8	56.0	69.6	72.1	100.4	60.8	89.9	120.0
Height(cm)	153.0	161.1	165.1	164.7	166.8	170.0	173.7	173.0	178.5	177.8
BMI	34.73	30.32	38.81	20.64	25.03	24.95	33.28	20.30	28.22	37.96
Tumor location	R/l/P	L/l/P	R/m/M	L/l/A	R/m/M	L/l/M	R/l/A	L/l/P	R/m/A	L/l/P
Waveform type	A	B	E	D	B	A	E	D	C	F
AAC of diaphragm	2.0	3.0	4.0	3.0	5.0	4.0	2.0	4.0	5.0	6.0
AAC of chest-wall	1.2	1.1	1.3	1.2	1.0	1.6	0.9	1.2	1.4	1.5

Note: Tumor location in the lung is presented in a way of R-Right, L-Left/ l-lower, m-middle (lobe) / P-Posterior, A-Anterior, M-Middle; AAC indicates amplitude amplification coefficient.

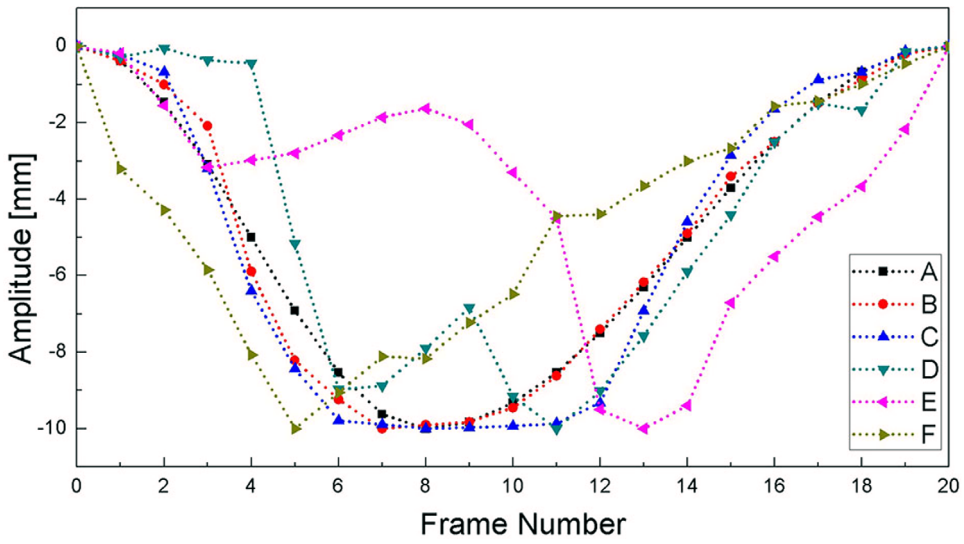

FIGURE 3 Diaphragm motion waveforms of the original sequences. The waveform is multiplied by an amplitude amplification coefficient (AAC) to characterize the original patient-specific respiratory motion pattern. The amplitude for the corresponding type of chest-wall expansion waveforms is obtained following the same curve but with a different AAC. Intra-frame motion trajectory is not considered in this process.

TABLE 2 Tumor motion data along the SI and AP direction for the simulated patients.

Patient ID		01	02	03	04	05	06	07	08	09	10
Peak-to-peak motion amplitude (mm)	SI	15.8	23.9	27.1	24.9	30.0	37.7	17.4	31.2	41.5	55.5
	AP	9.5	8.8	7.6	10.6	6.0	12.7	8.0	9.4	13.2	10.1
Average tumor motion speed (mm/s)	- -	7.4	10.2	11.2	10.8	12.2	15.9	7.7	13.1	17.4	22.6
Largest intra-frame motion displacement (mm)	SI	2.9	6.3	8.1	8.6	7.8	7.2	5.3	10.9	10.1	18.0
	AP	1.7	2.3	2.0	3.7	1.6	2.5	2.3	3.2	3.2	3.5
Largest intra-frame average motion speed (mm/s)	- -	13.6	26.8	33.3	37.4	31.8	30.2	23.2	45.3	42.2	73.3

intra-frame motion trajectories in the domain of all the potential anatomical structure positions, an intra-frame motion pattern perturbation scheme was proposed to determine the image of control points mentioned in Section 2.2. The scheme is generally divided into two steps:

Step1 involves defining the control point interval and the corresponding control point images. To achieve this, four additional frames were interpolated between two consecutive frames in the original sequence. Figure 4a illustrates a schematic view of this step: for the k^{th} frame in the original sequence (black dot in the figure, labeled

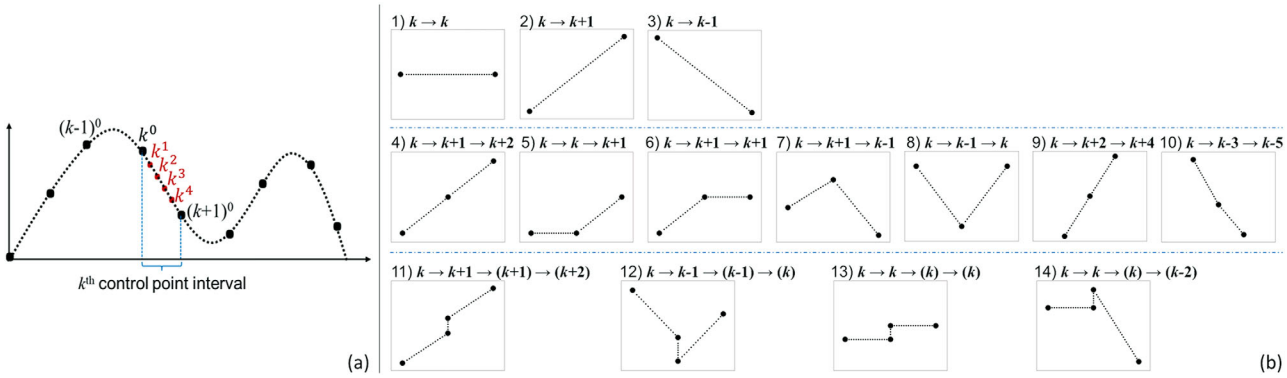

FIGURE 4 Schematic illustration of the intra-frame motion pattern perturbation scheme. (a) Step1: the definition of the k^{th} control point interval and the corresponding images; (b) Step2: the definition of the intra-frame motion trajectory. All k in (b) refer to randomly selecting one image from the k^{th} control point interval, and $\{k\}$ denotes the rigid transformation of the image at k, representing a sudden change in its position.

as k^0), images at k,[1] k^2, k^3, and k^4 were generated based on the linear interpolation of the DVF between the corresponding images of k^0 and $(k+1)^0$. In this way, five points were introduced and considered to fall within the k^{th} control point interval, comprising k^0 from the original sequence, and four interpolated frames k,[1] k^2, k^3, and k.[4] This process could also be interpreted as increasing the temporal resolution of the original cine-MR sequence by a factor 5: while it is technically possible to directly generate a digital phantom with a temporal resolution of 20 Hz, we opted to stick with 4 Hz and employed this additional process to augment the temporal resolution, considering the limitations in temporal resolution when constructing the training dataset with real clinical data. Briefly, by implementing this step, we are able to effectively enhance the position diversity of the reference final-position images.

Step2 involves defining the intra-frame motion trajectory leveraging the motion model proposed in Section 2.2, which will take the defined control point images in Step1 as inputs. In this step, motion patterns with 2, 3, or 4 control points were developed based on the scheme shown in Figure 4b, aiming at gradually increasing the degrees of freedom in breathing irregularity. All the k here referred to randomly selecting one out of five control point images from the k^{th} control point interval, except in the case of simulating a static scenario (pattern 1). After this process, the intra-frame motion displacement is no longer limited to the displacement between two sequential frames defined by the original sequence. Cases with 2 control points (patterns 1∼3) will take $i = 0$, $m = ns$-1 in Equation 5; while for cases with 3 or 4 control points (patterns 4∼14), a random insertion moment (denoted as pm) was chosen between [1, ns-2] for the middle control points, which would split the shot number into 2 intervals, [0, pm] and [pm+1, ns-1]. Cases with 4 control points were simulating a sudden position change of the target happening at the middle control points, represented by a rigid transformation; taking the image at k, for example, the rigid motion of it was denoted as $k \rightarrow \{k\}$. Parameters of the rigid motion were determined by taking a random number between [$-\pi/20$, $\pi/20$] as a rotation angle, and [−1, 1] mm as a translation distance along each axis. Extreme situations, which may never happen in reality, can be observed through patterns 9, 10 and 14 to offer lager differences between $\mathbf{I}_{motion}$ and $\mathbf{I}_{ns\text{-}1}$, forcing the potential network to pay more attention to the dynamic mechanism and keep robust to the variation in the motion amplitude. Briefly, three degrees of freedom were introduced in this process: the control point range, the insertion moment of the middle control points, and the rigid motion parameters. They introduce randomness in the database, thus allowing a thorough exploration in the domain of potential anatomical structure positions. This step also allowed us to customize any synthetic but realistic breathing motion pattern, including a dedicated intra-frame motion trajectory.

Consequently, the dataset included 11200 (10 patients × 4 slices × 20 frames × 14 patterns) input-output pairs, with data from randomly selected 8 patients (Patient 01, 03, 04, 05, 07, 08, 09, 10) used for training or validation and the remaining 2 patients (Patient 02, 06) for testing. Since the purpose of this work was to validate the feasibility of deep-learning based intra-frame motion compensation approaches, only a linearly and fully acquired Cartesian readout trajectory of k-space was considered; the k-space phase encoding direction was along the AP (anterior-posterior) direction for sagittal slices and the LR (left-right) direction for coronal slices, orthogonal to the main direction of intra-frame motion; the number of shots was set as 64 (ns = 64), that is, the target was considered to remain stationary (or the motion is negligible) while acquiring every 8 rows in phase encoding direction. To enable both the intra-frame motion compensation and denoising capabilities simultaneously, the input-output pair was the SNR = 10 dB motion-corrupted image $\mathbf{I}^*_{motion}$

and the corresponding noiseless final-position image $\mathbf{I}_{ns\text{-}1}$, which were both normalized by dividing by the maximum magnitude value of the input ($\mathbf{I^*}_{motion}$) before being fed into the network.

2.3.2 | U-Net model

As mentioned in Section 2.2, the intra-frame motion information is spatially and temporally encoded in the k-space of the motion-corrupted image. The intuitive idea of intra-frame motion compensation is using later acquired data to drive the process of filtering earlier acquired data. The properties of CNN make it a promising model for this application. In addition, with a Cartesian linear phase encoding ordering scheme, the later acquired data was corresponding to the higher frequency components and should be preserved while capturing the image features. The concatenative skip connections in the U-Net architecture could pass features from the encoder to the decoder of same dimensionality and enable the recovery of fine-grained details lost during the down-sampling process. Consequently, this study exploited the typical 5-level U-Net architecture,[44,45] with the real and imaginary parts of the image represented as separate channels.

Each level of the network consisted of a double convolution block using 3 × 3 convolution kernels, followed by batch normalization and ReLU activation. The first level had 64 feature channels, which were then sequentially doubled in the subsequent levels. 2 × 2 max pooling operation with stride two was applied for down-sampling in the contracting path, while "up-convolution" was implemented for up-sampling in the expansive path followed by concatenation. A 1×1 convolutional layer was set as the final layer of the network, which ultimately provided the output image. Three loss functions employing different metrics (mean absolute error [MAE] and MSE) across either the image or Fourier domain served to measure the L1 or L2 distance between the network-estimated final-position image/k-space ($\mathbf{I'}_{ns\text{-}1}$ / $\mathbf{F'}_{ns\text{-}1}$) and the corresponding reference ($\mathbf{I}_{ns\text{-}1}$ / $\mathbf{F}_{ns\text{-}1}$), which are: L1 loss in image domain, L1 loss in Fourier domain, and L2 loss. It is expected that, in accordance with Parseval's theorem, the L2 loss in the image and Fourier domains should exhibit equivalence, assuming uniformity in all other training configurations. The acquisition of the Fourier loss necessitated the conversion of both the network output channels and ground truth target image into the frequency domain, achieved through the utilization of the differentiable FFT operation.[46,47] Furthermore, we employed the Adam optimizer throughout the training, and the ratio of training to validation datasets was set at 7:1, with seven patients allocated for training and one patient (Patient10) for validation.

3 | RESULTS

3.1 | Intra-frame motion deterioration effects

This section aims to demonstrate how intra-frame motion deterioration affects GTV contouring accuracy in MRgRT. We chose the sagittal and coronal sequences where the tumor centroid is located for each patient, comparing the corresponding motion-corrupted sequence to the reference. The general quantitative results are summarized in Table 3. In the table, GTV variations due to intra-frame motion are evaluated by comparing GTV_0 (contoured on initial-position image $\mathbf{I}_0$) to GTV_{ref} (contoured on final-position image $\mathbf{I}_{ns\text{-}1}$). Meanwhile, GTV_{meas}, contoured on motion-corrupted image $\mathbf{I}_{motion}$, is compared to GTV_{ref} to evaluate the measured GTV errors. In the table, $\mathbf{I}_{motion}$ is simulated following a linear intra-frame motion from $\mathbf{I}_0$ to $\mathbf{I}_{ns\text{-}1}$ (motion pattern 2 in Figure 4), and the quantitative outcomes of GTV_{meas} including all of the other motion patterns can be found in Figures 8,10 and Table 4, evaluated among the network testing patients. The results for GTV_{meas} are presented separately for AP/LR (top) and SI (bottom) phase encoding directions. For each patient, the average value over all 40 frames (2 slices × 20 frames) and the results corresponding to the largest intra-frame motion (in the square brackets) are listed, while values at "Mean" show the average over all 400 frames across the 10 patients. It can be observed that in most cases, the centroid shift between GTV_{meas} and GTV_{ref} is around half of that between GTV_0 and GTV_{ref}. On average, 3.2 mm intra-frame tumor motion leads to a centroid positioning error of 1.5 mm. To understand it from another perspective, this means it causes an imaging latency of approximately 50% of the acquisition time. However, this ratio tends to be slightly higher in frames containing larger motion, except for Patient 10. When comparing the results of phase encoding directions, there is no noticeable difference. Although GTV errors are a bit higher with SI encoding direction in frames containing larger motion, DSC / COM shift differences are less than 2% / 1 mm (pixel). Some patients exhibit extreme intra-frame motion, resulting in considerable GTV contouring errors, with DSC values as low as 54% and HD_{95} values as high as 6.9 mm (Patient10), indicating severe deterioration effects.

To visually illustrate the errors in DVF derivation caused by cine-MR intra-frame motion, Figure 5 takes exemplary frames from Patient 02 (in sagittal) and Patient 08 (in coronal), who are respectively partitioned into the testing and training set in the subsequent deep learning framework, and displays several DVFs involved in the process of the frame acquisition and target tracking. The measured DVF_{meas} which is derived from the

TABLE 3 GTV quantitative assessment.

Patient ID		01	02	03	04	05	06	07	08	09	10	Mean
GTV_0	DSC (%)	92[84]	90[78]	83[58]	79[46]	80[58]	88[76]	93[77]	84[51]	69[48]	50[20]	81
	COM shift (mm)	1.8[3.8]	2.5[6.5]	2.8[8.3]	2.7[9.4]	3.0[7.8]	3.9[7.4]	1.8[5.8]	3.2[11]	4.3[10]	5.6[18]	3.2
	HD_{95} (mm)	2.7[3.6]	2.9[6.1]	3.6[8.5]	3.0[9.2]	3.4[7.2]	5.1[7.3]	2.4[7.1]	3.7[12]	6.1[11]	6.8[18]	4.0
GTV_{meas} AP/LR SI	DSC (%)	94[89]	94[89]	91[76]	87[71]	87[76]	93[84]	96[88]	91[73]	81[69]	72[54]	88
		94[89]	94[88]	90[74]	87[69]	86[76]	93[85]	96[86]	91[73]	79[67]	72[54]	87
	COM shift (mm)	0.9[2.3]	1.2[3.3]	1.5[4.4]	1.3[5.1]	1.3[3.5]	1.8[4.4]	0.9[3.3]	1.4[6.0]	2.0[5.6]	2.3[8.3]	1.5
		0.9[2.0]	1.2[3.6]	1.5[4.8]	1.2[5.3]	1.3[4.0]	1.7[4.9]	0.9[3.5]	1.4[6.1]	2.3[6.5]	2.5[8.2]	1.5
	HD_{95} (mm)	1.8[3.1]	2.0[4.2]	2.3[5.4]	2.1[4.5]	2.5[4.1]	3.8[5.0]	1.5[4.2]	2.5[7.0]	3.5[6.4]	3.4[6.9]	2.5
		1.9[2.8]	2.3[5.0]	2.9[7.0]	2.2[5.0]	2.7[4.8]	3.2[5.2]	2.0[6.0]	3.1[8.0]	3.8[8.4]	3.6[6.2]	2.8

Note: GTV intra-frame motion is represented by comparing the GTV of $\mathbf{I_0}$ (GTV_0) to that of $\mathbf{I_{ns\text{-}1}}$ (GTV_{ref}), while the measured GTV errors are evaluated by comparing the GTV obtained from $\mathbf{I_{motion}}$ (GTV_{meas}) to GTV_{ref}. For each patient, the average[largest motion] value over all 20 frames are listed. The "Mean" column is calculated by averaging the results over all 200 frames across the 10 patients. The results for GTV_{meas} are presented separately for AP/LR (top) and SI (bottom) phase encoding directions.

TABLE 4 Quantitative results evaluating the measured GTV contours before and after intra-frame motion compensation.

	DSC (%)				HD_{95} (mm)			
	Motion Corrupted	$NN_{Img\text{-}Loss\text{-}L1}$	$NN_{F\text{-}Loss\text{-}L1}$	$NN_{L2\text{-}Loss}$	Motion Corrupted	$NN_{Img\text{-}Loss\text{-}L1}$	$NN_{F\text{-}Loss\text{-}L1}$	$NN_{L2\text{-}Loss}$
Small	93 (4)	98 (2)	97 (2)	97 (2)	3.2 (1.9)	1.4 (1.1)	1.5 (1.3)	1.7 (1.5)
Medium	88 (3)	97 (2)	97 (2)	96 (2)	5.0 (2.2)	1.7 (1.3)	2.0 (1.6)	2.0 (1.5)
Large	80 (3)	95 (2)	95 (2)	95 (3)	7.3 (1.8)	2.5 (1.6)	2.5 (1.8)	2.5 (1.7)
Total	88 (7)	96 (3)	97 (2)	96 (3)	4.8 (3.1)	1.9 (1.7)	1.9 (1.7)	2.1 (1.8)

Note: Mean and (standard deviation) of DSC and HD_{95} for all the testing slices.

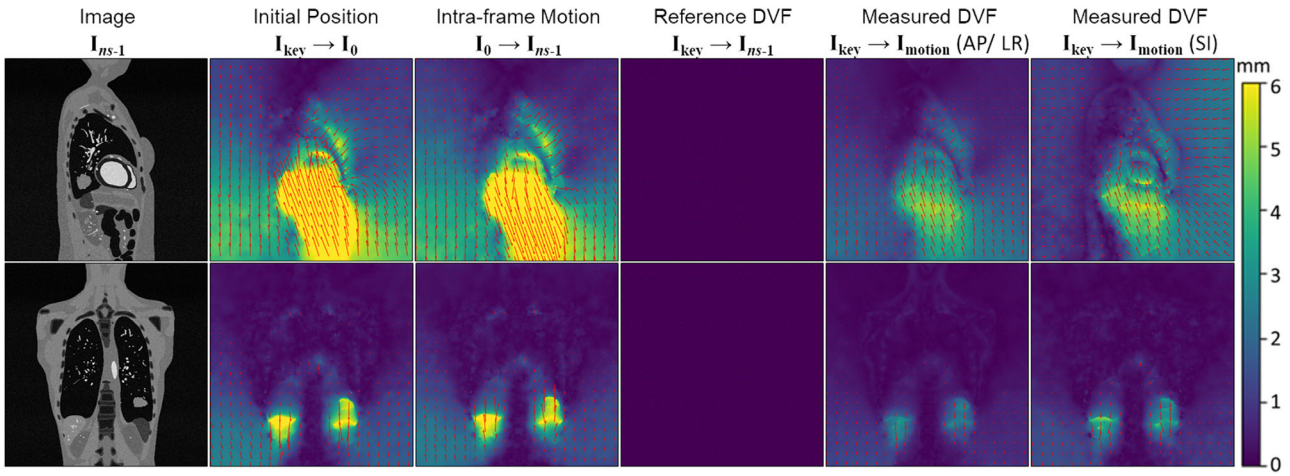

FIGURE 5 Displacement vector fields for Patient 02 (top) and Patient 08 (bottom). From left to right: final-position image $\mathbf{I_{ns\text{-}1}}$; DVF from key frame to initial-position image ($\mathbf{I_{key}} \rightarrow \mathbf{I_0}$); DVF of the intra-frame motion ($\mathbf{I_0} \rightarrow \mathbf{I_{ns\text{-}1}}$); reference DVF ($\mathbf{I_{key}} \rightarrow \mathbf{I_{ns\text{-}1}}$); measured DVF ($\mathbf{I_{key}} \rightarrow \mathbf{I_{motion}}$), with phase encoding direction orthogonal (AP/LR) or parallel (SI) to the main direction of intra-frame motion.

motion-corrupted image is compared to the reference DVF_{ref} (i.e., DVF derived from final-position image $\mathbf{I_{ns\text{-}1}}$). To facilitate a more intuitive comparison, the key frame was exactly selected as the final-position image: $\mathbf{I_{key}} = \mathbf{I_{ns\text{-}1}}$. In this case, DVF_{ref} is $\mathbf{0}$, and the DVF from the initial-position image $\mathbf{I_0}$ to $\mathbf{I_{ns\text{-}1}}$, which reflects the intra-frame motion, should have the same magnitude as the DVF from $\mathbf{I_{key}}$ to $\mathbf{I_0}$ but in the opposite direction. $\mathbf{I_{motion}}$ is also simulated following a linear motion from $\mathbf{I_0}$ to $\mathbf{I_{ns\text{-}1}}$, with different k-space phase encoding directions being considered. From the results, residual components of the intra-frame motion could be expected from the measured DVF, indicating the pronounced errors of DVF determination introduced by intra-frame motion

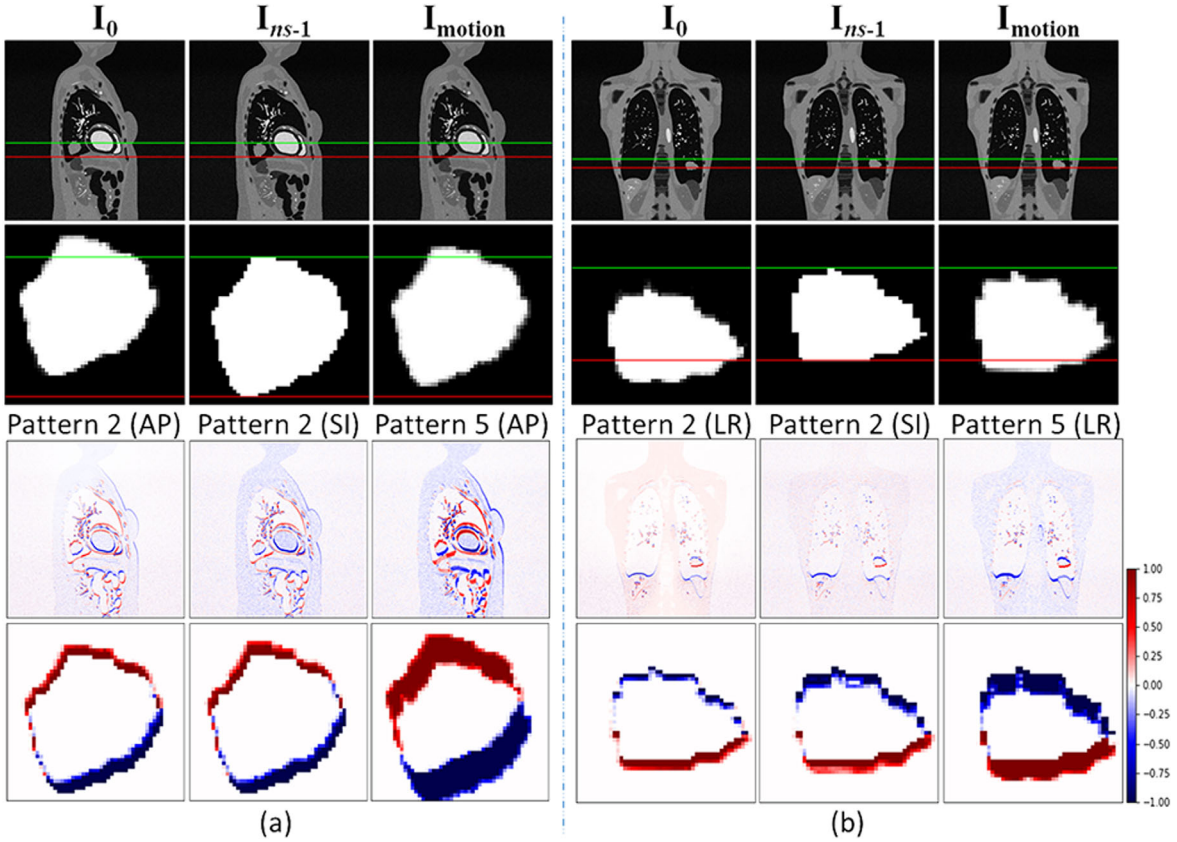

FIGURE 6 Intra-frame motion corrupted image and GTV contouring errors of exemplary frames from (a) Patient 02 and (b) Patient 08 cine-MR sequences. The top 2 rows are showing the images (1st row) as well as their corresponding GTV contours (2nd row) of initial-position image $\mathbf{I_0}$ (left), final-position image $\mathbf{I_{ns-1}}$ (middle) and motion-corrupted image $\mathbf{I_{motion}}$ with motion pattern 2 (right). The bottom 2 rows are showing the difference image between $\mathbf{I_{motion}}$ and $\mathbf{I_{ns-1}}$ (3rd row) and their corresponding GTV contouring errors (4th row), where $\mathbf{I_{motion}}$ is simulated with different motion patterns and phase encoding directions: motion pattern 2 and AP phase encoding direction (left), motion pattern 2 and SI phase encoding direction (middle), and motion pattern 5 with the insertion moment of 65% the acquisition time (pm = 65% ns) and AP/LR phase encoding direction (right).

deterioration effects. Qualitatively, the dominant component of the intra-frame anatomical changes is along the SI direction, and compared to the orthogonal phase encoded scheme, slightly greater errors are appreciable with phase encoded in the SI direction.

Figure 6 shows the motion-corrupted image and the resulting GTV contouring errors for the same frames as in Figure 5. During the frame acquisition time, the tumor is gradually moving from GTV_0 (GTV of $\mathbf{I_0}$) to GTV_{ref}. Since the frame in Figure 6a is taken during inhale, the tumor is generally moving downwards, while the acquisition of Figure 6b frame is happening during exhale and the tumor is moving upwards. A reference line was set at the upper and lower boundaries of GTV_{ref}. It can be observed that the contour of the target is well preserved in $\mathbf{I_{motion}}$, this is different from other imaging systems whose signals are acquired in image domain such as the fluoroscopic system, where all the passing positions (pixels) of the target during the acquisition time might be recorded; the GTV contour derived from $\mathbf{I_{motion}}$ lags behind the actual final position of the target, clearly indicating noticeable imaging latency; in comparison, the motion artifacts (or image blur) exhibited a negligible contribution to the overall geometric errors. Similar to the conclusions drawn from DVF analysis, qualitatively, the choice of k-space phase encoding direction slightly impacts the contouring accuracy, with the SI direction exhibiting slightly more motion artifacts compared to orthogonal directions. The intra-frame motion patterns can significantly contribute to motion degradation, which in turn, yield the errors in GTV contouring. Specifically, an insertion moment of 65% the acquisition time (pm = 65% ns) in motion pattern 5 ($\mathbf{I_0 \rightarrow I_0 \rightarrow I_{ns-1}}$) led to worse image quality, because in this case, the lower frequency components in k-space, which have much higher orders of magnitude, are mainly taken from $\mathbf{I_0}$.

3.2 | Motion compensation with U-Net

The model was built with the PyTorch library,[47] trained and tested on an NVIDIA Quadro P5000 GPU with the memory of 16GB. For the sake of convenience,

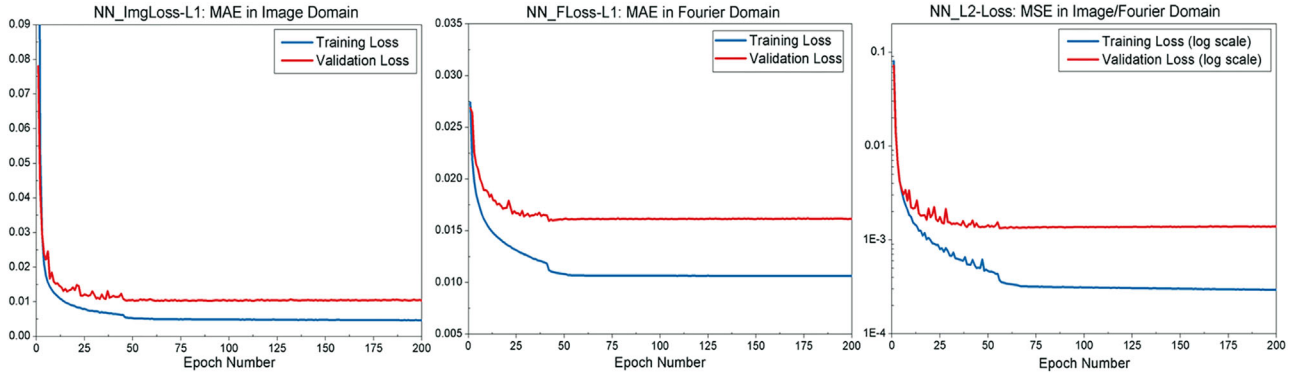

FIGURE 7 Training and validation losses for the U-Net with three distinct loss functions. Note: L2 loss is plotted on logarithmic scale.

the U-Net incorporating the three loss functions (L1 loss in image domain, L1 loss in Fourier domain, and L2 loss) will be indicated as $NN_{ImgLoss\text{-}L1}$, $NN_{Floss\text{-}L1}$, and $NN_{L2\text{-}Loss}$, respectively. To determine the optimal learning rate for each model, a hyper-parameter search was conducted sampling from the set $\{1\times10^{-3}, 1\times10^{-4}, 1\times10^{-5}\}$. The selected learning rates were 1×10^{-4} for $NN_{ImgLoss\text{-}L1}$, 1×10^{-3} for $NN_{Floss\text{-}L1}$, and 1×10^{-4} for $NN_{L2\text{-}Loss}$. All models were trained with a batch size of 6.

To measure the motion compensation speed of the network, we calculated the average final-position image estimation time across the testing dataset. The result was 6.3 ms per frame.

3.2.1 | Image comparison

Figure 7 shows the training and validation losses for the U-Net using three different loss functions, where L2 loss is shown on logarithmic scale to facilitate the distinction of subtle differences. Notably, $NN_{Floss\text{-}L1}$ exhibits a relatively larger discrepancy between the training and validation datasets compared to the other models. Nevertheless, all validation loss curves eventually converged to a steady and horizontal line at the end of the training process. During the final inference stage, the weights from epoch 100 of all three models were loaded for testing.

Two evaluation metrics, MSE and MAE, were used to quantitatively assess the effectiveness of the models in image quality enhancement through motion compensation. Figure 8 illustrates the results obtained for all the testing frames grouped by three types of intra-frame motion patterns. In general, the U-Net exhibited a remarkable reduction in image errors when compared to the reference image across the testing dataset. For cases involving type01 and type02 motion patterns, $NN_{ImgLoss\text{-}L1}$ demonstrated a slight tendency towards a superior performance over the others in terms of both MAE and MSE: on average, applying the network $NN_{ImgLoss\text{-}L1}$ / $NN_{Floss\text{-}L1}$/ $NN_{L2\text{-}Loss}$ resulted in a decrease of MSE (MAE) to 7.3% (11.6%) / 9.8% (17.6%) / 15.0% (23.5%) of the initial value, respectively. In the cases with type03 motion pattern, a wider range of MAE or MSE variations was observed in motion-corrupted images due to the sudden introduction of rigid motion. However, the performance of three models was comparably good, as evidenced by the average MSE (MAE) values dropping to approximately 9.8% (15.2%) of the initial value, indicating the effective mitigation of the intra-frame motion deterioration effect.

Figure 9 presents comparison results of motion-corrupted images ($\mathbf{I_{motion}}$) versus the network-estimated final-position images ($\mathbf{I'_{ns\text{-}1}}$) for the testing patients. The image quality has seen a great enhancement with the help of the U-Net: an advantage of the network estimated positions over the ones derived from the motion-corrupted image is visible for the tumor, the cardiac and the structures in the abdomen. Among the models, $NN_{ImgLoss\text{-}L1}$ exhibited better image contrast restoration than $NN_{Floss\text{-}L1}$ and $NN_{L2\text{-}Loss}$, showing closer pixel values to the reference in adipose and muscle tissues. By comparing the motion-corrupted image to the reference final-position image, it can be observed that the cardiac image could have undergone a large intra-frame deformation during the acquisition. However, all three compensation models were able to estimate the precise anatomic structure corresponding to the moment the acquisition is completed. Additionally, the analysis of k-space discrepancy images suggest that the U-Net effectively captures the position information from the later-acquired higher frequency components, and then utilizes it to guide the fine-tuning of the lower frequency components, which ultimately led to a successful compensation of intra-frame motion.

3.2.2 | GTV contour comparison

We performed a separate analysis of the slices in the testing dataset where the tumor centroid is located, to

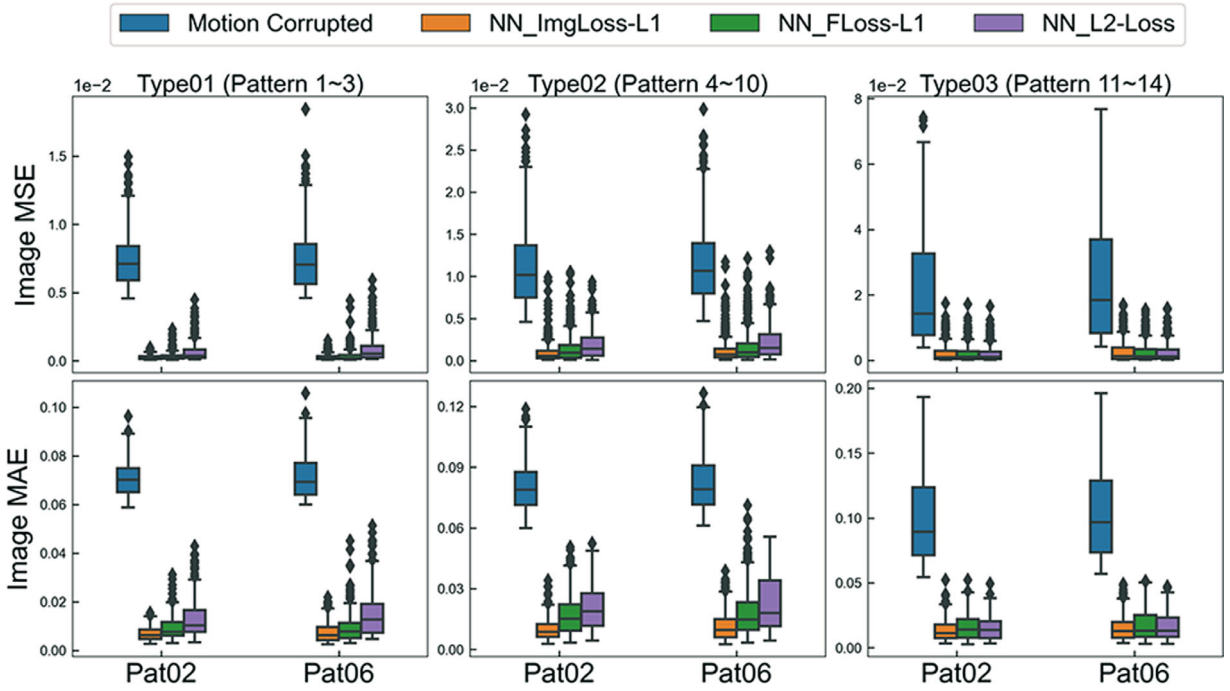

FIGURE 8 Box plots comparing the MSE (top) and MAE (bottom) for all the testing frames before and after intra-frame motion compensation. Testing frames are grouped by the type of intr-frame motion pattern. Note: all the images are normalized to [0, 1], and the y-axis has different scale limits among subfigures.

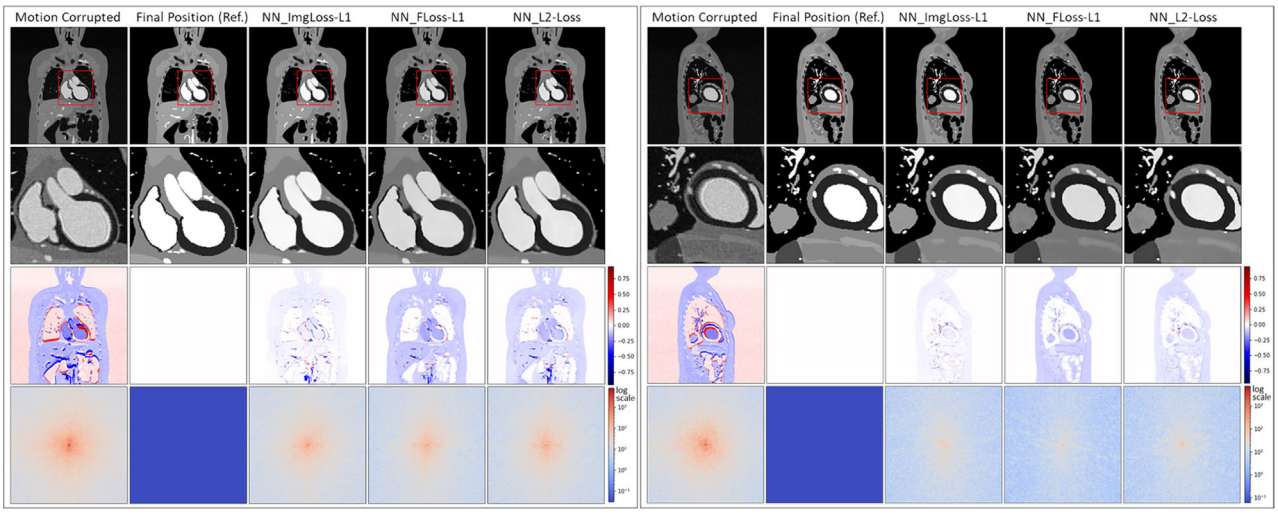

FIGURE 9 Image comparison before and after intra-frame motion compensation. Exemplary coronal (left) and sagittal (right) frames are selected from the testing patients. From top to bottom: original image, zoom-in cardiac image, image difference and the magnitude of k-space difference on logarithmic scale (calculated by subtracting the reference).

assess the impact of motion compensation on GTV contour estimation accuracy. Results were ordered in 3 categories according to the GTV centroid shift of the motion-corrupted image from the reference final-position image: Small, for COM shift $\leq$ 2 mm; Medium, for 2 mm < COM shift $\leqq$ 5 mm; Large, for 5 mm < COM shift < 8 mm. Cases where COM shift > 8 mm were ignored, as in these cases, the intra-frame tumor motion speed tends to be greater than the highest velocity observed in clinical studies.

Figure 10 and Table 4 list the evaluation results for all the testing slices containing tumors. The results indicate an evident benefit of applying the deep learning based intra-frame motion compensation. All models yielded a notable improvement in DSC, surpassing 95% for the median of each category. The mean DSC for all the evaluated cases increased by 8% from the initial value of 88%. Among the three models, $NN_{ImgLoss-L1}$ demonstrated better performance in the Small and Medium category, yet a slight tendency towards a higher

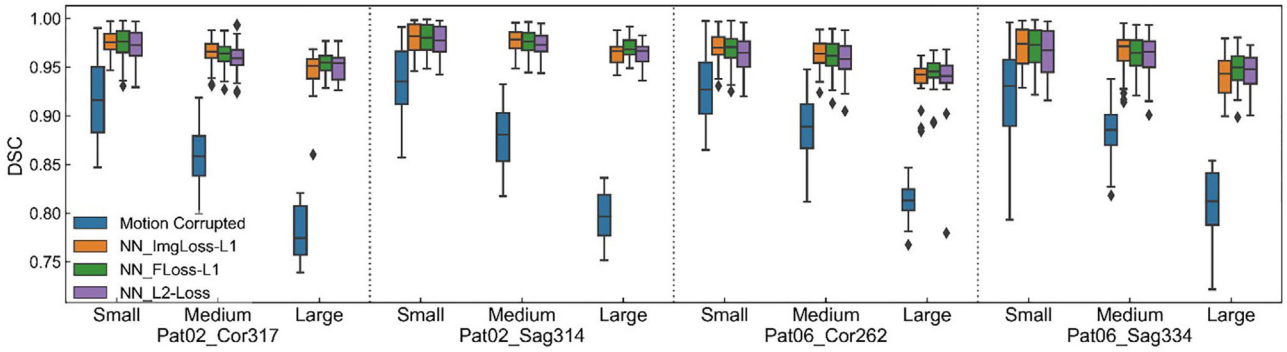

FIGURE 10 Box plots comparing the DSC of all the testing slices before and after intra-frame motion compensation. Based on the GTV centroid shift, the testing dataset was divided into 3 categories: Small, Medium and Large.

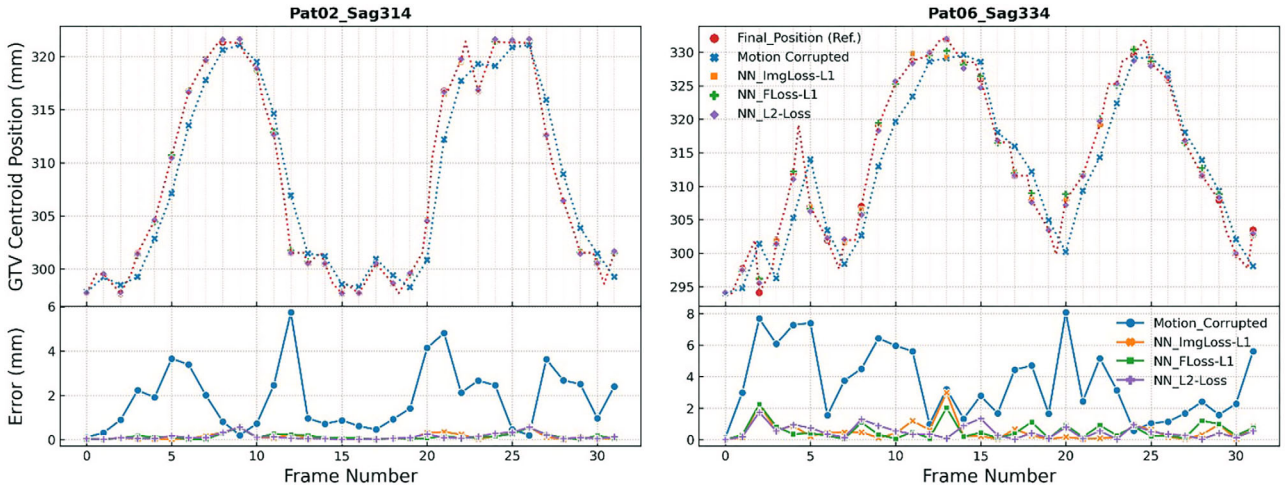

FIGURE 11 GTV Centroid Position comparison curve. The constructed breathing motion, including the intra-frame motion trajectory, is depicted as the red line, with the red dots serving as the reference GTV centroid position at the moment the frame acquisition is terminated. Results before and after motion compensation are displayed for motion-corrupted results in blue, $NN_{ImgLoss\text{-}L1}$ in yellow, $NN_{FLoss\text{-}L1}$ in green and $NN_{L2\text{-}Loss}$ in purple. The difference curves between them and the reference are shown in the panels below.

DSC median value with $NN_{Floss\text{-}L1}$ can be observed in the Large category. Moreover, $NN_{ImgLoss\text{-}L1}$ / $NN_{Floss\text{-}L1}$/ $NN_{L2\text{-}Loss}$ brought a decrease in the mean HD_{95} from 4.8 mm to 1.9/1.9/2.1 mm, respectively. Overall, these results suggest that while different models show slight performance variations in different intra-frame motion amplitude categories, all models are effective in eliminating the motion-related deterioration effects within cine-MR, improving the image quality and thereby enhancing the accuracy of GTV contour position estimation.

With the original sequences and intra-frame motion pattern perturbation scheme proposed in Section 2.3.1, it is feasible to customize any synthetic but realistic breathing motion pattern, including a dedicated intra-frame motion trajectory. Based on this, a GTV centroid motion curve was constructed for sagittal slices of Patient02 and Patient06 from the testing dataset, as shown in Figure 11, which serves as the ground truth. The absolute GTV centroid position derived from motion-corrupted images and the network estimated final-position images are compared to the reference. As can be seen from this figure, in the motion-corrupted results, most frames demonstrated an imaging latency that is approximately half of the frame acquisition time. However, a longer time delay was noticeable for certain frames of Patient06, particularly Frame 10, 11, and 13. This can be attributed to the potential degradation of image quality caused by motion artifacts and noise, which consequently affects the accuracy of the optical flow algorithm. The three network estimated results overlap well with the ground truth in all the cases, providing an effective GTV position offset. The only exception was Frame13 for Patient06, where the optical flow algorithm failed to precisely contour the tumor from the $NN_{ImgLoss\text{-}L1}$ and $NN_{Floss\text{-}L1}$ estimated image. The intra-frame motion deterioration effects are negligible or completely vanishing in cases with a very shallow breathing mode, such as in Frame13 ~ 18 for Patient 02.

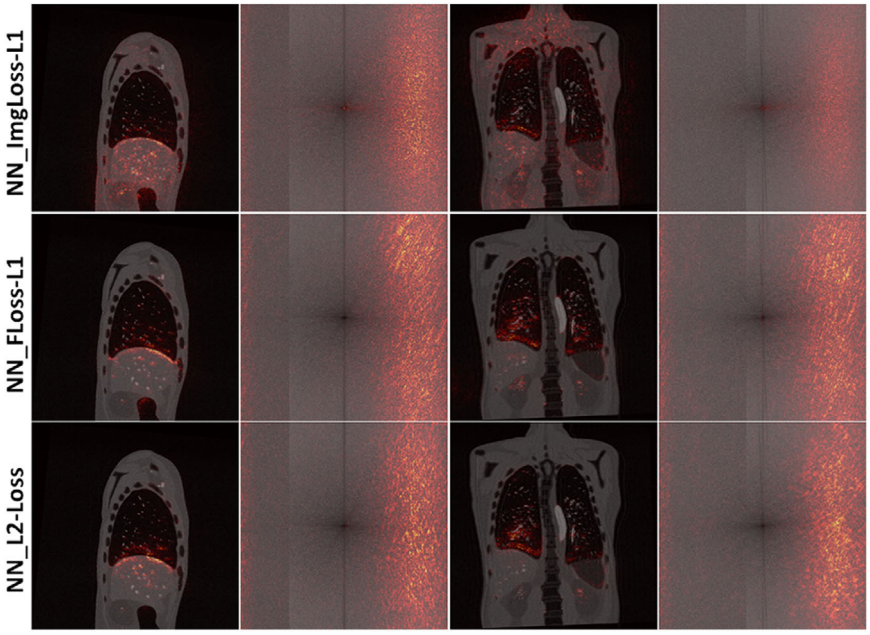

FIGURE 12 Overlaid saliency map in the image (left) and Fourier (right) domain for model $NN_{ImgLoss\text{-}L1}$ (top), $NN_{Floss\text{-}L1}$ (middle) and $NN_{L2\text{-}Loss}$ (bottom).

3.2.3 | Saliency map

To visualize which part of the motion-corrupted image / k-space contributes more to the inference, saliency maps were generated in both the image and Fourier domain for the three models. Specifically, saliency maps in the image domain calculated the gradient of the network loss function w.r.t the input motion-corrupted image utilizing the SmoothGrad[48] technique; to obtain saliency maps in the Fourier domain, input motion-corrupted image tensors were converted to the frequency domain and loaded onto the device (GPU) for gradient computation, which were then converted back to the image domain before being fed into the network.

The overlaid saliency maps of exemplary testing patients are shown in Figure 12. On the one hand, the right part of k-space corresponding to the later acquired data is highlighted in the heat map, representing a large contribution to the final results; on the other hand, saliency maps in the image domain indicate a primary focus on the edges of the moving structures; Particularly, U-Net can detect the edges in their final position, as evidenced by the coronal slice, where the model-highlighted liver edge deviates from the edge perceived by visual observation.

4 | DISCUSSION

From published results of respiration-related organ motion measurements, it is known that the motion varies markedly across patients and treatment sites, and both complex motion patterns and great tumor motion velocity have been observed.[21] Furthermore, despite the application of imaging acceleration techniques, cine-MR has limited temporal resolution and the imaging latency is found to be the largest latency contribution to the whole feedback chain of MRgRT,[19,20] indicating that the intra-frame motion can be appreciable, especially in the case of fast anatomical variations. This work investigated intra-frame motion deterioration effects in MRgRT by conducting a digital phantom study, where the acquisition of motion-corrupted 2D+t cine MR images was simulated based on the predefined intra-frame motion model. Currently, real-time beam gating is clinically realized based on the target deformation using the DVF from the key frame to the live cine MRI frames estimated by motion tracking algorithms.[5,49] Results in Figure 5 demonstrated that pronounced DVF errors may be measured in the motion-corrupted image compared to the reference. Additionally, quantitative results in Table 3 show that the GTV contouring errors can push DSC values as low as 54%, suggesting the necessity of implementing the intra-frame motion compensation for fast breathers.

In terms of the phase encoding directions, although slightly higher errors with an encoding direction parallel to the main tumor motion direction can be qualitatively appreciable (results in Figures 5 and 6) compared to an orthogonal direction, the results in Table 3 indicate a DSC / COM shift difference of less than 2% / 1 mm (pixel). This suggests that the choice of the phase encoding direction is less critical in 2D+t cine-MR compared to the 3D MR acquisition. The findings in Table 3 show that with a linearly and fully acquired Cartesian readout k-space trajectory, intra-frame motion results in an imaging latency of approximately 50% of the acquisition time. This is consistent with the findings in the work of Borman et al.[18] and Riederer et al.,[50] showing that target position is mostly determined by the moment of

acquiring the central k-space profile. However, the proportion can be a bit higher in certain cases, due to the higher k-space components and the interplay between target motion and k-space sampling. A few cases in our dataset exhibited such a behavior: in Patient06 the centroid position difference is 4.4 mm for a total intra-frame shift of 7.4 mm. Similar observations can be found from Figure 11: the GTV centroid position derived from the motion-corrupted image is in general corresponding to the position at half of the frame acquisition time, with the exception of some frames in Patient06, where a longer latency is observed. Compared to the imaging latency, which is found to be the primary manifestation of intra-frame motion deterioration effects, geometric errors caused by intra-frame motion artifacts can be assumed insignificant (see Figure 6).

By means of the implemented motion-corrupted image simulation, we were able to reveal how the intra-frame motion data has been spatially and temporally encoded in the k-space relative to the signal acquisition process. This presents the opportunity to develop a method for deriving the final-position image directly from the motion-corrupted image, provided that a model trained on realistic intra-frame motion patterns is available. Therefore, we validated the feasibility of deep-learning based motion compensation approaches: the intra-frame motion database was built based on our proposed motion pattern perturbation scheme, enabling a thorough exploration in the domain of potential anatomical structure positions; a U-Net with three types of loss functions was trained to estimate the exact final-position image, providing simultaneous intra-frame motion compensation and denoising. While the U-Net was initially designed for image segmentation,[44] it has recently been adapted for image reconstruction purposes.[51] In order to estimate the final-position image precisely, the network is expected to learn to drive the convergence of the earlier acquired components utilizing the later acquired data. With a Cartesian linear phase encoding ordering scheme, the later acquired data was corresponding to the higher frequency components, and the concatenative skip connections in the U-Net architecture could enable the preservation of them while extracting the image features. Moreover, since the intra-frame motion pattern has been identified as one of the dominant contributors to deterioration effects (see Figure 6), its diversity is crucial during the preparation of the training dataset. The proposed motion pattern perturbation scheme could also be regarded as the data augmentation process to prevent overfitting, and the design with 3 degrees of freedom proved to be sufficient for the generated patterns.

The network takes around 6.3 ms to complete the motion compensation, which is far less than the cine-MR frame acquisition time. Furthermore, this speed depends highly on the hardware configuration as well as the matrix size of the input image: with the advancement of GPU computing performance, the actual processing time is expected to be further reduced.

The analysis in Section 3.2 shows all $NN_{ImgLoss\text{-}L1}$, $NN_{Floss\text{-}L1}$ and $NN_{L2\text{-}Loss}$ models have achieved a successful motion compensation, with a better image MSE or MAE, as well as GTV contour DSC and HD_{95}. In the investigated testing dataset for GTV contouring, the average DSC improved from 88% to 96%, yet the HD_{95} dropped from 4.8 mm to 2.1 mm. Different models showed slight performance variations across different intra-frame motion amplitude categories: $NN_{ImgLoss\text{-}L1}$ excelled in the Small and Medium cases, whereas $NN_{Floss\text{-}L1}$ demonstrated a bit higher DSC median values in the Large category. This outcome aligns with expectations, as the input-output image pairs of the network are normalized, resulting in absolute prediction errors that are less than 1. Consequently, the L2 loss is less sensitive to the outliers compared to the L1 loss. Relatively large deformations in the cardiac region were observed within a single cine-MR frame acquisition as depicted in Figure 9. However, the models exhibited the capability to estimate the precise anatomic structure corresponding to the moment the acquisition is completed, presenting potential benefits in real-time MR imaging of cardiac function. Furthermore, in Figure 11, the network-estimated results overlap well with the ground truth, showing a clear benefit. Finally, the saliency maps of the motion corrupted image in Figure 12 highlight the far-right part of k-space as well as the edges of the moving anatomical structures, and these detected edges are at their final positions which could deviate from the edges perceived by visual observation. This makes it more transparent that the models have learned to detect and extract information from the later acquired data, which thereby serve as the image filter of processing the earlier acquired components.

This work has its limitations, one of which concerns the accuracy of the implemented optical flow algorithm. Equation 5 was applied to minimize the errors introduced by optical flow when simulating the motion-corrupted images, however, during the evaluation process of Table 3, we have found it not precise enough for GTV contouring in 2 or 3 frames of Patient 09 and 10. Although this does not significantly affect our results, and the algorithm is performing well when applied to the network testing patients (Patient 02 and 06), future research should strive to seek a more reliable way for contouring or to provide an additional uncertainty analysis on DIR algorithms.

Another limitation of this work is that it is currently limited to the simulation stage and has not yet been tested with clinical data. A key challenge in extending the approach to real clinical data is the difficulty in obtaining training pairs of motion-corrupted images and ground truth final-position images for patients. Further

investigation into the interpretability of the network and its generalization to out-of-distribution (OOD) data can help determine the model's robustness and potential to perform well on diverse datasets. Furthermore, we suggest that the forthcoming practical application can be divided into two primary aspects: Firstly, V&V (verification and validation) of the motion-corrupted image simulator is essential. For a specific facility with a specific MR acquisition sequence, it is crucial to ensure that the simulation is consistent with the machine, for example, featuring the same imaging latency and spatial resolution. This can be achieved with a motion phantom study where the imaging latency and noise are quantified. Other k-space trajectories like radial or spiral, and MRI acceleration techniques like partial Fourier or parallel imaging methods could be explored: the effects of such techniques are yet to be quantified, especially for acceleration methods that imply sharing of k-space data across different coil images, which might impact on the overall latency and image quality. Secondly, the original sequence, that is, the real patient image acquisition without motion corruption (motion-corruption-free) should be available for training. Based on our findings, patients undergoing breath hold or a very shallow breathing mode have negligible intra-frame motion, which can be regarded as motion-corruption-free. Therefore, images should be extracted from these stages with varying inhale/exhale amplitudes, for instance, patients or volunteers can be instructed to hold their breath at the middle/end of inhale or exhale; then with the presented motion pattern perturbation scheme, it is possible to create the library of intra-frame motion trajectories for training. More suitable alternatives would include anthropomorphic physical motion phantoms or other relatively complex phantoms such as the porcine lung phantom.[52] Moreover, transfer learning such as patient-specific learning[28] could be investigated.

Further research is necessary to explore the potential benefits of intra-frame motion compensation for improving the accuracy of the downstream tasks in MRgRT such as beam gating, motion prediction[53] or real-time tumor tracking. Following the correction of earlier acquired data, deep-learning-based image auto-segmentation or reconstruction may see notable improvements, particularly for models that rely heavily on higher frequency components.[54,55] Recent studies have shown that the prediction of tumor contours on cine MRI frames using a convolutional LSTM is challenging.[56] Our intra-frame motion compensation model works in 2D, providing an efficient time latency offset for anatomical structure changes, and the network can be trained to estimate the complete intra-frame motion trajectory by generating output images not only at the final position but also at intermediate positions, offering potential information for 2D motion prediction algorithms. The intra-frame motion compensation could bypass the issue of inadequate temporal resolution in MR imaging, a generalization of this technique from 2D+t cine MR to time-resolved volumetric (3D+t) MR[8,57] might be investigated in the future.

5 | CONCLUSIONS

In this work, we have shown that the intra-frame motion deterioration effects can be noteworthy in MRgRT, resulting in pronounced imaging latency and the subsequent errors in GTV contouring. Based on the presented motion pattern perturbation scheme, a U-Net with three types of loss functions was successfully trained to estimate the exact noiseless final-position image. The proposed models led to an evident image quality and GTV position accuracy enhancement in simulated images compared to the motion-corrupted image, confirmed by a decreased image MSE/MAE and an improvement in terms of GTV DSC and HD_{95}.

ACKNOWLEDGMENTS

This work was funded by the China Scholarship Council within the joint LMU-CSC program.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The Department of Radiation Oncology of the University Hospital of LMU Munich has research agreements with Brainlab, Elekta and ViewRay.

REFERENCES

1. Zou W, Dong L, Kevin Teo BK. Current state of image guidance in radiation oncology: implications for PTV margin expansion and adaptive therapy. *Semin Radiat Oncol*. 2018;28(3):238-247. 10.1016/j.semradonc.2018.02.008
2. Liney GP, Whelan B, Oborn B, Barton M, Keall P. MRI-linear accelerator radiotherapy systems. *Clin Oncol (R Coll Radiol)*. 2018;30(11):686-691. 10.1016/j.clon.2018.08.003
3. Winkel D, Bol GH, Kroon PS, et al. Adaptive radiotherapy: the elekta unity MR-linac concept. *Clin Transl Radiat Oncol*. 2019;18:54-59. 10.1016/j.ctro.2019.04.001
4. Kurz C, Buizza G, Landry G, et al. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol*. 2020;15(1):93. 10.1186/s13014-020-01524-4
5. Kluter S. Technical design and concept of a 0.35 T MR-Linac. *Clin Transl Radiat Oncol*. 2019;18:98-101. 10.1016/j.ctro.2019.04.007
6. Corradini S, Alongi F, Andratschke N, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiat Oncol*. 2019;14(1):92. 10.1186/s13014-019-1308-y
7. Kim T, Lewis B, Lotey R, Barberi E, Green O. Clinical experience of MRI(4D) QUASAR motion phantom for latency measurements in 0.35T MR-LINAC. *J Appl Clin Med Phys*. 2021;22(1):128-136. 10.1002/acm2.13118
8. von Siebenthal M, Szekely G, Gamper U, Boesiger P, Lomax A, Cattin P. 4D MR imaging of respiratory organ motion and its variability. *Phys Med Biol*. 2007;52(6):1547-1564. 10.1088/0031-9155/52/6/001
9. Glitzner M, Woodhead PL, Borman PTS, Lagendijk JJW, Raaymakers BW. Technical note: MLC-tracking performance on

the Elekta unity MRI-linac. *Phys Med Biol*. 2019;64(15):15NT02. 10.1088/1361-6560/ab2667
10. Green OL, Rankine LJ, Cai B, et al. First clinical implementation of real-time, real anatomy tracking and radiation beam control. *Med Phys*. 2018;45(8):3728-3740. 10.1002/mp.13002
11. Chun J, Lewis B, Ji Z, et al. Evaluation of super-resolution on 50 pancreatic cancer patients with real-time cine MRI from 0.35T MRgRT. *Biomed Phys Eng Express*. 2021;7(5):055020. 10.1088/2057-1976/ac1c51
12. Jackson S, Glitzner M, Tijssen RHN, Raaymakers BW. MRI B (0) homogeneity and geometric distortion with continuous linac gantry rotation on an Elekta Unity MR-linac. *Phys Med Biol*. 2019;64(12):12NT01. 10.1088/1361-6560/ab231a
13. Dorsch S, Mann P, Elter A, et al. Measurement of isocenter alignment accuracy and image distortion of an 0.35 T MR-Linac system. *Phys Med Biol*. 2019;64(20):205011. 10.1088/1361-6560/ab4540
14. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: a complex problem with many partial solutions. *J Magn Reson Imaging*. 2015;42(4):887-901. 10.1002/jmri.24850
15. Loot K. *Measurement of the modulation transfer function of a digital radiographic system in consideration of object motion*. Heinrich-Heine-Universität Düsseldorf; 2014.
16. Friedman SN, Cunningham IA. A moving slanted-edge method to measure the temporal modulation transfer function of fluoroscopic systems. *Med Phys*. 2008;35(6):2473-2484. 10.1118/1.2919724
17. Rowlands JA. Videofluorography: the role of temporal averaging. *Med Phys*. 1984;11(2):129-136. 10.1118/1.595488
18. Borman PTS, Tijssen RHN, Bos C, Moonen CTW, Raaymakers BW, Glitzner M. Characterization of imaging latency for real-time MRI-guided radiotherapy. *Phys Med Biol*. 2018;63(15):155023. 10.1088/1361-6560/aad2b7
19. Liu PZY, Dong B, Nguyen DT, et al. First experimental investigation of simultaneously tracking two independently moving targets on an MRI-linac using real-time MRI and MLC tracking. *Med Phys*. 2020;47(12):6440-6449. 10.1002/mp.14536
20. Glitzner M, Woodhead P, Borman P, Lagendijk J, Raaymakers B. MLC-tracking performance on the Elekta unity MRI-linac. *Phys Med Biol*. 2019;64(15):15NT02.
21. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys*. 2006;33(10):3874-3900. 10.1118/1.2349696
22. Shirato H, Suzuki K, Sharp GC, et al. Speed and amplitude of lung tumor motion precisely detected in four-dimensional setup and in real-time tumor-tracking radiotherapy. *Int J Radiat Oncol Biol Phys*. 2006;64(4):1229-1236. 10.1016/j.ijrobp.2005.11.016
23. Shan L, Li X, Wang W, Decouple the High-Frequency and Low-Frequency Information of Images for Semantic Segmentation. 2021.
24. Li X, Li X, Zhang L, et al. *Improving semantic segmentation via decoupled body and edge supervision*. Springer International Publishing; 2020:435-452.
25. Ghodrati V, Bydder M, Ali F, et al. Retrospective respiratory motion correction in cardiac cine MRI reconstruction using adversarial autoencoder and unsupervised learning. *NMR Biomed*. 2021;34(2):e4433. 10.1002/nbm.4433
26. Johnson PM, Drangova M. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. *Magn Reson Med*. 2019;82(3):901-910. 10.1002/mrm.27772
27. Fu Y, Mazur TR, Wu X, et al. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. *Med Phys*. 2018;45(11):5129-5137. 10.1002/mp.13221
28. Kawula M, Hadi I, Nierer L, et al. Patient-specific transfer learning for auto-segmentation in adaptive 0.35 T MRgRT of prostate cancer: a bi-centric evaluation. *Med Phys*. 2023;50(3):1573-1585. 10.1002/mp.16056
29. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*. 2018;45(8):3627-3636. 10.1002/mp.13047
30. Shen C, Nguyen D, Chen L, et al. Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning. *Med Phys*. 2020;47(6):2329-2336. 10.1002/mp.14114
31. Cusumano D, Boldrini L, Dhont J, et al. Artificial Intelligence in magnetic Resonance guided Radiotherapy: medical and physical considerations on state of art and future perspectives. *Phys Med*. 2021;85:175-191. 10.1016/j.ejmp.2021.05.010
32. Rahsepar AA, Saybasili H, Ghasemiesfe A, et al. Motion-corrected real-time cine magnetic resonance imaging of the heart: initial clinical experience. *Invest Radiol*. 2018;53(1):35-44. 10.1097/RLI.0000000000000406
33. Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. *Pattern Recognit*. 2018;77:354-377. 10.1016/j.patcog.2017.10.013
34. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010;37(9):4902-4915. 10.1118/1.3480985
35. Paganelli C, Summers P, Gianoli C, Bellomi M, Baroni G, Riboldi M. A tool for validating MRI-guided strategies: a digital breathing CT/MRI phantom of the abdominal site. *Med Biol Eng Comput*. 2017;55(11):2001-2014. 10.1007/s11517-017-1646-6
36. Paganelli C, Portoso S, Garau N, et al. Time-resolved volumetric MRI in MRI-guided radiotherapy: an in silico comparative analysis. *Phys Med Biol*. 2019;64(18):185013. 10.1088/1361-6560/ab33e5
37. Paganelli C, Portoso S, Garau N, et al. Time-resolved volumetric MRI in MRI-guided radiotherapy: an in-silico comparative analysis. *Phys Med Biol*. 2019.
38. Batchelor PG, Atkinson D, Irarrazaval P, Hill DL, Hajnal J, Larkman D. Matrix description of general motion correction applied to multishot images. *Magn Reson Med*. 2005;54(5):1273-1280. 10.1002/mrm.20656
39. Song J, Liu Y, Gewalt SL, Cofer G, Johnson GA, Liu QH. Least-square NUFFT methods applied to 2-D and 3-D radially encoded MR image reconstruction. *IEEE Trans Biomed Eng*. 2009;56(4):1134-1142. 10.1109/TBME.2009.2012721
40. Sha L, Guo H, Song AW. An improved gridding method for spiral MRI using nonuniform fast Fourier transform. *J Magn Reson*. 2003;162(2):250-258. 10.1016/s1090-7807(03)00107-1
41. Griswold MA, Jakob PM, Nittka M, Goldfarb JW, Haase A. Partially parallel imaging with localized sensitivities (PILS). *Magn Reson Med*. 2000;44(4):602-609. 10.1002/1522-2594(200010)44:4<602::aid-mrm14>3.0.co;2-5
42. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002;47(6):1202-1210. 10.1002/mrm.10171
43. Wedel A, Pock T, Zach C, Bischof H, Cremers D. *An improved algorithm for TV-L1 optical flow*. Springer; 2009:23-45.
44. Ronneberger O, Fischer P, Brox T. *U-Net: convolutional networks for biomedical image segmentation*. Springer International Publishing; 2015:234-241.
45. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408-1419. 10.1002/mp.12155
46. Singh NM, Iglesias JE, Adalsteinsson E, Dalca AV, Golland P. Joint frequency and image space learning for MRI reconstruction and analysis. *J Mach Learn Biomed Imaging*. 2022;2022:018.
47. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32:8026-8037.
48. Smilkov D, Thorat N, Kim B, Viégas F, Wattenberg M, Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:170603825*. 2017.

49. Ehrbar S, Braga Kaser S, Chamberlain M, et al. MR-guided beam gating: residual motion, gating efficiency and dose reconstruction for stereotactic treatments of the liver and lung. *Radiother Oncol*. 2022;174:101-108. 10.1016/j.radonc.2022.07.007
50. Riederer SJ, Tasciyan T, Farzaneh F, Lee JN, Wright RC, Herfkens RJ. MR fluoroscopy: technical feasibility. *Magn Reson Med*. 1988;8(1):1-15. 10.1002/mrm.1910080102
51. Dhengre N, Sinha S. Multiscale U-net-based accelerated magnetic resonance imaging reconstruction. *Signal, Image and Video Processing*. 2022;16(4):881-888. 10.1007/s11760-021-02030-0
52. Rabe M, Paganelli C, Riboldi M, et al. Porcine lung phantom-based validation of estimated 4D-MRI using orthogonal cine imaging for low-field MR-Linacs. *Phys Med Biol*. 2021;66(5):055006. 10.1088/1361-6560/abc937
53. Lombardo E, Rabe M, Xiong Y, et al. Offline and online LSTM networks for respiratory motion prediction in MR-guided radiotherapy. *Phys Med Biol*. 2022;67(9). 10.1088/1361-6560/ac60b7
54. He Z, Hong K, Zhou J, Liang D, Wang Y, Liu Q. Deep frequency-recurrent priors for inverse imaging reconstruction. *Signal Process*. 2022;190:108320.
55. Hu S, Liao Z, Xia Y. *Domain specific convolution and high frequency reconstruction based unsupervised domain adaptation for medical image segmentation*. Springer; 2022:650-659.
56. Lombardo E, Rabe M, Xiong Y, et al. Evaluation of real-time tumor contour prediction using LSTM networks for MR-guided radiotherapy. *Radiother Oncol*. 2023;182:109555. 10.1016/j.radonc.2023.109555
57. Han F, Zhou Z, Du D, et al. Respiratory motion-resolved, self-gated 4D-MRI using Rotating Cartesian K-space (ROCK): initial clinical experience on an MRI-guided radiotherapy system. *Radiother Oncol*. 2018;127(3):467-473. 10.1016/j.radonc.2018.04.029

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sui Z, Palaniappan P, Brenner J, et al. Intra-frame motion deterioration effects and deep-learning-based compensation in MR-guided radiotherapy. *Med Phys*. 2023;1-19. https://doi.org/10.1002/mp.16702