Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

4D-Precise: Learning-based 3D motion estimation and high temporal resolution 4DCT reconstruction from treatment 2D+t X-ray projections

Arezoo Zakeri [a,*], Alireza Hokmabadi [b,1], Michael G. Nix [c], Ali Gooya [d,e], Isuru Wijesinghe [f], Zeike A. Taylor [f,*]

[a] Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Computing, University of Leeds, Leeds, UK
[b] Department of Infection, Immunity & Cardio Disease, University of Sheffield, Sheffield, UK
[c] Leeds Cancer Centre, Leeds Teaching Hospitals NHS Trust, UK
[d] School of Computing Science, University of Glasgow, Glasgow, UK
[e] Alan Turing Institute, London, UK
[f] Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Mechanical Engineering, University of Leeds, Leeds, UK

ARTICLE INFO

Keywords:
Learning-based spatio-temporal deformation estimation
Treatment imaging 4DCT reconstruction
Respiratory motion modelling
Digitally reconstructed radiographs
Probabilistic motion modelling
Recurrent variational Bayes

ABSTRACT

Background and Objective: In radiotherapy treatment planning, respiration-induced motion introduces uncertainty that, if not appropriately considered, could result in dose delivery problems. 4D cone-beam computed tomography (4D-CBCT) has been developed to provide imaging guidance by reconstructing a pseudo-motion sequence of CBCT volumes through binning projection data into breathing phases. However, it suffers from artefacts and erroneously characterizes the averaged breathing motion. Furthermore, conventional 4D-CBCT can only be generated post-hoc using the full sequence of kV projections after the treatment is complete, limiting its utility. Hence, our purpose is to develop a deep-learning motion model for estimating 3D+t CT images from treatment kV projection series.

Methods: We propose an end-to-end learning-based 3D motion modelling and 4DCT reconstruction model named 4D-Precise, abbreviated from **P**robabilistic **rec**onstruction of **i**mage **se**quences from CBCT kV projections. The model estimates voxel-wise motion fields and simultaneously reconstructs a 3DCT volume at any arbitrary time point of the input projections by transforming a reference CT volume. Developing a Torch-DRR module, it enables end-to-end training by computing Digitally Reconstructed Radiographs (DRRs) in PyTorch. During training, DRRs with matching projection angles to the input kVs are automatically extracted from reconstructed volumes and their structural dissimilarity to inputs is penalised. We introduced a novel loss function to regulate spatio-temporal motion field variations across the CT scan, leveraging planning 4DCT for prior motion distribution estimation.

Results: The model is trained patient-specifically using three kV scan series, each including over 1200 angular/temporal projections, and tested on three other scan series. Imaging data from five patients are analysed here. Also, the model is validated on a simulated paired 4DCT-DRR dataset created using the Surrogate Parametrised Respiratory Motion Modelling (SuPReMo). The results demonstrate that the reconstructed volumes by 4D-Precise closely resemble the ground-truth volumes in terms of Dice, volume similarity, mean contour distance, and Hausdorff distance, whereas 4D-Precise achieves smoother deformations and fewer negative Jacobian determinants compared to SuPReMo.

Conclusions: Unlike conventional 4DCT reconstruction techniques that ignore breath inter-cycle motion variations, the proposed model computes both intra-cycle and inter-cycle motions. It represents motion over an extended timeframe, covering several minutes of kV scan series.

[*] Corresponding author at: Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), University of Leeds, Leeds, UK.
E-mail addresses: a.zakeri@leeds.ac.uk (A. Zakeri), z.taylor@leeds.ac.uk (Z.A. Taylor).

[1] The first two authors contributed equally.

https://doi.org/10.1016/j.cmpb.2024.108158
Received 8 December 2023; Received in revised form 23 March 2024; Accepted 29 March 2024
Available online 4 April 2024

1. Introduction

Respiratory motion is of great concern for image-guided radiotherapy (IGRT) when treating tumours in the abdominal and thoracic organs. Respiratory-correlated cone-beam computed tomography (CBCT) or 4D-CBCT imaging considers the respiratory motion by sorting the CBCT projections into several breathing phase bins, using diaphragm position as a surrogate signal and then independently constructing the CBCT volumes at different phases using the clinical Feldkamp-Davis-Kress (FDK) algorithm [1]. However, in this conventional method, an adequate number of projections, from a range of angles, must be made for each phase bin; otherwise, low-quality 4D-CBCT images with severe streaking artefacts will result. 4D-CBCT imaging is commonly impacted by irregular breathing, leading to inaccurate phase binning and image artefacts that limit the ability to correctly localize moving targets and organs at risk. Uncertainty in the delivered radiation dose results, limiting the safe deliverable dose and hence the treatment efficacy. There is a critical need to develop methods for estimating the 4D motion of internal anatomy using limited kV projection imaging, ideally in real time.

To reduce imaging dose, hardware-based methods of respiratory-triggered 4D-CBCT [2] and software-based novel reconstruction techniques [3–7] have been developed to improve the image quality from an under-sampled acquisition. In the studies proposed by Jia et al. [3], Zhang et al. [4], 4D-CBCT reconstruction is achieved by defining regularization techniques based on the temporal non-local mean energy. Several studies have used the total variation (TV) metric for iterative few-view image reconstruction [6], where improved results have been reported by incorporating prior images into the objective function of the iterative image reconstruction process [8,9]. However, in these approaches, as the forward and backprojection phases are carried out iteratively to determine the fidelity between measured and estimated projections, their utility is restricted by their high computational demands in actual applications [7].

Over the past decade, there has been a significant interest in developing motion models to deal with the problem of respiratory motion in 4DCT/CBCT reconstruction for IGRT [5,10–15]. Wang and Gu [10] introduced the SMEIR method for simultaneous motion estimation and image reconstruction via alternating steps of reconstructing a motion-compensated primary CBCT (m-pCBCT) image and deformation field estimation between the m-pCBCT image and other 4D-CBCT phases. However, for the reconstruction of the m-pCBCT image, all the measured projections from an entire set of 4D-CBCT are utilised, which is updated during forward and backprojection steps. The SMEIR technique updates the motion fields by matching the forward projection of the deformed m-pCBCT and measured projections from other phases of 4D-CBCT [10]. Later versions of the SMEIR algorithm termed SMEIR-Bio [16] and SMEIR-Unet [15] were proposed by this team, incorporating biomechanical modelling and deep learning into the base algorithm. Li et al. [11] have shown the feasibility of real-time 4DCT/CBCT reconstruction based on a single-view X-ray projection by optimizing a principal component analysis (PCA) motion model, given a prior set of 4DCT/CBCT volumetric images. The model updates the PCA coefficient such that the projection of a corresponding deformed reference volume matches the measured projection. PCA is frequently used to develop motion models for the lung [17,18] and liver [5]. However, this approach approximates the motion as a linear combination of a small number of eigenvectors, meaning some motion details inevitably are lost [12]. Other techniques such as free-form deformation [12] or biomechanical modelling [5] were utilised to fine-tune the motion fields obtained from PCA.

Surrogate-driven respiratory motion models [13,19] relate the motion of the internal anatomy to respiratory signals acquired externally, e.g. by tracking the displacement of markers on the thoraco-abdominal surface or measuring the pressure variation of an elastic belt in these regions. McClelland et al. [13] proposed a surrogate-driven approach that unifies image registration from a number of dynamic images and fitting a respiratory correspondence model into a single optimisation. Recently, Huang et al. [20] extended the general framework in [13] to fit a motion model directly on CBCT projections. This technique estimates the motion for every projection and uses that motion to reconstruct a single 3D motion-compensated image. This involves alternate and iterative processes of motion model fitting and motion-compensated reconstruction. An inherent assumption of surrogate-driven respiratory motion models is that the internal anatomical motion is well-correlated with the surrogate signals. Concerns about this assumption were raised for example by Yan et al. [21] and Gierga et al. [22].

These analytical solutions for respiration-induced motion modelling, although valuable, are constrained by some important challenges, such as parameter selection and initialization, difficulties in handling multiple objects, and poor single-view performance. Recent advances in deep learning (DL) techniques for statistical organ motion modelling (cf. our recent work on cardiac cine magnetic resonance imaging [23,24]) enable enhanced reliability in managing the complexity, diversity, and high-dimensionality of the data. DL potentially can deliver high-speed, high-quality reconstructions from low-dose images tailored to the specific needs of individual patients, while also producing a natural scan [25,26]. DL has been used for CT image denoising [27], or improving the accuracy of intra-lung deformation vector fields estimated by an analytical motion model [15]. Shao et al. [28] proposed a framework combining graph neural network-based deep learning and biomechanical modelling to track liver tumours in real-time from a single onboard X-ray projection. For the purpose of CT volume reconstruction from limited X-ray views, some DL approaches have been proposed recently [29–31]. Henzler et al. [29] suggested a convolutional neural network architecture to learn the mapping from an X-ray image to a 3D volume from many mammalian skull CT volumes paired with single-view 2D X-rays. Ying et al. [30] proposed a framework based on generative adversarial networks (GANs) (X2CT-GAN model) to reconstruct a CT volume from two orthogonal X-ray projections. However, these frameworks are supervised learning models and due to the absence of a paired X-ray and CT dataset, synthesized X-rays were utilised for training the model. The synthetic X-rays were generated using a CycleGAN [32] or Contrastive Unpaired Translation (CUT) [31], and Digitally Reconstructed Radiograph (DRR) technique [33]. Ying et al. [30] treat each X-ray image independently, assuming there is no data shift caused by patient motions. Hence, they do not model patient motion in the observed X-rays for reconstructing CT volumes corresponding to each time point.

To address these limitations, we propose a novel DL framework for high temporal resolution 3D+t CT reconstruction from treatment 2D+t kV projections (X-ray images) scanned the entire 360°around the patient. Our model incorporates temporal/angular dependencies across a sequence of images, enabling the estimation of dynamic changes over time. The current work is a proof of principle, employing patient specific models, trained on data from the first 3 RT fractions, and validated on the remaining fractions. Clinical translation of the proposed framework will require two further developments, namely population-level general models and intra-fraction kV projection acquisition (i.e. during MV beam delivery). Both these aspects will be addressed in future work. Here, a kV image refers to a single CBCT projection acquired from a linac gantry-mounted on-board X-ray imaging system. The proposed framework, named 4D-Precise, is composed of two main parts. The first part is a probabilistic motion model that uses UNet structures distributed over time, and whose latent space variables are fed into a long short-term memory (LSTM) network. The second part is responsible for generating DRRs from the reconstructed 4DCT in the forward pass of the model training. Voxel-wise displacement fields are modelled using a Gaussian distribution in a variational inference formulation and are used to deform a reference CT. Our main contributions are:

- We propose an end-to-end learning-based spatio-temporal motion model (4D-Precise) for 3DCT reconstruction at any arbitrary time points from treatment kV scan series. It is an explicit motion model that computes the probability distribution of the 3D motion fields using recurrent variational Bayes through time-distributed UNet structures in combination with recurrent neural networks for modelling temporal dependencies.

- We demonstrate the unsupervised estimation of parameters of the probability distribution function of motion fields at the voxel level, where employing a Torch-DRR module enables end-to-end training by quick extraction of DRRs from the deformed states and similarity measurements at the projection level with the input kVs.
- We introduce a novel loss function to regulate the spatio-temporal motion field variations across the entire CT scan, leveraging planning 4DCT for estimating the prior distribution of motion.
- To evaluate the model performance, we conducted both qualitative and quantitative assessments on over 19,000 3DCT and projection pairs estimated for five patients during multiple treatment sessions compared to the real treatment X-ray projections. We also validated the model's performance using a simulated paired 4DCT-DRR dataset. This simulated data was created solely to validate the method against a ground-truth, which is not available for real-world data.

The rest of this paper is organized as follows: Section 2 describes the proposed model, Section 3 presents the experimental analysis and results, and in Section 4, we discuss the results, draw conclusions, and suggest directions for future work.

2. Methodology

We introduce the 4D-Precise formulation, which aims to estimate spatio-temporal motion fields in 3D space from a sequence of kV projections measured at different angles. More formally, the model seeks to compute the probability distribution of the displacement vector fields (DVFs) per time step t, which spatially transform a reference CT volume, $\mathbf{CT}_{Ref}$, into a volume from which the extracted DRR well matches the input 2D kV projection. To accomplish this, we leverage planning 4DCT to establish a prior distribution for motion fields across ten distinct phases of a breathing cycle. Additionally, we utilize the Amsterdam Shroud (AS) method to assign each kV projection to a specific respiratory phase by analysing the respiratory signal extracted from the kV projections. These pre-processing steps are required for spatio-temporally controlling motion fields in the learning procedure. Fig. 1 illustrates the pipeline of the proposed method. The mathematical notation used throughout the paper is defined in Table 1.

Table 1
Definition of the variables used in our model.

Notation	Description
$\mathbf{kV}_t^a$	The kV projection at gantry angle of a and time t
$\mathbf{kV}_{\leq T}^{\leq A}$	An observed sequence of kV projections covering the gantry angles $[a_1, A]$ and time indices $[t_1, T]$ $(\{\mathbf{kV}_{t_1}^{a_1}, \cdots, \mathbf{kV}_T^A\})$
$\mathbf{CT}_{Ref}$	A reference CT volume
$\mathbf{D}_t$	Displacement field map at time t
p	Respiratory phase, $p \in [0, 9]$
$p(\mathbf{D}_{pi}^p)$	Prior probability distribution of displacement fields at phase p, $\mathbf{D}_{pi}^p$.
$\mathcal{N}$	Gaussian distribution
$\boldsymbol{\mu}_{\mathbf{D}_{pi}^p}$	Mean of the prior probability of displacements at phase p computed using NiftyReg
$\mathbf{h}_t$	LSTM hidden state variables at time t
$\boldsymbol{\mu}_{\mathbf{D}_t}$	Mean of the posterior probability distribution of displacements at time t
$\boldsymbol{\Sigma}_{\mathbf{D}_t}$	Covariance of the posterior probability distribution of displacements at time t
b_j	j^{th} bone voxel in the reference CT
v_i	i^{th} voxel in the reference CT
$\mathbf{DRR}_t^a$	Estimated projection by model at gantry angle of a and time t

2.1. Pre-processing procedure

2.1.1. Respiratory signal extraction and obtaining phase-binned kV projections

We improved local contrast in the kV projections using the Contrast Limited Adaptive Histogram Equalization (CLAHE) approach [34] with a clipping level of 5 and a tile grid size of 3×3. Amsterdam Shroud technique [21] is used to extract an image-based respiratory signal from the sequence of kV projections. We used the Reconstruction Toolkit [35] for the implementation. Once the signal was extracted, the Hilbert transform was applied to obtain instantaneous phase variations, which were then divided into 10 bins. Subsequently, each signal point (and its corresponding kV projection) was assigned to the appropriate phase bin ($p \in [0, 9]$).

2.1.2. Estimation of a prior distribution for motion fields

We utilized pre-treatment 4DCT data (planning 4DCT), comprising 10 CT volumes, and employed the NiftyReg deformable image registration method [36] to estimate displacements between a reference CT and the available CT volumes within the planning 4DCT (Fig. 1). The setting parameters for the NiftyReg are presented in Appendix A. Here, we used the CT image at the end-expiration phase as the reference volume due to its relative stability and minimal motion artefacts. Inspired by previous works [23,37], we assumed a Gaussian prior distribution for the motion fields. The estimated displacements using NiftyReg were considered the means of multivariate normal distributions with unit covariance (identity matrices $\mathbf{I}$) at each phase $p \in [0, 9]$:

$$p(\mathbf{D}_{pi}^p) = \mathcal{N}(\mathbf{D}_{pi}^p; \boldsymbol{\mu}_{\mathbf{D}_{pi}^p}, \mathbf{I}) \tag{1}$$

where the subscript pi here indicates the prior distribution, and $\boldsymbol{\mu}_{\mathbf{D}_{pi}^p}$ refers to the mean of the prior probability distribution of displacement fields at phase p. Using multivariate Gaussian offers a closed-form differentiable solution, and the choice of unit covariance for the prior distribution of the motion fields encourages the DVFs to be distributed evenly and smoothly, acting as a regularization mechanism. This helps in generating diverse and smooth samples during training and inference.

2.2. 4D-Precise model

Fig. 1 shows the structure of the proposed model. We aimed to model the joint distribution $p(\mathbf{kV}_{\leq T}^{\leq A}, \mathbf{D}_{\leq T})$ to solve spatio-temporal 3D motions $\{\mathbf{D}_{t_1}, \cdots, \mathbf{D}_T\}$ from a given sequence of kV projections $\{\mathbf{kV}_{t_1}^{a_1}, \cdots, \mathbf{kV}_T^A\}$, covering the gantry angles $[a_1, A]$ and time indices $[t_1, T]$ as:

$$p(\mathbf{kV}_{\leq T}^{\leq A}, \mathbf{D}_{\leq T}) = \prod_{t,a} p(\mathbf{kV}_t^a | \mathbf{kV}_{<t}^{<a}, \mathbf{D}_{\leq t}) p(\mathbf{D}_t | \mathbf{D}_{<t}, \mathbf{kV}_{<t}^{<a}) \tag{2}$$

Eq. (2) indicates that the likelihood of projection $\mathbf{kV}_t^a$, acquired at gantry angle $a \in [a_1, A]$ and time $t \in [t_1, T]$, depends on a set of preceding projections $\mathbf{kV}_{<t}^{<a}$ and the displacement field maps $\mathbf{D}_{\leq t}$. The model used an LSTM network [38] to learn spatio-temporal dependencies between projections, which were then encoded in the LSTM hidden state variables $\mathbf{h}_t$. Hence, we modelled the dependencies among the preceding projections and displacement field maps (i.e., $\mathbf{kV}_{<t}^{<a}, \mathbf{D}_{<t}$) through the hidden state variable $\mathbf{h}_{t-1}$ of the LSTM and obtained the following factorization replacing Eq. (2)

$$p(\mathbf{kV}_{\leq T}^{\leq A}, \mathbf{D}_{\leq T}) = \prod_{t,a} p(\mathbf{kV}_t^a | \mathbf{h}_{t-1}, \mathbf{D}_t) p(\mathbf{D}_t | \mathbf{h}_{t-1}) \tag{3}$$

Therefore, the joint distribution in Eq. (2) reduced to the factorization of likelihood $p(\mathbf{kV}_t^a | \mathbf{h}_{t-1}, \mathbf{D}_t)$ and the prior probability of $p(\mathbf{D}_t | \mathbf{h}_{t-1})$. Still, to keep the model simple, we obtained the prior distribution of

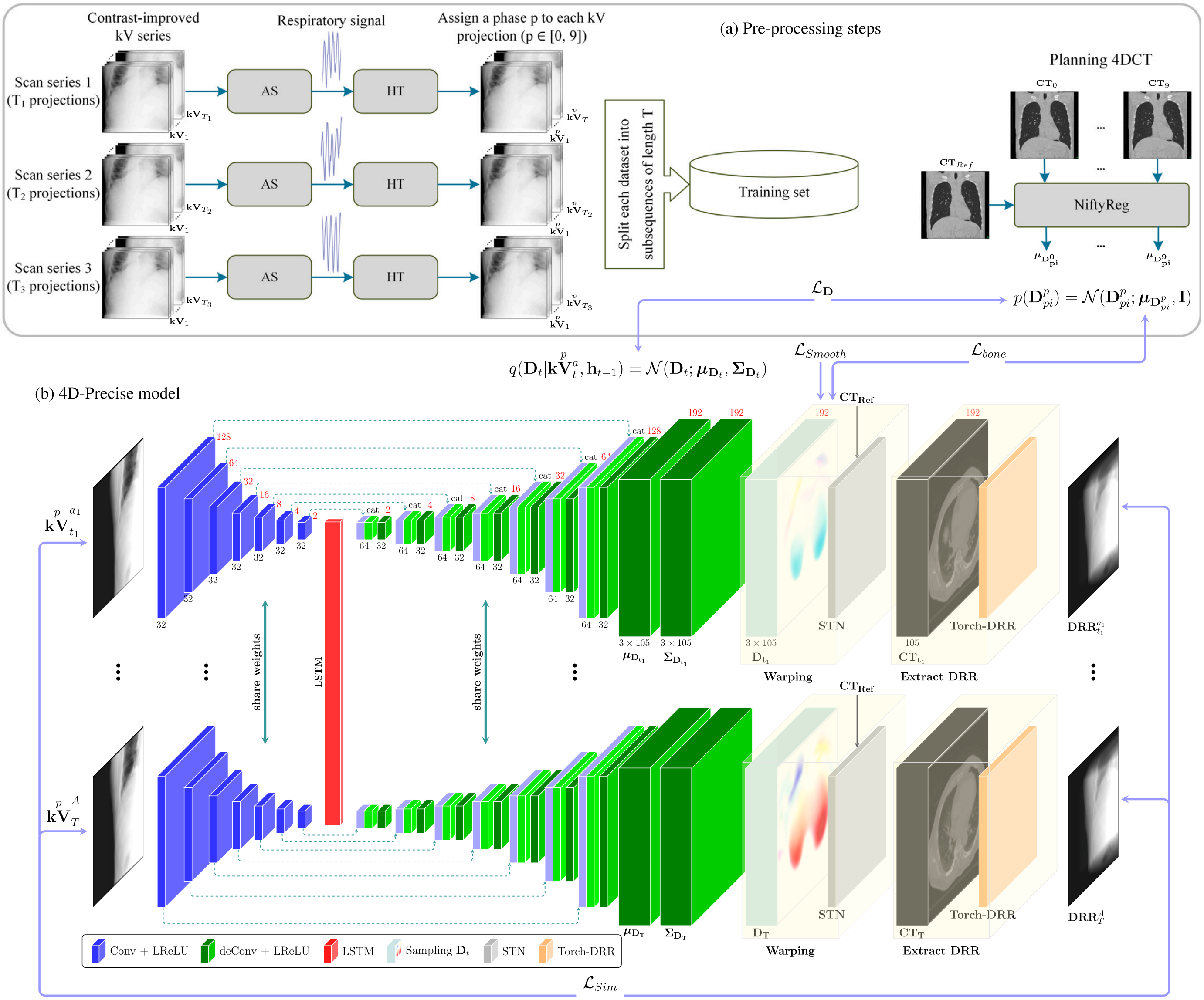


Fig. 1. Pipeline of the proposed method, including (a): Pre-processing steps include creating a training dataset from three datasets of kV projections after applying the Amsterdam Shroud (AS) method and Hilbert transform (HT) to obtain phase-binned projections. A prior motion distribution is estimated using the planning 4DCT and NiftyReg at each respiratory phase. (b): Structure of the 4D-Precise model. Given a sub-sequence of kV projections $\{\mathbf{kV}^{a_1}_{t_1}, \cdots, \mathbf{kV}^{A}_{T}\}$, the model inferred the spatio-temporal posterior distributions for displacement fields $\{\mathbf{D}_{t_1}, \cdots, \mathbf{D}_{T}\}$. This was achieved by also conditioning on the hidden states of an LSTM network that captured temporal dependencies in the latent space. Subsequently, at each time step, sampled motion fields were used to transform the reference CT volume into deformed states, denoted as $\{\mathbf{CT}_{t_1}, \cdots, \mathbf{CT}_{T}\}$, using STN. Next, Torch-DRR was used to extract DRRs from the deformed volumes that correspond to each input kV instance.

motion fields from Eq. (1) without conditioning it on history information $\mathbf{h}_{t-1}$, which is obtained from kVs.

Our goal was to estimate the posterior probability distribution of motion fields given the kV projections, which can be expressed as $p(\mathbf{D}_{\leq T}|\mathbf{kV}^{\leq A}_{\leq T})$. However, it is computationally infeasible to obtain an exact calculation of this posterior probability. We use a variational approach introduced by [37], and assume an approximate posterior probability $q(\mathbf{D}_{\leq T}|\mathbf{kV}^{\leq A}_{\leq T})$ as:

$$q(\mathbf{D}_{\leq T}|\mathbf{kV}^{\leq A}_{\leq T}) = \prod_{t,a} q(\mathbf{D}_t|\mathbf{D}_{<t}, \mathbf{kV}^{\leq a}_{\leq t})$$
$$= \prod_{t,a} q(\mathbf{D}_t|\mathbf{kV}^{a}_{t}, \mathbf{h}_{t-1}) \tag{4}$$

where the approximate posteriors $q(\mathbf{D}_t|\mathbf{h}_{t-1}, \mathbf{kV}^{a}_{t})$ were learned from a combined UNet-LSTM-based structure to capture image features and spatio-temporal dependencies. We modelled $q(\mathbf{D}_t|\mathbf{kV}^{a}_{t}, \mathbf{h}_{t-1})$ as a multivariate normal distribution with mean $\boldsymbol{\mu}_{\mathbf{D}_t}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{D}_t}$ (see Fig. 1):

$$q(\mathbf{D}_t|\mathbf{kV}^{a}_{t}, \mathbf{h}_{t-1}) = \mathcal{N}(\mathbf{D}_t; \boldsymbol{\mu}_{\mathbf{D}_t}, \boldsymbol{\Sigma}_{\mathbf{D}_t}) \tag{5}$$

Using the motion posteriors, we could obtain the most likely motion fields $\mathbf{D}_t$ that transformed a reference volume to a deformed volume with a DRR matching the given projection at time t and angle a (i.e., $\mathbf{kV}^{a}_{t}$). To achieve this, we employed a spatial transformer network (STN) [39], which is frequently used in learning-based registration models [23] for deforming the reference volume in a fully differentiable manner. In our model, image similarity optimization during training was performed at the projection level between the DRR extracted from the deformed volume and the input $\mathbf{kV}^{a}_{t}$. To this end, the forward pass of the training model needs to project DRRs from 3DCT volumes. To accomplish this, we have developed the Torch-DRR module, which implements the ray-tracing algorithm in PyTorch. This module enables our end-to-end learning model, allowing for the DRR extraction in a differentiable manner during training. Torch-DRR reformulates the DeepDRR approach introduced by Unberath et al. [40], implementing vectorized tensor-based operations within PyTorch, in contrast to DeepDRR's PyCUDA-based implementation. Using DeepDRR simultaneously with PyTorch raises some issues due to CUDA initialization conflicts [40]. Torch-DRR is fast, runs on the GPU, and simulates a forward projection in about 0.6 seconds for a volume of 192 × 192 × 105 voxels, resulting in a projection of 128 × 128 pixels.

2.3. Model objective function

The overall motion generative model $p(\mathbf{kV}_{\leq T}, \mathbf{D}_{\leq T})$, including the inference model $q(\mathbf{D}_{\leq T} | \mathbf{kV}_{\leq T})$ and the recurrent network were trained jointly by maximising a variational Evidence Lower Bound (ELBO) [37] with respect to their parameters using stochastic gradient methods. In addition, other constraints were defined in the model loss function to regulate motion fields in the bones ($\mathcal{L}_{bone}$) as well as over the entire volume, producing spatially smooth deformations ($\mathcal{L}_{Smooth}$). Therefore, the overall loss function was the sum of the terms given by:

$$\mathcal{L} = -\mathcal{L}_{ELBO} + \lambda_1 \mathcal{L}_{bone} + \lambda_2 \mathcal{L}_{Smooth} \tag{6}$$

where λ_1 and λ_2 represent the weighting coefficients for the corresponding loss terms. $\mathcal{L}_{ELBO}$ is defined as

$$\mathcal{L}_{ELBO} = \mathbb{E}_{q(\mathbf{D}_{\leq T} | \mathbf{kV}^{\leq A}_{\leq T})} \log \frac{p(\mathbf{kV}^{\leq A}_{\leq T}, \mathbf{D}_{\leq T})}{q(\mathbf{D}_{\leq T} | \mathbf{kV}^{\leq A}_{\leq T})} \tag{7}$$

Using Eqs. (3) and (4), the ELBO term can be written as

$$\mathcal{L}_{ELBO} = \mathbb{E}_{\prod_{t,a} q(\mathbf{D}_t | \mathbf{kV}^a_t, \mathbf{h}_{t-1})} \left[\sum_{t,a} \log p(\mathbf{kV}^a_t | \mathbf{h}_{t-1}, \mathbf{D}_t) \right. \\ \left. + \log \frac{p(\mathbf{D}_t | \mathbf{h}_{t-1})}{q(\mathbf{D}_t | \mathbf{kV}^a_t, \mathbf{h}_{t-1})} \right] \tag{8}$$

which can be decomposed into two terms as follows

$$\mathcal{L}_{ELBO} = \mathcal{L}_{Sim} + \mathcal{L}_{\mathbf{D}} \tag{9}$$

where $\mathcal{L}_{Sim}$ controls the similarity between the estimated projection by the 4D-Precise model (i.e., $\mathbf{DRR}^a_t$ in Fig. 1) and the input kV at time t associated with angle a, $\mathbf{kV}^a_t$. On the other hand, the probability distribution of the motion fields is controlled by $\mathcal{L}_{\mathbf{D}}$. We can compute the $\mathcal{L}_{Sim}$ as

$$\mathcal{L}_{Sim} \simeq \sum_{t,a} \frac{1}{L} \sum_{l=1}^{L} \log p(\mathbf{kV}^a_t | \mathbf{h}_{t-1}, \mathbf{D}^{(l)}_t) \tag{10}$$

where the expectation over $q(\mathbf{D}_t | \mathbf{kV}^a_t, \mathbf{h}_{t-1})$ in Eq. (8) was taken empirically using L Monte Carlo samples (i.e., $\mathbf{D}^{(l)}_t \sim q(\mathbf{D}_t | \mathbf{kV}^a_t, \mathbf{h}_{t-1})$). The log-likelihood in Eq. (10) is equivalent to a measure of similarity between the observed kV projections and estimated DRRs, which can also be represented by the mean square error metric. While the estimated DRRs and the kV projections share similarities in terms of the underlying anatomy, their intensity levels are different. Therefore, the mean square error, which is highly dependent on intensity, was not an appropriate metric for similarity measurement in our application. Instead, we approximated it with the structural similarity index measure (SSIM) [41] and compared the spatial gradient maps of the estimated and input projections:

$$\mathcal{L}_{Sim} \simeq \frac{1}{L} \sum_{t,a} \sum_{l=1}^{L} SSIM(\|\nabla \mathbf{kV}^a_t\|, \|\nabla \mathbf{DRR}^{a(l)}_t\|), \tag{11}$$

where the SSIM between two images $\mathbf{X}$ and $\mathbf{Y}$ was defined as follows:

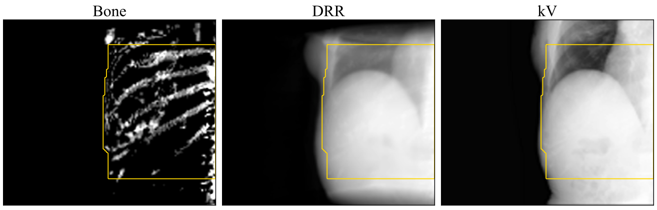


Fig. 2. Angle-dependent region of interest in a projection obtained at the angle of 32°. Left: The region of interest is specified by segmenting bones in the bone component of the ray-tracing algorithm, while the upper and lower parts of the image were discarded. Middle: Mapped ROI on the DRR; Right: Mapped ROI on the real kV.

$$SSIM(\mathbf{X}, \mathbf{Y}) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{12}$$

where μ_x and μ_y are the average values over all pixels in images $\mathbf{X}$ and $\mathbf{Y}$, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance of $\mathbf{X}$ and $\mathbf{Y}$, c_1 and c_2 are constants that are added to stabilize the division with weak denominator, respectively.

The simulated DRRs generated by the ray-tracing algorithm exhibited halo artefacts at the top and bottom regions of the DRRs, due to the 3DCT volume limits. These artefacts are known to cause discrepancies with the input KV projections in these regions. To mitigate the effect of these discrepancies on the abdominal organ motion modelling particularly for the liver, we implemented an adaptive region of interest (ROI) for the SSIM loss computation at each time step. This ROI was adjusted with the changing projection angle and included the internal anatomies within the rib cage while excluding the affected areas with halo artefacts. To define the ROI at each angle, we utilized the Torch-DRR module to extract the bone component of the DRR. Then, we discarded the upper and lower parts of the projections to remove the corresponding halos in DRRs. We segmented the bones in the remaining image using Otsu's thresholding method, resulting in a binary mask primarily indicating the bone regions. Next, a morphological closing operation with a 13 × 13 square structuring element was applied to close holes within the rib cage and connect components together. Finally, the largest contour was identified and filled in to obtain the ROI. Fig. 2 shows a sample ROI for a projection angle of 32°, derived from the bone component and mapped on the DRR and the kV images.

The term $\mathcal{L}_{\mathbf{D}}$ in Eq. (9) denotes the Kullback-Leibler (KL) divergence ($\mathcal{D}_{\mathcal{KL}}$) [42] between the approximate posterior $q(\mathbf{D}_t | \mathbf{kV}^a_t, \mathbf{h}_{t-1})$ and the prior distribution $p(\mathbf{D}_t | \mathbf{h}_{t-1})$ and is given by

$$\mathcal{L}_{\mathbf{D}} = -\sum_{t,a} \mathcal{D}_{\mathcal{KL}} \Big(q(\mathbf{D}_t | \mathbf{kV}^a_t, \mathbf{h}_{t-1}) \| p(\mathbf{D}_t | \mathbf{h}_{t-1}) \Big). \tag{13}$$

The derivation of Eq. (13) is presented in Appendix B.

The term $\mathcal{L}_{bone}$ in Eq. (6) was used to constrain the motion of the bones. To accomplish this, we segmented the bones in the reference CT volume using the Hounsfield Unit (HU) values in the range of [200, 3000] and morphological operations. The objective of $\mathcal{L}_{bone}$ was to minimize the amplitude difference between motion estimated in the bone voxels and the mean of the prior distribution of the motion fields ($\boldsymbol{\mu}_{\mathbf{D}_t, pi}$):

$$\mathcal{L}_{bone} = \sum_{t} \sum_{j=1}^{M} (\|\mathbf{D}_t(b_j)\| - \|\boldsymbol{\mu}_{\mathbf{D}_t, pi}(b_j)\|)^2 \tag{14}$$

where b_j indicates the j^{th} bone voxel in the reference CT volume. In other words, $\mathcal{L}_{bone}$ helped to ensure that the motion within the bones was consistent with the expected motion patterns based on the prior distribution of motion fields.

$\mathcal{L}_{Smooth}$ in Eq. (6) serves as a diffusion regularizer and encourages a smooth displacement field $\mathbf{D}_t$ by computing the spatial gradients of displacements:

$$\begin{array}{c}\text{1276 subsequences}\\ \Big\downarrow\end{array}\begin{pmatrix} 0 & 2 & 4 & 6 & 8 & 10 & 12 \\ 1 & 3 & 5 & 7 & 9 & 11 & 13 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 \\ & & & \vdots & & & \\ 1275 & 1277 & 1279 & 1281 & 1283 & 1285 & 1287 \end{pmatrix}$$

Fig. 3. An example illustrating a dataset index matrix for a patient's kV scan series, which contains 1287 projections within a treatment fraction; These projections are subdivided into 1276 subsequences, each with a length of 7, and their indices are shown in each row.

$$\mathcal{L}_{Smooth} = \sum_{t} \sum_{i=1}^{N} \|\nabla \mathbf{D}_t(v_i)\|^2 \tag{15}$$

where v_i indicates the i^{th} voxel in the volume consisting of N voxels, and $\nabla \mathbf{D}_t(v_i) = (\frac{\partial \mathbf{D}_t(v_i)}{\partial x}, \frac{\partial \mathbf{D}_t(v_i)}{\partial y}, \frac{\partial \mathbf{D}_t(v_i)}{\partial z})$.

3. Experiments and results

3.1. Training/testing datasets and implementation

We analysed data from five liver cancer patients at Leeds Cancer Centre to train and evaluate our proposed model. All patients gave informed consent for their retrospective data to be used for research. This data consists of kV projections and a set of pre-treatment phase-binned 4DCT images (planning 4DCT) including ten 3DCT images. For each patient, there were several scan series of kV projections acquired at different treatment fractions. Projections were acquired using an Elekta XVI (Elekta AB, Stockholm, Sweden) system under standard 4D CBCT protocol settings, producing 1287-1416 projections during 4^+ min scan (scan angle: 360°, source-to-isocenter distance (SID): 1000 mm, source-to-detector distance (SDD): 1536 mm). The kV projections were acquired with a temporal resolution of 0.2 s, spatial resolution of $0.8\,\text{mm} \times 0.8\,\text{mm}$, and dimensional size of 512×512 pixels. To reduce the computational burden of the model, we downsampled the original projections to an image size of 128×128 pixels. The use of kV projections from 4D CBCT protocols was driven by retrospective data availability and is not a prerequisite of the 4D-Precise approach.

The planning 4DCT data consisted of ten phase-binned CT volumes representing the liver and parts of the lung. The acquisition system was a clinical Philips Brilliance Big Bore RT scanner [Koninklijke Philips N.V.]. The volumes had a size of $512 \times 512 \times 105$ and a voxel spacing of $0.976 \times 0.976 \times 2$ mm. To further reduce the computational burden, we cropped the volumes by 64 pixels from the right and left on each side to decrease the image background. We then downsampled the images by a factor of two, resulting in a volume size of $192 \times 192 \times 105$. The resulting volumes include only minimal image background.

To create a dataset of kV sequences for training our model, we needed to subdivide the long kV scan series into shorter subsequences of T-projection length. In our experiments, we set $T = 7$. Fig. 3 illustrates subdividing a kV scan series with 1287 projections into 1276 subsequences of length 7 using a dataset index matrix. Here, we considered an image gap between consecutive projections within a subsequence. This arrangement allows seven projections to cover a longer time interval of breathing.

To train our patient-specific motion model, we utilised kV subsequences from three distinct scan series obtained from the first three treatment fractions. For each patient, at least a total of 3×1276 kV subsequences were used for training, presenting inter-fraction respiratory motion variabilities. To test the model for each patient, we used three other scan series, collectively comprising over 3×1276 kV subsequences.

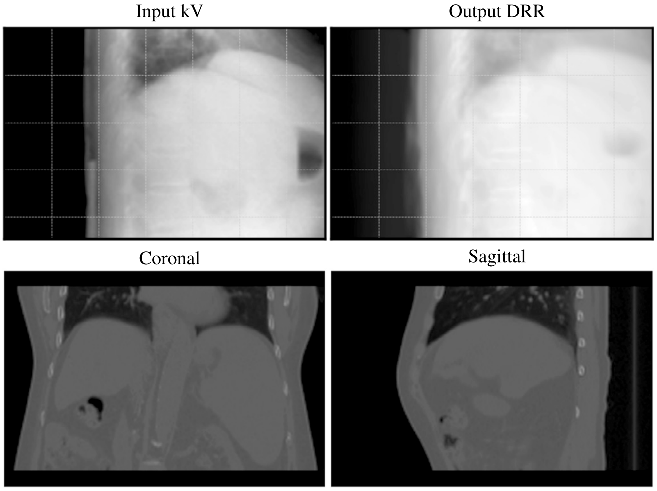


Fig. 4. A sample reconstructed 3DCT volume by the proposed model in the coronal and sagittal views corresponding to the observed kV image (input kV). The output DRR is obtained from the reconstructed volume at the same gantry angle as the input kV.

The whole framework is implemented using Python and PyTorch. The Adam optimiser [43] is used for optimising the loss function, with the learning rate of 0.001. The loss weight coefficients $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$ were set empirically. The model was trained for approximately 8 hours on an NVIDIA GeForce RTX 3080 GPU with 10 GB of memory. In contrast, it takes only 0.44 s seconds to compute the entire DVFs and construct seven 3DCT volumes for a test sequence containing seven X-ray projections.

3.2. Qualitative evaluation of the motion model

Fig. 4 shows a sample reconstructed CT volume by the proposed model corresponding to the observed kV projection. While the resulting DRR from this volume (output DRR) differs from the input kV in terms of intensity, it represents similar anatomical locations observed in the input kV image, such as the location of the diaphragm. This indicates that the reconstructed volume and output DRR have effectively learned the motion captured in the kV image.

We have also included in Supplementary Materials, movies for the full set of 15 predicted motion sequences, from the real-world validation data. While no ground truth is available (by definition) for these examples, they are useful to determine the subjective quality of the predicted motion. The observed motion predictions show realistic motions of the superior liver and associated anatomy, without obvious artefacts or discontinuous motion. However, certain aspects of the reconstructions appear less realistic. For example, there is some motion in the position of the ribs in the observation slice, which is not typically observed in 4DCT reconstructions. This indicates the model struggles to reconstruct the full 3D expansion of the rib-cage from single kV projections. This is both unsurprising and clinically of limited significance, as the additional attenuation of a rib is usually less than 1% of the target dose. Additionally, the motion expected at the inferior and extreme superior of the field of view (FoV) is often limited or absent. This is a direct consequence of the fact the kV FoV is smaller than the CT (prior) FoV, so deformations cannot be implied for this region, beyond the motion already contained in the 4DCT prior. Visual guide lines are presented on the 4DCT motion predictions to indicate the extent of the kV projection FoV. Again this limitation is of no clinical significance, as regions outside the treatment kV FoV are by definition also outside the MV treatment field, where RT doses are minimal.

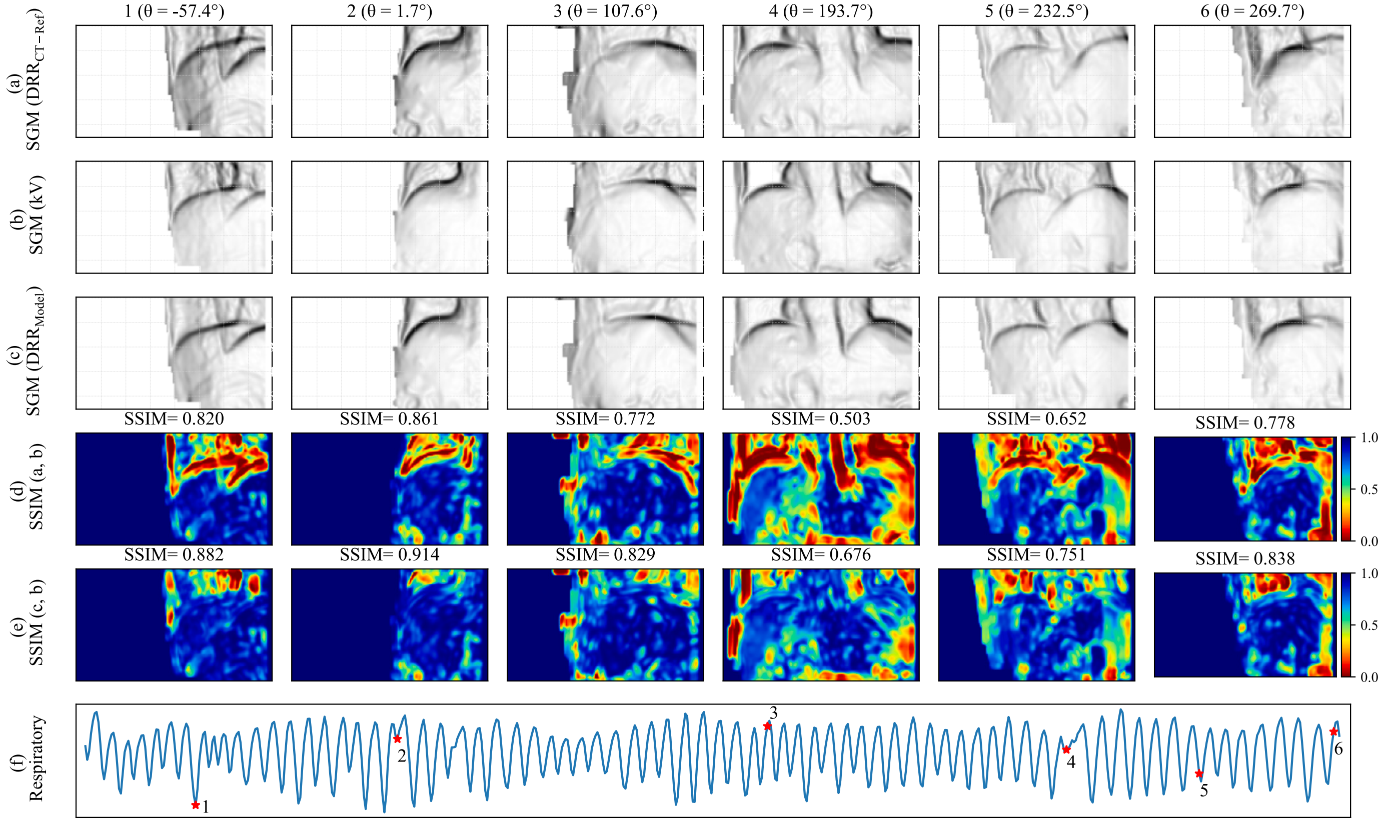


Fig. 5. Evaluation of anatomical similarities in the input kV series and the model's output DRRs. (a): Spatial gradient maps (SGMs) of the DRRs extracted from the reference CT at different angles corresponding to the input kVs, before applying the motion model; (b): SGMs of the input kV images; (c): SGMs of the corresponding model's output DRRs; (d): Local SSIM maps between images in rows (a) and (b); (e): Local SSIM maps between images in rows (b) and (c);(f): The respiratory signal obtained from the kV scan series, with six arbitrary time points marked for evaluating the model at each of them, as shown in each column. The average SSIM values are also reported for each time point, indicating improved SSIM values in the model outputs (i.e., row (e) vs row(d)). The SSIM values are particularly higher at the diaphragm locations in row (e) compared with row (d).

Table 2
Model evaluation in terms of the structural similarity metric (SSIM) between the observed kVs and estimated DRRs. The results are presented for two cases: DRRs of the reference CT computed before applying the motion model (MA: No) and after applying the motion model to the reference CT (MA: Yes). Average ± std values across the three test scan series are shown, considering gantry angles covering side/back views and front views. **Bold** values indicate that a significant difference between the two cases is observed (statistical significance p-value <0.001).

Subject	MA*	Test scan series 1		Test scan series 2		Test scan series 3	
		Side/back views	Front views	Side/back views	Front views	Side/back views	Front views
Patient 1	No	0.816 ± 0.03	0.642 ± 0.05	0.807 ± 0.03	0.621 ± 0.06	0.792 ± 0.07	0.628 ± 0.06
	Yes	**0.837 ± 0.02**	**0.661 ± 0.06**	**0.810 ± 0.02**	**0.635 ± 0.07**	**0.817 ± 0.07**	**0.647 ± 0.06**
Patient 2	No	0.709 ± 0.08	0.729 ± 0.07	0.720 ± 0.08	0.738 ± 0.07	0.709 ± 0.10	0.730 ± 0.07
	Yes	**0.741 ± 0.08**	**0.751 ± 0.07**	**0.751 ± 0.08**	**0.756 ± 0.07**	**0.742 ± 0.11**	**0.748 ± 0.08**
Patient 3	No	0.806 ± 0.06	0.502 ± 0.08	0.806 ± 0.05	0.517 ± 0.09	0.807 ± 0.06	0.502 ± 0.08
	Yes	**0.815 ± 0.06**	**0.517 ± 0.08**	**0.819 ± 0.05**	**0.533 ± 0.08**	0.815 ± 0.06	**0.516 ± 0.08**
Patient 4	No	0.808 ± 0.04	0.564 ± 0.08	0.813 ± 0.03	0.565 ± 0.08	0.806 ± 0.04	0.568 ± 0.08
	Yes	**0.826 ± 0.03**	**0.583 ± 0.08**	**0.834 ± 0.03**	**0.586 ± 0.08**	**0.826 ± 0.03**	**0.591 ± 0.08**
Patient 5	No	0.848 ± 0.04	0.602 ± 0.10	0.848 ± 0.05	0.609 ± 0.09	0.842 ± 0.05	0.591 ± 0.10
	Yes	**0.868 ± 0.03**	**0.685 ± 0.07**	**0.862 ± 0.04**	**0.676 ± 0.06**	**0.869 ± 0.04**	**0.672 ± 0.08**

* Model Applied (MA).

3.3. *Model evaluation in terms of SSIM between the input and estimated projections*

This section examines the model evaluation in terms of anatomical similarities between the estimated DRRs generated by the model and the input kV images, which are highlighted by computing the spatial gradients of the projections. Fig. 5 illustrates the gradient maps for the motion-corrected DRRs generated by the model (row (c)), as well as those from the reference CT before applying the motion model (row (a)) and the input kV projections (row (b)) at various gantry angles corresponding to some arbitrary time points, specified on the respiratory signal in row (f). This respiratory signal is obtained from the input kV scan series using the AS technique. Rows (d) and (e) of Fig. 5 display the local SSIM maps that compare the local similarities between the gradient maps in the first and third rows with the second row from the kV images, respectively. The results show that the model's output DRRs

Table 3
Average ± std of the phase differences (in degrees) between the respiratory signal extracted from the model outputs versus the one obtained from the input kVs.

Subject	Scan series 1	Scan series 2	Scan series 3
Patient 1	9.6 ± 12.9	10.0 ± 11.8	8.4 ± 10.0
Patient 2	2.4 ± 11.7	3.6 ± 6.5	1.9 ± 8.0
Patient 3	1.0 ± 6.6	3.6 ± 7.1	4.4 ± 8.8
Patient 4	2.3 ± 14.9	0.9 ± 12.4	1.8 ± 15.9
Patient 5	1.4 ± 10.1	1.2 ± 12.7	-0.4 ± 12.0

are closely aligned with the observed kVs, capturing motion in the kV images even at the irregular breathing time points such as 1, 2, and 4. Also, higher local SSIM values are observed in row (e) compared to row (d). Particularly, noticeable improvements are observed in specific target anatomical areas, such as in the diaphragm locations. This indicates that the model effectively learned the respiratory motion from the input kVs and generated DRRs aligned with them. Table 2 quantitatively shows the average and standard deviation results of the SSIM values between the observed kVs and estimated DRRs for three distinct test scan series in individual patients. The results are computed across the front views corresponding to the gantry angles between 100°-245°and for the side/back views at all other angles. The results for both cases, before and after applying the model to the reference CT, are included. A statistical paired t-test was applied to determine if there was a significant difference between the two groups of measured values before and after applying the model to the reference CT.

3.4. Evaluation of the learned motion in the model outputs

To evaluate the model's ability to learn the breathing motion from the input kV projections, we applied the AS technique to extract the respiratory signal from the model outputs too. Because of intensity differences between kVs and DRRs, respiratory signals with incomparable amplitudes were generated using the AS technique from the model's inputs and outputs, as illustrated in Fig. 6a. Instead, we used the Hilbert transform to assess the instantaneous phase for the resulting respiratory signals (Fig. 6b), and their differences are presented in Fig. 6c. Fig. 6d shows the distribution of the instantaneous phase differences across all evaluation scan series for individual patients. Analysing the instantaneous phase relationship between signals provides insights into how much the signals are synchronized or the presence of a time delay between them. As seen, the respiratory signals obtained from the model input and output show similar instantaneous phase characteristics. The average phase shift between the model's input and output was $3.6^\circ \pm 6.5^\circ$, which corresponds to the range of 0.1 ± 0.18 bin difference in a ten-frame binned 4DCT. Table 3 presents quantitatively the average instantaneous phase shift differences computed in test scans for each patient, indicating a small difference between the observed breathing pattern within the input kV projections and the corresponding reconstructed outputs by the model.

3.5. Model validation using a simulated dataset

In this section, we assess the model's performance in estimating spatio-temporal 3D motion fields by reconstructing ground-truth 4DCT images from a simulated paired 4DCT-DRR dataset using only DRRs as inputs to our model. The use of synthetic data for validation is of particular importance in this study, as ground-truth 4DCT is not available for real datasets.

3.5.1. Creating simulated paired 4DCT-DRR datasets

Here, we created a simulated paired 4DCT-DRR dataset by utilizing a real patient's breathing signal, planning 4DCT, and employing Surrogate Parametrised Respiratory Motion Modelling (SuPReMo) [13] as a state-of-the-art surrogate-driven respiratory motion model. The process of generating a simulated 4DCT dataset with SuPReMo involves two key steps: initially, fitting a respiratory correspondence model (RCM), followed by its application to generate 4DCT images.

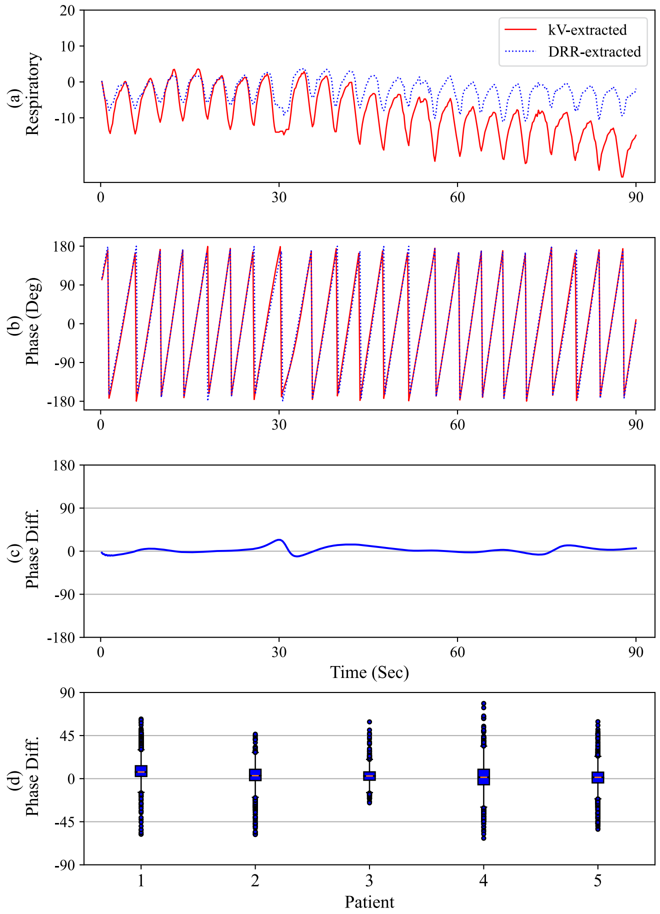


Fig. 6. (a): Respiratory signal extracted from the input kVs (shown as a solid red line) versus the signal obtained from the model output (depicted as a dashed blue line) using the AS technique. Since AS is an intensity-dependent approach, the respiratory signal extracted from the kVs exhibits a different amplitude compared to the one obtained from the model outputs. (b) Instantaneous phase variability of the respiratory signals obtained from the model inputs versus the model outputs; (c) Instantaneous phase differences between the model input and output in row (b); (d) The distribution of the instantaneous phase shifts between the model input and output across all test scan series for the analysed patients.

To fit a motion model, we utilized planning 4DCT data from a real patient. Two respiratory surrogate signals, S_1 and S_2, were essential for the process. S_1 was derived from the planning 4DCT by tracking the diaphragm across ten respiratory phases using cv2.TrackerCSRT_create() function in Python. This process was executed on a mid-slice in the sagittal view, encompassing diaphragm motion across all ten time points. Subsequently, the temporal gradient of S_1 was used as S_2. The surrogate signals were normalized to have zero mean and unit standard deviation. We also used Motion-compensated image reconstruction (MCR) functionality of SuPReMo [13] to create a super resolution MCR image and utilized that as the reference image in the SuPReMo's application phase. The hyperparameters for fitting the motion model are presented in Appendix C. The fitting step returns the RCM and MCR to be utilized in the application phase.

For the SuPReMo's application phase, we used the patient's real respiratory signal. This data was recorded using a belt during the 4DCT fan beam image acquisition using Anzai Respiratory Gating System, AZ-733VI, with the Load Cell sensor. Depending on scan field of view length it comprised between 2 and 4 mins of data, corresponding to approximately 20 to 40 breaths. The extended data was produced by repeating breath cycles from the acquired data randomly, to generate realistic but diverse sets of breathing motion in the application phase, for the training of the 4D-Precise model. The amplitudes of the breaths were adjusted to avoid discontinuities in the generated traces, which resulted in different motion for each SuPReMo generation run. This signal was downsampled to represent each breathing cycle with ten samples after applying the Hilbert transform and splitting each cycle into ten phases. The limitation of the original 4DCT phase-binned data to 10 phases enforced this temporal resolution on the surrogate signals. However, this process does not enforce equal breathing cycle length in time. The temporal resolution is variable as whilst there are always 10 samples per breathing cycle, the length of these samples in time is not fixed. The required surrogate signals S_1 and S_2 for running SuPReMo in this step were each patient's breathing cycle and their temporal gradient, respectively. Hence, in the application phase, SuPReMo was utilised with these inputs: the MCR as the reference image, RCM from the fitting step, the patient's planning 4DCT, and two surrogate respiratory signals. It was also required to set the maximum number of RCM fitting and MCR iterations to zero. Running SuPReMo for all breathing cycles yields dynamic images (4DCT images) and corresponding DVFs at all time points of the respiratory signal. The simulated dynamics are the deformation of the MCR image using the estimated DVFs at each time point. More details about SuPReMo can be found in McClelland et al. [13]. To create simulated paired 4DCT-DRR datasets, we initially split the long reconstructed 4DCT for each patient into six shorter sequences. Each sequence comprises 48 breathing cycles, resulting in a total of 480 temporal CT volumes. Subsequently, we extracted DRRs from the dynamic images in each sequence using the CBCT geometry from the real patient dataset and at gantry angle steps of 0.75°to cover the entire 360°. In this way, we created paired 4DCT-DRR data consisting of six datasets for training and testing the proposed model. This simulated data was only created to enable validation of the method against a ground-truth, which is not available for real world data. Therefore, any limitations of the lower sample rate would only affect these validation results and not those of the real kV trained model, as the real kV model was not trained using simulated data.

3.5.2. Training the model and evaluation metrics

We assessed the model's performance by evaluating its capability to reconstruct the source 4DCT volumes from the given DRR sequences in the simulated datasets. In addition the estimated motion fields are compared to the ones from SuPReMo. Worth noting that the 4D-Precise model employs unsupervised learning to estimate 3D motion fields. Hence, the ground-truth 4DCT images in the paired 4DCT-DRR datasets and their corresponding DVFs were not seen during training of our model. The DRR series from three fractions were used as inputs for training the 4D-Precise model, while the remaining three fractions were utilised for testing. In this experiment, we used the same reference image as in SuPReMo (i.e., the MCR image) for our model to enable a comparison of reconstructed volumes and estimated deformations with those from the SuPReMo method. It is worth noting that SuPReMo could also use an existing CT from the planning 4DCT as the reference image. However, using MCR allowed us to evaluate the generalizability of the proposed model and its independence from a specific phase for the reference image. The deformations between the MCR image and ten phase-binned images of the planning 4DCT were used to establish the prior distribution of motion fields in our model based on Eq. (1).

To evaluate the volumes generated by the 4D-Precise model in comparison to the source volumes from SuPReMo, and also as a way to assess the amplitude of the estimated motion fields, we computed Dice scores and volume similarity metrics for the liver and lungs. To accomplish this, an expert clinician segmented the liver and lungs in the same reference volume, used by both approaches (i.e., MCR image). We deformed the masks using the DVFs from 4D-Precise and SuPReMo respectively. Subsequently, the Dice score (DSC) was computed for each organ by

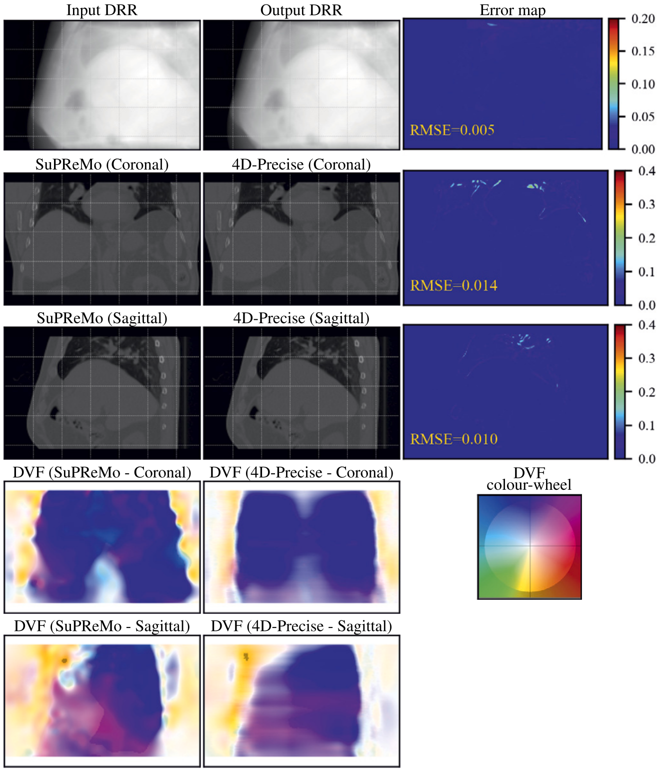


Fig. 7. Validation of the reconstructed 4DCT volumes using a simulated 4DCT-DRR dataset. Top row: Model input and output projections are compared by a pixel-wise error map; Rows two and three: Reconstructed volumes by SuPReMo and 4D-Precise are compared at the coronal and sagittal views respectively and the average RMSE is reported on the corresponding error map; Rows four and five: Corresponding colour-coded DVF maps showing the direction of displacements, estimated by the two approaches in the coronal and sagittal views respectively.

$$DSC = \frac{|\mathbf{X} \cap \mathbf{Y}|}{\frac{1}{2}(|\mathbf{X}| + |\mathbf{Y}|)} \tag{16}$$

where **X** was the deformed organ mask obtained from the SuPReMo model, and **Y** is the corresponding one from 4D-Precise. Dice measures the overlap between two masks. On the other hand, the volume similarity (VS) metric only considers the size of the masks, regardless of their overlap. VS is defined by

$$VS = \frac{|\mathbf{X}| - |\mathbf{Y}|}{\frac{1}{2}(|\mathbf{X}| + |\mathbf{Y}|)} \tag{17}$$

If masks have the same size, VS is zero. We also computed the Mean Contour Distance (MCD) and the Hausdorff distance (HD) between the corresponding organ contours from the two approaches in various slices of coronal, sagittal, and axial views. MCD measures the mean distance between two contours of $\partial\mathbf{X}$ and $\partial\mathbf{Y}$ by

$$MCD = \frac{1}{2|\partial\mathbf{X}|} \sum_{x \in \partial\mathbf{X}} d(x, \partial\mathbf{Y}) + \frac{1}{2|\partial\mathbf{Y}|} \sum_{y \in \partial\mathbf{Y}} d(y, \partial\mathbf{X}) \tag{18}$$

where $d(x, \partial)$ denotes the minimal distance from point x to the contour ∂. Finally, HD indicates the maximum distance between two contours $\partial\mathbf{X}$ and $\partial\mathbf{Y}$ by

$$HD = \max\left(\max_{x \in \partial\mathbf{X}} d(x, \partial\mathbf{Y}), \max_{y \in \partial\mathbf{Y}} d(y, \partial\mathbf{X})\right) \tag{19}$$

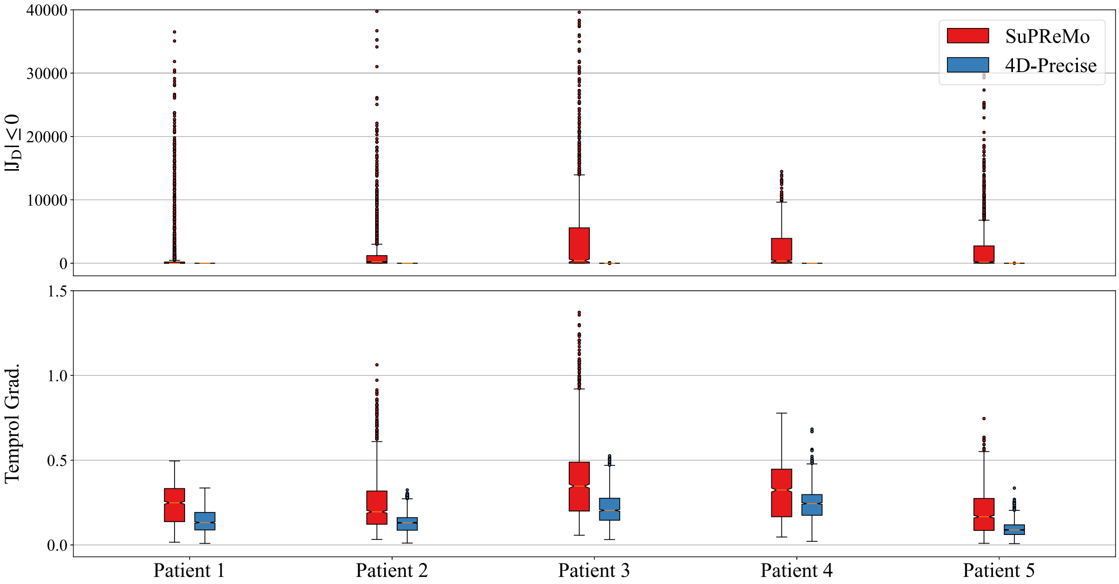


Fig. 8. The results of the model evaluation in terms of spatio-temporal characteristics of estimated DVFs, including the number of negative Jacobian determinant and temporal gradients of the displacement fields compared with the SuPReMo model. The results indicate that the 4D-Precise presents spatio-temporally smoother DVFs in all subjects.

3.5.3. Results on the simulated dataset

Fig. 7 illustrates the quantitative similarity between the model's estimation and SuPReMo's CT volume by displaying pixel-wise intensity error maps in the coronal and sagittal views, corresponding to the DRR at an angle of 79.2°. An intensity error map between the input DRR and the estimated output DRR generated by the 4D-Precise model is also displayed in Fig. 7, demonstrating the model's accuracy in learning the input projections resulted in a low root-mean-square error (RMSE). The corresponding DVFs estimated by the two approaches are depicted in Fig. 7 at the coronal and sagittal views. We have presented a video in the Supplementary Materials (video 16) that shows the moving 4DCT estimated by our model in comparison to that of SuPReMo, together with angular/temporal variations in the input/output DRRs, estimated DVFs, and corresponding differences in the coronal and sagittal views. These results indicate that while the reconstructed CT volumes by 4D-Precise are highly similar to SuPReMo's volumes, the estimated DVFs are slightly different. There is generally a good match between DVFs from the two approaches; however, 4D-Precise presents spatially smoother deformations. The differences are noticeable in the back of the lungs due to sliding motion, and our model presents smoother deformations in these regions. Fig. 8 compares the spatio-temporal characteristics of the estimated displacements by the two approaches, including the number of negative Jacobian determinant elements and the temporal gradients of the DVFs. The quantitative results indicate that the DVFs estimated by 4D-Precise have fewer foldings (negative Jacobian determinants) and are temporally smoother than the DVFs estimated by SuPReMo, which show higher temporal gradients. 4D-Precise incorporates various elements to enhance motion smoothness. These include spatio-temporal learning from training sets of kV sequences facilitated by a recurrent neural network (LSTM), a smoothness loss term, utilization of prior motion distribution from the planning 4DCT, and regularization through the KL divergence loss term. These considerations collectively enable smoother motion estimation compared to DVFs from SuPReMo.

Table 4 reports the RMSE between the volumes generated by 4D-Precise and SuPReMo, as well as between the input and output DRRs to the 4D-Precise model for individual patients. The results demonstrate that our model is capable of unsupervised estimating motion fields and reconstructing output volumes that closely resemble SuPReMo's volumes when given only DRRs generated from SuPReMo's 4DCT. Additionally, the output DRRs from the 4D-Precise model demonstrate good agreement with the input DRRs, both in terms of intensity variations and the ability to track motion within them.

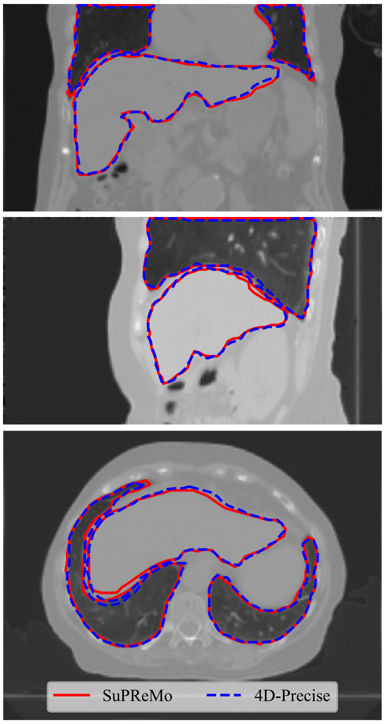
Table 4
Model validation in terms of RMSE (mean±std) between the estimated and source volumes and between in/out projections are presented.

Subject	4D-Precise/SuPReMo volumes	in/out DRRs
Patient 1	0.038 ± 0.002	0.006 ± 0.003
Patient 2	0.040 ± 0.002	0.008 ± 0.005
Patient 3	0.046 ± 0.006	0.014 ± 0.009
Patient 4	0.045 ± 0.004	0.012 ± 0.006
Patient 5	0.043 ± 0.004	0.009 ± 0.003

Fig. 9 shows the distance metric results obtained from a sample 3D volume reconstructed using 4D-Precise, compared with the deformed 3D mask generated by SuPReMo's DVFs, showing a significant resemblance between them across various organs. Fig. 10 displays distributions of the computed metrics for different patients measured from reconstructed volumes across time. The results quantitatively indicate high degrees of similarity between the reconstructed volumes by 4D-Precise and the source volumes from SuPReMo in the liver and lungs. These findings suggest that the model can efficiently reconstruct dynamic volumetric data in an unsupervised manner from the measured projections in a sequence. Similar reconstructed CT volumes in terms of RMSE and other computed distance metrics for several organs coupled with lower temporal gradients of the displacements (shown in Fig. 8) indicate that 4D-Precise possesses the capability to enhance the temporal regularity of the deformation fields.

4. Discussion and conclusion

The methodology presented in this paper addresses the challenge of estimating spatio-temporal 3D motions from a sequence of kV projections by leveraging deep learning techniques. Traditional CT reconstruction algorithms fail to handle this problem due to the gap between the projection domain and the image domain. Our approach tackles this issue by employing a deep learning framework trained patient-specifically on a large dataset to learn the mapping from a series of X-rays to 3D motion fields. Specifically, we utilize a reference CT volume



	MCD (mm)	HD (mm)
Liver	0.06	4.45
Right Lung	0.07	5.59
Left Lung	0.14	5.59

	MCD (mm)	HD (mm)
Liver	0.07	8.39
Right Lung	0.06	7.09
Left Lung	—	—

	MCD (mm)	HD (mm)
Liver	0.12	8.29
Right Lung	0.15	7.04
Left Lung	0.07	4.37

	DSC	VS
Liver	0.96	0.02
Right Lung	0.97	0.03
Left Lung	0.97	0.05

Fig. 9. The distance metric results of MCD and HD together with the volumetric DSC and VS obtained for a 3D volume reconstructed by 4D-Precise (with dashed blue contours in the liver and lungs) in the three views of coronal (top), sagittal (middle), and axial (bottom) in comparison with the deformed 3D mask by SuPReMo (with solid red contours). The results illustrate the similarity of the volumes reconstructed by the two models across different organs.

that undergoes deformation according to the estimated dense motion fields at each time point, effectively generating a CT volume from which the extracted DRR closely matches the input X-ray projection in terms of the anatomical structures. This process is guided by a structural similarity index (SSIM) loss term operating at the projection level that controls the process of mapping from 2D X-rays to 3D motion fields. Our model is enriched with mechanisms to enhance the spatio-temporal smoothness of the estimated motions. These include spatio-temporal learning from training sets of kV sequences via a recurrent neural network (LSTM), along with incorporating domain-specific knowledge by utilizing prior motion distribution obtained from the planning 4DCT, and computing the spatial gradients of the displacements as a loss term.

Our approach involved formulating an unsupervised motion model based on recurrent variational Bayes. By explicitly modelling the mean and variance of the displacement fields across both space and time, the model learned to efficiently sample motion and reconstruct a cine-like temporal sequence of 3DCT volumes, corresponding to motion observed via high-temporal resolution kV projections. In this manner, unlike conventional 4DCT reconstruction techniques that ignore inter-cycle motion modelling, the proposed model is designed to compute both intra-cycle and inter-cycle motions and represents the motion over an extended timeframe for several minutes of scan series. During the training phase, a training set of kV projection series taken from the same patient at various fractions is used to learn inter-fraction motion variabilities. Once this rigorous training process is complete, the trained network can swiftly reconstruct 4DCT sequences in a forward pass when presented with a previously unseen kV series, regardless of the fraction. This capability is beneficial for motion estimation in radiation therapy.

The current approach enables rapid motion predictions based on a sequence of kV projections as they are acquired, rather than relying on post-hoc offline analysis of the full image sequence. This ability is crucial to enabling real-time adaptive radiotherapy methods [44], which can translate the power of this method into patient benefit, through more accurate and personalised radiation treatment for cancers of the thorax and abdomen [45]. In order to clinically realise this advantage, it will be necessary to further develop 4D-Precise to operate at a population level with generalisable models, and to validate it on kV projection data acquired intra-fraction, during MV treatment beam delivery. This constitutes future work.

While other approaches to the 4D, real time, motion challenge exist (e.g. MRI enabled linear accelerators) they are prohibitively expensive for most patients, and currently unable to achieve the spatio-temporal resolution required [46], whilst also having much lower patient throughput than conventional linac systems. Indeed, rather than projection images, intra-fraction MRI usually consists of a few orthogonal slices of the 3D volume, and a similar approach to that developed here could be used to predict complete temporal sequences of 3D MRI volumes, also improving the utility of that technology. In any case, for the foreseeable future, the majority of cancer patients will be treated on conventional equipment with kV X-ray-based imaging. The proposed model could allow a low-cost motion modelling, based entirely on existing hardware.

The applications of this approach are potentially wide-ranging, opening the possibility of personalised, motion-compensated dose accumulation [47] after each treatment fraction. This would enable re-planning of subsequent fractions to mitigate the effects of unexpected motion, lowering toxicity and improving patient outcomes. Somewhat more ambitiously, it may be possible to adapt to motions in real-time, during radiation delivery. Dose gating [48] and tumour tracking [49] have been extensively investigated in the past, with image quality and motion uncertainty being limiting factors. Estimated motions across an extended time frame could be used for dose planning in radiation therapy, replacing conventional 4DCT (which is limited to ∼10 phase- or amplitude-binned images, representing an average breathing cycle). This would allow personalised treatment margins based on a patient's actual breathing motion [50], improving the conformality of their radiation dose to the tumour. Additionally, the reconstructed 4DCT could potentially provide diagnostic information on abnormal organ shape and motion, bringing a dynamic component into disease categorisation and prognosis estimation. This would be of particular interest in light of recent developments along the lines of multi-messenger predictive models for treatment stratification [51].

4.1. Limitations

Our design choices for the downsampling of input data were due to the limitations of the utilized GPU memory. Including more projections in a sequence can enhance model accuracy by updating its parameters based on more observed data in a sequence. Nonetheless, elevating the spatio-temporal resolution of the data, while refining model precision, increases the number of model parameters. This higher count can lead to complexity, longer training times, and overfitting risks. Deploying such models in resource-constrained settings poses challenges. We aimed to balance model complexity and efficiency for practical viability and generalizability.

In the current study, we made the assumption that there are no static changes in the patient's anatomy between treatment sessions, which might potentially be a limitation if tumour growth or regression, or inflammation were significant during treatment. Further research would be needed to investigate this aspect. The current model is patient-specific and requires training on early treatment data, precluding clinical use until late in treatment. It is also trained and validated on treatment (not intra-fraction) kV projection images. For clinical translation population level models, validated on intra-fraction kV projections will be required, and both these aspects are the subject of ongoing work. Our future work will also focus on extending this model by incorporating Physics-Informed Neural Networks to model physical constraints in

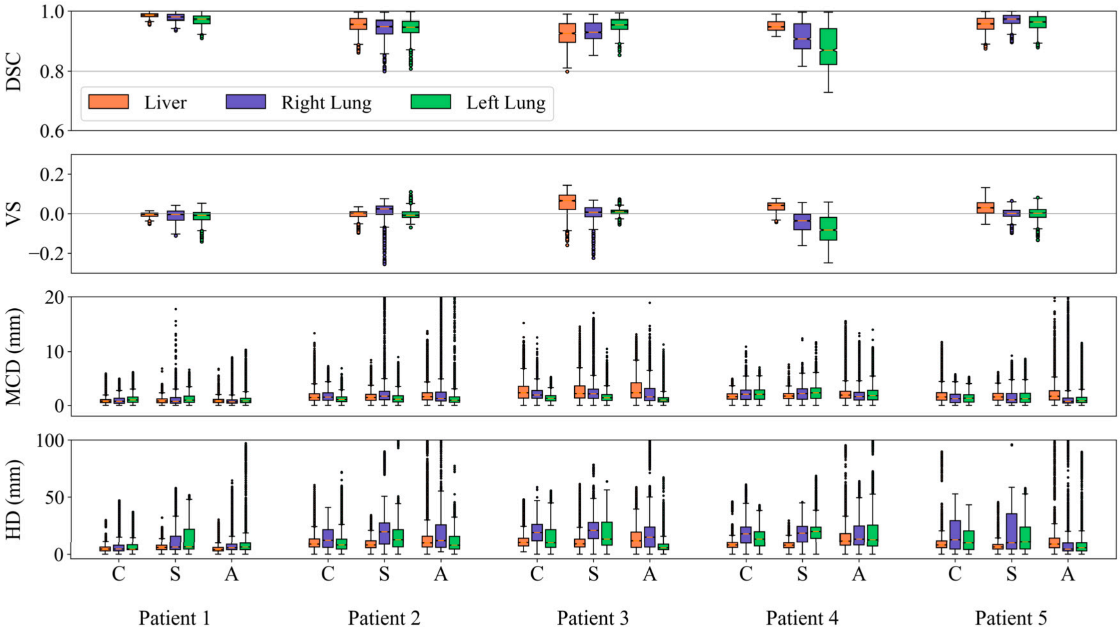


Fig. 10. The results of model validation in the unsupervised reconstruction of 4DCT volumes on a simulated paired 4DCT-DRR dataset in terms of DSC, VS, MCD, and HD were computed over time for the liver, left and right lung in different patients. MCD and HD values are computed in various slices of coronal (C), sagittal (S), and axial (A) views of the three organs. The results quantitatively indicate that 4D-Precise can efficiently reconstruct dynamic volumetric data in an unsupervised manner from a sequence of input DRRs.

the estimation of motion in specific organs and volumetric mesh generation over time given the kV projections.

Funding

This research was supported by Cancer Research UK funding (grant C19942/A28832) for the Leeds Radiotherapy Research Centre of Excellence (RadNet). ZAT is supported by an RAEng / Leverhulme Trust Research Fellowship (DIADEM-ART LTRF2021-17115).

Ethical statement

The project involved retrospective patient data, approved for research use under Yorkshire & The Humber - Leeds East Research Ethics Committee (REC)(REC Reference: 19/YH/0300). All patients gave informed consent for their retrospective data to be used for research.

CRediT authorship contribution statement

Arezoo Zakeri: Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Alireza Hokmabadi:** Software, Methodology, Investigation, Formal analysis, Conceptualization. **Michael G. Nix:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Ali Gooya:** Writing – review & editing, Supervision, Conceptualization. **Isuru Wijesinghe:** Writing – review & editing, Conceptualization. **Zeike A. Taylor:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors gratefully acknowledge the support and advice of Dr Jamie McClelland on using SuPReMo.

Appendix A

We used the NiftyReg package based on the Free-Form Deformation algorithm for non-rigid registration [36] and utilised the CPU version. Specifically, we used the Localized Normalized Cross Correlation (LNCC) objective function, grid spacing of 5, the number of iterations of 300, and the weight of the bending energy (second derivative of the transformation) penalty term of 0.015.

Appendix B

Computing $\mathcal{L}_D$:

$$\begin{aligned}\mathcal{L}_{\mathbf{D}} &= \mathbb{E}_{\prod_{t,a} q(\mathbf{D}_t|\mathbf{kV}_t^a,\mathbf{h}_{t-1})} \sum_{t,a} \log \frac{p(\mathbf{D}_t|\mathbf{h}_{t-1})}{q(\mathbf{D}_t|\mathbf{kV}_t^a,\mathbf{h}_{t-1})} \\ &= -\sum_{t,a} \int_{\mathbf{D}_t} q(\mathbf{D}_t|\mathbf{kV}_t^a,\mathbf{h}_{t-1}) \log \frac{q(\mathbf{D}_t|\mathbf{kV}_t^a,\mathbf{h}_{t-1})}{p(\mathbf{D}_t|\mathbf{h}_{t-1})} d\mathbf{D}_t \\ &\simeq -\sum_{t,a} \mathcal{D}_{\mathcal{KL}}\Big(q(\mathbf{D}_t|\mathbf{kV}_t^a,\mathbf{h}_{t-1}) \| p(\mathbf{D}_t|\mathbf{h}_{t-1})\Big) \qquad \text{(A.1)}\end{aligned}$$

Appendix C

We used the open-source software SuPReMo (https://github.com/UCL/SuPReMo). The hyperparameters used for this study are: grid spacing of 10 voxels; the maximum number of respiratory correspondence model fitting iterations was 300; the maximum number of times to iterate between motion compensate image reconstruction and fitting the respiratory correspondence model was 10; the MCR function was set to be super resolution; and the maximum number of iterations to use with iterative reconstruction methods was 5.

Appendix D. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cmpb.2024.108158.

References

[1] J.-J. Sonke, L. Zijp, P. Remeijer, M. Van Herk, Respiratory correlated cone beam CT, Med. Phys. 32 (2005) 1176–1186.

[2] B.J. Cooper, R.T. O'Brien, S. Balik, G.D. Hugo, P.J. Keall, Respiratory triggered 4D cone-beam computed tomography: a novel method to reduce imaging dose, Med. Phys. 40 (2013) 041901.

[3] X. Jia, Z. Tian, Y. Lou, J.-J. Sonke, S.B. Jiang, Four-dimensional cone beam CT reconstruction and enhancement using a temporal nonlocal means method, Med. Phys. 39 (2012) 5592–5602.

[4] H. Zhang, D. Zeng, H. Zhang, J. Wang, Z. Liang, J. Ma, Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: a review, Med. Phys. 44 (2017) 1168–1185.

[5] Y. Zhang, M.R. Folkert, X. Huang, L. Ren, J. Meyer, J.N. Tehrani, R. Reynolds, J. Wang, Enhancing liver tumor localization accuracy by prior-knowledge-guided motion modeling and a biomechanical model, Quant. Imag. Med. Surg. 9 (2019) 1337.

[6] J.J. Sohn, C. Kim, D.H. Kim, S.-R. Lee, J. Zhou, X. Yang, T. Liu, Analytical low-dose CBCT reconstruction using non-local total variation regularization for image guided radiation therapy, Front. Oncol. 10 (2020) 242.

[7] H. Lee, J. Park, Y. Choi, K.R. Park, B.J. Min, I.J. Lee, Low-dose cbct reconstruction via joint non-local total variation denoising and cubic B-spline interpolation, Sci. Rep. 11 (2021) 1–15.

[8] Z. Hu, H. Zheng, Improved total variation minimization method for few-view computed tomography image reconstruction, Biomed. Eng. Online 13 (2014) 1–10.

[9] H. Lee, L. Xing, R. Davidi, R. Li, J. Qian, R. Lee, Improved compressed sensing-based cone-beam ct reconstruction using adaptive prior image constraints, Phys. Med. Biol. 57 (2012) 2287.

[10] J. Wang, X. Gu, Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone-beam CT, Med. Phys. 40 (2013) 101912.

[11] R. Li, X. Jia, J.H. Lewis, X. Gu, M. Folkerts, C. Men, S.B. Jiang, Real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image for lung cancer radiotherapy, Med. Phys. 37 (2010) 2822–2826.

[12] Y. Zhang, F.-F. Yin, W.P. Segars, L. Ren, A technique for estimating 4d-cbct using prior knowledge and limited-angle projections, Med. Phys. 40 (2013) 121701.

[13] J.R. McClelland, M. Modat, S. Arridge, H. Grimes, D. D'Souza, D. Thomas, D. O'-Connell, D.A. Low, E. Kaza, D.J. Collins, M.O. Leach, D.J. Hawkes, A generalized framework unifying image registration and respiratory motion models and incorporating image reconstruction, for partial image data or full images, Phys. Med. Biol. 62 (2017) 4273.

[14] M. Guo, G. Chee, D. O'Connell, S. Dhou, J. Fu, K. Singhrao, D. Ionascu, D. Ruan, P. Lee, D.A. Low, J. Zhao, J.H. Lewis, Reconstruction of a high-quality volumetric image and a respiratory motion model from patient CBCT projections, Med. Phys. 46 (2019) 3627–3639.

[15] Y. Zhang, X. Huang, J. Wang, Advanced 4-dimensional cone-beam computed tomography reconstruction by combining motion estimation, motion-compensated reconstruction, biomechanical modeling and deep learning, Vis. Comput. Ind. Biomed. Art 2 (2019) 1–15.

[16] X. Huang, Y. Zhang, J. Wang, A biomechanical modeling-guided simultaneous motion estimation and image reconstruction technique (SMEIR-Bio) for 4D-CBCT reconstruction, Phys. Med. Biol. 63 (2018) 045002.

[17] S. Dhou, J. Lewis, W. Cai, D. Ionascu, C. Williams, Quantifying day-to-day variations in 4DCBCT-based PCA motion models, Biomed. Phys. Eng. Expr. 6 (2020) 035020.

[18] W. Harris, Y. Zhang, F.-F. Yin, L. Ren, Estimating 4D-CBCT from prior information and extremely limited angle projections using structural PCA and weighted free-form deformation for lung radiotherapy, Med. Phys. 44 (2017) 1089–1104.

[19] A. Fassi, J. Schaerer, M. Fernandes, M. Riboldi, D. Sarrut, G. Baroni, Tumor tracking method based on a deformable 4D CT breathing motion model driven by an external surface surrogate, Int. J. Radiat. Oncol. Biol. Phys. 88 (2014) 182–188.

[20] Y. Huang, K. Thielemans, G. Price, J.R. McClelland, Surrogate-driven respiratory motion model for projection-resolved motion estimation and motion compensated Cone-Beam CT reconstruction from unsorted projection data, Phys. Med. Biol. 69 (2024) 025020.

[21] H. Yan, X. Wang, W. Yin, T. Pan, M. Ahmad, X. Mou, L. Cerviño, X. Jia, S.B. Jiang, Extracting respiratory signals from thoracic cone beam CT projections, Phys. Med. Biol. 58 (2013) 1447.

[22] D.P. Gierga, J. Brewer, G.C. Sharp, M. Betke, C.G. Willett, G.T. Chen, The correlation between internal and external markers for abdominal tumors: implications for respiratory gating, Int. J. Radiat. Oncol. Biol. Phys. 61 (2005) 1551–1558.

[23] A. Zakeri, A. Hokmabadi, N. Bi, I. Wijesinghe, M.G. Nix, S.E. Petersen, A.F. Frangi, Z.A. Taylor, A. Gooya, DragNet: learning-based deformable registration for realistic cardiac MR sequence generation from a single frame, Med. Image Anal. 83 (2023) 102678.

[24] A. Zakeri, A. Hokmabadi, N. Ravikumar, A.F. Frangi, A. Gooya, A probabilistic deep motion model for unsupervised cardiac shape anomaly assessment, Med. Image Anal. 75 (2022) 102276.

[25] T.P. Szczykutowicz, G.V. Toia, A. Dhanantwari, B. Nett, A review of deep learning CT reconstruction: concepts, limitations, and promise in clinical practice, Current Radiol. Rep. 10 (2022) 101–115.

[26] N. Yuan, B. Dyer, S. Rao, Q. Chen, S. Benedict, L. Shang, Y. Kang, J. Qi, Y. Rong, Convolutional neural network enhancement of fast-scan low-dose cone-beam CT images for head and neck radiotherapy, Phys. Med. Biol. 65 (2020) 035003.

[27] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss, IEEE Trans. Med. Imaging 37 (2018) 1348–1357.

[28] H.-C. Shao, J. Wang, T. Bai, J. Chun, J.C. Park, S. Jiang, Y. Zhang, Real-time liver tumor localization via a single x-ray projection using deep graph neural network-assisted biomechanical modeling, Phys. Med. Biol. 67 (2022) 115009.

[29] P. Henzler, V. Rasche, T. Ropinski, T. Ritschel, Single-Image Tomography: 3D Volumes from 2D Cranial x-Rays, Computer Graphics Forum, vol. 37, Wiley Online Library, 2018, pp. 377–388.

[30] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, Y. Zheng, X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10619–10628.

[31] J. Dai, G. Dong, C. Zhang, W. He, L. Liu, T. Wang, Y. Jiang, W. Zhao, X. Zhao, Y. Xie, et al., Volumetric tumor tracking from a single cone-beam x-ray projection image enabled by deep learning, Med. Image Anal. (2023) 102998.

[32] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[33] N. Milickovic, D. Baltas, S. Giannouli, M. Lahanas, N. Zamboglou, Ct imaging based digitally reconstructed radiographs and their application in brachytherapy, Phys. Med. Biol. 45 (2000) 2787.

[34] K. Zuiderveld, Contrast limited adaptive histogram equalization, Graph. Gems (1994) 474–485.

[35] S. Rit, M.V. Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, G.C. Sharp, The Reconstruction Toolkit (RTK), an Open-Source Cone-Beam CT Reconstruction Toolkit Based on the Insight Toolkit (ITK), J. Phys. Conf. Ser. 489 (2014) 012079, IOP Publishing.

[36] M. Modat, G.R. Ridgway, Z.A. Taylor, M. Lehmann, J. Barnes, D.J. Hawkes, N.C. Fox, S. Ourselin, Fast free-form deformation using graphics processing units, Comput. Methods Programs Biomed. 98 (2010) 278–284.

[37] D.P. Kingma, M. Welling, Auto-encoding variational bayes, preprint, arXiv:1312.6114, 2013.

[38] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.

[39] M. Jaderberg, K. Simonyan, A. Zisserman, K. Kavukcuoglu, Spatial transformer networks, Adv. Neural Inf. Process. Syst. 28 (2015).

[40] M. Unberath, J.-N. Zaech, S.C. Lee, B. Bier, J. Fotouhi, M. Armand, N. Navab, DeepDRR– a catalyst for machine learning in fluoroscopy-guided procedures, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 98–106.

[41] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.

[42] J.R. Hershey, P.A. Olsen, Approximating the Kullback Leibler divergence between Gaussian mixture models, in: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, volume 4, IEEE, 2007, pp. IV–317.

[43] D.P. Kingma, J. Ba Adam, A method for stochastic optimization, preprint, arXiv:1412.6980, 2014.

[44] S. Lim-Reinders, B.M. Keller, S. Al-Ward, A. Sahgal, A. Kim, Online adaptive radiation therapy, Int. J. Radiat. Oncol. Biol. Phys. 99 (2017) 994–1003.

[45] R. De Jong, K. Crama, J. Visser, N. Van Wieringen, J. Wiersma, E. Geijsen, A. Bel, Online adaptive radiotherapy compared to plan selection for rectal cancer: quantifying the benefit, Radiat. Oncol. 15 (2020) 1–9.

[46] J. Ng, F. Gregucci, R.T. Pennell, H. Nagar, E.B. Golden, J.P. Knisely, N.J. Sanfilippo, S.C. Formenti, MRI-LINAC: a transformative technology in radiation oncology, Front. Oncol. 13 (2023).

[47] S. Bharat, P. Parikh, C. Noel, M. Meltsner, K. Bzdusek, M. Kaus, Motion-compensated estimation of delivered dose during external beam radiation therapy: implementation in Philips' Pinnacle3 treatment planning system, Med. Phys. 39 (2012) 437–443.

[48] H.D. Kubo, B.C. Hill, Respiration gated radiotherapy treatment: a technical study, Phys. Med. Biol. 41 (1996) 83.

[49] I.R. de Vries, M. Dahele, H. Mostafavi, B. Slotman, W. Verbakel, Markerless 3D tumor tracking during single-fraction free-breathing 10MV flattening-filter-free stereotactic lung radiotherapy, Radiother. Oncol. 164 (2021) 6–12.

[50] P. Trémolières, A. Gonzalez-Moya, A. Paumier, M. Mege, J. Blanchecotte, C. Theotime, D. Autret, S. Dufreneix, Lung stereotactic body radiation therapy: personalized PTV margins according to tumor location and number of four-dimensional CT scans, Radiat. Oncol. 17 (2022) 1–9.

[51] Z. Zhou, H. Deng, W. Yang, Z. Wang, L. Lin, J. Munasinghe, O. Jacobson, Y. Liu, L. Tang, Q. Ni, F. Kang, Y. Liu, G. Niu, R. Bai, C. Qian, J. Song, X. Chen, Early stratification of radiotherapy response by activatable inflammation magnetic resonance imaging, Nat. Commun. 11 (2020) 3032.