Physics in Medicine & Biology

PAPER

Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR)

You Zhang1,* Hua-Chieh Shao1, Tinsu Pan2 and Tielige Mengke1

1 Advanced Imaging and Informatics in Radiation Therapy (AIRT) Laboratory, Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX, 75235, United States of America
2 Department of Imaging Physics, University of Texas MD Anderson Cancer Center, Houston, TX, 77030, United States of America

* Author to whom any correspondence should be addressed.

E-mail: you.zhang@utsouthwestern.edu

Keywords: image reconstruction, dynamic imaging, cone-beam CT, implicit neural representation, principal component analysis, motion modeling

Abstract

Objective. Dynamic cone-beam CT (CBCT) imaging is highly desired in image-guided radiation therapy to provide volumetric images with high spatial and temporal resolutions to enable applications including tumor motion tracking/prediction and intra-delivery dose calculation/accumulation. However, dynamic CBCT reconstruction is a substantially challenging spatiotemporal inverse problem, due to the extremely limited projection sample available for each CBCT reconstruction (one projection for one CBCT volume). Approach. We developed a simultaneous spatial and temporal implicit neural representation (STINR) method for dynamic CBCT reconstruction. STINR mapped the unknown image and the evolution of its motion into spatial and temporal multi-layer perceptrons (MLPs), and iteratively optimized the neuron weightings of the MLPs via acquired projections to represent the dynamic CBCT series. In addition to the MLPs, we also introduced prior knowledge, in the form of principal component analysis (PCA)-based patient-specific motion models, to reduce the complexity of the temporal mapping to address the ill-conditioned dynamic CBCT reconstruction problem. We used the extended-cardiac-torso (XCAT) phantom and a patient 4D-CBCT dataset to simulate different lung motion scenarios to evaluate STINR. The scenarios contain motion variations including motion baseline shifts, motion amplitude/frequency variations, and motion non-periodicity. The XCAT scenarios also contain inter-scan anatomical variations including tumor shrinkage and tumor position change. Main results. STINR shows consistently higher image reconstruction and motion tracking accuracy than a traditional PCA-based method and a polynomial-fitting-based neural representation method. STINR tracks the lung target to an average center-of-mass error of 1–2 mm, with corresponding relative errors of reconstructed dynamic CBCTs around 10%. Significance. STINR offers a general framework allowing accurate dynamic CBCT reconstruction for image-guided radiotherapy. It is a one-shot learning method that does not rely on pre-training and is not susceptible to generalizability issues. It also allows natural super-resolution. It can be readily applied to other imaging modalities as well.

1. Introduction

X-ray computed tomography (CT) is widely used in radiotherapy practices, providing volumetric images of high spatial resolution and geometric accuracy to guide radiotherapy planning and delivery (Jaffray et al 2002, Pan et al 2004, Létourneau et al 2005, Pereira et al 2014). Modern radiotherapy linear accelerators (LINACs) are commonly equipped with onboard x-ray imaging sources and flat panel detectors, which can acquire pre-delivery cone-beam CTs (CBCTs) for patient setup, plan adaptation, and dose accumulation (Borst et al 2007, Topolnjak et al 2010, Zhang et al 2015a, Kong et al 2016, Sibolt et al 2021); and intra-delivery CBCTs for
treatment positioning verification (Li et al. 2019). To acquire a fully-sampled CBCT image, the LINAC gantry needs to rotate at least 200° for a full-fan acquisition, or 360° when the detector is offset to increase the axial field-of-view (half-fan mode) (Song et al. 2008). Considering potential collision risks between the gantry and the patients, currently, the gantry rotation speed is mostly limited to 6° per second (s), which requires a substantial image acquisition time on the order of 1 min. For cardiac and respiratory motion-impacted anatomical sites including thoracic and upper abdominal regions, the slow imaging speed results in CBCTs being affected by motion artifacts, which manifest as blurred anatomies and poorly defined structure boundaries (Sonke et al. 2005). Such artifacts introduce substantial uncertainties to the localization of moving tumors and surrounding organs-at-risk (OARs) for radiotherapy planning and treatment. The motion-blurred CBCTs fail to capture the motion trajectories of the anatomies, and may substantially underestimate and under-dose the radiotherapy target volume (Vergalasova et al. 2011). To address the motion challenges, respiratory-correlated CBCT, also named four-dimensional CBCT (4D-CBCT), has been developed (Sonke et al. 2005, Sweeney et al. 2012, Zhang et al. 2013, 2015b, Thengumpallil et al. 2016). The 4D-CBCT technique assigns each acquired cone-beam projection to a respiratory phase based on tracked surrogate motion signals (surface motion, diaphragm motion, etc), with each phase corresponding to a motion state along an assumed periodic motion cycle. It then reconstructs a semi-static CBCT volume at each phase bin, and the CBCT volumes from all bins are stacked to represent the motion kinematics during the nominal, averaged motion cycle. To address the under-sampling issues caused by the retrospective phase sorting, the 4D-CBCT phase number is usually limited to $<10$. In addition, the cone-beam projections are often intentionally over-sampled in number at a cost of imaging dose and scan time, to ensure an adequate amount of projections exist in each phase bin after phase sorting (Li and Xing 2007, Bergner et al. 2009, Thengumpallil et al. 2016). To reduce the imaging dose and scan time, different reconstruction algorithms, based on various a priori assumptions and motion models, were also developed to use limited projections within each phase bin to reconstruct high-quality, artifact-free 4D-CBCT images (Leng et al. 2008, Wang and Gu 2013, Zhang et al. 2013, Yan et al. 2014, Zhang et al. 2015b, Harris et al. 2017, Shieh et al. 2019, Huang et al. 2020).

However, the 4D-CBCT imaging technique is essentially built upon the assumption that anatomical motion is periodical and regular, such that the projections acquired at different angles and time stamps can be sorted into the same phase bin. Although the motion of the thoracic and upper abdominal regions of real patients presents cardiac and pulmonary function-related periodicity, irregular and non-periodic motions, like those with amplitude/frequency variations or baseline shifts, are commonly observed as well (Huang et al. 2010, Clements et al. 2013, Pan et al. 2019, Yasue et al. 2022). Such irregularity may lead to substantial intra-phase motion variations and strong residual motion artifacts after sorting (Cooper et al. 2015). The nominal cycle resolved by 4D-CBCT fails to capture the irregularity and non-periodicity, which may provide crucial information on motion statistics and trends to guide patient immobilization, set-up, and treatment monitoring (Poulsen et al. 2014, Li et al. 2018). The ultimate solution to such a challenge is time-resolved CBCT imaging, or dynamic CBCT (Li et al. 2010, Cai et al. 2014, Gao et al. 2018, Jailin et al. 2021). Dynamic CBCT, in contrast to the phase-resolved 4D-CBCT, reconstructs a continuous time series of volumetric images reflecting the spatial and temporal kinematics of patient anatomy without the phase-binning process. Dynamic CBCT essentially treats each CBCT projection as an individual phase and reconstructs a CBCT volume out of every single projection. However, the extreme under-sampling challenges the current reconstruction methods, as they require at least tens or hundreds of projections spanning over a large scan angle to reconstruct a high-quality volume. Some previous studies tried to address the ill-posed, spatiotemporal CBCT reconstruction problem via different strategies. Cai et al. introduced low-rank matrix factorization into solving the dynamic CBCT, by viewing each temporal CBCT volume as a linear combination of a few image basis (Cai et al. 2014). The linear coefficients and the basis images were solved simultaneously under a pre-defined matrix rank number (20). However, the study only reconstructed a single CBCT slice rather than the full 3D volume and only evaluated regular breathing scenarios. The low-rank assumption and the chosen rank number also remain to be further validated. Gao et al. viewed the 4D CBCT sequences as a product of spatial principal components and temporal motion coefficients (Gao et al. 2018). Instead of solving the temporal motion coefficients directly from the angle-varying CBCT projections, their method proposes to learn the temporal motion coefficients from a previously-acquired 2D fluoroscopy sequence at a fixed gantry angle. The learned a priori temporal motion coefficients from 2D projections were found to improve the CBCT reconstruction accuracy by reducing the degree of freedom in the spatiotemporal inverse problem. However, the described method relies on a motion trajectory learned from fixed-angle 2D projections, which may fail to represent the complex 3D motion and motion variations that occurred in the following CBCT acquisition. Taking 2D fluoroscopy images at fixed angles also incurs additional costs of imaging time and dose. The solved CBCT images are not fully time-resolved but are limited to 50 phases as well. Another study tried to solve dynamic CBCTs by combining projection-based motion estimation and motion-compensated reconstruction (Jailin et al. 2021). The method models the time kinematics via a series of time functions including surrogate motion signals. The motion irregularity and the poor representation/
correlation of surrogate signals, however, may render the time regularization less effective and lead to motion estimation errors.

Another category of methods introduces prior CT/CBCT images into solving the dynamic, time-resolved CBCTs (Li et al 2010). They view each time-resolved CBCT as a deformed prior image via a deformation vector field (DVF). The principal component analysis (PCA) based method uses prior 4D-CT/4D-CBCT images to extract a patient-specific model of principal motion components and solves the DVF as a linear combination of the components (Li et al 2010, Zhang et al 2013). The substantial dimension reduction from PCA allows the linear coefficients to be solved from a single x-ray projection. However, a potentially major drawback of the pure DVF-driven CBCT estimation technique is that the variations between prior and new images may not be deformation alone (Zhang et al 2017). The shading changes from different acquisition hardware (fan-beam CT versus cone-beam CT), various imaging protocols, and distinct noise/scatter patterns lead to errors when solving the motion fields. Non-deformation-induced anatomical changes and intensity variations cannot be recovered by the DVFs (Zhang et al 2017). Inter-scan deformations, such as tumor shrinkage, may not be captured by an intra-scan motion model like PCA either. A deep learning-based method was also developed to map cone-beam projections directly to PCA coefficients without explicit optimization (Wei et al 2020). It however suffers from similar issues as the conventional PCA-based techniques. Recently, another deep learning-based technique was proposed to directly convert single 2D projections into 3D volumes via a patient-specific encoder-decoder framework (Shen et al 2019). However, the 2D to 3D conversion technique is extremely ill-conditioned and its performance can be unstable to image intensity variations due to shading changes or noises. It also requires a model to be pre-trained for each patient and each scan angle and thus is more intended for reconstructions from a fixed scan angle rather than a rotating scan geometry of the normal CBCT imaging.

Recently, implicit neural representation (INR) learning has gathered much interest in the artificial intelligence field (Esrami et al 2018, Sitzmann et al 2020, Peng et al 2021). INR uses the power of neural networks, mostly multi-layer perceptrons (MLPs) (Heidari et al 2020), to construct and map complex objects including natural structures and medical images into continuous and differentiable functions. The MLPs can accept query inputs, for instance, the coordinates of image voxels, and output the physical properties like image intensities at queried voxels, to implicitly represent a complex medical image without specifying the details of the constitutive functions in advance. It offers a new way to reconstruct and represent volumetric objects and has recently been applied toward novel view synthesis, CT/MR reconstructions, and dose map compression (Lombardi et al 2019, Sitzmann et al 2019, Shen et al 2022, Vasudevan et al 2022). A recent study also tried to use the INR to reconstruct a reference fan-beam CT volume, while using polynomial-based motion fields to generate dynamic CT images by deforming the reference CT volume (Reed et al 2021). In this study, we proposed to use the representation capability of INR to develop a new dynamic CBCT reconstruction technique via simultaneous spatial and temporal INR learning (STINR). By STINR, we decoupled the complex spatiotemporal inverse problem of dynamic CBCT reconstruction into solving a spatial INR to represent a reference CBCT image, and several temporal INRs to represent the DVFs that characterize the time-resolved motion along different Cartesian directions. To reduce the complexity and leverage the inherent redundancy of DVFs, STINR combined PCA-based motion modeling with INR-based PCA coefficient learning to represent complex motion characteristics observed in each angle-variation projection. Compared to conventional machine/deep learning methods, STINR is a ‘one-shot’ learning technique which directly uses available cone-beam projection data to construct a patient-specific spatial and temporal imaging model that fits the specific projection set. In other words, the dynamic CBCT sequence is encoded by STINR as a neural network, which is solved on-the-fly in a self-supervised fashion with no prior training and no ‘ground-truth’ dynamic CBCTs required. Correspondingly, STINR does not suffer from the generalizability issues encountered by conventional pre-trained deep learning models (Reed et al 2021). In this study, we used the extended-cardiac-torso (XCAT) phantom and a patient 4D-CBCT dataset to simulate dynamic volumetric images and projections of lung patients (Segars et al 2010), featuring different regular and irregular motion scenarios, including motion amplitude/frequency variations, motion baseline shifts, and non-periodical motion. We also simulated different anatomical variation scenarios in the XCAT study to represent inter-scan deformation, including tumor size shrinkage and tumor position change. We used STINR to reconstruct dynamic lung CBCTs, which were compared with the known ‘ground-truth’ lung CBCTs from the simulations. We compared STINR with the conventional PCA-based method (Li et al 2010, Zhang et al 2013) and the polynomial fitting-based INR method (Reed et al 2021).

2. Materials and methods

STINR decoupled the dynamic CBCT reconstruction problem into the reconstruction of a reference CBCT volume \( \text{CBCT}_{0} \), and the simultaneous motion estimation by solving time-resolved DVFs \( \text{D}(t) \) to deform \( \text{CBCT}_{0} \) to dynamic CBCTs \( \text{CBCT}_{t} \) at each time frame:
\[ \text{CBCT}_{\text{dyn}} = \text{CBCT}_{\text{Ref}}(x + D(t)), \]  

(1)

where \( x \) denotes the voxel coordinates of \( \text{CBCT}_{\text{dyn}} \), which were mapped to those of \( \text{CBCT}_{\text{Ref}} \) through time-varying \( D(t) \) by trilinear interpolation (Bourke 1999). Previous studies have found using a single MLP to map images both spatially and temporally to be particularly challenging (Shrestha and Hirano 2021), which yielded inaccurate results.

STINR breaks a spatiotemporal INR into specialized partial INRs to represent the spatial volume and temporal kinematics independently, which reduces the overall complexity of the network and allows each component to be customized to fit the representation needs:

\[ \text{CBCT}_{\text{Ref}} = \Phi^t(\theta) \]  

(2)

\[ D(t) = \overline{\text{PC}}_{\text{dim},0} + \Phi_{\text{dim},n}(\varphi) \ast \overline{\text{PC}}_{\text{dim},n}, \text{dim} = i^1, i^2, i^3. \]  

(3)

As shown in equation (2), the reference CBCT volume was represented by spatial INR \( \Phi^t \) parameterized by the to-be-optimized coefficients \( \theta \). In equation (3), \( \text{dim} \) indicates the three Cartesian directions \( (i^1, i^2, i^3) \). Along each direction, a deformation matrix was defined. \( n \) indicates the number of principal motion components along each dimension to model the deformation matrix. In this study, we used \( n = 3 \) as they were found sufficient to model the lung motion (Li et al 2011), while more can be readily applied. The temporal DVF\( s \) deformation matrices) were constructed as scaled principal motion components \( (\overline{\text{PC}}_{\text{dim},n}) \) by time-varying coefficients learned and represented as the temporal INRs \( \Phi_{\text{dim},n} \), which were parameterized by the to-be-optimized coefficients \( \varphi \). \( \overline{\text{PC}}_{\text{dim},0} \) denotes the average DVF extracted from the PCA (Zhang et al 2013).
By decoupling the dynamic CBCT reconstruction problem into solving \( \text{CBCT}_{\text{Ref}} \) and \( D(t) \) separately, we reduced the complexity of the spatiotemporal inverse problem. The use of PCA-based motion modeling also introduces prior knowledge into solving the intra-scan motion, which helps to address the challenges of extreme under-sampling in dynamic CBCT reconstruction. By reconstructing \( \text{CBCT}_{\text{Ref}} \) directly from onboard cone-beam projections, we avoided the challenges of shading mismatches and non-deformation-induced changes as encountered by methods that directly use prior images for registration (Li et al. 2016, Zhang et al. 2013, Wei et al. 2020). The PCA motion model is only used to represent intra-scan motion while the inter-scan motion/deformation is implicitly solved via \( \text{CBCT}_{\text{Ref}} \). The STINR thus enjoys unique advantages as compared to previously described methods. Below we first introduced the INR learning of \( \text{CBCT}_{\text{Ref}} \), which is followed by the details of the INR learning of \( D(t) \) that maps \( \text{CBCT}_{\text{Ref}} \) into each temporal frame.

### 2.1. Details of implicit neural representation learning of \( \text{CBCT}_{\text{Ref}} \)

#### 2.1.1. General methodology

As shown in equation (2) and figure 1, the INR-based reconstruction solves a function \( \Phi^t \) to represent \( \text{CBCT}_{\text{Ref}} \), in the form of a multi-layer perceptron. The MLP maps the 3-dimensional query voxel coordinates \( x_i \) to their intensity distribution \( \sigma(x_i) \in \mathbb{R}^3 \), where \( i \) indicates the voxel number. The MLP-based function \( \Phi^t \) is continuous, differentiable, and not limited to a specific voxel spatial resolution. It can be further defined as:

\[
\Phi^t(x_i|\theta) = \tilde{\sigma}(x_i), \quad x_i \in [-1, 1]^3, \quad \tilde{\sigma}(x_i) \in \mathbb{R}^3,
\]

where \( \tilde{\sigma}(x_i) \) denotes the output of the mapping function \( \Phi^t \), which serves as the reconstruction and approximation of the true target property \( \sigma(x_i) \). For the CBCT reconstruction problem, \( \sigma(x_i) \) represents the attenuation coefficients of the scanned patient volume. The coordinates \( x_i \) were normalized to the range \([-1, 1]\) along each Cartesian direction, which was found to improve the INR learning accuracy (Vasudevan et al. 2022).

#### 2.1.2. Fourier feature encoding for the query coordinates

The vanilla coordinates-based MLPs were found difficult to learn high-frequency functions, as the neural tangent theory suggests the MLPs resemble kernels with rapid falloffs at high-frequency regions (Basri et al. 2020, Tancik et al. 2020). To capture the high-frequency features in the images, the input query coordinates \( x_i \) can be encoded by a large set of scalar functions before feeding into the MLP. Sinusoidal functions are commonly used to encode the query coordinates with Fourier features to fit the high-frequency signals. In this study, we used random Fourier feature (GRFF) encoding (Tancik et al. 2020) which maps the input coordinates vector \( x_i \) as:

\[
\gamma(x_i) = [\sin(2\pi B x_i), \cos(2\pi B x_i)],
\]

where the matrix \( B \sim \mathcal{N}(0, \sigma^2) \) is randomly sampled from a Gaussian distribution with width \( \sigma \), which was determined empirically. In our implementation, we used 128 Fourier features for each input coordinate, with \( \sigma = 2.5 \) for the XCAT study and \( \sigma = 4 \) for the patient study. The encoded coordinates \( \gamma(x_i) \) were fed subsequently into the MLP.

#### 2.1.3. Solving MLP for \( \text{CBCT}_{\text{Ref}} \)

The function of the MLP \( \Phi^t \) is to map the coordinates \( x_i \) to the true image intensity \( \sigma(x_i) \) of \( \text{CBCT}_{\text{Ref}} \), such that

\[
\Phi^t(x_i|\theta) = \tilde{\sigma}(x_i) = \sigma(x_i).
\]

Based on equation (6), the MLP parameters \( \theta \) can be solved by minimizing a loss function defined as:

\[
\theta = \arg\min_{\theta} L \left( \sum_{i=1}^{N} \Phi^t(x_i|\theta), \sigma(x_i) \right),
\]

where \( L \) indicates the loss function between the reconstructed CBCT and the true CBCT. \( N \) denotes all voxels within the CBCT volume. For reconstruction, the true attenuation coefficient map \( \sigma(x_i) \) is not available to directly optimize the MLP, as only x-ray projections are provided. We can first reconstruct the x-ray projections into CBCT volumes using conventional analytical algorithms like the Feldkamp–Davis–Kress (FDK) algorithm (Feldkamp et al. 1984), or other iterative algorithms (Andersen and Kak 1984, Wang et al. 2009), and use the reconstructed volumes to replace \( \sigma(x_i) \). Alternatively, the data fidelity can be optimized through a loss function directly defined on cone-beam projections like conventional iterative reconstruction algorithms:

\[
\theta = \arg\min_{\theta} P \left( \sum_{i=1}^{N} A \Phi^t(x_i|\theta), P \right),
\]

\( P \) denotes the acquired cone-beam projections. \( A \) denotes the system matrix which generates cone-beam projections from the CBCT volume, with the acquisition geometry identical to \( P \). The loss function \( L \) measures the distance between the forward projections of the INR-reconstructed CBCT volume and the true projections. In this study, we used the sum of squared differences as the distance metric.
\[ \theta = \arg \min_{\theta} \left\| \sum_{i=1}^{N} \Phi(x_i|\theta) - F \right\|_2^2. \]  

The parameters \( \theta \) of the MLP can be conveniently optimized by minimizing the loss function. In our implementation, the MLP for \( CBCT_{\text{ref}} \) was constructed as four layers, with each layer containing 256 neurons. Except for the last layer, each layer was followed by a Swish activation function (Ramachandran et al. 2017):

\[ \text{actv}(y) = \frac{y \cdot \text{sigmoid}(y)}{1 + e^y}. \]  

We compared Swish against ReLU (Agarap 2018) and Siren (Sitzmann et al. 2020), and found Swish provided slightly better results, which was chosen as the activation function in this study.

The above reconstructions and INR learning implicitly assume that the cone-beam projections \( P \) contain only a static patient volume without anatomical motion. For dynamic CBCT projections \( P_i \), the underlying anatomy varies with time due to the physiological motion. Directly using these projections to reconstruct the spatial INR \( \Phi^e \) will lead to motion artifacts-compromised images. To address this issue, the intra-scan DVFs \( D(t) \), represented via the temporal INRs \( \Phi^f \), are needed (figure 2).

### 2.2. Details of implicit neural representation learning of \( D(t) \)

Incorporating the intra-scan motion turns equation (9) into:

\[ \theta, \varphi = \arg \min_{\theta, \varphi} \sum_t \left\| A \Phi^t(x + D(t) | \theta, \varphi) - P_t \right\|_2^2 \]

\[ = \arg \min_{\theta, \varphi} \sum_t \left\| A \Phi^t(x + \tilde{P}C_{\text{dim}, 0} + \Phi^f_{\text{dim}, n} + \tilde{P}C_{\text{dim}, 0} | \theta, \varphi) - P_t \right\|_2^2. \]  

In equation (11), we removed the subscript \( i \) from \( x_i \) to simplify the notation. With the principal components available, we can reconstruct \( D(t) \) by estimating the weightings from the onboard projections and representing the weightings via the temporal INRs \( \Phi^f_{\text{dim}, n} \). To obtain the principal components, we followed the previous works by performing inter-phase deformable registration within a planning 4D-CT volume (Li et al. 2010, Zhang et al. 2013). In radiotherapy, 4D-CTs are routinely acquired for sites impacted by respiratory motion and are widely available to provide high-quality prior knowledge (Pan et al. 2004). We used the end-expiration (EE) phase volume as the reference volume and deformed it to the other phases to extract the inter-phase DVFs. The EE phase was selected due to its relative stability (Agarap 2018) and limited intra-phase motion.

The registration was performed using the open-source Elastix toolbox, of which the accuracy has been validated in many previous publications (Klein et al. 2010). From these inter-phase DVFs, the eigenvectors (principal components) were extracted as \( \tilde{P}C_{\text{dim}, n} \) along each Cartesian dimension \( (v^1, v^2, v^3) \). As previously mentioned, \( \tilde{P}C_{\text{dim}, 0} \) was also extracted as the average of the inter-phase DVFs, denoting the DC component of the motion. For \( n \), the first three PCA eigenvectors corresponding to the largest three eigenvalues were used since they are the most de-correlated and proved sufficient to model the lung motion (Ruan and Keall 2010).

In contrast to the spatial INR \( \Phi^e \) which was fitted by one MLP, the temporal INRs \( \Phi^f_{\text{dim}, n} \) were composed of 9 sub-MLPs, each of which represented one weighting for the three principal motion components along the three Cartesian directions. The temporal input, \( t \), was normalized to [0, 1] and encoded by the GRFF (equation (5)), like the spatial coordinate input, to enhance the INRs’ capability to map high-frequency motion variations. Similar to the encoding used for spatial coordinates, we used 128 Fourier features with \( \sigma = 2.5 \). Since the temporal MLP only takes \( t \) as input and the representation complexity is lower than that of the spatial MLP, each temporal MLP used 3 layers, with the first layer composed of 256 neurons, and the subsequent layers of 100 neurons each.

### 2.3. The detailed workflow of STINR

Minimizing the objective function in equation (11) solves the spatiotemporal reconstruction problem of dynamic CBCTs. To further accelerate the INR learning process, and to reduce the possibility of the optimization being trapped at a local optimum for the ill-posed problem, we took a three-stage approach as shown in figure 3: (1) We initialized the reference volume INR \( \Phi^e \) using a CBCT volume directly reconstructed by the FDK algorithm. Since the PCA motion model in Sec. 2.2 was derived based on the EE phase of the 4D-CT volume, we extracted the EE phase cone-beam projections from the full onboard projection set (figure 3) and reconstructed a coarse FDK volume. Using a reconstructed CBCT volume to directly solve the spatial INR \( \Phi^e \) avoids the iterative forward-and-backward projection process and quickly pre-conditions the INR. (2) We further fine-tuned the reference volume INR \( \Phi^e \) solved in (1) using the extracted cone-beam projections at the EE phase. To train the INR to better fit the true \( CBCT_{\text{ref}} \), this stage directly uses the EE projections to fine-tune the learned representation. Using projections directly can remove the artifacts introduced from the FDK reconstruction process. In both stage (1) and stage (2), the motion model was not introduced, assuming no
intra-phase motion. In stage (3), the temporal INRs, along with the corresponding DVFs, were introduced into mapping the CBCT$_{Ref}$ to the dynamic CBCT$_{dyn}$. Digitally reconstructed radiographs (DRRs) were projected from CBCT$_{dyn}$ and compared with the acquired cone-beam projections $P_t$ to assess the representation learning loss (equation (11)). At this stage, all the acquired cone-beam projections were used to compute and optimize the loss function. The spatial INR $\Phi^s$ and the temporal INRs $\Phi^t$ were optimized jointly and simultaneously during this stage. Through stage (3), the reference CBCT volume representation $\Phi^v$ was further corrected, to remove potential errors in stage (1) and stage (2) that assume no intra-phase motion. The time kinematics, as represented by $\Phi^v$, were fitted in free-form via the temporal INR learning towards all kinds of motion trajectories and scenarios.

In this study, we implemented the overall STINR framework based on the Pytorch backend (ver. 1.11.0). The optimization was performed automatically through the Pytorch framework. The initial learning rate was set to 0.002 for all the INRs. We used 500 iteration steps for stage 1, 500 iteration steps for stage 2, and 4000 iteration steps for stage 3.

---

### Table 1. Details of the eight simulated motion/anatomy scenarios.

| Scenarios | Motion/anatomy features |
|-----------|-------------------------|
| S1        | Small motion baseline shift |
| S2        | Large motion baseline shift |
| S3        | Inter-scan motion frequency variation (from baseline) |
| S4        | Motion amplitude and baseline variations |
| S5        | Simultaneous motion frequency/amplitude variations |
| S6        | Non-periodic motion (or fast gantry rotation) |
| S7        | Same as S1 but with the tumor diameter reduced by 50% from prior |
| S8        | Same as S1 but with the tumor shifted from the prior position by 6 mm in each Cartesian direction |
2.4. Experimental design
2.4.1. XCAT digital phantom study—data curation
To quantitatively evaluate the accuracy of the reconstructed dynamic CBCT volumes by STINR, in this study we used the 4D extended cardiac-torso (XCAT) phantom to simulate different breathing patterns and variations (Segars et al. 2010). With XCAT, the reconstructed CBCT volumes can be directly compared with the ‘ground-truth’ simulated images for evaluation. The tracked target motion by the dynamic CBCTs can also be directly compared against the simulated motion trajectories. We simulated a total of eight motion/anatomy scenarios, featuring different patterns and degrees of motion variations/irregularities and anatomical changes (table 1, figure 4).

The baseline motion curve used by the XCAT simulation was a sinusoidal curve with a 5 s cycle, which was used to generate the 10-phase prior 4D-CT for the PCA motion modeling. For the simulated lung patient, we inserted a 15-mm spherical tumor in radius into the lower lobe of the right lung. As shown in table 1 and figure 4, different onboard motion/anatomy variation scenarios during the CBCT acquisition were simulated, including motion baseline shift, motion frequency/amplitude variations, non-periodic motion, and tumor size and positional changes (simulating inter-scan anatomical variations). The non-periodic motion scenario (S6) is also similar to scenarios with an extremely fast gantry rotation speed, such that the motion captured is not repeated (non-periodic). The ‘ground-truth’, dynamic onboard CBCT volumes were simulated of 3.0 mm × 3.0 mm × 3.0 mm spatial resolution per voxel, and of 128 × 128 × 128 voxels in dimension. For each motion/anatomy scenario, a dynamic CBCT volume was simulated every $\frac{1}{11}$ s, to match with the common frame rate (11 frames s$^{-1}$) used in clinical cone-beam projection acquisition scenarios (Ling et al. 2011). We used a gantry rotation speed of 60° s$^{-1}$, which translates into a scan time of 60 s for a 360° scan angle. In total, 660
'ground-truth’ dynamic CBCT volumes were simulated for each motion/anatomy scenario. From each simulated dynamic volume, a corresponding cone-beam projection was simulated via the ray-tracing technique, using a gantry angle $\alpha$ defined in equation (12) based on the assumed gantry rotation speed and x-ray frame rate:

$$\alpha = \frac{1}{N} \times 6 \times (N - 1), \quad N = 1, 2, \ldots, 660.$$  \hfill (12)

Here $N$ denotes the projection frame number under simulation. The CBCT projection was simulated with $512 \times 512$ pixels, with each pixel measuring $1.17 \times 1.17$ mm in dimension. The source-to-detector distance was $1500$ mm and the source-to-isocenter distance was $1000$ mm, and the gantry rotation axis was defined along the superior–inferior direction.

To implement STINR (figure 3), the cone-beam forward-projection layer and FDK back-projection layer were realized by using the differentiable ASTRA projectors (van Aarle et al 2016) provided by the Operator Discretization Library (ODL) (Adler et al 2017). The computations were performed on an NVIDIA GeForce RTX 2080 super graphic processing unit (GPU) card with 8 GB memory. Due to the memory limit of the GPU card, the immediately-reconstructed CBCT volumes were down-sampled to $64 \times 64 \times 64$ in dimension during the optimization of STINR.

2.4.2. Comparison methods
To benchmark the STINR technique against other currently available methods, we also evaluated the reconstruction results of the conventional PCA-based, single-projection-driven CBCT estimation technique (PCA$_{cv}$) (Li et al 2010, Zhang et al 2013). The objective of the conventional PCA method was formulated as:

$$w_{\text{dim}, n}^t = \arg \min_{w_{\text{dim}, n}} \| A\mu(x + D(t)) - P_i \|^2$$

$$= \arg \min_{w_{\text{dim}, n}} \| A\mu(x + \mathbf{\bar{P}}_{\text{dim}, 0} + w_{\text{dim}, n} \circ \mathbf{\bar{P}}_{\text{dim}, n}) - P_i \|^2.$$  \hfill (13)

Similar to STINR (equation (11)), PCA$_{cv}$ uses the same principal motion components ($\mathbf{\bar{P}}_{\text{dim}, 0}$ $\mathbf{\bar{P}}_{\text{dim}, n}$) for DVF derivation. However, instead of using a spatial INR to represent the reference CBCT volume, such a volume ($\mu$) of PCA$_{cv}$ was extracted from either a prior 4D-CT or could be reconstructed online from onboard projections. In this study, we used the EE phase of cone-beam projection subset (figure 3), extracted from all projections, to reconstruct a CBCT volume to serve as $\mu$. Since the PCA model was derived based on the EE phase of the prior 4D-CT (II.B), using the EE phase of cone-beam projections can maximize their similarity to the motion model, while also allowing inter-scan anatomical changes to be reconstructed and avoiding the potential shading variations among different imaging systems. To improve the reconstruction quality of the reference volume $\mu$, we used an algebraic reconstruction technique (ART) with the total variation regularization (Ouyang et al 2013). The ART-based image update (equation (14)) and TV (equation (15)) minimization steps were alternated until convergence is achieved:

$$\mu_{i+1} = \mu_i + \lambda a_{ij} \left[ P_j - \sum_i a_{ij} \mu_i \right]$$  \hfill (14)

$$TV(\mu) = \| \nabla(\mu) \|.$$  \hfill (15)

Different from STINR which solves temporal INRs to model the temporal kinematics, PCA$_{cv}$ uses a scalar $w_{\text{dim}, n}^t$ to fit each principal component weighting at each temporal stamp (each projection), such that the optimization was performed independently per projection (equation (13)). The objective function of equation (13) was analytically optimized using the nonlinear conjugate gradient algorithm, of which the details could be found in our previous publications (Zhang et al 2013).

In addition to PCA$_{cv}$, we also compared STINR with the INR and polynomial fitting-based dynamic CT study (INR$_{poly}$) (Reed et al 2021). The INR$_{poly}$ method was developed to reconstruct dynamic CT images from limited-angle projections, with each temporal dynamic occupying a partial scan angle. The temporal DVFs were derived as voxel-wise motion coefficients weighted by temporal polynomials. To fit our reconstruction needs, we modified the INR$_{poly}$ method by introducing the cone-beam projection geometry and benchmarked its reconstruction results against STINR.

2.4.3. Evaluation metrics
To quantitatively assess STINR and compare it against PCA$_{cv}$ and INR$_{poly}$, we evaluated the reconstructed dynamic CBCTs (CBCT$_{est}$) of each method by comparing with the ‘ground-truth’ simulations via the relative error (RE) metric (equation (16)). We also evaluated the solved intra-scan DVFs $D(t)$, by comparing $D(t)$-propagated lung tumor motion with the ‘ground-truth’ tumor motion. We used the DICE coefficients (equation (16)) and the center-of-mass errors (COMEs) (Zhang et al 2019) of the tumor contours.
Table 2. Summary of CBCT scanning parameters applied in the XCAT and the patient studies. The XCAT and the patient studies shared most parameters except for the projection matrix/pixel sizes and the CBCT matrix/voxel sizes. The projection and CBCT sizes of the patient study were kept the same as the original 4D-CBCT scan used for simulation. For the XCAT study, two sets of the CBCT matrix/voxel sizes were applied: the intermediate sizes were used during STINR optimization due to the memory limit of the GPU card; and the final output sizes were used during STINR inference. Since STINR uses MLPs to represent the images, CBCTs of arbitrary resolutions can be output during inference.

| CBCT imaging parameters                        | Value                                      |
|-----------------------------------------------|--------------------------------------------|
| Source-imager-distance                        | 1500 mm                                    |
| Source-axis-distance                          | 1000 mm                                    |
| Scan rotation speed                           | 6° s⁻¹                                     |
| Scan rotation angle                           | 0°~360°                                    |
| Projection acquisition frame rate             | 11 frames s⁻¹                               |
| Number of projections per scan                | 660                                        |
| Projection acquisition mode                   | Full-fan                                    |
| Projection matrix size                        | 512 × 512 (XCAT) and 512 × 384 (Patient)  |
| Pixel size                                    | 1.17 mm × 1.17 mm (XCAT) and 0.776 mm × 0.776 mm (Patient) |
| Intermediate CBCT matrix size (during STINR optimization) | 64 × 64 × 64 (XCAT) and 128 × 128 × 56 (Patient) |
| Output CBCT matrix size (STINR inference)     | 128 × 128 × 128 (XCAT) and 128 × 128 × 56 (Patient) |
| Intermediate CBCT voxel size (STINR optimization) | 6.0 mm × 6.0 mm × 6.0 mm (XCAT) and 3.9 mm × 3.9 mm × 3.9 mm (Patient) |
| Output CBCT voxel size (STINR inference)      | 3.0 mm × 3.0 mm × 3.0 mm (XCAT) and 3.9 mm × 3.9 mm × 3.9 mm (Patient) |

\[
RE = \sqrt{\frac{\sum (\mu_{\text{recon}} - \mu_{\text{GT}})^2}{\sum \mu_{\text{GT}}^2}}
\]  
\[
DICE = 2 \frac{|V_{\text{recon}} \cap V_{\text{GT}}|}{|V_{\text{recon}} \cup V_{\text{GT}}|},
\]

In equation (16), \(\mu_{\text{recon}}\) denotes the reconstructed dynamic CBCTs by different methods. \(\mu_{\text{GT}}\) denotes the corresponding ‘ground-truth’ images. The voxel-wise attenuation coefficient differences were computed and summed up to assess the overall reconstruction errors relative to the ‘ground-truth’ attenuation coefficients. Equation (17) defines the DICE coefficient, which measures the match between the dynamically-resolved tumor volumes \(V_{\text{recon}}\) and the ‘ground-truth’ tumor volumes \(V_{\text{GT}}\). For implementation, the tumors were manually segmented from the reconstructed reference CBCT volumes \(V_{\text{ref}}\) for each combination of motion/anatomy scenario and reconstruction technique. The manual segmentations were propagated onto each dynamic volume using the DVFs \(D(t)\) solved by each technique, and then compared with the ‘ground-truth’ tumor contours segmented via automatic intensity thresholding from the ‘ground-truth’ dynamic CBCT volumes. A DICE coefficient of 1 indicates a perfect match and 0 indicates non-overlapping volumes. The COME metric, on the other hand, measures the distance between the dynamically-resolved tumor location and the ‘ground-truth’ tumor location, which also serves as an important metric to evaluate the accuracy of image guidance in radiotherapy.

2.4.4. Patient study
To further assess the accuracy of STINR, we also performed a patient study in addition to the XCAT study. Compared with the XCAT phantom, the real patient images contained more detailed structures and degrading signals including scatter and noise. The patient imaging data were acquired under an IRB-approved protocol. A 4D-CBCT set was acquired for a lung cancer patient via an adaptive-speed slow-gantry rotation setting using 120 kVp, 80 mA, and 25 ms (Lu et al 2007). The projections were acquired in full-fan mode, with each projection measuring 512 × 384 pixels in dimension and each pixel measuring 0.776 mm × 0.776 mm. Each respiratory phase of the 4D-CBCT dataset has ~200 projections after sorting and can reconstruct good-quality, phase-specific CBCT volumes via the standard FDK algorithm for reference and target contouring. To best preserve the scatter/noise in the cone-beam projections and the correspondingly-reconstructed CBCTs, no additional scatter/noise corrections were applied, i.e. the CBCTs were directly reconstructed using the basic FDK algorithm from the raw projections without additional projection-domain pre-processing or CBCT-domain image regularization.

To evaluate the accuracy and robustness of the STINR method by different breathing scenarios, we used a PCA-driven motion augmentation technique (Shao et al 2022) to create motion scenarios featuring: (S1-P). A regular sinusoidal motion curve; (S2-P). A motion curve featuring both amplitude variations and baseline drifts; (S3-P). A motion curve featuring amplitude and frequency variations; and (S4-P). A slow-, non-periodic motion.

\[
\frac{\sum (\mu_{\text{recon}} - \mu_{\text{GT}})^2}{\sum \mu_{\text{GT}}^2}
\]
In detail, from the 4D-CBCT, we chose the end-expiration phase as the reference phase and performed inter-phase deformable registrations to extract a PCA-based motion model. Based on the PCA-based motion model, we customized the weightings of the principal components along the three cartesian directions to simulate the four motion scenarios. The reference phase was deformed using the DVFs derived from the weighting-customized PCA model, to generate 660 dynamic, time-resolved CBCTs as 'gold-standard' for further simulation and evaluation. Cone-beam projections were re-projected from these motion-augmented CBCTs, at scan angles calculated via equation (12), assuming a gantry rotation speed of $6^\circ \, s^{-1}$ and a frame rate of 11 frames $s^{-1}$. Detailed cone-beam scan simulation parameters were summarized below in table 2 for both the XCAT and the patient studies.

Similar to the XCAT study, we applied STINR and $PCA_{cv}$ to solve the dynamic CBCT images and the motion curves, and compared the results with the known 'ground-truth'. A lung target was segmented out of the CBCT volumes, and the accuracy of its motion tracked via STINR or $PCA_{cv}$ was assessed in terms of DICE and COME. The reconstructed dynamic CBCT images were also evaluated against the 'ground-truth' via the RE metric.

**Figure 5.** Comparison of the reconstructed CBCT volumes between $INR_{poly}$, $PCA_{cv}$, and STINR of the XCAT study. The difference images between the reconstructed images and the 'ground-truth' XCAT images were also presented. The display window for the CBCT images is $[0,0.05] \, mm^{-1}$, and the display window for the difference images is $[-0.025,0.025] \, mm^{-1}$. The images shown here correspond to the XCAT motion/anatomy scenario S2.
3. Results

3.1. Evaluation of reconstructed CBCTs by different methods: XCAT study

Figure 5 compares the STINR-reconstructed CBCT volume against those reconstructed by the PCA\textsubscript{cv} method and the INR\textsubscript{poly} method. The images in figure 5 correspond to the motion scenario S2 (table 1), where a large intra-scan baseline shift was introduced. Due to the intra-scan baseline shift, the reconstructed CBCT of PCA\textsubscript{cv} shows pronounced motion blurriness, since the EE phase projections extracted for the reference CBCT (CBCT\textsubscript{Ref}) reconstruction contain two different motion baselines. In comparison, STINR achieves substantially more accurate reconstruction with the motion blurriness successfully subdued. The STINR images show sharper edges and well-defined tumor boundaries. Although the CBCT\textsubscript{Ref} volume of STINR was also initialized using the extracted EE phase projections via the 3-stage optimization scheme (figure 3), the simultaneous spatial and temporal INR learning at stage 3 helps to incorporate the solved dynamic motion of each projection to correct the residual motion contained in CBCT\textsubscript{Ref} and the resulting dynamic CBCT volumes. It also addresses potential motion mismatches between prior and new scans. Some degrees of blurriness remain in the STINR images, which are caused by the intrinsic low spatial resolution of reconstruction (6 mm × 6 mm × 6 mm for the down-sampled 64 × 64 × 64 volume during optimization) due to the limitation of GPU memory. Compared
CAcv and STINR, generally reconstructs the worst-quality CBCT image with amplified noise and motion blurriness. Without introducing prior knowledge from the PCA motion model, INRpoly’s accuracy was inferior due to the complexity of the spatiotemporal inverse problem, as the solution can be easily trapped at a local optimum.

3.2. Evaluation of dynamic CBCT reconstruction: XCAT study

Figure 6 shows the x-ray projections acquired at different time spots (frames, 1st row), the correspondingly-reconstructed STINR dynamic CBCTs (axial view-3rd row, coronal view-6th row, and sagittal view-9th row), the ‘ground-truth’ dynamic CBCTs (axial view-2nd row, coronal view-5th row, and sagittal view-8th row), and the difference images (axial view-4th row, coronal view-7th row, and sagittal view-10th row). It can be observed that STINR can reconstruct dynamic CBCTs to match well with the ‘ground-truth’, in terms of both the image quality (boundary sharpness, intensity variations, noise level, etc) and the fidelity of reconstructed anatomy (tumor location/shapes, lung, heart, air pockets, etc).

Table 3. Mean and standard deviation of the relative errors (REs) for all motion/anatomy scenarios and three different methods of the XCAT study. The RE was calculated between the reconstructed dynamic CBCT volume and the ‘ground-truth’ CBCT volume. Each motion scenario comprises 660 dynamic CBCT volumes.

| Motion scenarios | INRpoly | PCAcv | STINR |
|------------------|---------|-------|-------|
| S1               | 18.25 ± 3.26% | 14.06 ± 0.72% | 9.85 ± 0.47% |
| S2               | 22.39 ± 3.64% | 16.75 ± 0.83% | 10.57 ± 0.84% |
| S3               | 16.54 ± 2.63% | 12.58 ± 1.23% | 9.97 ± 0.53% |
| S4               | 17.75 ± 5.18% | 14.34 ± 0.77% | 10.08 ± 0.92% |
| S5               | 15.01 ± 5.19% | 12.88 ± 0.65% | 10.38 ± 0.42% |
| S6               | 10.40 ± 3.19% | 42.06 ± 2.53% | 9.85 ± 0.65% |
| S7               | 18.12 ± 3.39% | 14.05 ± 0.72% | 9.99 ± 0.43% |
| S8               | 18.25 ± 3.46% | 14.08 ± 0.72% | 9.79 ± 0.46% |

Figure 7 and table 3 present the calculated REs between the reconstructed dynamic CBCTs by different methods. Each sub-boxplot of figure 7 corresponds to one motion/anatomy scenario (table 1) by one reconstruction method. Similarly, STINR offers consistently higher reconstruction accuracy as compared to the other two methods. The INRpoly performed poorly in general except for scenario S6. In contrast, PCAcv performed the worst for scenario S6, where non-periodic motion was simulated. Since PCAcv relies on the extracted EE-phase projections to reconstruct the reference CBCT volume (CBCTref), non-periodic motion (or equivalently fast gantry rotation) clustered these projections into a limited, partial scan angle. It led to substantial anatomical and geometric distortions in the reconstructed CBCTref. Unlike STINR, PCAcv does not have the mechanism to simultaneously and continuously update the CBCTref and the intra-scan motion model during.

Figure 7. Boxplots of the relative error results for three different methods and all eight motion/anatomy scenarios of the XCAT study.

Table 3. Mean and standard deviation of the relative errors (REs) for all motion/anatomy scenarios and three different methods of the XCAT study. The RE was calculated between the reconstructed dynamic CBCT volume and the ‘ground-truth’ CBCT volume. Each motion scenario comprises 660 dynamic CBCT volumes.

| Motion scenarios | INRpoly | PCAcv | STINR |
|------------------|---------|-------|-------|
| S1               | 18.25 ± 3.26% | 14.06 ± 0.72% | 9.85 ± 0.47% |
| S2               | 22.39 ± 3.64% | 16.75 ± 0.83% | 10.57 ± 0.84% |
| S3               | 16.54 ± 2.63% | 12.58 ± 1.23% | 9.97 ± 0.53% |
| S4               | 17.75 ± 5.18% | 14.34 ± 0.77% | 10.08 ± 0.92% |
| S5               | 15.01 ± 5.19% | 12.88 ± 0.65% | 10.38 ± 0.42% |
| S6               | 10.40 ± 3.19% | 42.06 ± 2.53% | 9.85 ± 0.65% |
| S7               | 18.12 ± 3.39% | 14.05 ± 0.72% | 9.99 ± 0.43% |
| S8               | 18.25 ± 3.46% | 14.08 ± 0.72% | 9.79 ± 0.46% |
optimization. The under-sampling artifacts remained in the CBCTs of PCAcv and got carried over into the reconstructed dynamic CBCTs, which also led to amplified errors in PCA coefficient optimization, resulting in substantially larger motion tracking errors. In contrast, for INRpoly, the one-cycle motion of scenario S6 is easier to be fitted via polynomials as compared to the other scenarios, which resulted in relatively better performance. For STINR, since it could use all projections for simultaneous image and motion optimization, it was not susceptible to the limited-angle reconstruction errors of PCAcv.

Figure 8. Comparison of tracked tumor motion along the superior–inferior (SI) direction for all reconstruction methods and motion/anatomy scenarios of the XCAT study. The curves were style-coded and color-coded for differentiation (green solid: ‘ground-truth’; blue solid: STINR; red dot-dashed: PCAcv; and black dashed: INRpoly).
Table 4. The DICE and COME results of dynamically-resolved tumor volumes for all motion scenarios of the XCAT study, by different methods. The DICEs and COMEs were calculated between the DVF-propagated tumors and the ‘ground-truth’ tumor contours. Each motion scenario comprises 660 dynamic CBCT volumes.

| Motion scenarios | Evaluation metrics | INR\(poly\) | PCA_{cv} | STINR |
|------------------|--------------------|------------|----------|--------|
| S1 DICE          | 0.73 ± 0.10        | 0.86 ± 0.03 | 0.87 ± 0.04 |
| S1 COME [mm]     | 5.6 ± 3.0          | 1.5 ± 0.7  | 1.4 ± 0.6  |
| S2 DICE          | 0.69 ± 0.10        | 0.65 ± 0.06 | 0.85 ± 0.04 |
| S2 COME [mm]     | 6.9 ± 3.4          | 2.9 ± 1.8  | 2.4 ± 1.0  |
| S3 DICE          | 0.68 ± 0.10        | 0.84 ± 0.05 | 0.86 ± 0.03 |
| S3 COME [mm]     | 6.4 ± 2.9          | 2.1 ± 1.0  | 2.0 ± 0.7  |
| S4 DICE          | 0.72 ± 0.09        | 0.83 ± 0.06 | 0.88 ± 0.04 |
| S4 COME [mm]     | 5.6 ± 3.1          | 3.2 ± 1.8  | 1.6 ± 0.8  |
| S5 DICE          | 0.79 ± 0.09        | 0.78 ± 0.40 | 0.84 ± 0.04 |
| S5 COME [mm]     | 4.0 ± 2.4          | 3.1 ± 0.8  | 1.2 ± 0.6  |
| S6 DICE          | 0.78 ± 0.07        | 0.24 ± 0.25 | 0.84 ± 0.04 |
| S6 COME [mm]     | 3.3 ± 2.3          | 30.8 ± 20.1 | 1.6 ± 0.6  |
| S7 DICE          | 0.47 ± 0.13        | 0.75 ± 0.06 | 0.77 ± 0.06 |
| S7 COME [mm]     | 9.2 ± 5.3          | 2.8 ± 0.7  | 1.6 ± 0.8  |
| S8 DICE          | 0.77 ± 0.07        | 0.84 ± 0.04 | 0.87 ± 0.04 |
| S8 COME [mm]     | 4.5 ± 2.3          | 1.6 ± 0.8  | 1.3 ± 0.5  |

3.3. Evaluation of dynamic motion reconstruction: XCAT study

In addition to the images, we also evaluated the solved tumor motion by different methods. The comparison of tracked tumor motion curves along the superior–inferior (SI) direction was presented in figure 8, as the motion along the SI direction was dominant. The corresponding DICE coefficient and COME results were reported in table 4. As shown in figure 8 and table 4, the STINR-solved tumor motion matched closely with the motion that was tracked from the ‘ground-truth’ dynamic CBCT images. In general, the average DICE was above 0.8 for all scenarios except for scenario S7 (0.77) where the tumor diameter was reduced by 50%. The lower DICE values for S7 were expected due to the increased sensitivity of DICE to decreasing volumes. Similarly, the average COME was no larger than 2 mm for all scenarios, except for scenario S2 (2.4 mm) where a large baseline shift occurred. In comparison, INR\(poly\) failed almost completely to recover the high-frequency motion signals due to the limitation of the polynomial fitting approach. PCA_{cv}, on the other hand, also failed to correctly capture the true motion, especially for CBCTs close to motion peaks/valleys. It can be caused by the inferior quality of CBCT_{Ref} for PCA_{cv} (figure 5), as the poorly defined tumor/organ boundaries and shapes led to uncertainties in solving the PCA motion scaling factors. Echoing figure 7, PCA_{cv} failed in solving scenario S6, due to the poor quality of CBCT_{Ref} from limited-angle sampling.

3.4. Patient study results

As shown in figures 9, 10, and table 5, similar to the XCAT study, STINR consistently showed an advantage over PCA_{cv} for the patient study, even though the patient dataset contained more intensity features and also degrading scatter and noise signals as compared to the XCAT dataset. The INR\(poly\) method was not evaluated in the patient study due to its poor performance in the XCAT study. Similar to the XCAT study, PCA_{cv} is affected by the compromised quality of the reconstructed reference volume, due to factors including intra-phase motion averaging (caused by baseline drifts, motion amplitude variations, etc) and under-sampling artifacts (caused by partial angle sampling such as the slow-motion scenario S4-P). The sagittal view comparison of figure 9 shows an example that intra-phase motion leads to a blurred target region for PCA_{cv}, which is caused by the baseline drift (figure 10: motion scenario S2-P) contained within the reference EE phase. In comparison, STINR is able to use simultaneous spatial and temporal learning to continuously update the reference CBCT image using all available projections, which in turn helps to further optimize the motion coefficients at each temporal frame to solve the dynamic motion more accurately.

4. Discussion

4.1. Dynamic CBCT Imaging and STINR

Dynamic CBCT imaging generates volumetric images with superior temporal and spatial resolutions and is highly desired in clinical applications including radiotherapy targeting verification, tumor motion monitoring and prediction, treatment dose tracking and accumulation, and robust treatment planning. Our study developed a simultaneous spatial and temporal implicit neural representation learning (STINR) framework to address the extremely challenging problem of reconstructing dynamic CBCTs from singular x-ray projections.
In comparison to traditional voxel-wise representations of volumetric images, the STINR method uses different multi-layer perceptrons to represent the spatial features and temporal evolutions of imaged anatomical structures. MLP is powerful in representing images or motions with complex, continuous, and differentiable functions, while needless to specify/constrain the detailed function forms in advance. Compared to conventional voxel-wise image representations, MLP maps the image/motion to neural networks and in theory can generate voxelized moving images of arbitrary spatial or temporal resolutions to achieve inherent super-resolution.

Due to the difficulty of mapping the full spatiotemporal imaging series into a single MLP, we developed STINR to use individual MLPs to fit the spatial and temporal INRs separately, which allows the MLPs to be customized to tailor to the inherent complexity variations of different representation tasks (imaging/motion). We introduced patient-specific prior knowledge, the PCA-based motion models, into solving the temporal INRs to further reduce the complexity of the ill-conditioned spatiotemporal inverse problem. PCA-based motion modeling allows substantial dimension reductions to represent complex motion scenarios accurately and effectively.

4.2. Comparison with the PCA-based methods

Conventional PCA-based dynamic CBCT reconstruction methods solve the motion based on a fixed reference volume, and the reference volume can be reconstructed from onboard projections (Dhou et al 2022) or extracted from patient-specific prior 4D-CTs (Zhang et al 2013). The reference volume is subsequently used to solve the PCA motion coefficients to construct dynamic motion. In comparison, STINR reconstructs the reference CBCT volume as a spatial INR and optimizes the reference CBCT volume (spatial INR) and the dynamic DVFs (temporal INRs) simultaneously. The simultaneous solution of spatial and temporal INRs allows STINR to correct the residual intra-phase motions of the reference CBCT volume, which can come from sources including projection sorting errors, motion baseline shifts, or other irregular motion (figures 5, 8–10). For non-periodic motion or fast gantry rotation scenarios, it can be rather challenging to reconstruct such a reference volume for conventional PCA-based methods, since the projections of a certain motion phase may only occupy a restricted
small angle instead of scattering around the full scan angle as in periodic motion or slow-gantry rotation scenarios. The limited scan angle severely distorts the reconstructed reference volume and significantly reduces the accuracy of dynamic CBCT reconstruction for conventional PCA-based methods (S6 in tables 3, 4, figures 7, 8; S4-P in table 5, figure 10). In comparison, STINR uses all dynamic projections to resolve the reference volume.

**Table 5.** Mean and standard deviation of the relative error (RE), the DICE coefficient, and the center-of-mass-error (COME) metrics for all motion scenarios evaluated in the patient study. Each motion scenario comprises 660 dynamic CBCT volumes.

| Scenario | Metric | STINR | PCA<sub>cv</sub> |
|----------|--------|-------|-----------------|
| S1-P     | RE     | 14.71 ± 0.57% | 23.58 ± 0.59% |
|          | DICE   | 0.90 ± 0.08   | 0.59 ± 0.07    |
|          | COME (mm) | 1.1 ± 0.9   | 3.4 ± 0.9      |
| S2-P     | RE     | 12.71 ± 0.43% | 21.69 ± 0.55% |
|          | DICE   | 0.92 ± 0.07   | 0.64 ± 0.10    |
|          | COME (mm) | 0.9 ± 0.8   | 2.3 ± 1.1      |
| S3-P     | RE     | 13.98 ± 0.81% | 23.84 ± 0.69% |
|          | DICE   | 0.81 ± 0.05   | 0.67 ± 0.11    |
|          | COME (mm) | 1.3 ± 0.6   | 2.5 ± 1.5      |
| S4-P     | RE     | 15.53 ± 0.55% | 68.77 ± 4.06% |
|          | DICE   | 0.85 ± 0.08   | 0.15 ± 0.23    |
|          | COME (mm) | 1.3 ± 0.9   | 19.5 ± 21.3    |

**Figure 10.** Comparison of tracked target motion along the superior–inferior (SI) direction for different reconstruction methods and motion scenarios of the patient study. The curves were style-coded and color-coded for differentiation (green solid: ‘ground-truth’; blue solid: STINR; and red dot-dashed: PCA<sub>cv</sub>).
and the intra-scan motion simultaneously, which allows the reference volume to be reconstructed based on information from all projection angles while accounting for the intra-scan motion. The full-angle coverage successfully removes the limited-angle distortions to generate a high-quality reference volume, which in return helps to solve intra-scan DVFs more accurately and reliably. Recent advancements in technology, such as the developments of ring-gantry-based radiotherapy modalities, achieve much faster gantry rotation and imaging speed than conventional C-shaped linear accelerators. The push for faster imaging helps to boost efficiency and potentially reduce motion-related artifacts. It, however, poses additional challenges to conventional phase-sorted 4D imaging techniques, as the motion periodicity assumption inherently adopted by these techniques may not stand anymore. STINR can be a potential solution in these scenarios, which is agnostic and robust to the motion patterns observed during image acquisitions.

In this study, we used the onboard projections at the EE phase to reconstruct reference CBCT volumes for the conventional PCA-based reconstruction technique, since the PCA motion model is as well built on the EE phase volume of the prior 4D-CT. Alternatively, previous studies also directly used the EE phase volume of the prior 4D-CT as the reference volume (Li et al 2010, Zhang et al 2013). The challenge of this approach is that it cannot fully resolve the inter-scan anatomical variations due to factors like treatment response (tumor shrinkage, lung inflammation, etc) or disease progression, which cannot be learned from an intra-scan motion model of PCA as these long-term variations do not exist or happen within a single 4D-CT scan. The approach cannot address non-deformation-induced anatomical changes either. Another challenge of this approach is from the shading variations between CT/CBCT, as the prior 4D-CT is usually acquired in a fan-beam geometry while the latter CBCT is acquired in the cone-beam geometry. The cone-beam geometry usually suffers from amplified degradation signals including photon scatter and electronic noise (Ouyang et al 2013, Chen et al 2017). The difference in x-ray source, energy, mA, ms, and other hardware variations also introduce difficulties/inaccuracy in directly using the CT volume as the reference for onboard dynamic CBCT reconstruction.

4.3. Comparison with the INR-based method with polynomial fitting

In addition to the conventional PCA-based dynamic CBCT reconstruction methods, in this study, we also compared STINR with another INR method using polynomial fitting to generate the intra-scan DVFs (Reed et al 2021). In comparison to STINR, the polynomial fitting-based INR method was originally developed for limited-angle reconstruction problems, where the motion is expected to be slow and gradual (such that each motion state will occupy at least a small scan angle). Our dynamic CBCT reconstruction problem, however, is much more challenging due to the motion scenarios evaluated (much faster and more volatile breathing motion). Results show that the polynomial fitting-based method fails to accurately capture the breathing motion that occurred within most dynamic CBCT scans. Compared to the temporal INRs, the polynomials showed very limited accuracy in representing periodical motion and its variations. Another advantage of STINR over the polynomial fitting-based method is the introduction of PCA-based motion modeling as prior knowledge. The polynomial fitting-based INR method needs to solve PCA-like motion matrices from the dynamic projections directly, which incurred substantially-elevated complexity and trapped the reconstructions into obvious sub-optimal solutions in our evaluations.

In addition to the PCA-based method and the polynomial-fitting-based INR method, multiple other methods as mentioned in the introduction are also potentially available for dynamic CBCT reconstruction. These other methods are challenging to be directly compared, due to multiple factors including the need for code redevelopment and adaptation to 3D cone-beam geometry (Cai et al 2014, Jailin et al 2021), special pre-training (Gao et al 2018), and angle-agnostic model development (Shen et al 2019). The substantial amount of work needed to evaluate and compare these methods is beyond the scope and capacity of the current study. Limited by the authors’ knowledge, there also might be other dynamic CBCT reconstruction works not referred to in this study. Future studies are warranted to compare STINR against other methods for comprehensive evaluations under different patient motion scenarios.

4.4. Limitations and further directions

In our study, we used XCAT to simulate dynamic patient volumes and cone-beam projections, which allowed us to control and customize the motion/anatomical scenarios to test the accuracy and robustness of STINR with known ‘ground-truth’. Previous studies also frequently used XCAT to generate dynamic volumes to test their algorithms (Cai et al 2014, Gao et al 2018). In addition to the XCAT study, we also used a real patient 4D-CBCT dataset to simulate different onboard motion scenarios within a 60 s CBCT scan, and the results similarly demonstrated the advantage of STINR in accurate spatial and temporal reconstructions. Future experimental evaluations are warranted to further validate STINR and compare it with other available methods on real patient data featuring different motion scenarios. The challenge of using real patient data without controlled simulations, however, is attaining ‘ground-truth’ dynamic volumetric images and motion to validate our
methods. One potential workaround is to compare the re-projected image features from the reconstructed dynamic CBCTs to those extracted from the dynamic cone-beam projections, for instance, the locations of radiopaque fiducial markers or trackable anatomical features (Wei et al 2020). In addition, studies using anthropomorphic physical phantoms can also be performed.

In the XCAT and patient studies, we have to convert the continuously-defined INRs into intermediate voxelized images to feed into forward/backward ASTRA operators to drive optimization. Although the INRs allow CBCTs to be represented at arbitrary resolutions due to their continuous nature, due to the limit of the GPU memory, we have to down-sample the volume matrix dimensions during the optimization, which yielded lower spatial resolutions along each Cartesian direction. Although the final output CBCTs can be up-scaled to arbitrarily high spatial resolution, the intermediate down-sampling may adversely affect the final reconstruction accuracy (figures 5, 9). In our study, we found a residual DVF error of around 1–2 mm (tables 4, 5), which could be partially contributed by the above-mentioned spatial resolution limitations. Further investigations quantifying the effects of such under-sampling operations, and the improvements from using larger memory GPUs or low memory footprint operators, are warranted to examine the potential and limitations of STINR.

The current STINR framework solves the reference CBCTs and dynamic DVFs without implementing additional spatial and temporal regularizations. Introducing further a priori knowledge, for instance, the total variation regularization on the spatial domain (Zhang et al 2017), may further improve the reconstruction accuracy. Regularizations in the motion domain, like those encouraging the smoothness of motion, may also be applied in caution (not to substantially suppress the reconstruction of irregular motion). Another way is to further condition the STINR framework and initialize the temporal INRs via tracked motion curves, for instance, those from the external surface marker tracking or internal diaphragm tracking (Jailin et al 2021). The pre-conditioning may help to further accelerate the reconstruction speed and reduce the chances of the reconstructions being trapped at local optima. In addition to the temporal INR initialization, the spatial INR can also be initialized by prior CT/CBCT images, which was found effective in improving image reconstruction accuracy while reducing the onboard imaging sampling needs (Shen et al 2022). However, potential biases from the spatial and temporal INR initializations, especially due to large anatomical variations (spatial) or the imperfect correlations between the surrogate motion and the real volumetric motion (temporal), may negatively affect the reconstruction accuracy and remain to be further evaluated.

5. Conclusion

We developed a spatial and temporal implicit neural representation learning method (STINR) to reconstruct dynamic CBCTs from singular x-ray projections acquired during a normal 60 s CBCT scan. The ‘one-shot’ learning STINR method uses the powerful representation capability of multi-layer perceptrons and adopts prior knowledge from PCA-based motion modeling. The combination successfully addresses the challenging spatiotemporal inverse problem of dynamic CBCT reconstruction. STINR can be easily adapted and implemented for other imaging modalities including magnetic resonance imaging. The spatially- and temporally-resolved images can benefit many clinical applications, including target motion tracking/prediction, intra-treatment dose calculation/accumulation, and robust planning/delivery. Future studies based on more real patient data with various motion/anatomy scenarios are warranted for a comprehensive evaluation and potential methodology fine-tuning.

Acknowledgments

The study was supported by funding from the National Institutes of Health (R01CA240808, R01CA258987). We would like to thank Dr. Paul Segars at Duke University for providing the XCAT phantom for our study.

Ethical statements

The patient dataset used in this study was retrospectively collected from an IRB-approved study at MD Anderson Cancer Center in 2007. Individual patient consent was signed for the anonymized use of the imaging data for retrospective analysis. The study was conducted in accordance with the principles embodied in the Declaration of Helsinki. This is a retrospective analysis study and not a clinical trial. No clinical trial ID number is available.
References

Adler J, Kohr H and Oktem O 2017 Operator discretization library (ODL) (https://github.com/odlgroup/odl)

Agarap A F 2018 Deep learning using rectified linear units (relu) at arXiv:1803.08375

Andersen A H and Kak A C 1984 Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm Ultrason. Imaging 6 81–94

Basri R et al 2020 Frequency bias in neural networks for input of non-uniform density Int. Conf. on Machine Learning vol 119, p 119

Bergner F et al 2009 Autoadaptive phase-correlated (AAPC) reconstruction for 4D CBCT Med. Phys. 36 5695–706

Borst G R et al 2007 Kilo-voltage cone-beam computed tomography setup measurements for lung cancer patients; first clinical results and comparison with electronic portal imaging device Int. J. Radiat. Oncol. * Biol. Phys. 68 555–61

Bourke P 1999 Interpolation methods Miscellaneous: Projection, Modelling, Rendering vol 1, p 10

Cal J-F et al 2014 Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study IEEE Trans. Med. Imaging 33 1581–91

Chen X et al 2017 Optimization of the geometry and speed of a moving blocker system for cone-beam computed tomography scatter correction Med. Phys. 44 e215–29

Clements N et al 2013 The effect of irregular breathing patterns on internal target volumes in four-dimensional CT and cone-beam CT images in the context of state-of-the-art radiotherapy Med. Phys. 40 021904

Cooper B J et al 2015 Quantifying the image quality and dose reduction of respiratory triggered 4D cone-beam computed tomography with patient-measured breathing Phys. Med. Biol. 60 4943–513

Dhou S et al 2018 Principal component reconstruction (PCR) for cine CBCT with motion learning from 2D fluoroscopy Med. Phys. 45 167–77

Harries W et al 2017 Estimating 4D-CBCT from priorit information and extremely limited angle projections using structural PCA and weighted free-form deformation for lung radiotherapy Med. Phys. 44 1089–104

Heidari A A et al 2020 Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks Nat. Inspired Opt. 8 23–46

Huang L et al 2010 A study on the dosimetric accuracy of treatment planning for stereotactic body radiation therapy of lung cancer using average and maximum intensity projection images Radiat. Oncol. 96 68–54

Huang X et al 2020 U-net-based deformation vector field estimation for motion-compensated 4D-CBCT reconstruction Med. Phys. 47 3000–12

Jaffray D A et al 2002 Flat-panel cone-beam computed tomography for image-guided radiation therapy Int. J. Radiat. Oncol. * Biol. Phys. 53 1337–49

Jalin C et al 2021 Projection-based dynamic tomography Phys. Med. Biol. 66 215018

Klein S et al 2010 Elastic: a toolbox for intensity-based medical image registration IEEE Trans. Med. Imaging 29 196–205

Kong V C, Marshall A and Chan H B 2016 Cone beam computed tomography: the challenges and strategies in its application for dose accumulation J. Med. Imaging Radiat. Sci. 47 92–7

Leng S et al 2008 High temporal resolution and streak-free four-dimensional cone-beam computed tomography Phys. Med. Biol. 53 5635–73

Léontoureaux D et al 2005 Cone-beam–CT guided radiation therapy: technical implementation Radiat. Oncol. 75 279–86

Li R et al 2010 Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy Med. Phys. 37 2822–6

Li R et al 2011 On a PCA-based lung motion model Phys. Med. Biol. 56 6009–30

Li T and Xing L 2007 Optimizing 4D cone-beam CT acquisition protocol for external beam radiotherapy Int. J. Radiat. Oncol. * Biol. Phys. 67 1211–9

Li X et al 2018 Effects of irregular respiratory motion on the positioning accuracy of moving target with free breathing cone-beam computed tomography Int. J. Med. Phys. Clin. Eng. Radiat. Oncol. 7 173–83

Liang J et al 2019 Intrafraction 4D-cone beam CT acquired during volumetric arc radiotherapy delivery: kV parameter optimization and 4D motion accuracy for lung stereotactic body radiotherapy (SBRT) patients J. Appl. Clin. Med. Phys. 20 10–24

Ling C et al 2011 Acquisition of MV-scatter-free kilovoltage CBCT images during RapidArc™ or VMAT Radiat. Oncol. 100 145–9

Lombardi S, Simon T, Saragh, J, Schwartz G, Lehrmann A and Sheikh Y 2019 Yeulul approved learning: learning dynamic renderable volumes from images ACM Trans Graph 38 65

Lu J et al 2007 Four-dimensional cone-beam CT with adaptive gantry rotation and adaptive data sampling Med. Phys. 34 3520–9

Ouyang L, Song K and Wang J 2013 A moving blocker system for cone-beam computed tomography scatter correction Med. Phys. 40 071903

Pan C H et al 2019 The irregular breathing effect on target volume and coverage for lung stereotactic body radiotherapy J. Appl. Clin. Med. Phys. 20 109–20

Pan T et al 2004 4D–CT imaging of a volume influenced by respiratory motion on multi-slice CT Med. Phys. 31 333–40

Peng S et al 2021 Neural body: implicit neural representations with structured latent codes for novel view synthesis of dynamic humans Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition

Pereira G C, Traugher M and Muizc R F 2014 The role of imaging in radiation therapy planning: past, present, and future BioMed Res. Int. 2014 231090

Poulsen P R et al 2014 Kilovoltage intrafraction motion monitoring and target dose reconstruction for stereotactic volumetric modulated arc therapy of tumors in the liver Radiat. Oncol. 111 424–30

Ramachandran P, Zoph B and Le Q V 2017 Searching for activation functions arXiv:1710.05941

Orcid iDs

You Zhang @ https://orcid.org/0000-0002-8033-2755

Tielige Mengke @ https://orcid.org/0000-0003-1836-9829
Adhikari S, Wang J, Li T and Xing L 2009 Iterative image reconstruction for CBCT using edge-preserving prior
Wang J and Gu X 2013 Simultaneous motion estimation and image reconstruction
Yan H, Yasue K, Zhang Y, Zhang Y, Zhang Y, Zhang Y, Sitzmann V, Zollhöfer M and Wetzstein G 2019 Scene representation networks: continuous 3D-structure-aware neural scene reconstruction IEEE Trans. Neural Netw. Learn. Syst. (https://doi.org/10.1109/TNNLS.2022.3177134)
Shen L, Zhao W and Xing L 2019 Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning Nat. Biomed. Eng. 3 880–8
Shieh C C et al 2019 SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1 min scan Med. Phys. 46 3799–811
Sibolt P et al 2021 Clinical implementation of artificial intelligence-driven cone-beam computed tomography-guided online adaptive radiotherapy in the pelvic region Phys. Imaging Radiat. Oncol. 17 1–7
Shrestha P and Hirano Y 2021 Four Dimensional CT Using Implicit Neural Representation (https://tech.preferred.jp/en/blog/four-dimensional-cbct-using-implicit-neural-representation/)
Sitzmann V et al 2020 Implicit neural representations with periodic activation functions Adv. Neural Inf. Process. Syst. 33 7462–73
Sitzmann V, Zollhofer M and Wetzstein G 2019 Scene representation networks: continuous 3D-structure-aware neural scene representations 33rd NeurIPS 2019 (https://proceedings.neurips.cc/paper/2019/file/b5dc4e3d9b495d0196f61d45b26ef33e-Paper.pdf)
Song W Y et al 2008 A dose comparison study between XVI® and OBI® CBCT systems Med. Phys. 35 480–6
Sonke J J et al 2005 Respiratory correlated cone beam CT Med. Phys. 32 1176–86
Sweeney R A et al 2012 Accuracy and inter-observer variability of 3D versus 4D cone-beam CT based image-guidance in SBRT for lung tumors Radiat. Oncol. 7 1–8
Tancik M, Srinivasan P, Mildenhall B, Fridovich-Keil S, Raghavan N, Singhal U, Ramamoorthi R, Barron J and Ng R 2020 Fourier features let networks learn high frequency functions in low dimensional domains Adv. Neural Inf. Process. Syst. 33 7537–47
Thengumpillil S et al 2016 Difference in performance between 3D and 4D CBCT for lung imaging: a dose and image quality analysis J. Appl. Clin. Med. Phys. 17 99–106
Topolnjak R et al 2010 Breast patient setup error assessment: comparison of electronic portal image devices and cone-beam computed tomography matching results Int. J. Radiat. Oncol. Biol. Phys. 78 1235–43
van Aarle W et al 2016 Fast and flexible x-ray tomography using the ASTRA toolbox Opt. Express 24 25129–47
Vasudevan V et al 2022 Implicit neural rendering for radiation therapy dose distribution Phys. Med. Biol. 67 125014
Vergalasova L, Maurer J and Yin F F 2011 Potential underestimation of the internal target volume (ITV) from free-breathing CBCT Med. Phys. 38 4689–99
Wang J and Gu X 2013 Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone-beam CT Med. Phys. 40 101912
Wang J, Li T and Xing L 2009 Iterative image reconstruction for CBCT using edge-preserving prior Med. Phys. 36 252–60
Wei R et al 2020 Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN) Phys. Med. Biol. 65 085012
Yan H et al 2014 A hybrid reconstruction algorithm for fast and accurate 4D cone-beam CT imaging Med. Phys. 41 071903
Yasue K et al 2022 Quantitative analysis of the intra-breath respiratory motion with baseline drift for respiratory-gating lung stereotactic body radiation therapy J. Radiat. Res. 63 137–47
Zhang Y et al 2013 A technique for estimating 4D-CBCT using prior knowledge and limited-angle projections Med. Phys. 40 121701
Zhang Y et al 2015b Preliminary clinical evaluation of a 4D-CBCT estimation technique using prior information and limited-angle projections Radiother. Oncol. 115 22–9
Zhang Y et al 2017 A new CT reconstruction technique using adaptive deformation recovery and intensity correction (ADRIC) Med. Phys. 44 2223–41
Zhang Y et al 2019 4D liver tumor localization using cone-beam projections and a biomechanical model Radiother. Oncol. 133 183–92
Zheng Y, Yin F F and Ren L 2015a Dosimetric verification of lung cancer treatment using the CBCTs estimated from limited-angle on-board projections Med. Phys. 42 4783–95