A probabilistic deep learning model of inter-fraction anatomical variations in radiotherapy

Pastor-Serrano, Oscar; Habraken, Steven; Hoogeman, Mischa; Lathouwers, Danny; Schaart, Dennis; Nomura, Yusuke; Xing, Lei; Perkó, Zoltán

DOI
10.1088/1361-6560/acc71d

Publication date
2023

Document Version
Final published version

Published in
Physics in medicine and biology

A probabilistic deep learning model of inter-fraction anatomical variations in radiotherapy

IOP Publishing Phys. Med. Biol. 68 (2023) 085018 https://doi.org/10.1088/1361-6560/acc71d

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
20 September 2022

REVISED
20 February 2023

ACCEPTED FOR PUBLICATION
23 March 2023

PUBLISHED
10 April 2023

PAPER

A probabilistic deep learning model of inter-fraction anatomical variations in radiotherapy

Oscar Pastor-Serrano[1,2,*], **Steven Habraken**[3,4], **Mischa Hoogeman**[3,4], **Danny Lathouwers**[1], **Dennis Schaart**[1,4], **Yusuke Nomura**[2], **Lei Xing**[2] **and Zoltán Perkó**[1]

[1] Delft University of Technology, Department of Radiation Science & Technology, Delft, The Netherlands
[2] Stanford University, Department of Radiation Oncology, Stanford, CA, United States of America
[3] Erasmus University Medical Center, Department of Radiotherapy, Rotterdam, The Netherlands
[4] HollandPTC, Department of Medical Physics and Informatics, Delft, The Netherlands
[*] Author to whom any correspondence should be addressed.

E-mail: o.pastorserrano@tudelft.nl

Keywords: deep learning, intra-fraction motion, radiation therapy, deformable registration, latent variable model, diffeomorphic transformation

Abstract

Objective. In radiotherapy, the internal movement of organs between treatment sessions causes errors in the final radiation dose delivery. To assess the need for adaptation, motion models can be used to simulate dominant motion patterns and assess anatomical robustness before delivery. Traditionally, such models are based on principal component analysis (PCA) and are either patient-specific (requiring several scans per patient) or population-based, applying the same set of deformations to all patients. We present a hybrid approach which, based on population data, allows to predict patient-specific inter-fraction variations for an individual patient. *Approach.* We propose a deep learning probabilistic framework that generates deformation vector fields warping a patient's planning computed tomography (CT) into possible patient-specific anatomies. This daily anatomy model (DAM) uses few random variables capturing groups of correlated movements. Given a new planning CT, DAM estimates the joint distribution over the variables, with each sample from the distribution corresponding to a different deformation. We train our model using dataset of 312 CT pairs with prostate, bladder, and rectum delineations from 38 prostate cancer patients. For 2 additional patients (22 CTs), we compute the contour overlap between real and generated images, and compare the sampled and 'ground truth' distributions of volume and center of mass changes. *Results.* With a DICE score of 0.86 ± 0.05 and a distance between prostate contours of 1.09 ± 0.93 mm, DAM matches and improves upon previously published PCA-based models, using as few as 8 latent variables. The overlap between distributions further indicates that DAM's sampled movements match the range and frequency of clinically observed daily changes on repeat CTs. *Significance.* Conditioned only on planning CT values and organ contours of a new patient without any pre-processing, DAM can accurately deformations seen during following treatment sessions, enabling anatomically robust treatment planning and robustness evaluation against inter-fraction anatomical changes.

1. Introduction

Modern radiotherapy techniques such as intensity modulated proton therapy (IMPT) have the potential to deliver highly conformal doses to tumors while maximally sparing organs at risk (OARs). Although offering dosimetric advantages with respect to conventional modalities, such treatments are particularly sensitive to geometrical uncertainties arising from setup errors before delivery or range errors caused by organ movements between or during treatment sessions. In the presence of uncertainties, planned doses are delivered to anatomies different from the 3D computed tomography (CT) scan used during treatment planning, which may translate into shifting high dose regions away from clinical target volumes (CTVs) into critical OARs. Being one of the main sources of error in, e.g. prostate cancer treatments (van Herk *et al* 2002), the magnitude of the deformations and their effect on the final dose distribution must be quantified to ensure robust delivery. Ideally, treatments could be real-time adapted via image guidance, or alternatively adjusted before each treatment session (Jagt *et al* 2017, 2018), but such adaptive workflows are constrained by the speed of the CT acquisition, delineation, dose calculation and treatment re-optimization processes in practice.

An efficient alternative currently used in the clinic consists of including setup and range uncertainties during treatment planning optimization to design robust treatment plans that withstand positioning and range errors (van der Voort *et al* 2016, Unkelbach and Paganetti 2018, Rojo-Santiago *et al* 2021). Similarly, inter-fractional movement information could be incorporated during treatment planning or treatment evaluation to make treatment plans robust against complex geometrical variations. To account for such anatomical changes, some published works propose computing expected dose distributions using weighted scenarios, where each scenario corresponds to the dose deposited in a patient geometry generated by an anatomy model. Typically, such models extract the main eigenmodes of organ deformation—groups of correlated movements— via principal component analysis (PCA) (Söhn *et al* 2005, Jeong *et al* 2010, Budiarto *et al* 2011, Szeto *et al* 2017). During the last decades, linear PCA models have been successfully employed to quantify and understand the effect of organ deformations in different treatment sites and modalities (Thörnqvist *et al* 2013a, Rios *et al* 2017, Magallon-Baro *et al* 2019); to extend clinical volumes with extra margins and compensate for anatomical changes (Thörnqvist *et al* 2013b, Bondar *et al* 2014); to characterize respiratory deformations (Zhang *et al* 2007, Badawi *et al* 2010, Dhou *et al* 2015, 2015); and to simulate dosimetric outcomes of delivery in the presence of geometrical uncertainties (Nie *et al* 2012, Söhn *et al* 2012, Xu *et al* 2014, Tilly *et al* 2017). Focusing on conventional photon-based modalities, most of these studies are based only on organ contours without including CT intensity values, and require time-consuming image registrations as pre-processing to find corresponding points across a population of patients before being usable for learning generic deformations. Furthermore, all previously introduced models are either patient-specific (requiring several CTs per patient) or population-based (applying the same set of deformations to all patients), which limits their accuracy and applicability. For widespread adoption of anatomically robust treatment planning, we require accurate probabilistic models quickly generating patient-specific treatment anatomies.

All published PCA models learn correlated organ movements from a dataset of 3D deformation vector fields (DVFs), where each vector indicates the magnitude and direction of displacement for each point in a voxelized volume. Such DVFs can be obtained via image registration algorithms finding a nonlinear correspondence between, e.g. two CT scans (Ashburner 2007, Vásquez Osorio *et al* 2009, Bruveris and Holm 2015). While traditional not data-driven algorithms require minutes to solve a registration task, recent deep learning based methods reduce computing times down to few seconds and additionally increase registration accuracy (de Vos *et al* 2017, Balakrishnan *et al* 2019), typically using 2D (Ronneberger *et al* 2015) or 3D (Cicek *et al* 2016) U-net convolutional architectures in combination with spatial transformer networks (Jaderberg *et al* 2015). Several architectures generating DVFs and warping pairs of images have been proposed and applied to radiotherapy problems such as automated contour propagation in adaptive workflows (Liang *et al* 2021) or 4D image registration and generation of moving images due to breathing in lung geometries (Lei *et al* 2020, Romaguera *et al* 2020, Chang *et al* 2021), modeling intra-fraction breathing movements that occur during the delivery of the patient.

Our objective, however, is to generate a set of DVFs to warp a single planning CT into different repeat CTs that are likely to be observed during the course of a radiotherapy treatment. Ideally, a suitable model would be able to implicitly capture the relative likelihood of correlated groups of movements depending on the input patient geometry. Probabilistic frameworks based on variational inference (Kingma and Welling 2014, Rezende *et al* 2014, Blei *et al* 2017) have been successfully applied to model uncertainty in organ segmentation tasks (Kohl *et al* 2018, Baumgartner *et al* 2019, Hu *et al* 2019, Kohl *et al* 2019), making use of auxiliary latent variables that represent the main factors of variation behind the model's predictions. Similar probabilistic U-net based architectures have also been proposed for pure image registration tasks (Dalca *et al* 2019, Krebs *et al* 2019), with applications to unsupervised contouring problems (Dalca *et al* 2019) and breathing movement prediction based on motion surrogates (Romaguera *et al* 2021).

Extending on these recent architectures, we present a probabilistic deep learning framework that represents common anatomical movements and deformations in a population of patients using few latent variables. The proposed daily anatomy model (DAM) first generates DVFs conditioned on an input planning CT scan and latent variables, where each combination of latent variables corresponds to a different group of movements; and subsequently warps the planning CT with the generated DVFs into a set of artificial repeat scans. We train the model using a dataset containing planning and repeat CTs recorded at different stages of prostate cancer treatments in three different institutions, evaluating whether DAM is able to learn realistic movements with two external patients. Compared to previous methods, DAM does not require any pre-processing registration step and can in principle be applied to quickly simulate patient anatomies for treatment adaptation and robustness evaluation purposes.

2. Methods and materials

Here we describe the fundamentals of the variational framework used to capture anatomical variations, including the different parametric models and the procedure used to tune their parameters. Subsequently, we describe the model architecture in detail, together with the data and the evaluation metrics used in each experiment.

2.1. Proposed framework

During the course of a radiotherapy treatment, the internal structures and organs of the patient change between fractions/days. As a result, the anatomy captured in the planning image $\boldsymbol{x} \in \mathbb{R}^M$ and organ structures $\boldsymbol{s}_x \in \mathbb{R}^M$ (both represented as 3D matrices) can significantly differ from the repeat images $\boldsymbol{y} \in \mathbb{R}^M$ and structures $\boldsymbol{s}_y \in \mathbb{R}^M$ taken during following treatment sessions. M voxels comprise the entire volume, where the voxels in $\boldsymbol{x}$ and $\boldsymbol{y}$ represent image intensity values, and the voxels in $\boldsymbol{s}_x$ and $\boldsymbol{s}_y$ contain an integer corresponding to the organ present in the voxel.

As demonstrated in previous studies (Budiarto *et al* 2011) for treatment sites like prostate, common anatomical variations such as volume and contour changes are observed across an entire population. Based on the existence of such generic movements we assume that, given a planning image $\boldsymbol{x}$ and structures $\boldsymbol{s}_x$, there is an unknown patient-specific generative distribution $P^*(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}_x)$ of repeat scans that can be approximated via a probabilistic model with learned parameters. Given a planning image from a new patient, we can sample the resulting model distribution $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}_x)$ parametrized by $\boldsymbol{\theta}$ to generate a set of artificial anatomies observed at future treatment stages.

In this case, $\boldsymbol{\theta}$ corresponds to the parameters of the U-net neural network that is used to compute a DVF $\Phi: \mathbb{R}^3 \to \mathbb{R}^3$ mapping coordinates between images. We model Φ as a diffeomorphic transformation, which is invertible, preserves topology, and in our case practically allows obtaining the forward and inverse transformations in a very simple manner. Such diffeomorphic transformation is represented by a stationary velocity field $v: \mathbb{R}^3 \to \mathbb{R}^3$, as $\Phi = \exp v$. As for the inputs, we discretize the velocity field $\boldsymbol{v} \in \mathbb{R}^{M\times 3}$ and DVF $\boldsymbol{\Phi} \in \mathbb{R}^{M\times 3}$ into M voxels, using $\boldsymbol{\Phi}(\boldsymbol{p})$ to denote the displacement applied to the voxel centered at location $\boldsymbol{p} \in \mathbb{R}^3$. Following previous work (Dalca *et al* 2019), the U-net predicts $\boldsymbol{v}$, which is exponentiated via scaling and squaring using a spatial transformer network (Jaderberg *et al* 2015) (details in appendix A) to obtain the final DVF $\boldsymbol{\Phi}$ used to warp planning images into artificial repeats $\boldsymbol{y} = \boldsymbol{\Phi} \circ \boldsymbol{x}$.

Generative model. We use a probabilistic model that conditions the generated DVFs (and thus also the repeat images) on N unobserved latent variables $\mathbf{z} \in \mathbb{R}^N$ capturing the main factors of variation in the data, i.e. the main groups of anatomical deformations. The latent variables distribute following a multivariate Gaussian prior probability distribution that depends on the input planning anatomy

$$P(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x) = \mathcal{N}(\boldsymbol{z}; \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{s}_x), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{s}_x)), \tag{1}$$

where the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ are deterministic functions calculated by a neural network referred to as *Encoder* (figure 1), which corresponds to the down-sampling path of a U-net. The prior dependence on the input results in a different distribution over latent variables per patient, which allows the model to select the groups of movements that are likely to be observed for each specific input image. The Encoder additionally outputs a volume $\boldsymbol{r} = g_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{s}_x)$, which is the results of several deterministic convolution operations containing features from the input. Since $\boldsymbol{r}$ is a deterministic function of the input, we substitute any conditioning on $\boldsymbol{r}$ with $\boldsymbol{x}$ and $\boldsymbol{s}_x$ in the remainder of the paper.

The relationship between the input planning image and latent variables and the output warped repeat images is computed in the up-sampling path of the U-net, which takes sampled latent variables and the low-dimensional features $\boldsymbol{r}$ to generate a velocity field $\boldsymbol{v}_{\boldsymbol{z},\boldsymbol{\theta}} = f_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x)$, where the subscripts denote the deterministic dependence to $\boldsymbol{z}$ and $\boldsymbol{\theta}$. After exponentiating $\boldsymbol{v}_{\boldsymbol{z},\boldsymbol{\theta}}$ to obtain the DVF $\boldsymbol{\Phi}_{\boldsymbol{z},\boldsymbol{\theta}}$, the output repeat image $\boldsymbol{y} \in \mathbb{R}^M$ is obtained by warping the input as $\boldsymbol{y} = \boldsymbol{\Phi}_{\boldsymbol{z},\boldsymbol{\theta}} \circ \boldsymbol{x}$.

Different latent variable samples $\boldsymbol{z}$ result in different repeat images given the same input planning scan, and the modeled distribution of repeat images can be recovered as a function of the prior $P(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x)$ and a likelihood $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x)$ distributions as

$$P_{\theta}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}_x) = \int P_{\theta}(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x) P(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x) d\boldsymbol{z}. \tag{2}$$

The choice of the likelihood distribution affects the final loss function. Based on previous work (Krebs *et al* 2019), we model the likelihood distribution as a function of the *cross-correlation* (CC) between predicted $\boldsymbol{y}$ and

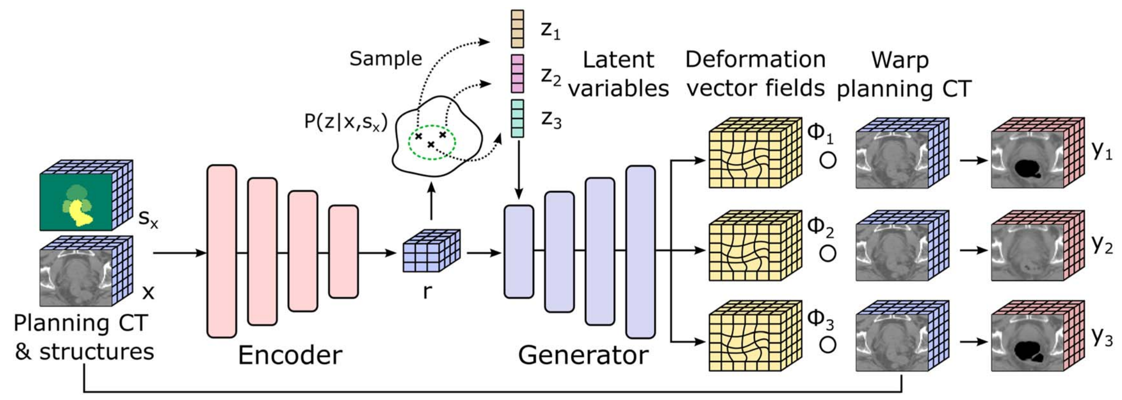


Figure 1. Proposed generative framework. The probabilistic models are embedded within a U-net, where the down-sampling path is referred to as Encoder, and the up-sampling path is the Generator. The Encoder takes the planning CT and structures and outputs both a compressed representation of the input $\boldsymbol{r}$ as well as a distribution $P(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x)$ over the region of the latent space containing variables corresponding to plausible patient-specific movements. Given $\boldsymbol{r}$ and any sample $\boldsymbol{z}$ from the latent space distribution, the Generator outputs a deformation vector field that is used to warp the planning CT into an artificial repeat CTs.

ground-truth $\hat{\boldsymbol{y}}$ images, scaled by a constant λ as

$$P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x) \propto \exp(\lambda \mathrm{CC}(\hat{\boldsymbol{y}}, \boldsymbol{y} = \boldsymbol{\Phi}_{\boldsymbol{z},\boldsymbol{\theta}} \circ \boldsymbol{x})). \tag{3}$$

The CC has been empirically found to yield better similarity than other metrics such as the mean squared error (Balakrishnan *et al* 2019), with larger CC values corresponding to more alike images. Let $y(\boldsymbol{p})$ and $\hat{y}(\boldsymbol{p})$ denote the intensity values for each voxel at position $\boldsymbol{p}$ in the predicted and ground-truth images, respectively. If $w(\boldsymbol{p})$ and $\hat{w}(\boldsymbol{p})$ are images where each voxel is the local mean of the n^3 neighbouring voxels, e.g. $w(\boldsymbol{p}) = \frac{1}{n^3}\sum_{j=1}^{n^3} y(\boldsymbol{p}_j))$ and $\hat{w}(\boldsymbol{p}) = \frac{1}{n^3}\sum_{j=1}^{n^3} \hat{y}(\boldsymbol{p}_j))$, the CC is defined as

$$\mathrm{CC}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \sum_{\boldsymbol{p}\in\Omega} \frac{\left[\sum_{i=1}^{n^3}(\hat{y}(\boldsymbol{p}_i) - \hat{w}(\boldsymbol{p}))(y(\boldsymbol{p}_i) - w(\boldsymbol{p}))\right]^2}{\left[\sum_{i=1}^{n^3}(\hat{y}(\boldsymbol{p}_i) - \hat{w}(\boldsymbol{p}))\right]\left[\sum_{i=1}^{n^3} y(\boldsymbol{p}_i) - \hat{w}(\boldsymbol{p}))\right]}. \tag{4}$$

As in previous work (Krebs *et al* 2019), instead of sampling the likelihood $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x)$ each time during inference to generate anatomies, we always use the mode of the distribution $\boldsymbol{\Phi}_{\boldsymbol{z},\boldsymbol{\theta}} \circ \boldsymbol{x}$.

Learning. With the presented probabilistic formulation, the goal is to maximize equation (2) by learning the parameters $\boldsymbol{\theta}$ from a dataset containing planning $\boldsymbol{x}$ and repeat $\boldsymbol{y}$ pairs. However, estimating the integral over the latent space would require sampling a large number of latent variables, being intractable in practice. Instead, we resort to a variational framework and define an *approximate posterior* distribution $Q_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y)$, parametrized by an *Inference Network* with parameters ψ. During training, the Inference Network has access to the real repeat scans and predicts the parameters of Gaussian distribution covering a small region of the latent space containing variables that are likely to explain the deformation between $\boldsymbol{x}$ and $\boldsymbol{y}$ scans. Thus, the predicted Gaussian is

$$Q_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y) = \mathcal{N}(z; \boldsymbol{\mu}_\psi(\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y), \boldsymbol{\Sigma}_\psi(\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y)), \tag{5}$$

with deterministic mappings $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$ computed by the Inference neural network. Our formulation allows estimating the model parameters $\boldsymbol{\theta}$ and ψ by minimizing the negative *evidence lower bound* as

$$\log(P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}_x)) \leqslant -\mathbb{E}_{\boldsymbol{z}\sim Q_\psi(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{s}_x,\boldsymbol{y},\boldsymbol{s}_y)}[\log(P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x))] + D_{KL}(Q_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y)||P_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x)). \tag{6}$$

The lower bound balances two terms: the $D_{KL}(\cdot||\cdot)$ term—Kullback–Leibler (KL) divergence—forces the approximated posterior to be close to the prior distribution, while the first term corresponds to maximizing the CC, encouraging similarity between real and generated images. Further details about deriving the lower bound are included in appendix B.

Explicit regularization terms. The current form of the likelihood enforces image similarity regardless of structure overlap or DVF quality. We modify the lower bound and add two regularization terms to enforce realistic predicted anatomies. To encourage smooth and realistic DVFs, we introduce a *spatial* regularization term that penalizes large unrealistic spatial gradients $\nabla\boldsymbol{\Phi}_{z,\theta}(\boldsymbol{p}) = \left(\frac{\partial\boldsymbol{\Phi}_{z,\theta}(\boldsymbol{p})}{\partial x}, \frac{\partial\boldsymbol{\Phi}_{z,\theta}(\boldsymbol{p})}{\partial y}, \frac{\partial\boldsymbol{\Phi}_{z,\theta}(\boldsymbol{p})}{\partial z}\right)$ of the DVF $\boldsymbol{\Phi}_{z,\theta}$, which is multiplied by a constant κ as

$$R(\boldsymbol{\Phi}_{z,\theta}) = -\kappa \sum_{\boldsymbol{p}\in\Omega} \|\nabla \boldsymbol{\Phi}_{z,\theta}(\boldsymbol{p})\|_2. \tag{7}$$

A *segmentation* regularization term is added to improve the overlap between propagated and ground-truth structures, using the DICE score (defined between 0 and 1, where 1 denotes perfect overlap). For K structures, let $\hat{\boldsymbol{s}}_y^k$ be the voxels in the ground-truth scan with structure number $k \in [1, K]$, $\boldsymbol{s}_y^k = \boldsymbol{\Phi}_{z,\theta} \circ \boldsymbol{s}_x^k$ the predicted voxels with structure number k, and $|\hat{\boldsymbol{s}}_y^k|$ the cardinality of structure $\hat{\boldsymbol{s}}_y^k$, i.e, the number of elements in $\hat{\boldsymbol{s}}_y^k$. The DICE score is defined as

$$\mathrm{DICE}(\hat{\boldsymbol{s}}_y^k, \boldsymbol{s}_y^k) = 2\frac{|\hat{\boldsymbol{s}}_y^k \cap \boldsymbol{s}_y^k|}{|\hat{\boldsymbol{s}}_y^k| + |\boldsymbol{s}_y^k|}. \tag{8}$$

With these two terms multiplying the likelihood in the lower bound of equation (9), the final optimization problem becomes

$$\begin{aligned}\boldsymbol{\theta}^*, \boldsymbol{\psi}^* = \underset{\boldsymbol{\theta},\boldsymbol{\psi}}{\operatorname{argmin}}\ \mathbb{E}_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{s}_x,\boldsymbol{s}_y \sim P_D(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s}_x,\boldsymbol{s}_y)} \Bigg[\mathbb{E}_{\boldsymbol{z}\sim Q_\psi(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{s}_x,\boldsymbol{y},\boldsymbol{s}_y)} \Bigg[\lambda \mathrm{CC}(\hat{\boldsymbol{y}}, \boldsymbol{y}) - \frac{1}{K}\textstyle\sum_{k=1}^{K} \mathrm{DICE}(\hat{\boldsymbol{s}}_y^k, \boldsymbol{\Phi}_{z,\theta} \circ \boldsymbol{s}_x^k)\Bigg] \\ + \kappa \textstyle\sum_{\boldsymbol{p}\in\Omega} \|\nabla \boldsymbol{\Phi}_{z,\theta}(\boldsymbol{p})\|_2 \Bigg] + D_{KL}(Q_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y) || P_\theta(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x)) \Bigg],\end{aligned} \tag{9}$$

with $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}_x$ and $\boldsymbol{s}_y$ sampled from the real data distribution $P_D(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}_x, \boldsymbol{s}_y)$.

2.2. Dataset

To learn the model parameters in a training stage, we use a dataset with 369 CTs from 40 prostate cancer patients, including prostate, seminal vesicles, bladder and rectum delineations with no overlap. For each of the patients, 3-11 repeat CTs were recorded at different points during their treatment at 3 different institutions: Erasmus University Medical Center (Rotterdam, Netherlands), Haukeland Medical Center (Bergen, Norway) and the Netherlands Cancer Institute (Amsterdam, Netherlands) (Deurloo *et al* 2005, Sharma *et al* 2012, Xu *et al* 2014). In total, 329 planning-repeat CT pairs are available, 312 of which are used for training and validation, while the remaining 22 CTs—corresponding to 2 independent test patients, as in previous studies (Budiarto *et al* 2011)—serve to evaluate performance on unseen geometries. After rigidly aligning each repeat to the planning CT, we crop the volumes to a region of $64 \times 64 \times 48$ voxels around the prostate with a voxel resolution of 2 mm, resulting in sub-volumes of $128 \times 128 \times 96$ that in all cases covers the prostate, seminal vesicles, rectum and a large portion the bladder. As a result, we obtain $\boldsymbol{x} \in \mathbb{R}^{64\times64\times48}$ and $\boldsymbol{y} \in \mathbb{R}^{64\times64\times48}$ with the original CT intensity values rescaled to the range [0, 1], and $\boldsymbol{s}_x \in \mathbb{R}^{64\times64\times48}$ and $\boldsymbol{s}_y \in \mathbb{R}^{64\times64\times48}$ with categorical labels depending on the organ present in each voxel. Given the stochasticity in the density of the rectum fillings, we adhere to clinical practice and mask all voxels in the rectum and set their intensity to -1000 (vacuum).

2.3. Model architecture

As shown in figure 2, the proposed variational framework comprises two different models, parametrized by artificial neural networks: the Inference network and the probabilistic U-net with down-sampling and up-sampling paths denoted as Encoder and Generator, respectively. Based on the input planning CT and structures, the Encoder computes (i) a low-dimensional volume of input image features $\boldsymbol{r}$, and (ii) the parameters $\boldsymbol{\mu}_\theta$ and $\boldsymbol{\Sigma}_\theta$ of the prior distribution $P_\theta(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x)$ over a region of the latent space containing movements that are likely to be observed for the patient. The prior depends on the input, thus one of the functions of the Encoder is selecting primary groups of movements for each patient based on planning CT anatomy. The Generator takes the features $\boldsymbol{r}$ and sampled latent variables $z \sim P_\theta(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{s}_x)$ and produces the velocity field $\boldsymbol{v}_{z,\theta}$ that is exponentiated to obtain a diffeomorphic transformation $\boldsymbol{\Phi}_{z,\theta}$.

During training, the Inference network takes a pair of planning and repeat CTs and outputs the parameters $\boldsymbol{\mu}_\psi$ and $\boldsymbol{\Sigma}_\psi$ of the distribution $Q_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{s}_x, \boldsymbol{y}, \boldsymbol{s}_y)$ over a much smaller region of the latent space containing latent variables that explain the deformation between both images. The DVF resulting from such latent variables is used to warp the planning CT into artificial repeat CTs $\boldsymbol{y}$ and structures $\boldsymbol{\Phi} \circ \boldsymbol{s}_x$. The distributions $Q_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$ from the Inference network and $P_\theta(\boldsymbol{z}|\boldsymbol{x})$ from the Encoder are forced to overlap via the KL divergence in equation (9), while the artificial CT and structures are forced to match the ground-truth repeat CTs via the CC and DICE terms in the likelihood.

For the model with the lowest validation loss, the Encoder and Inference network are identical: three consecutive convolutional blocks, where each block contains a 3D convolutional layer with 32 channels and a $3 \times 3 \times 3$ kernel followed by Group Normalization (Wu and He 2020), a rectified linear (ReLU) activation and a max pooling down-sampling operation. At the lowest level, an additional 3D convolution with 4 channels results in the low-dimensional feature volume $\boldsymbol{r} \in \mathbb{R}^{4\times8\times8\times6}$, which is mapped to the means and variances of the prior distribution via two different fully-connected layers. Conversely, the Generator first concatenates the latent

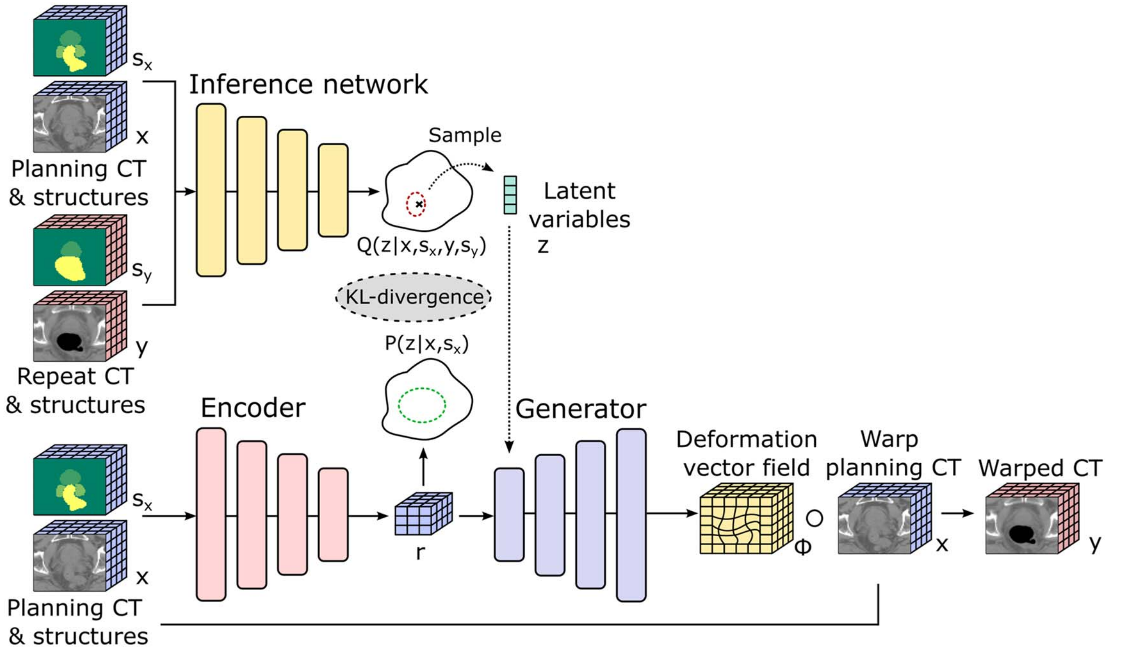


Figure 2. Learning the model parameters. An additional Inference Network takes a pair of planning and repeat CT and outputs the parameters of a distribution over a smaller region of the latent space that is likely to capture the deformation between the two images. The prior distribution predicted by the Encoder is forced to the distribution produced by the Inference Network via a KL-divergence term in the loss. Additionally, a reconstruction term encourages the resulting artificial CT (obtained after warping the planning scan with the predicted deformation) to be similar to real repeat CT.

variables to $\boldsymbol{r}$ as an additional channel, and then applies three up-sampling convolutional blocks with 32 channels. Two additional 3D convolution operations with 16 and 3 channels result in the final velocity field $\boldsymbol{v}_{\boldsymbol{z},\boldsymbol{\theta}}$. All models are trained for 1000 epochs using a learning rate of 0.001, hyper-parameters $\kappa = 0.1$ and $\lambda = 1000$, and the Adam optimizer (Kingma and Ba 2017) with default parameters.

2.4. Experiments

We assess the model's accuracy in both generating feasible groups of deformations and reconstructing the ground-truth repeat scans. Additional experiments aim at exploring the structure of the latent space and the types of movements triggered by different latent variables.

- Reconstruction accuracy. Given a planning and one of its repeat CTs in the test set, the Inference network can be used to obtain the latent variables corresponding to the deformation between both images, which are in turn used to get the DVF and warp the planning scan. For all 22 test planning/repeat pairs, we compare such generated repeat CTs to the ground truth repeats via computing the CC (equation (4)) and the DICE score (equation (8)). Additionally, we warp points $\boldsymbol{\pi}_i \in \mathbb{R}^3$ on the surface of the planning prostate and calculate their distance to corresponding points $\hat{\boldsymbol{\pi}}_i \in \mathbb{R}^3$ on the surface of the repeat prostates via the mean surface error as

$$\boldsymbol{e} = \frac{1}{L}\sum_{i=1}^{L} \|\hat{\boldsymbol{\pi}}_i - \boldsymbol{\Phi}\circ\boldsymbol{\pi}_i\|_2. \tag{10}$$

  To allow for a fair comparison with PCA-based methods, we compute the mean and standard deviation across the same $L = 5864$ randomly chosen points as in previous studies (Budiarto *et al* 2011). Finally, we evaluate the effect of the latent space dimensionality by comparing all accuracy metrics for different models trained with a varying number of latent variables.

- Generative performance. To finally be applied in clinical settings, the generated movements must match those from the recorded CT scans. Based on a previous study quantifying anatomical changes in prostate patients (Antolak *et al* 1998), we compute the volume changes and center of mass shifts between planning and repeat scans, and compare their distributions obtained using real and artificial repeat CTs. To be able to compare to the reference values (Antolak *et al* 1998), we reduce center of mass shifts to a single value by computing the average of absolute differences across coordinates.

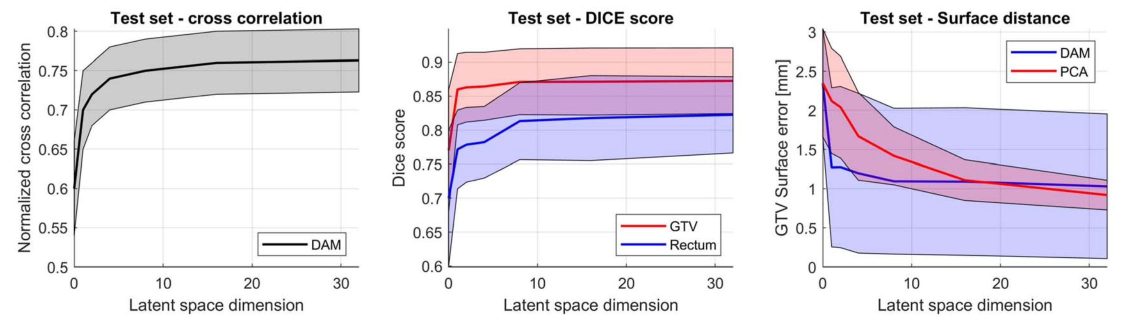


Figure 3. Reconstruction accuracy metrics. All figures show the mean (solid line) and standard deviation across all test planning-repeat pairs of the different metrics for a different number of latent variables, where 0 latent variables refers to using no model (always using the planning CT as a prediction). The left plot shows the cross-correlation between the real and reconstructed repeat CTs. In the middle plot, we show the DICE score measuring overlap between the warped planning structures and the organs delineated in the repeat CTs. Finally, the right figure shows the error between surface points in the prostate, compared to reference PCA values directly taken from Budiarto *et al* (2011).

- Latent space analysis. By individually varying the values of each latent variable while keeping the other fixed, we numerically and visually assess the volume changes and center of mass shifts triggered by each variable. Finally, to understand the structure of the latent space, we obtain the latent variables from all pairs in the dataset and classify them according to the magnitude of their induced center of mass shifts and volume changes. Ideally, similar latent variables should correspond to similar deformations. which can be verified by plotting a 2D representation of the N latent variables using t-SNE (van der Maaten and Hinton 2008) together with their associated label to determine the presence of clusters.

3. Results

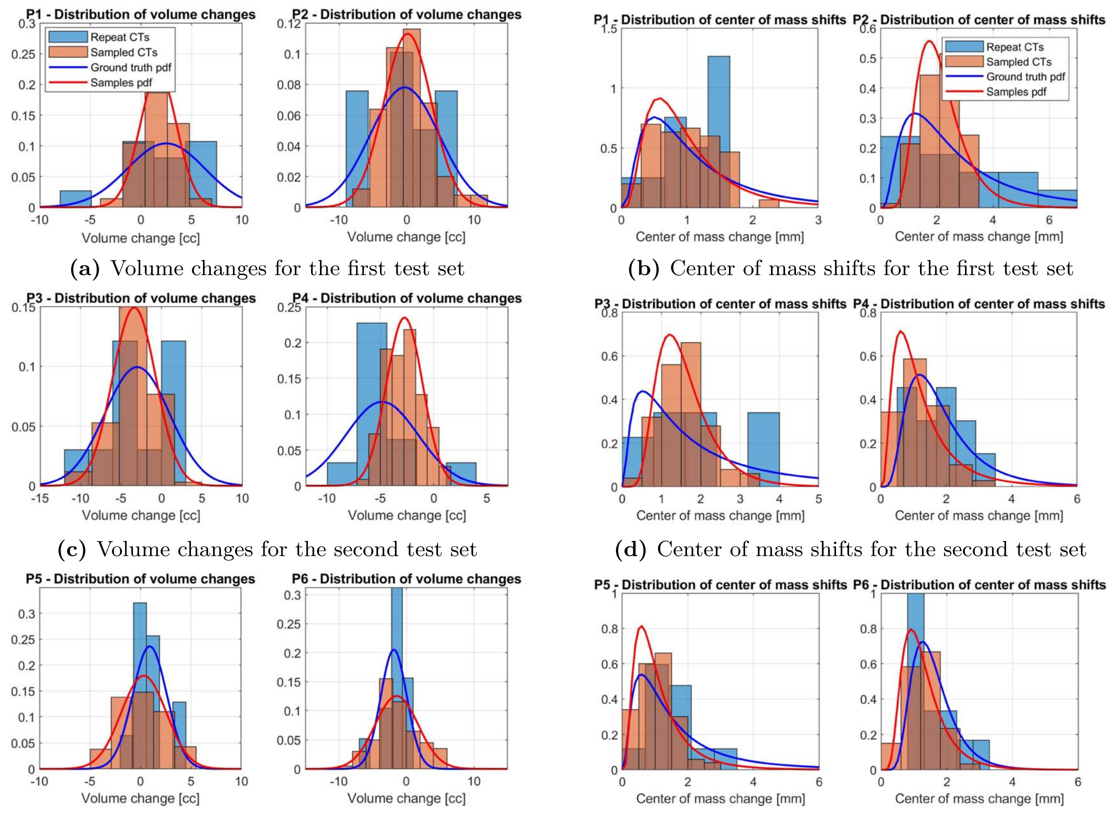
In this section, we evaluate DAM's performance in generating realistic CTs with anatomical changes that match those of the real recorded repeat CTs. First, the reconstruction accuracy of real CTs is assessed, followed by an analysis of the latent space, and the types of deformations captured by the latent variables.

3.1. Reconstruction accuracy

Given a planning-repeat pair of CT scans and structures in the test set, a repeat scan can be reconstructed via the same framework as used during training: sampling latent variables with the Inference network that are used by the Generator to generate a DVF. To verify the similarity between DAM's reconstructions and the real repeat CTs, we compute three metrics assessing CT and structure overlap: the CC, DICE score, and surface error $\boldsymbol{e}$. All three metrics in figure 3 are computed for different models trained with a varying number of latent variables, from 1 to 32. The values shown for 0 latent variables correspond to using the planning CTs as a prediction, which is equivalent to disregarding any model. First, the cross correlation between the real and reconstructed repeat CT is shown in the left plot of figure 3, indicating that the model significantly improves when adding the first few variables, whereas no substantial is observed beyond 10 variables. As seen in DICE scores for the prostate and rectum from the middle plot in figure 3, DAM can model prostate deformations with high accuracy even with a single latent variable, while representing rectum movements generally requires a slightly larger latent space with $\approx$8 variables. The relative simplicity in capturing prostate movements is further confirmed from the right plot in figure 3, showing that most surface error (equation (10)) reduction results from adding the first latent variable. On average, DAM matches—and even outperforms in the low-dimensional regime—the accuracy of countour-based PCA models (Budiarto *et al* 2011). The larger spread in error values is likely caused by the fact that, unlike for the values reported in the PCA study, all surface points are not equidistant but randomly sampled over the surface, increasing the distance between correspondent points in under-sampled areas.

3.2. Generative performance

Besides generating realistic CT scans, DAM should produce patient-specific movements whose distribution approximately matches those observed in the clinics, as reported in previous work (Antolak *et al* 1998). Given the limited dataset size, we train 3 models using different training-test dataset splits, i.e. changing the 2 test patients for each model. For these 2 test patients and each model, figure 4 displays the distribution of the anatomical variations seen in the 11 recorded repeat CTs (blue), compared to the deformations seen in 100 randomly sampled CTs (orange). Except for the large center of mass movements seen for the second patient figures 4(b) and (d), the ranges of values for both volume changes and center of mass shifts in figure 4 are



(a) Volume changes for the first test set

(b) Center of mass shifts for the first test set

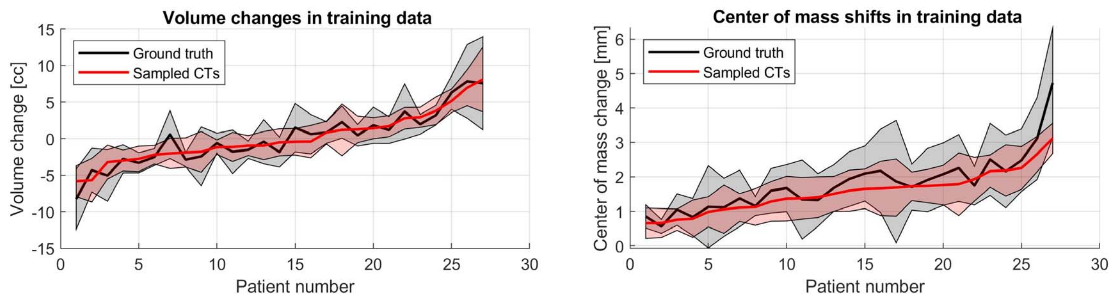
(c) Volume changes for the second test set

(d) Center of mass shifts for the second test set

(e) Volume changes for the third test set

(f) Center of mass shifts for the third test set

Figure 4. Test set histograms of anatomical variations. For two independent test patients in each of the 3 models trained with a different training-test dataset split, we plot histograms of prostate (a), (c), (e) volume changes and (b), (d), (f) center of mass shifts. Blue histograms correspond to changes between the planning CT and the 11 available repeat CTs, for which we additionally show their corresponding fitted normal and log-normal distributions in the same colors. Orange histograms are calculated using 100 randomly sampled CTs, obtained from 100 different latent variable combinations.



(a) Volume changes

(b) Center of mass shifts

Figure 5. Training set distribution of anatomical variations. For all the patients in the training set with 5 or more repeat CTs, we plot the mean (solid line) and standard deviation of prostate (a) volume changes and (b) center of mass shifts. Black lines are computed using the available planning-repeat pairs of CT. The red curves are calculated using 100 randomly sampled CTs, obtained from 100 different latent variable combinations.

approximately equal. Similarly, figure 5 shows the center of mass shift and volume changes distributions for all training patients with more than 5 repeat CTs. To compress all the information into one plot, we plot the mean and standard deviation, instead of the full histogram. The good overlap between distributions demonstrates that DAM captures the correct frequency and range of movements. As for the test patients, the biggest differences between both distributions occur for the last patient in figure 5(b) with large center of mass shifts, which is aggravated by the fact that this patient has three big outliers of >7 mm shift. Finally, figure 6 displays generated and real anatomies for one of the patients, showing high quality images and contours with similar features and shapes. Figure 7 displays representative DFVs from 3 patients, with overall smooth deformations that match the

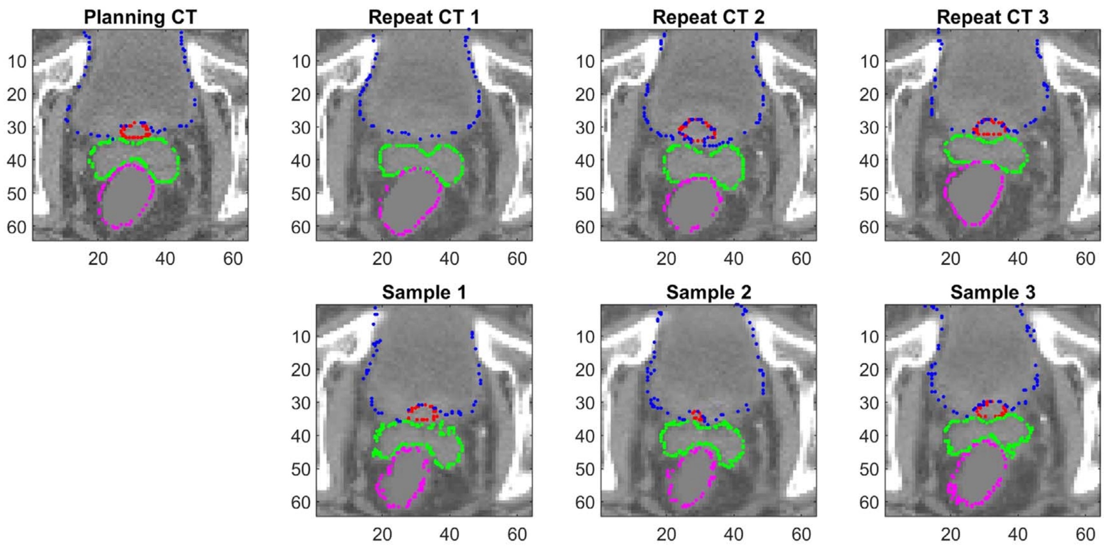


Figure 6. Real versus sampled anatomies. Three recorded repeat CTs (top row), and three anatomies generated by the model (bottom row) are shown for one of the planning CTs, including prostate (red), seminal vesicles (green), bladder (blue) and rectum (pink) contours. The images correspond to a perpendicular slice in the cranial-caudal axis, showing the top of the prostate.

movements observed between the planning and repeat CTs. To evaluate the realism of the deformations, we calculate determinant of the Jacobian of the DVF, and specifically the fraction of voxels whose Jacobian determinant is negative, indicating unrealistic 'folding' in the DVF. In our case, the diffeomorphic DVFs generated by the model prevent any folding from occurring, resulting in smooth, invertible deformations with Jacobian determinants that are always positive.

3.3. Latent space analysis

To investigate the deformations captured by the latent variables, we compute the center of mass shifts and volume changes triggered by each variable independently, while keeping the rest fixed. Figure 8 displays such changes for 4 randomly picked variables from the model with 8 latent variables, whose value was modified between −1.5 and 1.5 times the standard deviation of the prior distribution. The results show magnitudes and correlations between changes as can be expected: smaller prostate volume changes, and large bladder and rectum variations shifting the center of mass of the prostate and seminal vesicles. To further demonstrate DAM's learned correlated groups of movements, in figure 9 we plot a grid of structures corresponding to simultaneously varying two latent variables. Individual changes in the horizontal and vertical axis mainly control the bladder and rectum volumes, respectively. Correlated deformations arise: the increase of bladder volume above the seminal vesicles, together with the decrease of rectum filling below the prostate, cause a prostate and vesicles shift and rotation.

We analyze the structure of the latent space by determining if similar deformations (shifts and volume changes) or anatomical features (organ volume) result in similar latent variables. Figure 10 shows a two-dimensional t-SNE representation of the latent variables, where only samples with the smallest and largest movements or volumes are included, i.e. samples whose with center of mass shifts or volumes that fall above the 90% percentile or below the 10% percentile. Most of the latent space information seems to concern center of mass shifts and bladder/rectum volume changes, since their 2D representations can be clearly separated. Ideally, similar latent variables that are clustered together will correspond to different anatomical deformations, and will not carry information about anatomical features of the patient such as absolute organ volume. Instead, the Encoder is in charge to mapping deformations to anatomical traits observed in the planning CT or structures. Prostate and bladder volume seem to have no effect in how the latent space is organized, since similar latent variables correspond to very different sizes. To some extent, the effect of rectum size is also limited, resulting from the possible correlation between rectum fillings and volume changes.

4. Discussion

In this study, we developed a probabilistic framework to model patient-specific inter-fraction movements based on population data. The presented DAM captures deformation patterns, generating DVFs only based on the planning CT scan and delineations. Based on the metrics obtained in figure 3 for the 22 scans from two

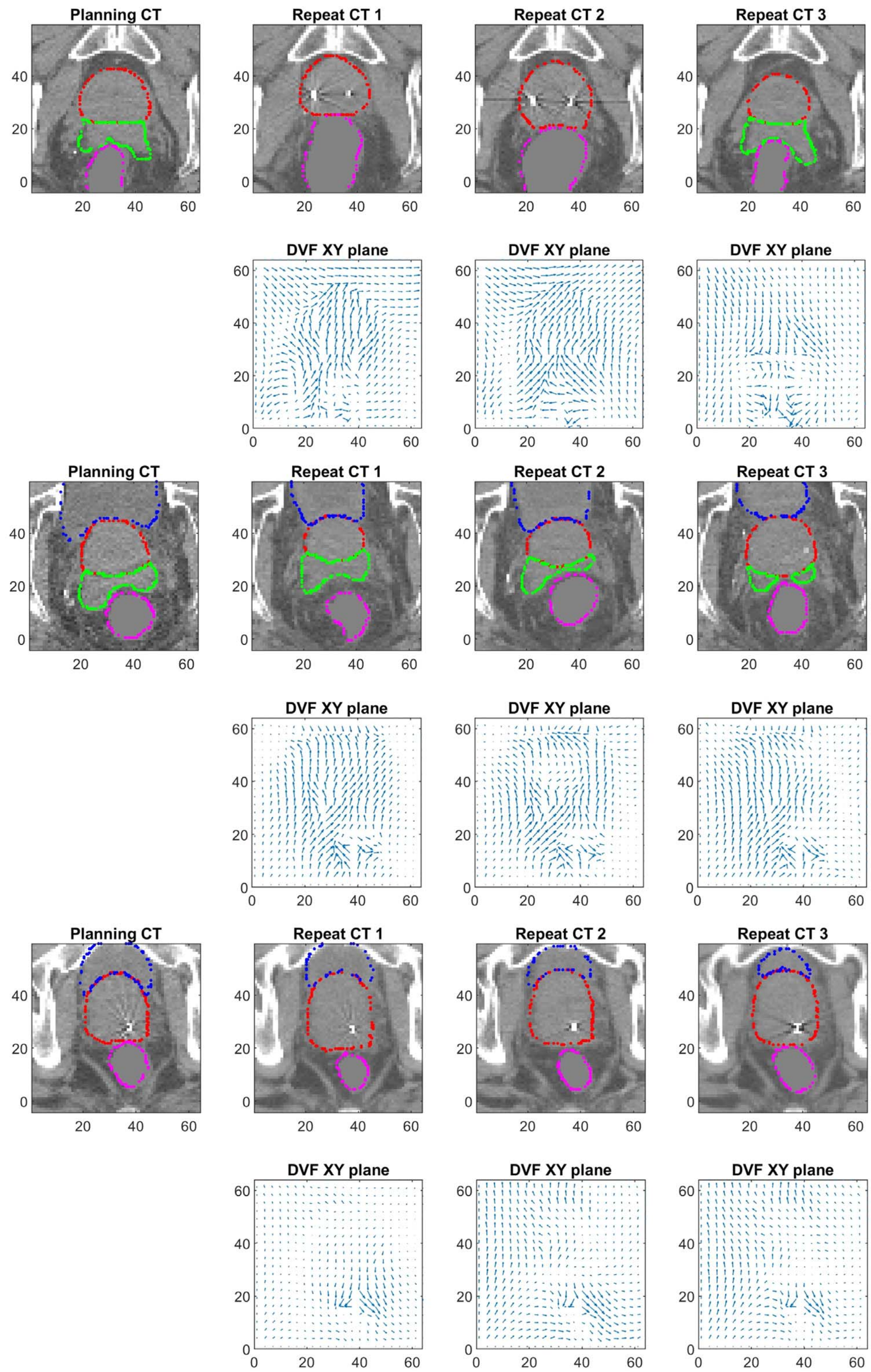


Figure 7. Representative deformation vector fields. The planning CT, three repeat CTs and the corresponding DVF warping the planning CT to each repeat CT are shown for three patients. including prostate (red), seminal vesicles (green), bladder (blue) and rectum (pink) contours. The images correspond to a perpendicular slice in the cranial-caudal axis (*XY* plane).

independent test patients, DAM can generate realistic CTs with anatomical variations that resemble those recorded in the clinics using a small number of latent variables. The structure overlap of a model with a single variable, measured as a DICE score of 0.856 ± 0.058, agrees with that of previous state-of-the-art pure segmentation/registration (non-generative) deep learning studies (Elmahdy *et al* 2019, Yuan *et al* 2019, Elmahdy *et al* 2021, Liang *et al* 2021). Compared to linear PCA models where each eigenvector captures an independent mode of motion, the non-linearities in DAM allow representing different groups of correlated movements using different values of only one latent variable. Given that a single latent variable practically

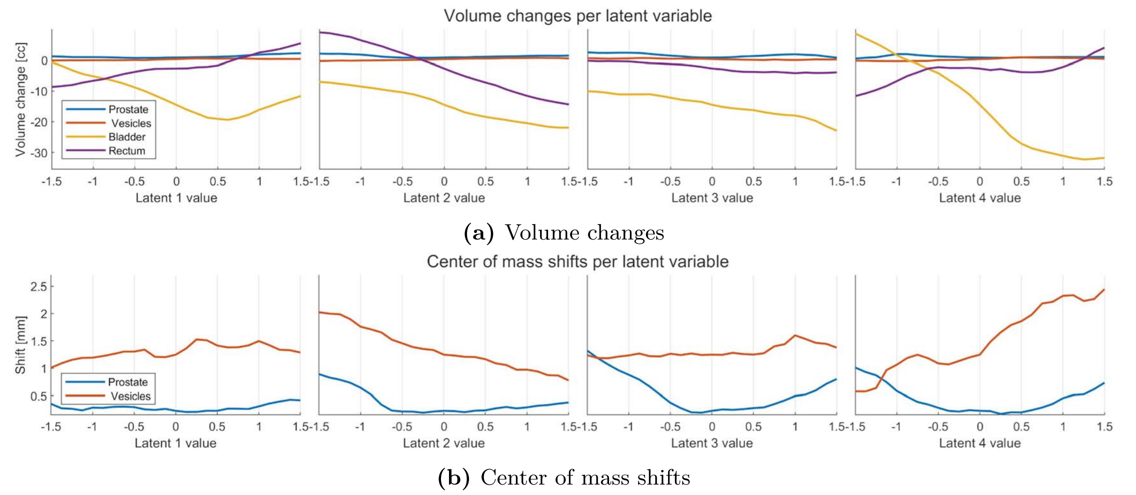


Figure 8. Effect of individual latent variables on deformations. (a) Volume changes and (b) center of mass shifts triggered by independently varying latent variables. For a model with $N = 8$ latent variables, four randomly selected variables are varied between values within -1.5 and 1.5 of their standard deviation, while keeping the remaining seven variables fixed and equal to their mean.

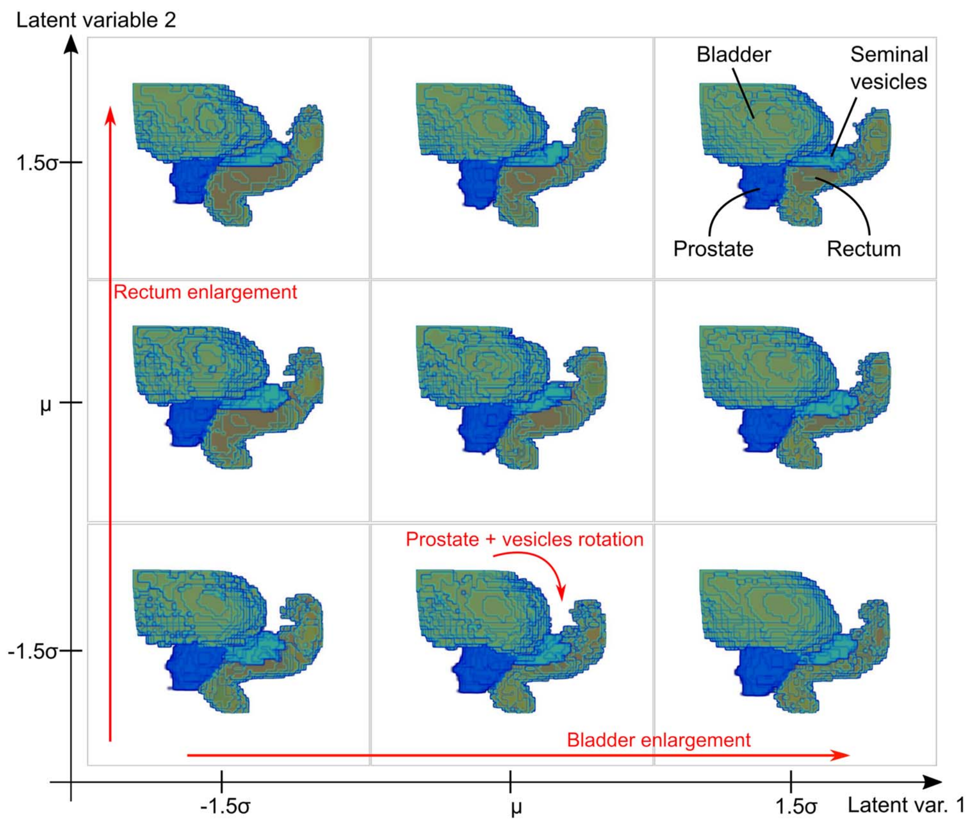


Figure 9. Latent space visualization. Grid plot of the prostate (blue), seminal vesicles (green), bladder (yellow) and rectum (orange) volumes. Each box corresponds to a different combination of latent variables in a 2D plane of the latent space, where the values for each variable are shown on the axes, with σ being the standard deviation and μ the mean. Changes in the horizontal axis translate into bladder enlargements, while the vertical axis controls rectum volume. Correlated groups of movements are observed, e.g. as prostate rotations triggered by an enlarged bladder and smaller rectum.

suffices to capture prostate movements, and that both the CC and rectum DICE score keep increasing with larger latent spaces, we can conclude that most of the computational effort is dedicated to modeling rectum deformations. Prostate IMPT treatments typically solely involve lateral beams, for which the impact of error due to rectum movement is small. In some cases, models with as little as 4–8 variables may be accurate enough, while 8–16 variables additionally ensure accurate rectum deformations for plans requiring more precision.

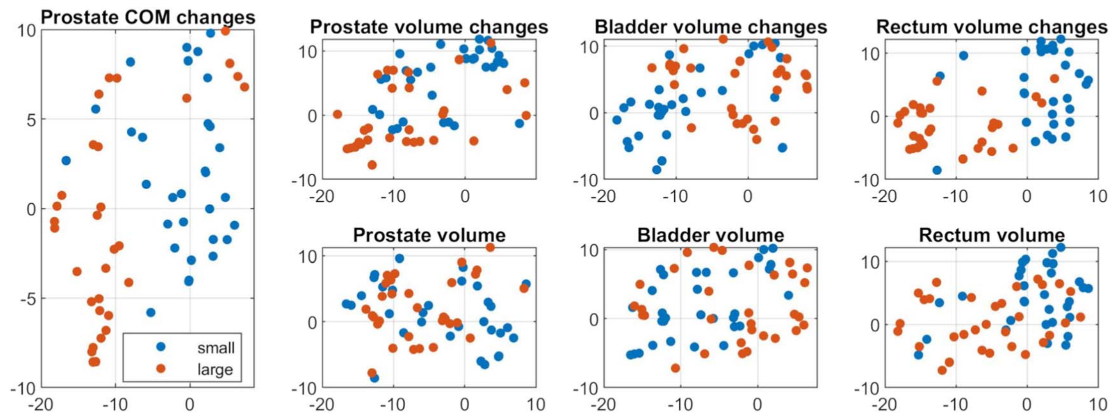


Figure 10. Latent space structure. Each latent variable is reduced to 2D space t-SNE representation and classified, from left to right, according to whether they correspond to small (blue) or large (orange): prostate center of mass shifts, prostate, bladder and rectum volume changes, or prostate, bladder and rectum sizes. 'Small' samples fall below the 10% percentile of all values, while 'large' samples include all values above the 90% percentile.

For clinical application, it is critical that the model generates realistic shifts and deformations of the volume to be irradiated/treated (in this case, the prostate). Overall, based on the results in figures 4 and 5, the center of mass shifts and volume changes produced by DAM show good overlap to the deformations and shifts recorded in the clinic, matching previously reported values (Antolak *et al* 1998). One reason why DAM struggles in simulating the most extreme shifts or slides is the regularization term of the loss, which limits large deformations. Despite this limitation, such large anatomical variations are typically taken care of by adapting the treatment plan to the new anatomy, whereas robust treatment planning and evaluation—the main potential applications of DAM—are in principle oriented to incorporating average, frequent deformations into treatment design and evaluation, and we expect DAM to be useful for such purposes.

Comparison to other methods. All the previously published approaches are either patient-specific or population models based on PCA. Patient-specific methods (Söhn *et al* 2005, Zhang *et al* 2007, Nie *et al* 2012, Thörnqvist *et al* 2013a, 2013b) require at least a few CTs recorded during a patient's treatment, and therefore they are unfeasible for pre-delivery robust treatment planning and evaluation, being restricted to post-treatment analysis. Conversely, population models (Budiarto *et al* 2011, Rios *et al* 2017, Szeto *et al* 2017, Tilly *et al* 2017, Magallon-Baro *et al* 2019) use a set of planning-repeat CT/contour pairs from previous patients, but simulate the same type of deformations for all patients regardless of their anatomy. In contrast, as seen in figures 4 and 5, DAM is able to retrieve patient-specific magnitude and frequency of movements from the entire population-based on the planning CT anatomy, making the model suitable for a wider range of applications.

Most previous studies (Söhn *et al* 2005, Budiarto *et al* 2011, Thörnqvist *et al* 2013b, Magallon-Baro *et al* 2019) model only the surface of the organs and not the intensities values in the CT. Without CT values the dose distributions are always calculated on the same planning CT with varying contours, which limits its applicability, especially in IMPT given the protons' finite range and tissue sensitivity. Conversely, PCA-based models modeling full DVFs require 7 (Tilly *et al* 2017) or up to 100 principal components (Szeto *et al* 2017) to capture 90% of the variance in the training data. A large number of components (equivalent to DAM's latent variables) adds more variation, increases the chance of sampling unrealistic deformations and limits their applicability as reduced order models. Most importantly, all previous population-based methods require a time-consuming pre-processing step involving multiple deformable image registration steps between scans and patients to an organ or CT template. The accuracy of such registration calculation degrades the final accuracy and generative performance of the model, with previous studies (Szeto *et al* 2017, Tilly *et al* 2017) showing surface errors of around 1.5 ± 1.0 mm introduced in their pre-processing step alone that are comparable the DAM's total errors reported in the right plot of figure 3. Given the lack of uniformity in treatment site and evaluation metrics in previous studies—where most focus on evaluating the variance captured by the PCA model components and the errors on the DVFs caused by truncating the number of eigenmodes—we compare DAM's performance to a PCA model of the prostate (Budiarto *et al* 2011) in the right plot of figure 3. Even without adding any pre-processing errors, DAM matches the overall performance and is to capture prostate motion with a lower number of modeling parameters. Being trained directly on CT images in an unsupervised manner, DAM bypasses any performance or time losses from any pre-processing step, and can be easily applied to generate new anatomies in few milliseconds, compared to the tens of minutes or hours needed to obtain accurate enough registrations using conventional clinical software.

Like PCA-based models, DAM assigns realistic correlated deformations to different values of the latent variables. Figures 8 and 9 show that variables control shifts, volume changes and rotations similar to those reported in previous studies (Budiarto *et al* 2011, Magallon-Baro *et al* 2019). Figure 10 demonstrates that the latent variables almost exclusively carry information about deformations, and not about anatomical traits from the patients. Instead, the Encoder is in charge of independently mapping planning anatomies to a subset of latent variables. Furthermore, unlike all previous approaches not evaluating the generative performance of their proposed models, we demonstrate the DAM also generates the adequate range and frequency of deformations for each patient.

Applicability. DAM's main application in robust treatment planning and robust evaluation against inter-fraction movements involves sampling patient anatomies and calculating the corresponding dose distributions. With prediction times of few milliseconds per generated anatomy, DAM offers huge speed-up possibilities for plan evaluation when coupled to fast dose calculation algorithms (Perkó *et al* 2016, Pastor-Serrano and Perkó 2022a, 2022b, Pastor-Serrano *et al* 2022). Few (3–5) representative scenarios corresponding to points around mean of the posterior distribution can be sampled to be used for scenario based robust optimization, which may translate into a dosimetric advantage or be used for margin reduction. In principle, the same modeling framework could be applied to any other treatment site, with additional applications involving obtaining weighted dose scenarios (Tilly *et al* 2017) to formulate anatomical robustness margin recipes (van der Voort *et al* 2016). Straightforward extensions include adapting the DAM framework to other types of movement (e.g. intra-fraction movements) by adding temporal dependence for treatments where patients' anatomies significantly change following a clear pattern during (e.g. breathing) or between the different fractions of the treatment (e.g. modeling tumor shrinkage). Such time-dependent model could be coupled to breathing interplay effect simulation tools (Pastor-Serrano *et al* 2021) to design plans based on breathing signals (Pastor-Serrano *et al* 2021) that mitigate the detrimental effect of movement during delivery.

Limitations. Like PCA-based models, DAM will struggle to generate deformations that are not represented in the training data, for which continuously updating the model (e.g. using cone beam CTs) can be a solution. We refer to this as having limited extrapolation capabilities, e.g. generating bladder emptying or filling when no bladder movement is included in the data, which is a problem shared with linear PCA models. Likewise, low resolution images with poor contrast can also affect performance by masking small movements of structures, especially in areas with similar organ tissue densities. DAM's implementation in the clinic thus requires a quality assurance protocol that evaluates robustness in predictions e.g. by training several models using different data, and evaluating result similarity on a same test dataset.

As for many other deep learning algorithms, DAM's generalization capabilities depend on the size and variability of the data in the dataset, as well as on the quality and resolution of the CT images. Due to the rather small size of the dataset in this preliminary study—caused by the scarcity of recorded sets of planning and repeat CTs—and based on the initial positive results, further testing appears warranted. Although DAM seems to learn representative deformations from the rather small dataset used during training, larger datasets will likely improve model performance, as typically observed in deep learning approaches. Once applied in clinical environments, DAM can be re-trained using more data, including the CTs collected in the center and possible new publicly available datasets. Other options to mitigate the (possible) negative impact of small datasets include generating extra data, .e.g. via a much more computationally expensive mechanistic model (such as the ones developed for heart motion (Gerach *et al* 2021)).

DAM's accuracy in generating reasonable patient-specific movements depends on the extent to which movements can be predicted only from the planning CT and structures. As with other classical and deep learning registration algorithms, DAM would struggle to register rectum structures due to the randomness in their intensity values. Following clinical practice, we opted for masking the rectum voxels with air. As a result, all deformed CTs have air-filled rectum structures, which can affect the accuracy in the dose calculation, especially for beams delivered in the anterior-posterior direction. Possible solutions include adding an additional generative model that generates rectum voxel intensities based on the organ mask shape.

5. Conclusion

We presented DAM, a deep learning-based DAM to simulate patient-specific deformations that may be observed during the course of a prostate cancer radiotherapy treatment. DAM captures groups of correlated movements via few auxiliary latent variables, where few variables are able to model prostate deformations and shifts with similar accuracy as state-of-the-art models based on PCA. Compared to previous population models, DAM can generate realistic CT images and contours in less than a second without any pre-processing, with volume changes and center of mass shifts that match in frequency and range those reported in the clinics and in previous studies. Given its simplicity and speed to generate CTs based on a single planning scan and delineations, DAM can be tested in treatment planning and evaluation to design treatment plans that are robust against inter-fraction variations.

Acknowledgments

This work is supported by KWF Kanker Bestrijding [Grant Number 11711] and is part of the KWF research project PAREL. Zoltán Perkó would like to thank the support of the NWO VENI grant ALLEGRO (016. Veni.198.055) during the time of this study. We would like to thank Haukeland University Hospital (Bergen, Norway), responsible oncologist Svein Inge Helle, physicist Liv Bolstad Hysing and Dr Jeffrey Siebers for providing the CT-data with contours. Lei Xing wishes to acknowledge the supports of the National Institutes of Health (NIH) (1R01CA223667, 1R01CA176553, and 1R01CA227713) and Varian Medical Systems (Palo Alto, CA).

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Code availability

The code, weights and results are publicly available at https://github.com/opaserr/dam.

CRediT authorship contribution statement

Oscar Pastor-Serrano: Conceptualization, Data Curation, Methodology, Investigation, Formal Analysis, Visualization, Software, Writing—original draft, Writing—review and editing. **Steven Habraken**: Resources, Formal Analysis, Writing—review and editing. **Mischa Hoogeman**: Conceptualization, Formal Analysis, Project administration, Funding acquisition, Writing—review and editing. **Danny Lathouwers**: Formal Analysis, Writing—review and editing. **Dennis Schaart**: Formal Analysis, Writing—review and editing. **Yusuke Nomura**: Formal Analysis, Writing—review and editing. **Lei Xing**: Supervision, Formal Analysis, Writing—review and editing. **Zoltán Perkó**: Conceptualization, Supervision, Methodology, Formal Analysis, Funding acquisition, Project administration, Writing—review and editing.

Appendix A. Diffeomorphic transformations

In this section, we provide details about the type of diffeomorphic transformation used in our model, based on the seminal works in (Ashburner 2007, Dalca *et al* 2019). The chosen diffeomorphic transformation is represented via the ordinary differential equation

$$\frac{\partial \Phi^{(t)}}{\partial t} = v(\Phi^{(t)}) \tag{11}$$

describing the evolution of the deformation over time, where $t \in [0, 1]$ is time, $\Phi^{(0)}$ is the identity transformation and v: $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ is the stationary velocity field. To generate a DVF, we start from the identity transformation $\Phi^{(0)}$, integrating over time to obtain $\Phi^{(1)}$. In our case, we scaling and squaring (Moler and Van Loan 2003, Arsigny *et al* 2006), which involves recursively updating the DVF in T successive small time steps

$$\Phi^{(1/2^T)} = \boldsymbol{p} + v(\boldsymbol{p})/2^T \tag{12}$$

$$\Phi^{(1/2^{t-1})} = \Phi^{(1/2^t)} \circ \Phi^{(1/2^t)} \tag{13}$$

$$\Phi^{(1)} = \Phi^{(1/2)} \circ \Phi^{(1/2)}, \tag{14}$$

where $\boldsymbol{p}$ are spatial locations. Typically, T is chosen so that $v(\boldsymbol{p})/2^T$ is small, with higher T leading to more accurate solutions. In Group theory, the velocity field $\boldsymbol{v}$ is a member of the Lie algebra, which is exponentiated to produce the member of the Lie group $\Phi^{(1)} = \exp v$, establishing the connection between the exponentiation and the integration of the ordinary differential equation.

Appendix B. Evidence lower bound

The lower bound (LB) derivation is bassed on Jensen's inequality. For concave functions such as the natural logarithm and a random variable x, Jensen's inequality states that

$$\log\,(\mathbb{E}[x]) \geqslant \mathbb{E}\,[\log(x)]. \tag{15}$$

Starting from the marginal likelihood of the probabilistic model in equation (2), the lower bound is obtained as

$$\log\,(P_\theta(\boldsymbol{y}|\boldsymbol{x},\,s_x)) = \log\int P_\theta(\boldsymbol{y}|\boldsymbol{z},\,\boldsymbol{x},\,s_x)P(\boldsymbol{z}|\boldsymbol{x},\,s_x)d\boldsymbol{z} \tag{16}$$

$$=\log\int P_\theta(\boldsymbol{y}|\boldsymbol{z},\,\boldsymbol{x},\,s_x)P(\boldsymbol{z}|\boldsymbol{x},\,s_x)\frac{Q_\phi(\boldsymbol{z}|\boldsymbol{x},\,s_x,\,\boldsymbol{y},\,s_y)}{Q_\phi(\boldsymbol{z}|\boldsymbol{x},\,s_x,\,\boldsymbol{y},\,s_y)}d\boldsymbol{z} \tag{17}$$

$$=\log\,\mathbb{E}_{\boldsymbol{z}\sim Q_\phi(\boldsymbol{z}|\boldsymbol{x},s_x,\boldsymbol{y},s_y)}\left[\frac{P_\theta(\boldsymbol{y}|\boldsymbol{z},\,\boldsymbol{x},\,s_x)P(\boldsymbol{z}|\boldsymbol{x},\,s_x)}{Q_\phi(\boldsymbol{z}|\boldsymbol{x},\,s_x,\,\boldsymbol{y},\,s_y)}\right] \tag{18}$$

$$\geqslant\mathbb{E}_{\boldsymbol{z}\sim Q_\phi(\boldsymbol{z}|\boldsymbol{x},s_x,\boldsymbol{y},s_y)}\left[\log\left(\frac{P_\theta(\boldsymbol{y}|\boldsymbol{z},\,\boldsymbol{x},\,s_x)P(\boldsymbol{z}|\boldsymbol{x},\,s_x)}{Q_\phi(\boldsymbol{z}|\boldsymbol{x},\,s_x,\,\boldsymbol{y},\,s_y)}\right)\right] \tag{19}$$

$$=\mathbb{E}_{\boldsymbol{z}\sim Q_\phi(\boldsymbol{z}|\boldsymbol{x},s_x,\boldsymbol{y},s_y)}[\log P_\theta(\boldsymbol{y}|\boldsymbol{z},\,\boldsymbol{x},\,s_x)] - D_{KL}(Q_\phi(\boldsymbol{z}|\boldsymbol{x},\,s_x,\,\boldsymbol{y},\,s_y)||P(\boldsymbol{z}|\boldsymbol{x},\,s_x)), \tag{20}$$

where the KL-divergence D_{KL} is defined as

$$D_{KL}(P(x)||Q(x)) = \int\log\left(\frac{P(x)}{Q(x)}\right)P(x)\,dx = \mathbb{E}_{x\sim P(x)}\log\left(\frac{P(x)}{Q(x)}\right). \tag{21}$$

ORCID iDs

Oscar Pastor-Serrano https://orcid.org/0000-0002-2328-1429
Dennis Schaart https://orcid.org/0000-0002-3199-5608
Yusuke Nomura https://orcid.org/0000-0002-2180-568X
Zoltán Perkó https://orcid.org/0000-0002-0975-4226

References

Antolak J A *et al* 1998 Prostate target volume variations during a course of radiotherapy *Int. J. Radiat. Oncol., Biol., Phys.* **42** 661–72
Arsigny V, Commowick O, Pennec X and Ayache N 2006 A log-euclidean framework for statistics on diffeomorphisms *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006* vol 4190 (*Lecture Notes in Computer Science*) ed R Larsen, M Nielsen and J Sporring (Berlin, Heidelberg: Springer) pp 924–31
Ashburner J 2007 A fast diffeomorphic image registration algorithm *Neuro. Image* **38** 95–113
Badawi A M *et al* 2010 Optimizing principal component models for representing interfraction variation in lung cancer radiotherapy *Med. Phys.* **37** 5080–91
Balakrishnan G *et al* 2019 VoxelMorph: a learning framework for deformable medical image registration *IEEE Trans. Med. Imaging* **38** 1788–800
Baumgartner C F *et al* 2019 PHiSeg: Capturing Uncertainty in Medical Image Segmentation *MICCAI 2019: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* pp 119–127
Jagt T *et al* 2017 Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations in prostate cancer *Phys. Med. Biol.* **62** 4254–72
Jagt T *et al* 2018 An automated planning strategy for near real-time adaptive proton therapy in prostate cancer *Phy. Med. Biol.* **63** 135017
Jeong Y, Radke R J and Lovelock D M 2010 Bilinear models for inter-and intra-patient variation of the prostate *Phys. Med. Biol.* **55** 3725–39
Kingma D P and Ba J 2017 Adam: a method for stochastic optimization arXiv:1412. 6980
Kingma D P and Welling M 2014 Auto-encoding variational bayes *2nd Int. Conf. on Learning Representations, ICLR 2014—Conf. Track Proc.* pp 1–14
Kohl S A A *et al* 2018 A probabilistic U-net for segmentation of ambiguous images *Advances in Neural Information Processing Systems* pp 6965–75
Kohl S A A *et al* 2019 A hierarchical probabilistic u-net for modeling multi-scale ambiguities
Krebs J *et al* 2019 Learning a probabilistic model for diffeomorphic registration *IEEE Trans. Med. Imaging* **38** 2165–76
Lei Y *et al* 2020 4D-CT deformable image registration using multiscale unsupervised deep learning *Phys. Med. Biol.* **65** 085003
Liang X *et al* 2021 Automated contour propagation of the prostate from pCT to CBCT images via deep unsupervised learning *Med. Phys.* **48** 1764–70
Magallon-Baro A *et al* 2019 Modeling daily changes in organ-at-risk anatomy in a cohort of pancreatic cancer patients *Radiother. Oncol.* **134** 127–34
Moler C and Van Loan C 2003 Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later *SIAM Rev.* **45** 3–49
Nie X, Liang J and Yan D 2012 Organ sample generator for expected treatment dose construction and adaptive inverse planning optimization *Med. Phys.* **39** 7329–37

Pastor-Serrano O, Lathouwers D and Perkó Z 2021 A semi-supervised autoencoder framework for joint generation and classification of breathing *Comput. Methods Programs Biomed.* **209** 106312

Pastor-Serrano O and Perkó Z 2022a Learning the physics of particle transport via transformers *Proc. of the AAAI Conf. on Artificial Intelligence* vol 36, pp 12071–9

Pastor-Serrano O and Perkó Z 2022b Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy *Phys. Med. Biol.* **67** 105006

Pastor-Serrano O *et al* 2021 How should we model and evaluate breathing interplay effects in IMPT? *Phys. Med. Biol.* **66** 235003

Pastor-Serrano O *et al* 2023 Sub-second photon dose prediction via transformer neural networks arXiv:2209.08923Medical Physics 10.1002/mp.16231

Perkó Z *et al* 2016 Fast and accurate sensitivity analysis of impt treatment plans using polynomial chaos expansion *Phys. Med. Biol.* **61** 4646–64

Rezende D J, Mohamed S and Wierstra D 2014 Stochastic backpropagation and approximate inference in deep generative models arXiv:1401.4082

Rios R *et al* 2017 Population model of bladder motion and deformation based on dominant eigenmodes and mixed-effects models in prostate cancer radiotherapy *Med. Image Anal.* **38** 133–49

Rojo-Santiago J *et al* 2021 Accurate assessment of a Dutch practical robustness evaluation protocol in clinical PT with pencil beam scanning for neurological tumors *Radiother. Oncol.* **163** 121–7

Romaguera L V *et al* 2020 Prediction of in-plane organ deformation during free-breathing radiotherapy via discriminative spatial transformer networks *Med. Image Anal.* **64** 101754

Romaguera L V *et al* 2021 Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy *Med. Image Anal.* **74** 102250

Ronneberger O, Fischer P and Brox T 2015 *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Cham: Springer International Publishing) pp 234–41

Sharma M, Weiss E and Siebers J V 2012 Dose deformation-invariance in adaptive prostate radiation therapy: Implication for treatment simulations *Radiother. Oncol.* **105** 207–13

Söhn M, Birkner M, Yan D and Alber M 2005 Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation *Phys. Med. Biol.* **50** 5893–908

Söhn M, Sobotta B and Alber M 2012 Dosimetric treatment course simulation based on a statistical model of deformable organ motion *Phys. Med. Biol.* **57** 3693–709

Szeto Y Z, Witte M G, van Herk M and Sonke J J 2017 A population based statistical model for daily geometric variations in the thorax *Radiother. Oncol.* **123** 99–105

Thörnqvist S *et al* 2013a Adaptive radiotherapy in locally advanced prostate cancer using a statistical deformable motion model *Acta Oncol.* **52** 1423–9

Thörnqvist S *et al* 2013b Treatment simulations with a statistical deformable motion model to evaluate margins for multiple targets in radiotherapy for high-risk prostate cancer *Radiother. Oncol.* **109** 344–9

Tilly D *et al* 2017 Dose coverage calculation using a statistical shape model - applied to cervical cancer radiotherapy *Phys. Med. Biol.* **62** 4140–59

Unkelbach J and Paganetti H 2018 Robust proton treatment planning: physical and biological optimization *Semin. Radiat. Oncol.* **28** 88–96

van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605

van der Voort S *et al* 2016 Robustness recipes for minimax robust optimization in intensity modulated proton therapy for oropharyngeal cancer patients *Int. J. Radiat. Oncol. Biol. Phys.* **95** 163–70

van Herk M, Remeijer P and Lebesque J V 2002 Inclusion of geometric uncertainties in treatment plan evaluation *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1407–22

Vásquez Osorio E M *et al* 2009 A novel flexible framework with automatic feature correspondence optimization for nonrigid registration in radiotherapy *Med. Phys.* **36** 2848–59

Wu Y and He K 2020 Group normalization *Int. J. Comput. Vision* **128** 742–55

Xu H *et al* 2014 Coverage-based treatment planning to accommodate deformable organ variations in prostate cancer treatment *Med. Phys.* **41** (10) 101705

Yuan Y *et al* 2019 Prostate segmentation with encoder-decoder densely connected convolutional network (ed-densenet) *2019 IEEE 16th Int. Symp. on Biomedical Imaging (ISBI 2019)* (IEEE) pp 434–7

Zhang Q *et al* 2007 A patient-specific respiratory model of anatomical motion for radiation treatment planning *Med. Phys.* **34** 4772–81