A deep learning model for inter-fraction head and neck anatomical changes in proton therapy

Burlacu, T.; Hoogeman, M.S.; Lathouwers, D.; Perko, Z.

DOI
10.1088/1361-6560/adba39

Publication date
2025

Document Version
Final published version

Published in
Physics in Medicine and Biology

Citation (APA)
Burlacu, T., Hoogeman, M. S., Lathouwers, D., & Perko, Z. (2025). A deep learning model for inter-fraction head and neck anatomical changes in proton therapy. *Physics in Medicine and Biology*, *70*(6), Article 065011. https://doi.org/10.1088/1361-6560/adba39

A deep learning model for inter-fraction head and neck anatomical changes in proton therapy

IOP Publishing Phys. Med. Biol. 70 (2025) 065011 https://doi.org/10.1088/1361-6560/adba39

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
12 November 2024

REVISED
6 February 2025

ACCEPTED FOR PUBLICATION
21 February 2025

PUBLISHED
10 March 2025

PAPER

A deep learning model for inter-fraction head and neck anatomical changes in proton therapy

Tiberiu Burlacu[1,3,*], Mischa Hoogeman[1,2,3], Danny Lathouwers[1,3] and Zoltán Perkó[1,3]

[1] Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands
[2] Department of Radiotherapy, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands
[3] HollandPTC Consortium[4], Delft, The Netherlands
[*] Author to whom any correspondence should be addressed.

E-mail: t.burlacu@tudelft.nl

Keywords: proton therapy, deep learning, variational autoencoder, anatomy changes

Abstract

Objective. To assess the performance of a probabilistic deep learning based algorithm for predicting inter-fraction anatomical changes in head and neck patients. *Approach.* A probabilistic daily anatomy model (DAM) for head and neck patients DAM ($\mathrm{DAM_{HN}}$) is built on the variational autoencoder architecture. The model approximates the generative joint conditional probability distribution of the repeat computed tomography (rCT) images and their corresponding masks on the planning CT images (pCT) and their masks. The model outputs deformation vector fields, which are used to produce possible rCTs and associated masks. The dataset is composed of 93 patients (i.e. 315 pCT–rCT pairs), 9 (i.e. 27 pairs) of which were set aside for final testing. The performance of the model is assessed based on the reconstruction accuracy and the generative performance for the set aside patients. *Main results.* The model achieves a DICE score of 0.83 and an image similarity score normalized cross-correlation of 0.60 on the test set. The generated parotid glands, spinal cord and constrictor muscle volume change distributions and center of mass shift distributions were also assessed. For all organs, the medians of the distributions are close to the true ones, and the distributions are broad enough to encompass the real observed changes. Moreover, the generated images display anatomical changes in line with the literature reported ones, such as the medial shifts of the parotids glands. *Significance.* $\mathrm{DAM_{HN}}$ is capable of generating realistic anatomies observed during the course of the treatment and has applications in anatomical robust optimization, treatment planning based on plan library approaches and robustness evaluation against inter-fractional changes.

1. Synthetic CT uses in proton therapy (PT)

PT has desirable dose characteristics, such as similar target coverage and lower organs at risk (OAR) doses, when compared to traditional photon based radiotherapy (RT) (Chen *et al* 2023). However, the increased dose conformality implies an increased susceptibility to dose degradation by uncertainties such as setup errors, range errors and anatomical changes over the course of the typically month long treatment duration (van Kranen *et al* 2009). To diminish the dose degradation, robust optimization and evaluation (Unkelbach and Paganetti 2018) with isotropic setup and range settings (Liu *et al* 2013) and offline adaptive replanning (Deiter *et al* 2020) is performed in clinical practice. This results in a high dose region that surrounds the target, which in the case of the head and neck (H&N) region where OARs are in close proximity to the target, could result in high chances of side effects. Moreover, there are certain anatomical changes (e.g. tumor shrinkage Cubillos-Mesías *et al* 2019) that are not effectively accounted for by robust optimization only

[4] HollandPTC Consortium—Erasmus Medical Center, Rotterdam, Holland Proton Therapy Centre, Delft, Leiden University Medical Center (LUMC), Leiden and Delft University of Technology, Delft, The Netherlands.

taking setup and range errors into account. One proposed option (Van de Water *et al* 2018) is the inclusion of additional (synthetic) CT images in the (anatomical) robust optimization process. While this provided increased target coverage and lower OAR doses for the specific H&N patients in the cohort, compared to conventional robust optimization, it still created a high dose region surrounding the target.

To reduce this region to its minimum and counter long and short-term inter-fraction occurring anatomical variations, online adaptive PT (OAPT) has been proposed. In this workflow, a new CT is acquired for each fraction and within a short time a new fully re-optimized plan is generated. The resulting plan would only need minimal robustness settings to counter the effects of range uncertainties, machine related setup uncertainties and remaining intra-fraction uncertainties. The short time available and the limited computational resources imply that fully robust reoptimization in the online setting still requires research (Oud *et al* 2024) and is not feasible clinically. The plan library (PL) approach was proposed as an intermediate solution (van de Schoot *et al* 2016, Oud *et al* 2022). This approach used the planning CT image (pCT) to generate multiple plans with varying robustness settings. On the given day, it administers an appropriately chosen plan, therefore resulting in NTCP reductions or sometimes in increased robustness that ensures adequate target coverage. In this approach, synthetic CT images could be used to expand the pre-compiled library of plans, by generating optimal plans for the future patient anatomies predicted by the model. An additional use case for synthetic CT images could be for plan QA, in the scenario in which an adapted or refined (e.g. by using yesterday's optimal plan) is generated with the patient on the table. Specifically, several CT images with associated truly optimal plans, could be generated *a priori*. On the given day, a fast dosimetric check can be performed between the adapted and refined plan and the truly optimal pre-generated plan.

Thus, models of inter-fractional anatomical changes have applications in several PT related workflows such as anatomical robust optimization, plan quality assurance in OAPT or expanding the PL approach. Multiple approaches to synthetic CT generation have been employed, such as principal component analysis (PCA) or deep learning. An overview of the different possible approaches is given by the work of Smolders *et al* (2024). Deep learning models have been shown to outperform PCA based ones in the case of prostate anatomies (Pastor-Serrano *et al* 2023) and denoising diffusion probabilistic models (DDPMs) (Smolders *et al* 2024) were successfully applied for artificial CT generation for the H&N site where they were additionally shown to increase robustness to anatomical changes. This work builds upon the previous publication of Pastor-Serrano *et al* (2023) on a generative deep learning daily anatomy model (DAM) for prostate inter-fractional anatomical changes. The model architecture and the data processing pipeline are changed and thereafter applied to a H&N RT cohort. The model is referred to from here on as $\mathrm{DAM}_{\mathrm{HN}}$. Section 2 details the probabilistic framework of the model. Section 3 provides details on the dataset generation and the specific architecture configuration used for training. Section 4 contains the results and their discussion. The performance of the model was assessed via several tests. The results of a reconstruction accuracy test are shown in section 4.1. The generative performance was assessed in terms of the model's capability to predict realistic anatomical changes. To this end, an overview of the typical changes in H&N patients reported by literature studies is given in section 4.2. The anatomical changes present on the training set are discussed in section 4.3. Section 4.4 presents and discusses the anatomical changes predicted by the model. Section 4.5 compares these anatomical changes with the ones presented in the recently published DDPMs DiffuseRT model (Smolders *et al* 2024). Lastly, a latent space analysis is presented in section 4.6. Section 5 concludes this work and discusses some improvement points.

2. Model architecture

This section provides only the main details of this model's architecture. An in-depth exposition can be found in Pastor-Serrano *et al* (2023). The patient anatomy at a certain point in time is described by the CT image and the associated RT structures (masks), which are both taken as random variables. On the pCT, an image with N voxels is denoted by $\mathbf{x} \in \mathbb{R}^N$ (defined as floats due to the need to normalize the data prior to further processing) and the corresponding structures (pM) are denoted by $\mathbf{s}_x \in \mathbb{R}^N$. On the repeat CT images (rCTs), the image is denoted by $\mathbf{y} \in \mathbb{R}^N$ and the corresponding masks (rM) by $\mathbf{s}_y \in \mathbb{R}^N$. Generally, pCTs and rCTs do not have the same dimensionality and to achieve this, the images are resampled and cropped.

The presence of anatomical deformations over the course of treatment, e.g. the systematic medial translation of the lateral regions of the parotid glands, the shrinkage of the parotid and submandibular glands (Fiorentino *et al* 2012), the change in the parotid shape from convex to flat or concave (dos Santos *et al* 2020) and the center of mass (COM) shifts towards the medial side (Vásquez Osorio *et al* 2008) motivates the existence of an unknown generative joint conditional probability distribution $P^*(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ of the voxel CT HU values $\mathbf{y}$ and the structure masks $\mathbf{s}_y$ conditioned on the planning CT $\mathbf{x}$ and structures $\mathbf{s}_x$. If such a distribution was known, given a new pCT and pM, it could be sampled to generate future possible

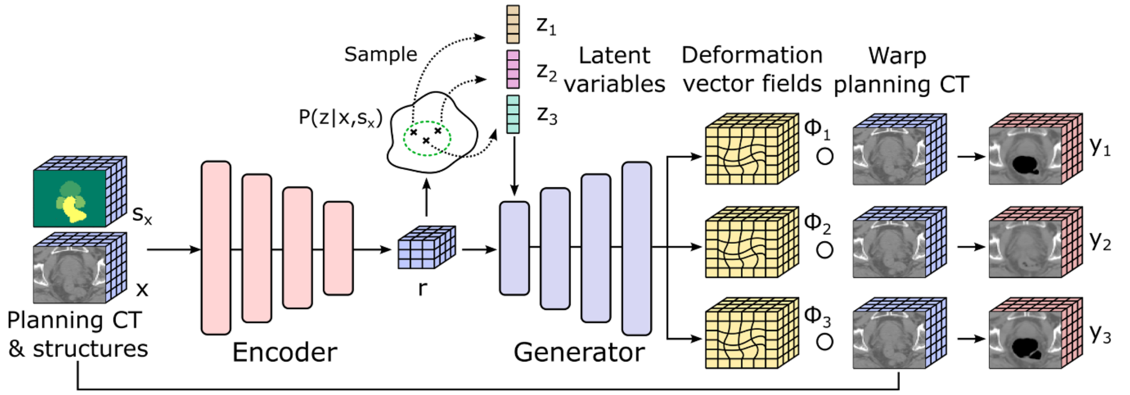


Figure 1. The proposed variational autoencoder model. The Encoder represents the down-sampling part of a U-net that computes the parameters of $P(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)$. The up-sampling part of the U-net, denoted by Generator, takes samples from $P(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)$ together with a reduced representation of the inputs $\mathbf{x}, \mathbf{s}_x$ and computes artificial CT images $\mathbf{y}, \mathbf{s}_y$. Figure reproduced from Pastor-Serrano *et al* (2023). The Author(s). CC BY 4.0.

anatomies, denoted by $\mathbf{y}$ and $\mathbf{s}_y$. In general it is impossible to find such a distribution, and a good approximation $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x) \approx P^*(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ is sought instead. The distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ is parametrized by a vector of parameters $\boldsymbol{\theta}$ that is learned during training.

The dataset $\mathbb{D}$ consists of elements $\mathbf{s}^i \in \mathbb{R}^{4N}$, which are the concatenation of a given pCT and rCT and their associated structures, i.e. $\mathbb{D} = \{\boldsymbol{\tau}_i = (\mathbf{x}_i, \mathbf{s}_{x_i}, \mathbf{y}_i, \mathbf{s}_{y_i})\,|\, i = 1, \ldots, N_D\}$ with N_D the number of elements in the dataset. Moreover, the dataset $\mathbb{D}$ is assumed to be independent and identically distributed (i.i.d.). As the dataset $\mathbb{D}$ is i.i.d. the log-probability assigned to the data is

$$\log P_{\boldsymbol{\theta}}\left(\mathbb{D}\right) = \sum_{\boldsymbol{\tau}\in\mathbb{D}} \log P_{\boldsymbol{\theta}}\left(\boldsymbol{\tau}\right). \tag{1}$$

The framework of maximum likelihood (ML) searches for the parameters $\boldsymbol{\theta}$ that maximize the sum, or equivalently the average, of the log-probabilities assigned to the data by the model in equation (1) (Kingma and Welling 2019).

As most explicitly parametrized generative distributions are too simplistic to model inter-fractional anatomical variations, implicitly parametrized distributions are considered instead. Therefore, a joint conditional probability distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x)$ that also depends on latent variables $\mathbf{z}$ is constructed. Latent variables are variables that are not observed, and therefore they are not part of the dataset of images and associated structures. They are meant to encode (represent in a lower dimensional space) the information between the pCT and the rCT. The marginal distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x)$ over the observed variables $\mathbf{y}, \mathbf{s}_y$ is recovered by marginalizing, namely

$$\begin{aligned} P_{\boldsymbol{\theta}}\left(\mathbf{y}, \mathbf{s}_y|\mathbf{x}, \mathbf{s}_x\right) &= \int \mathrm{d}\mathbf{z}\, P_{\boldsymbol{\theta}}\left(\mathbf{y}, \mathbf{s}_y, \mathbf{z}|\mathbf{x}, \mathbf{s}_x\right) \\ &= \int \mathrm{d}\mathbf{z}\, P_{\boldsymbol{\theta}}\left(\mathbf{y}, \mathbf{s}_y|\mathbf{z}, \mathbf{x}, \mathbf{s}_x\right) P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{x}, \mathbf{s}_x\right). \end{aligned} \tag{2}$$

This is also referred to as the (single datapoint) marginal likelihood, or model evidence, when taken as a function of $\boldsymbol{\theta}$ (Ghojogh *et al* 2022). The distribution $P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x}, \mathbf{s}_x)$ is called the prior distribution, which in the case of this work is taken as a multivariate Normal distribution with mean and variance that depend on the pCT and pM and on the vector of learned parameters $\boldsymbol{\theta}$, namely

$$P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{x}, \mathbf{s}_x\right) = \mathcal{N}\left(\mathbf{z}; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\mathbf{x}, \mathbf{s}_x\right), \Sigma_{\boldsymbol{\theta}}\left(\mathbf{x}, \mathbf{s}_x\right)\right). \tag{3}$$

The dependence of the parameters of the prior distribution on the pCT and pM results in a different distribution for each patient (insofar as a patient is identified with a single image). The mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}_x)$ and the covariance matrix $\Sigma_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{s}_x)$ are computed in the down-sampling part of a U-net neural network and the parameters $\boldsymbol{\theta}$ of the prior are the weights of the encoder, as illustrated in figure 1.

The up-sampling part of the U-net, denoted by Generator in figure 1, outputs a deformation vector field (DVF) $\Phi : \mathbb{R}^{N\times 3} \to \mathbb{R}^{N\times 3}$ used to map coordinates $\mathbf{p} \in \mathbb{R}^3$ between images. The DVF Φ is used to obtain the prediction of the model $\mathbf{y} = \Phi \circ \mathbf{x}$ (Jaderberg *et al* 2016). Based on work by Krebs *et al* (2019), the distribution $P_{\boldsymbol{\theta}}(\mathbf{y}, \mathbf{s}_y|\mathbf{z}, \mathbf{x}, \mathbf{s}_x)$ (referred to as the likelihood) is taken as a function of the normalized cross-correlation (NCC) between the ground truth image $\hat{\mathbf{y}}$ and the predicted image $\mathbf{y}$ with an additional scaling factor $w_{NCC} \in \mathbb{R}^{+}$, namely

$$P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y|\mathbf{z},\mathbf{x},\mathbf{s}_x\right) = \exp\left(-w_{\mathrm{NCC}}\mathrm{CC}\left(\mathbf{y},\hat{\mathbf{y}}\right)\right), \tag{4}$$

where the CC term is defined as

$$\mathrm{CC}\left(\mathbf{y},\hat{\mathbf{y}}\right) = \sum_{\mathbf{p}\in\Omega} \frac{\left[\sum_{i=1}^{n^3}\left(\hat{\mathbf{y}}\left(\mathbf{p}_i\right)-\hat{w}\left(\mathbf{p}\right)\right)\left(\mathbf{y}\left(\mathbf{p}_i\right)-w\left(\mathbf{p}\right)\right)\right]^2}{\left[\sum_{i=1}^{n^3}\left(\hat{\mathbf{y}}\left(\mathbf{p}_i\right)-\hat{w}\left(\mathbf{p}\right)\right)\right]\left[\sum_{i=1}^{n^3}\left(\mathbf{y}\left(\mathbf{p}_i\right)-w\left(\mathbf{p}\right)\right)\right]}, \tag{5}$$

and $w(\mathbf{p})$ and $\hat{w}(\mathbf{p})$ are the local mean over a small cube Ω with side length n voxels of the generated and true images, namely

$$w\left(\mathbf{p}\right) = \frac{1}{n^3}\sum_{j=1}^{n^3}\mathbf{y}\left(\mathbf{p}_j\right), \text{ and } \hat{w}\left(\mathbf{p}\right) = \frac{1}{n^3}\sum_{j=1}^{n^3}\hat{\mathbf{y}}\left(\mathbf{p}_j\right).$$

The vector of parameters $\boldsymbol{\theta}$ of the likelihood distribution $P_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{s}_y|\mathbf{z},\mathbf{x},\mathbf{s}_x)$, that stores in part of it the weights of the Encoder network, also stores the weights of the Generator network.

The main difficulty of this proposed framework is that the marginal probability of the data, or the model evidence, given in equation (2) is intractable due to not having an analytic solution or an efficient estimator. In turn, this makes optimization of such a model computationally expensive.

2.1. Learning the optimal parameters

To overcome the previously mentioned intractability of the framework, the posterior distribution $P_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x)$ is approximated by a multivariate Normal distribution $Q_{\boldsymbol{\psi}}(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x)$ parametrized by a vector of parameters $\boldsymbol{\psi}$ with mean and variance that depend on both the planning and repeat images and masks, namely

$$Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right) = \mathcal{N}\left(\mathbf{z};\boldsymbol{\mu}_{\boldsymbol{\psi}}\left(\mathbf{x},\mathbf{s}_x,\mathbf{y},\mathbf{s}_y\right),\Sigma_{\boldsymbol{\psi}}\left(\mathbf{x},\mathbf{s}_x,\mathbf{y},\mathbf{s}_y\right)\right). \tag{6}$$

The parameters $\boldsymbol{\psi}$ are the weights of the down-sampling part of a U-net, referred to as Inference network at the top of figure 2.

Regardless of the choice of the approximating posterior distribution $Q_{\boldsymbol{\psi}}$, the log-likelihood of the data can be written as

$$\begin{aligned}\log P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y|\mathbf{x},\mathbf{s}_x\right) &= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y|\mathbf{x},\mathbf{s}_x\right)\right]\\ &= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log\frac{P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y,\mathbf{z}|\mathbf{x},\mathbf{s}_x\right)}{P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)}\right]\\ &= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log\frac{P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y,\mathbf{z}|\mathbf{x},\mathbf{s}_x\right)}{Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)}\frac{Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)}{P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)}\right]\\ &= \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log\frac{P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y,\mathbf{z}|\mathbf{x},\mathbf{s}_x\right)}{Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)}\right] \qquad (7)\\ &\quad + D_{KL}\left(Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)||P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)\right). \qquad (8)\end{aligned}$$

The D_{KL} term in equation (8) defines the Kullback–Leibler divergence between the approximated posterior distribution and the true posterior distribution. The term is non-negative, measures the distance between the shapes of the two distributions, and is zero if, and only if, the approximated posterior equals the true posterior. The expectation term in equation (7), defines the evidence lower bound (ELBO) as

$$\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}} = \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y,\mathbf{z}|\mathbf{x},\mathbf{s}_x\right) - \log Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)\right],$$

which can also be re-written as

$$\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}} = \mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}\left[\log P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y|\mathbf{z},\mathbf{x},\mathbf{s}_x\right)\right] - D_{KL}\left(Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)||P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{x},\mathbf{s}_x\right)\right). \tag{9}$$

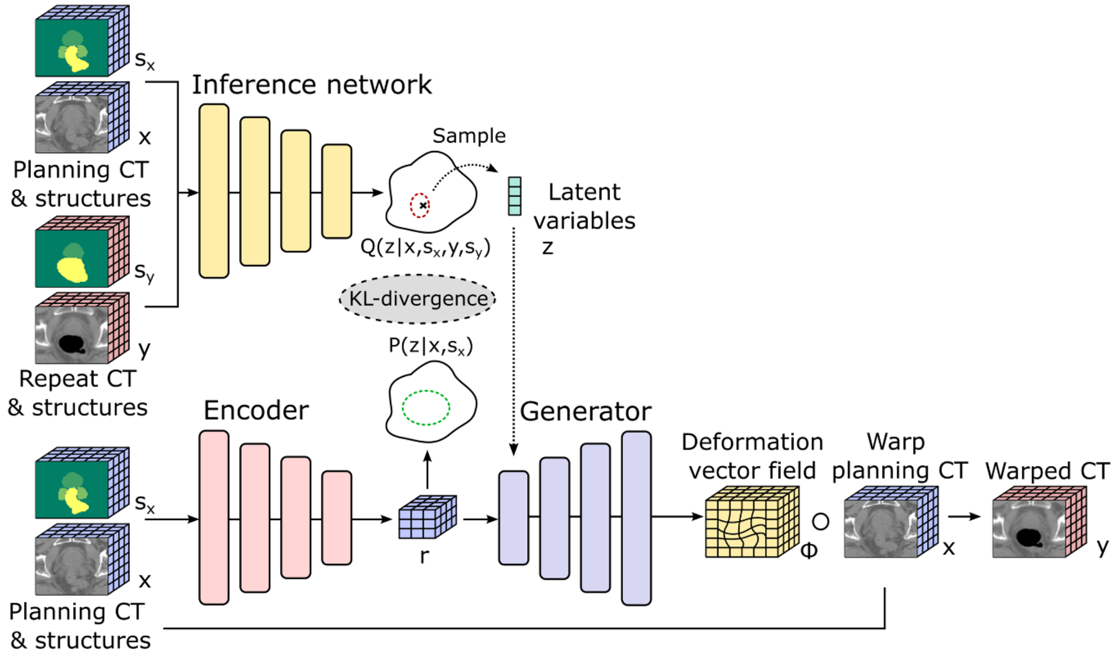


Figure 2. Architecture for finding the optimal parameters $\boldsymbol{\theta}$, $\boldsymbol{\psi}$ of the network. Figure reproduced from Pastor-Serrano *et al* (2023). The Author(s). CC BY 4.0.

As the D_{KL} term is non-negative, it is clear that the ELBO is a lower bound on the log-likelihood of the data, i.e.

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}} &= \log P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y|\mathbf{x},\mathbf{s}_x\right) - D_{KL}\left(Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)||P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)\right)\\ &\leqslant \log P_{\boldsymbol{\theta}}\left(\mathbf{y},\mathbf{s}_y|\mathbf{x},\mathbf{s}_x\right).\end{aligned}$$

Thus, by maximizing the ELBO $\mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\psi}}$ from equation (9) with respect to the parameters of the model $\boldsymbol{\theta},\boldsymbol{\psi}$, the marginal likelihood $P_{\boldsymbol{\theta}}$ is approximately maximized resulting in a better generative model and the KL divergence between the approximated posterior and the true posterior is lowered.

To improve model performance, the ELBO is expanded with two additional terms which are included via multiplication to the likelihood from equation (9) (Pastor-Serrano *et al* 2023). The first is a spatial regularization term,

$$R(\Phi) = -w_{\text{REG}}\sum_{\mathbf{p}\in\Omega}\|\boldsymbol{\nabla}\Phi(\mathbf{p})\|_2, \tag{10}$$

where w_{REG} is a multiplication constant. This term penalizes large and unrealistic gradients in the deformation and encourages neighboring voxels to deform somewhat similarly.

The second is a segmentation regularization term using the DICE score is added, which is also multiplied by a constant w_{DICE}. This aims to improve the overlap between the propagated and ground truth structures, and is written as

$$\text{DICE}\left(\mathbf{s}_y^k,\hat{\mathbf{s}}_y^k\right) = 2\,w_{\text{DICE}}\frac{\left|\mathbf{s}_y^k\cap\hat{\mathbf{s}}_y^k\right|}{\left|\mathbf{s}_y^k\right|+\left|\hat{\mathbf{s}}_y^k\right|}, \tag{11}$$

where k denotes the index of the structure present in the CT image, $k=1,\ldots,K$, with K the total number of structures present, and $\mathbf{s}_y^k$ and $\hat{\mathbf{s}}_y^k$ are the kth generated and ground truth structures respectively.

Including these two additional terms, and minimizing the negative ELBO instead, results in the following optimization problem

$$\begin{aligned}\min_{\boldsymbol{\theta},\boldsymbol{\psi}}\mathbb{E}_{\mathbf{z}\sim Q_{\boldsymbol{\psi}}}&\left[-w_{\text{NCC}}\text{NCC}(\mathbf{y},\hat{\mathbf{y}}) - w_{\text{DICE}}\frac{1}{K}\sum_{k=1}^{k}\text{DICE}\left(\mathbf{s}_y^k,\hat{\mathbf{s}}_y^k\right) + w_{\text{REG}}\sum_{\mathbf{p}\in\Omega}\|\boldsymbol{\nabla}\Phi(\mathbf{p})\|_2\right]\\ &+ w_{KL}D_{KL}\left(Q_{\boldsymbol{\psi}}\left(\mathbf{z}|\mathbf{y},\mathbf{s}_y,\mathbf{x},\mathbf{s}_x\right)||P_{\boldsymbol{\theta}}\left(\mathbf{z}|\mathbf{x},\mathbf{s}_x\right)\right).\end{aligned}$$

3. Dataset generation and training details

This retrospective dataset was acquired from the Holland PT Center and came from 93H&N patients with planning, rCTs and associated RT structures for each image. This resulted in 342 pCT–rCT pairs from which 10%, corresponding to 9 patients, were set aside for final testing. The remaining part was divided into 5% for validation and 95% for training. The training dataset consisted of patients with a number of rCTs ranging from 1 to 6, with most patients having 3 (24 patients) and 4 (25 patients) rCTs taken. This creates a bias in the dataset for the anatomical changes present in patients that are more likely to be re-imaged. All the rCTs were rigidly registered to the pCTs using the Simple ITK library (Beare *et al* 2018) with the resulting deformation vector fields used to register the RT masks. After this, all scans were interpolated to a $2 \times 2 \times 2$ mm grid and cropped around the COM of the present RT masks (the left and right parotid glands, the spinal cord and the constrictor muscle) into a shape of $96 \times 96 \times 64$ voxels. This resulted in volumes of $192 \times 192 \times 128$ mm^3 which were found to adequately cover the anatomical regions of interest.

The model was implemented in PyTorch (Paszke *et al* 2017). The down-sampling path of the U-net (Encoder) and the Inference network were identical, and consisted of 4 blocks, where each block is composed of a 3D convolution layer, a Group Normalization layer, a rectified linear activation and a max pooling down-sampling operation. All convolution layers had a kernel of dimensions $3 \times 3 \times 3$. The convolution layer in the first block had 16 channels while the remaining blocks had 32. At the lowest level, a last convolution with 4 channels results in the encoded volume $\mathbf{r} \in \mathbb{R}^{4\times4\times4\times3}$. This volume is mapped to the means and variances via two different fully-connected layers. The up-sampling part of the U-net (Generator) concatenates the sampled latent variables to the volume $\mathbf{r}$ after a linear layer. Next, 7 blocks (with up-sampling as opposed to down-sampling max pooling operations) are applied, where for the first 5 the convolutional layer has 32 channels and for the last 2, the convolutional layer has 16 channels. This is followed by a last convolution with 3 channels. The model was trained using a batch size of 32, on a A40 NVIDIA GPU, for 1500 epochs with an early stopping patience of 300 epochs and the Adam optimizer with a learning rate of 1.0×10^{-4}.

The constants $w_{\mathrm{NCC}}, w_{\mathrm{DICE}}, w_{\mathrm{REG}}$ together with the constant w_{KL} that multiplied the D_{KL} loss term were considered as hyparparameters to be optimized. These hyperparameters were optimized on the validation set using a grid search method with the validation loss defined as the sum of the NCC from equation (4) and DICE from equation (11) with unity weights. Thus, for a given latent space dimension, ranges of allowed values were defined for each hyperparameter (w_{NCC} and w_{DICE} from 1000 to 5000 in steps of 1000, w_{REG} from 1.0×10^{-5} to 1.0×10^{-1} in multiples of 10 and w_{KL} from 1.0×10^{-3} to 1.0×10^{1} in multiples of 10). After each combination was tested, the model with the lowest validation loss was chosen. This resulted in the model with $w_{\mathrm{NCC}} = 5000$, $w_{\mathrm{DICE}} = 3000, w_{\mathrm{REG}} = 1.0 \times 10^{-4}$ and $w_{\mathrm{KL}} = 1$.

4. Results and discussion

This section presents and discusses the performance of the model in a series of tests. The section starts by presenting and discussing in section 4.1 the performance of the model on the test set (a reconstruction accuracy test). Next, a baseline is set through a literature study for the expected anatomical changes in H&N patients in section 4.2. The anatomical changes displayed by the training set are compared to the expectations set out by literature, in section 4.3, in order to assess the degree to which the dataset used by this model is representative of the broader population. Given this framework, the generative performance of the model is presented and discussed in section 4.4. To gain insight into the model, a latent space analysis is presented and discussed in section 4.6. Lastly, a comparison to the recent diffusion model proposed by Smolders *et al* (2024) is given in section 4.5.

4.1. Test set accuracy

The reconstruction accuracy of the model on the test set was assessed. The accuracy was defined by two metrics, namely the normalized cross correlation (NCC) loss from equation (4) and the DICE loss from equation (11). Thus, each record in the test set (i.e. pair of pCT and rCT with associated masks) was used to generate through the inference network latent variables, which ultimately result in generated CTs and associated structures. The results were averaged over all records in the test set and the dimension of the latent space was varied between 2 and 256 in multiples of 2. The results of the two scores can be seen in figure 3, which shows the mean of the individual scores and a band of one standard deviation (SD) around the mean. Both figures show considerable improvement in both metrics as the latent space is increased from 2 to 32, and thereafter a plateau occurring between 64 and 256. The same behavior is observed in both metrics (a rise in accuracy up to ≈32 latent variables and thereafter a plateau). It should be noted that the metrics are sensitive to different features (as the NCC metric is computed based on the HU values of the voxels while the

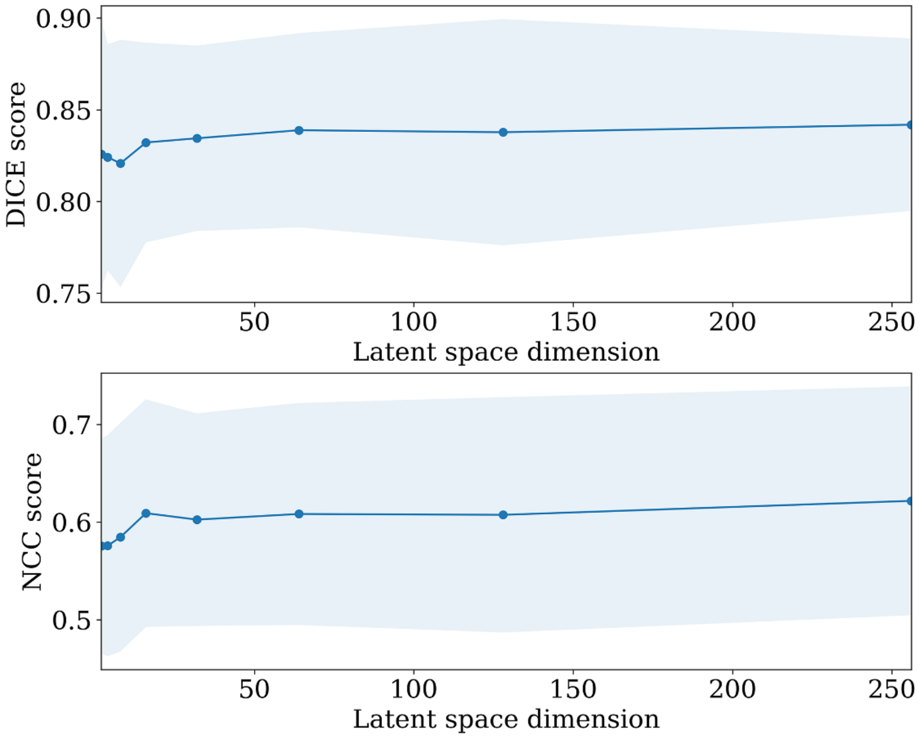


Figure 3. Reconstruction accuracy versus latent space dimensionality. The figure displays on the first row the DICE score and on the second row the NCC evaluated on the test set for models with the same hyperparameters and varying latent space dimensions.

DICE score is computed based on the overlap of the binary masks). Thus, the similar behavior that is observed is likely due to the chosen loss functions that include both the NCC and the DICE score. The model performs particularly well with regard to the DICE score, where it achieves a score of 0.82 with just 2 latent variables. The reconstruction accuracy values obtained were not directly comparable with the ones previously published in the work of Pastor-Serrano *et al* (2023), but exhibit the same behavior. The increased input size of this model (96 × 96 × 64 versus 64 × 64 × 48), the more complex anatomical site (H&N versus prostate) and a different configuration of the layers in the Inference, Encoder and Generator networks likely explain the need for additional latent variables to achieve good performance.

4.2. Expected anatomical changes of parotid glands

This subsection details the anatomical changes in H&N RT patients that literature studies report on. An overview of these changes can be seen in table 1. This overview is used in section 4.3 to assess the degree to which the changes observed in the training set, and therefore the changes that the $\mathrm{DAM_{HN}}$ learns to predict, correspond to the ones in the broader population.

The work of Bhide *et al* (2010) used repeat CT scans at weeks 2, 3, 4, and 5 during RT and compared the parotids and the target at succesive time points, i.e. pretreatment with week 2, week 2 with week 3, and so on. The greatest absolute and percent reduction in the volume of the parotid glands was 4200 $\mathrm{mm^3}$ or 14.7%, and occurred between week 0 and week 2. The absolute and percent reduction in the next two-week period was 4000 $\mathrm{mm^3}$ or 16%. The study found a significant medial shift of the parotid glands through the course of treatment, starting at week 2, with the highest mean movement of the COM being 2.3 mm at week 4. No significant movements of the COM in the anteroposterior and the inferosuperior directions were found.

In the work of Vásquez Osorio *et al* (2008) the impact of 46 Gy delivered to the tumor was assessed based on the planning and rCTs. They report that the parotids shrunk on average by 14% and that the shrinkage occurred by keeping the regions nearby to bony anatomy as an anchor. Moreover, the parotids exhibited a tendency to move inward (right parotid leftward and left parotid rightward) with the largest displacements being in the lateral and inferior regions. The region that moved the least was the medial region (partially adjacent to the bony structure). The study of Barker *et al* (2004) found a median medial shift of 3.1 mm for the COM of the parotid glands. They observed asymmetric shifts in parotid gland surfaces, with average displacements of 1 ± 3 mm and 3 ± 3 mm for the medial and lateral regions of the irradiated glands, respectively.

Table 1. Overview of documented quantitative and qualitative anatomical changes in the parotid glands. The table displays the study, the number of CTs used, the reported volumetric change (absolute, relative or both), the absolute shifts in the COM and its direction and qualitative notes on the reported changes.

Study	CT number	Volumetric loss	COM shift	Morphological alterations and notes
Barker *et al* (2004)	⩾2	Median 190 mm^3 per day Range of 40–840 mm^3 per day	Median 3.1 mm Range 0–9.9 mm in medial direction	Shrinkage correlated with patient weight loss
Vásquez Osorio *et al* (2008)	2	Average 14%	1 or 3 mm	Bony anatomy kept as anchor during shrinkage
Bhide *et al* (2010)	⩾2	14% or 4200 mm^3 between week 0 and 2 16% or 4000 mm^3 between week 2 and 4 35% over the course of chemoradiotherapy	2.3 mm by week 4 in the medial direction	COM shift insignificant in the anteroposterior and inferosuperior directions
dos Santos *et al* (2020)	2	Average 20.5% or 6560 mm^3 between CTs	N.A.	Shape shift from convex to concave COM shift towards the medial and cranial directions

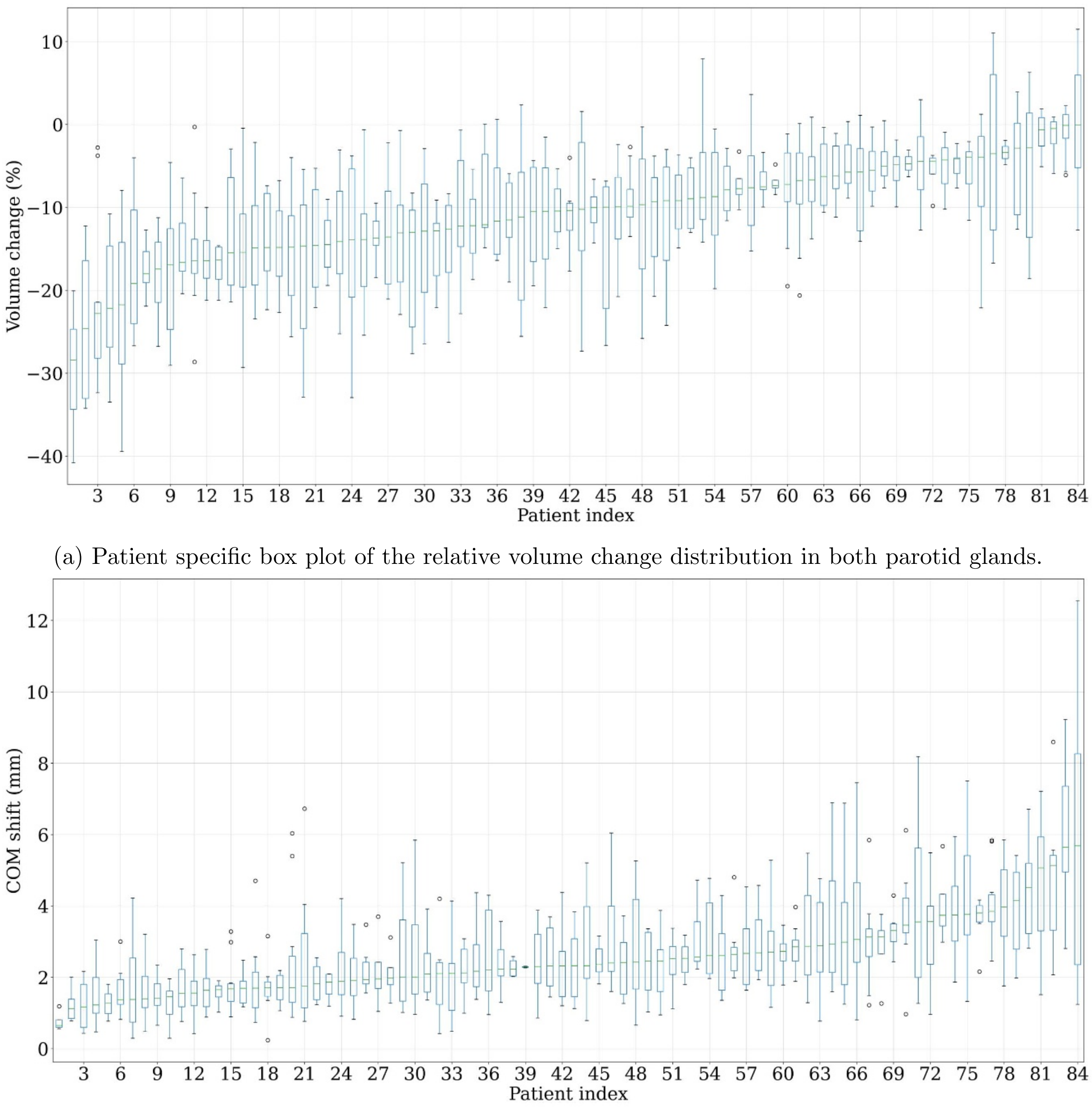
Table 2. Training set statistics. The table displays for both parotid glands the mean, standard deviation, minimum, median and maximum of the volume on the planning and repeat CT images, the difference between these volumes (absolute and relative) and the center of mass shifts.

		Statistic				
Organ	Metric	Mean	SD	Min.	Median	Max.
Parotid L	Planning volume (mm^3)	35 878	11 290	16 984	33 280	83 520
	Repeat volume (mm^3)	31 571	10 161	12 976	29 816	76 632
	Difference (mm^3)	−4307	3880	−30 456	−3548	4256
	Relative difference (%)	−12	9	−41	−11	10
	COM shift (mm)	3	2	0.2	2	13
Parotid R	Planning volume (mm^3)	35 447	12 568	11 344	33 024	87 352
	Repeat volume (mm^3)	31 507	11 160	7496	29 896	79 136
	Difference (mm^3)	−3941	3955	−29 112	−3320	4584
	Relative difference (%)	−11	8	−41	−10	10
	COM shift (mm)	3	2	0.4	3	12

4.3. Training set anatomical changes of parotid glands

The generative performance of the model is tied to the data provided during training in the training set. Therefore, the anatomical changes in the training set and the literature reported changes from table 1 were compared to assess the degree to which the training set is representative of the broader PT H&N patients population. The anatomical changes presented in section 4.2 come from studies in which uni or bilateral photon-based RT or a combination of chemotherapy and RT was delivered. In contrast, the dataset of this work comes exclusively from PT patients treated with mostly bilateral fields. The training set contained anonymized data and was composed of pairs of pCTs and consecutive rCTs (pCT–rCT$_1$, pCT–rCT$_2$, and so on). For each such pair and patient, the volume loss and COM shift in each parotid was computed and averaged over both parotid glands. Figure 4 displays, for each patient in the training set, boxplots of the distributions of percentage parotid glands volume changes and parotids COM shifts.

Figure 4 shows that the median of the volumetric loss in the parotids is ≈11% and the median of the COM shift is ≈3 mm. While the many patients have relatively unskewed volumetric change and COM shifts distributions, there are also patients (e.g. 3, 10, 60 and 72) that display skewed distributions with outliers. To facilitate comparison to previous publications, the data presented in figure 4 is summarized in table 2 where statistics on an individual parotid level are displayed. Specifically, the absolute volumes on the planning and rCTs, their difference (absolute and relative) and the COM shifts are characterized through their mean, SD, minimum, median and maximum.



(a) Patient specific box plot of the relative volume change distribution in both parotid glands.

(b) Patient specific box plot of the COM shift distribution in both parotid glands.

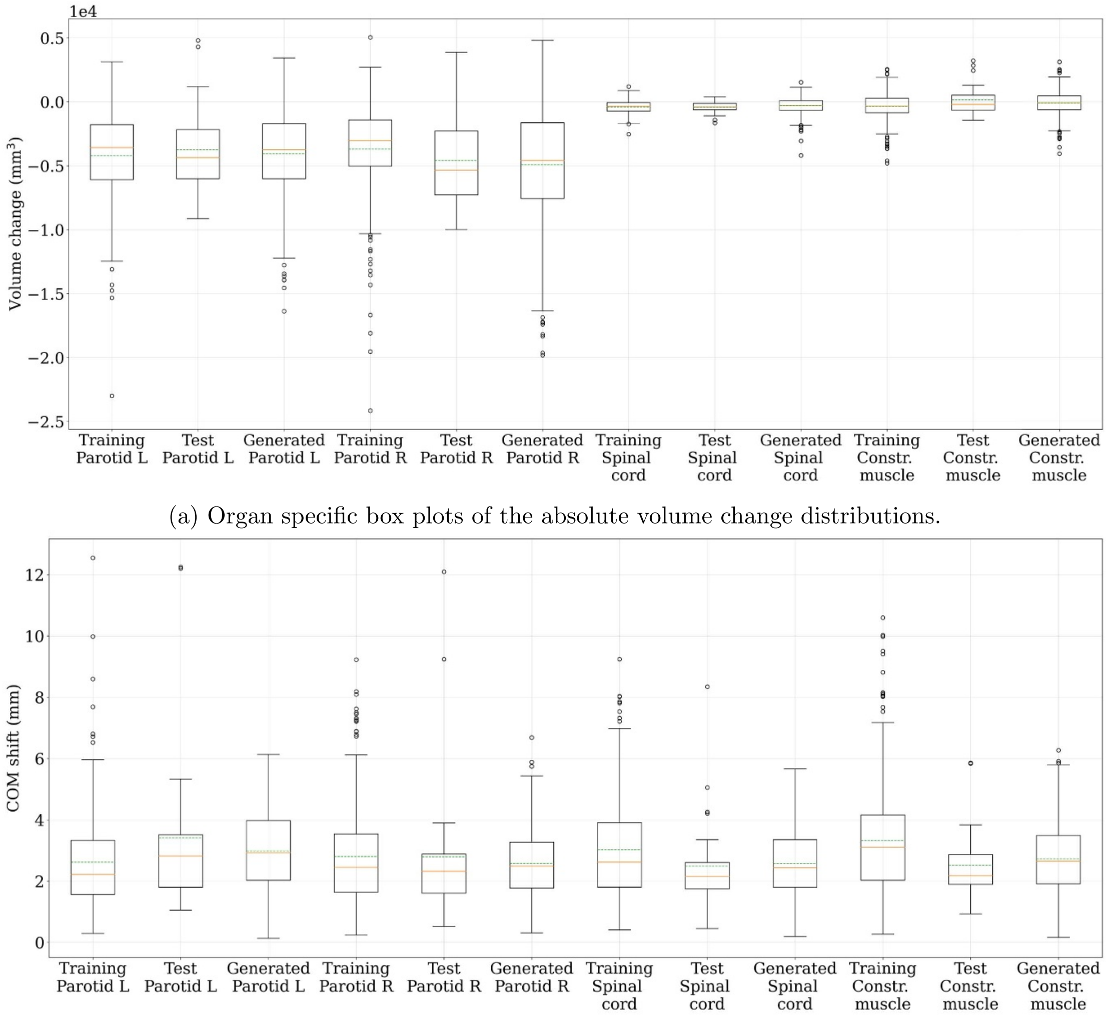
Figure 4. Training set characterization. The figures display median sorted boxplots (with the whiskers extending up to $1.5 \cdot \mathrm{IQR}$, $\mathrm{IQR} = Q_3 - Q_1$) with the *x* axis giving the patient identifying number and the *y* axis giving either the relative volumetric changes or the COM shifts distributions in both the parotid glands.

The absolute volumes of the parotids on the pCT images are a mean of 35 878 mm^3 with a range of 16 984–83 520 mm^3 for the left parotid and a mean of 35 447 mm^3 with a range of 11 344–87 352 mm^3 for the right parotid. Both mean parotid volumes are roughly 23% larger than the volumes reported by dos Santos *et al* (2020), namely 28 477 mm^3 for the left parotid and 29 274 mm^3 for the right parotid.

The differences between the parotid volumes in the training set are, a mean of −4307 mm^3 with a range of −30 456–4256 mm^3 corresponding to a mean of −12% with a range of −41%–10% for the left parotid and a mean of −3941 mm^3 with a range of −29 112–4584 mm^3 corresponding to a mean of −11% with a range of −41%–10% for the right parotid. This is slightly smaller but in line with previous studies, considering the averaging effect caused by the pCT–rCT pairings from the training set.

The COM shifts observed in the dataset are a median of 2 mm with a range of 0.2–13 mm for the left parotid and a median of 3 mm with a range of 0.4–12 mm for the right parotid. These values are in agreement with the median of 3.1 mm in a range of 0–9.9 mm reported by Barker *et al* (2004).

To conclude, the distributions from the training set are deemed in line with the expectations set out by previous studies. Differences between the data presented here and the one from previous studies, such as Medbery *et al* (2000) and dos Santos *et al* (2020) can be attributed to several factors. First, the pCT–rCT composition of the training set is bound to underestimate the changes when compared to studies based on only pCT-final CT pairs. Second, differences are expected due to the anonymization of the training set and the differences between the compared cohorts. Previous studies such as the ones of Ericson (1970), Vásquez



(a) Organ specific box plots of the absolute volume change distributions.

(b) Organ specific box plots of COM shift distributions.

Figure 5. Organ specific generative performance. The figures display boxplots (with the whiskers extending up to $1.5 \cdot \mathrm{IQR}$, $\mathrm{IQR} = Q_3 - Q_1$) for the training, test and generated sets with the *x* axis showing the organ and the *y* axis giving either the absolute volume change or the COM shift distribution for the given organ. The median of the distributions is displayed with the continuous orange line while the mean is displayed with the dotted green line.

Osorio *et al* (2008), dos Santos *et al* (2020)showed differences in parotid volumes depending on age, sex, weight, smoker status, planned doses, degree of parotid sparing and treatment modality, which are impossible to study in our current case. Third, a small effect could be expected due to inter-observer variability and systematic errors introduced by interpolating the original images on a new, coarser grid could also influence the observed absolute volumes.

4.4. Generative performance

To assess the generative performance of the model, the test set, that contained 9 patients, was input into the final trained model and 100 samples were drawn for each record (pair of pCT–rCT) in the test set. Figure 5(a) displays for all present organs (left and right parotids, the spinal cord and the constrictor muscle) boxplots of the volume changes on the training, test and generated sets. Figure 5(b) displays for all present organs boxplots of the COM shifts on the training, test and generated sets. In terms of volumetric change distributions, shown in figure 5(a), the figure shows that the parotid distributions on the training and test set are different. For example, the mean (indicated by the dotted green line) of the left parotid volume change distribution is below its median (indicated by the continuous orange line), while it is above it on the test set. A similar situation occurs for the right parotid. The same figure shows that the model generates volume change ranges that are broad enough to encompass the training and test sets, with means and medians in reasonable agreement (defined as a value within 20% of either the training or test set value) to the training or test set ones. The COM shift distributions, shown in figure 5(b), also display differences between the training and test sets. For example, the distribution of the constrictor muscle COM shifts on the test set has a

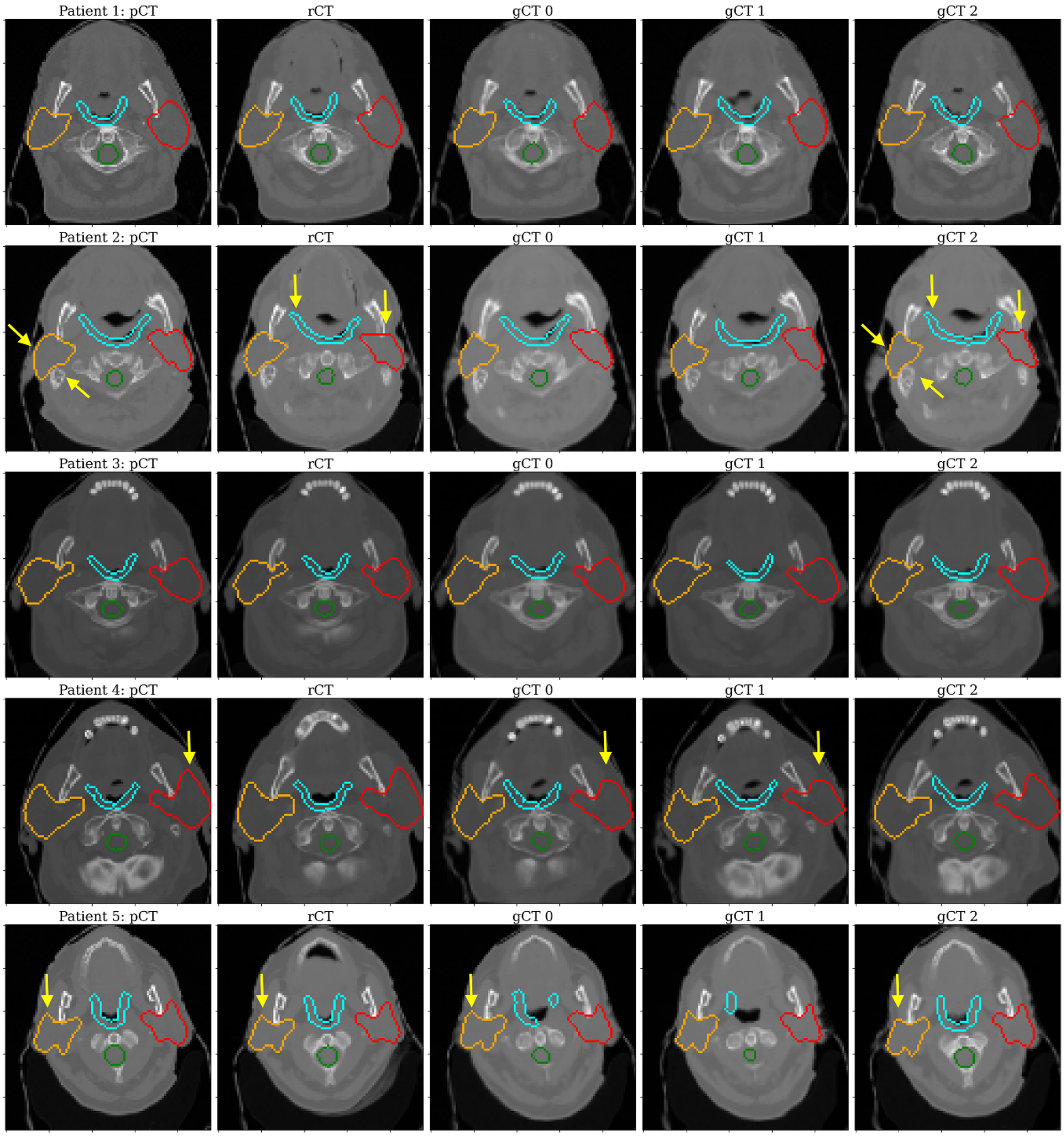
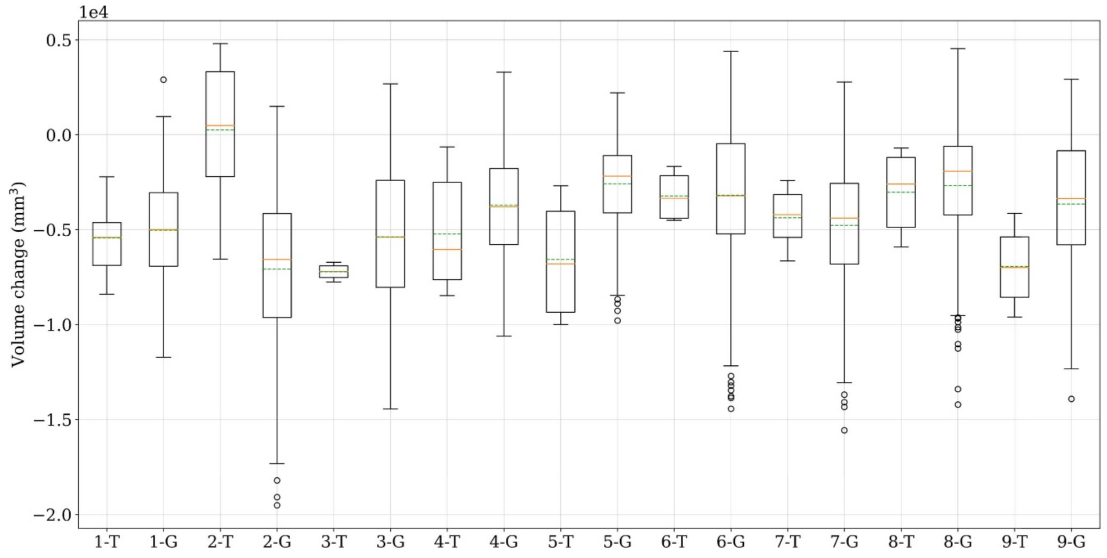


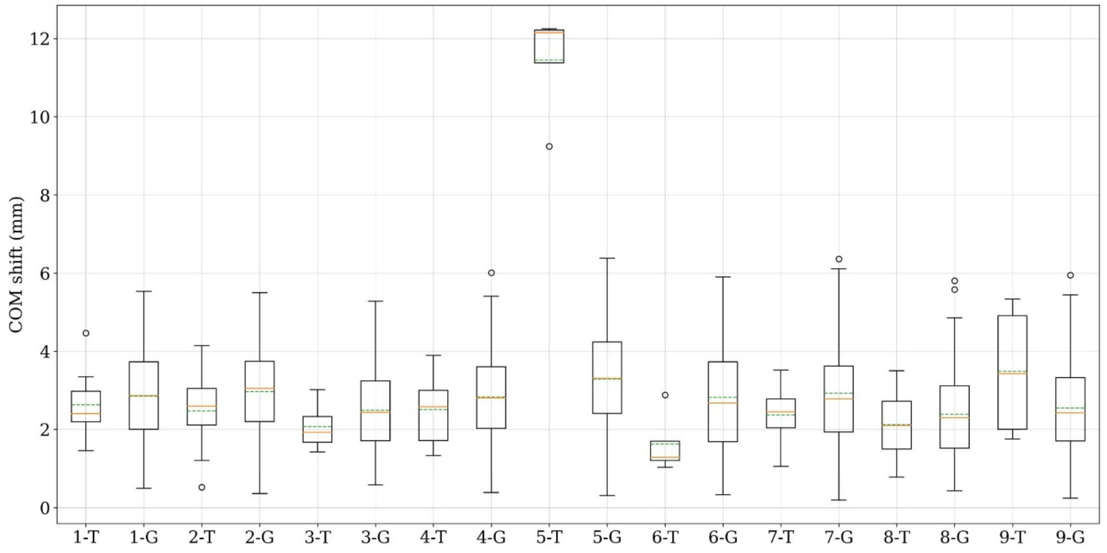
Figure 6. Example of generated images. The figure displays, for 5 randomly selected patients from the test set, in the first column the true pCT, in the second column one of the true rCTs and in the remaining columns generated CT images. Overlaid on all images are the left parotid (red), the right parotid (orange), the spinal cord (green) and the constrictor muscle (blue). Noteworthy anatomical changes are indicated with yellow arrows.

considerably smaller range of values, with smaller mean and median values. As was the case for figure 5(a), figure 5(b) also shows that the model predicts distributions of COM shifts that are broad enough to encompass the test set ones, with means and medians in reasonable agreement. Some discrepancies can also be observed, for example in the difference between the median of the distribution of COM shifts of the constrictor muscle on the test and generated sets. Given the overall good agreement presented by both figures 5(a) and (b), it can be concluded that $\mathrm{DAM}_{\mathrm{HN}}$ is capable of modelling volume and COM shift distributions present in the training and test set.

An illustration of the generative capabilities of the model is shown in figure 6. The figure displays for 5 patients in the test set, in the first column the pCT, in the second column one of the rCTs and in the following 3 columns three patient specific generated CT images with corresponding contours (the left parotid colored in red and the right parotid colored in orange, the spinal cord in green and the constrictor muscle in blue). As already mentioned in table 1, the flattening and medial movement of the parotids is expected. This feature is illustrated for patient 4 through the yellow arrows in the planning and generated images shown in columns 3–5. Patient 2 displays shrinking in the right parotid (in orange) and flattening of the left parotid (in red) as illustrated by the yellow arrows. The model also appears to predict neck pose shifts, as illustrated by the changing air gap in the oral cavity of patient 1 in the second generated image or by the change in the shown



(a) Patient specific box plot of absolute volume change distributions in both parotid glands.



(b) Patient specific box plot of COM shift distributions in both parotid glands.

Figure 7. Patient specific generative performance. The figures display boxplots (with the whiskers extending up to $1.5 \cdot \mathrm{IQR}$, $\mathrm{IQR} = Q_3 - Q_1$) with the *x* axis showing the organ and the *y* axis giving either the absolute volumetric changes or the COM shifts for the given organ. The median of the distributions is plotted using the continuous orange line while the mean is plotted using the dotted green line.

dentition of patient 5 in the generated image 1. Weight loss, which is usually observed in RT patients, is prominent in the comparison between the pCT and the generated images for patient 4. Minimal overlap between the parotid glands and the mandible bone is visible for patient 1 on the pCT and the rCT. The generated images also display this feature, which illustrates the anatomical coherence of the generated anatomies. While it is difficult to definitively assert the feasibility of the generated image, the figure supports the conclusion that the model is capable of generating realistic anatomies that are coherent and involve posture shifts, shifting air gaps, weight loss and the typical expected anatomical changes in the parotid glands.

To further test the population based model, figure 7 shows patient-specific boxplots of the anatomical changes in the parotid glands. Figure 7(a) displays for each patient in the test set, the true volumetric change (denoted by the patient number and -T) and the generated volumetric changes by drawing 100 samples (denoted by the patient number and -G). In terms of the volume change distributions illustrated in figure 7(a), the model largely predicts broad enough distributions that encompass the true ones. This is the case for patients 1, 3, 4, 6, 7, 8 and 9. Moreover, the means and medians are in reasonable agreement for patients 1, 6, 7 and 8. Discrepancies in the means and medians can be observed for patients 2, 4 and 9. In terms of COM shift distributions, the model produces distributions with large enough ranges to encompass the test set ones, except for patient 5. The means and medians are in reasonable agreement for most patients, with the exception of 5, 6 and 9. The discepancies on a per-patient level could be explained by an insufficient number of recorded rCTs for those patients but also by the non-patient specific nature of the model. While the model attempts to provide patient specificity by allowing the parameters of the prior distribution to depend on the pCT and associated masks, the model optimizes the log likelihood of the full dataset, therefore resulting in a sample (or population) based model.

4.5. Comparison to DiffuseRT

The generative performance of DAM with respect to PCA based models has already been documented in the previous work of Pastor-Serrano *et al* (2023), where it was shown to outperform them. Thus, the generative performance of this model was compared with the recently published DDPM of Smolders *et al* (2024). DDPM is also a generative deep learning model that approximates a data distribution, by inverting a gradual multi-step noise addition process. Similarly to the results shown by DDPM, figure 8 displays for all organs, the true (training set) and generated volume change distributions (in figures 8(a), (b), (e) and (f)) and COM shift distribution (in figures 8(c), (d), (g) and (h)) together with a kernel density estimate for each. The kernel density estimate was computed using the Scikit library (Pedregosa *et al* 2011) and a kernel bandwidth defined as one tenth of the range of values in the distribution. Both volume change and COM shift distributions that the $\mathrm{DAM_{HN}}$ training set exhibits are qualitatively different than the ones reported by DDPM, displaying less bimodality. This difference is likely attributable to the differences in the patient cohort and the specifics of treatment delivery (e.g. the chosen number and direction of beams). The kernel density estimates for the training and generated sets are generally in agreement, with disagreement occurring at the ends of the distributions, as is visible in figures 8(g) and (h).

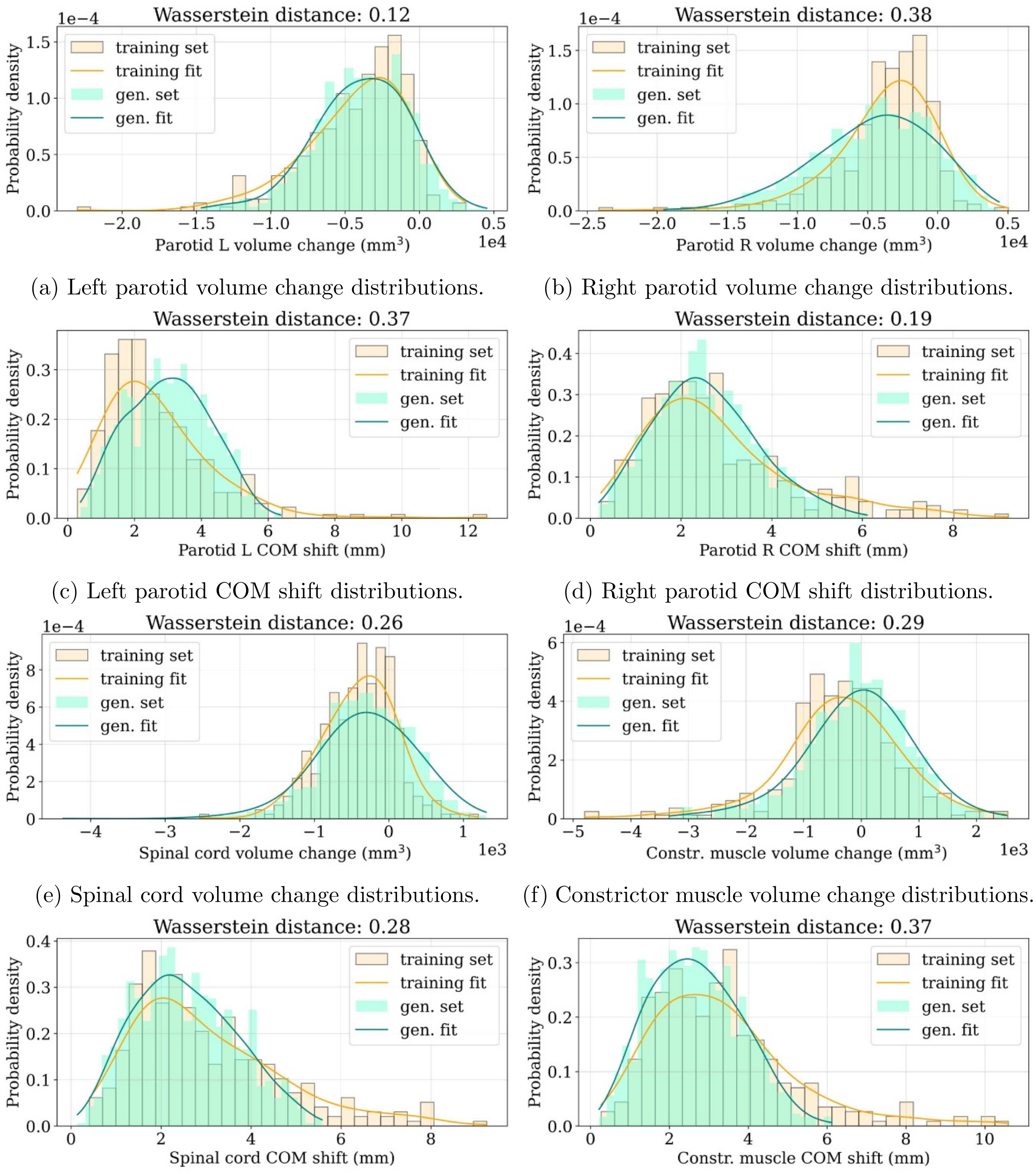
$\mathrm{DAM_{HN}}$ and DDPM were also compared in terms of the Wasserstein distance (WD) between the true (training set) and generated anatomical changes distributions. The WD is a metric for probability distribution similarity, with a value of zero occurring when the distributions are the same and larger values indicating more different distributions. To compute it, the volume changes and COM shifts in the organs for both training and generated sets were normalized by the mean and SD of the true (training set) values (to counter the scaling effect of the WD based on the range of the data) and thereafter input into the SciPy implementation (Virtanen *et al* 2020). Table 3 shows the comparison between DDPM and $\mathrm{DAM_{HN}}$. The qualitative agreement observed in figure 8 is also illustrated by the low WDs achieved by $\mathrm{DAM_{HN}}$, which is comparable to the ones obtained by DDPM for all metrics.

4.6. Latent space analysis

Given that $\mathrm{DAM_{HN}}$ encodes the information between the planning and rCTs into the latent space, the effect of varying individual latent variables while keeping the others fixed on organ volume changes and COM shifts was investigated. Figure 9 illustrates the volume changes for each organ (left parotid, right parotid, spinal cord and the constrictor muscle) that occur when an individual latent variable is varied from -5σ to 5σ, while the others are kept fixed to 0. Similarly, figure 10 displays the effect of varying individual latent variables on the COM shift.

Figure 9 shows consistently larger volumetric lossess in the parotid glands in comparison to spinal cord and constrictor muscle. This is expected, given that the spinal cord is smaller in volume than the parotid glands and is usually avoided during irradiation. Figure 9 also shows the relatively smooth latent space that the model learns and that the parotid glands volume changes are comparable, indicated the largely bilateral nature of the patient cohort. Variables that induce larger volumetric losses in one of the two parotids, could point to the presence of patients with unilateral fields, as non-irradiated parotids were shown to shrink less during treatment than radiated ones (Vásquez Osorio *et al* 2008). Figure 10 shows that for both parotids, the COM deformations are roughly similar in absolute value. This is in line with the expectation, set by the work of Vásquez Osorio *et al* (2008), that both parotids move in the medial direction with similar amplitudes. Moreover, figure 10 also shows that the learned latent space is smooth.

Volume and COM shifts are just one measure of latent space variations. Figure 11 shows, for a patient in the test set, a cut of the images produced when latent variables with numbers 1, 7, 20, 21 and 32 are varied. The particular latent variables were chosen due to the large changes they induce, as visible in figures 9 and 10. The first column of figure 11 displays the pCT, while the remaining columns display the image, the associated contours (as before the left parotid in red, the right parotid in orange, the spinal cord in green and the constrictor muscle in blue) and the overlaid deformation vector field that is created by the individual latent variables (with the value it was set to given in the title of the figure). As was already visible in figures 9 and 10, latent variable 7 induces large changes in the right parotid for extreme values of the latent variable. This effect is also observed through the deformation vector field around this structure. Latent variable 21 displays a similar behavior, for both the left and right parotids. Figures 9 and 10 also show that latent variable 1 and 32 generate deformation fields in the oral cavity, perhaps pointing to shifting patient poses. A limitation of



(a) Left parotid volume change distributions.

(b) Right parotid volume change distributions.

(c) Left parotid COM shift distributions.

(d) Right parotid COM shift distributions.

(e) Spinal cord volume change distributions.

(f) Constrictor muscle volume change distributions.

(g) Spinal cord COM shift distributions.

(h) Constrictor muscle COM shift distributions

Figure 8. Comparison of true and generated anatomical change distributions. The figures display for all organs (left and right parotid, constrictor muscle and spinal cord) the true and generated anatomical change (volume change and COM shift) distributions, their corresponding kernel density estimates and the WD between the true and generated distributions in the title.

Table 3. Wasserstein distance comparison between the best performing DDPM model of Smolders *et al* (2024) and DAM$_{\mathrm{HN}}$. The table displays the Wasserstein distance between the true (training set) and generated volume loss and COM shift distributions in the left and right parotids.

		Model	
Metric	Structure	DDPM	DAM$_{\mathrm{HN}}$
Δ Volume	Left parotid	0.60	0.12
	Right parotid	0.28	0.38
COM shift	Left parotid	0.31	0.37
	Right parotid	0.22	0.19

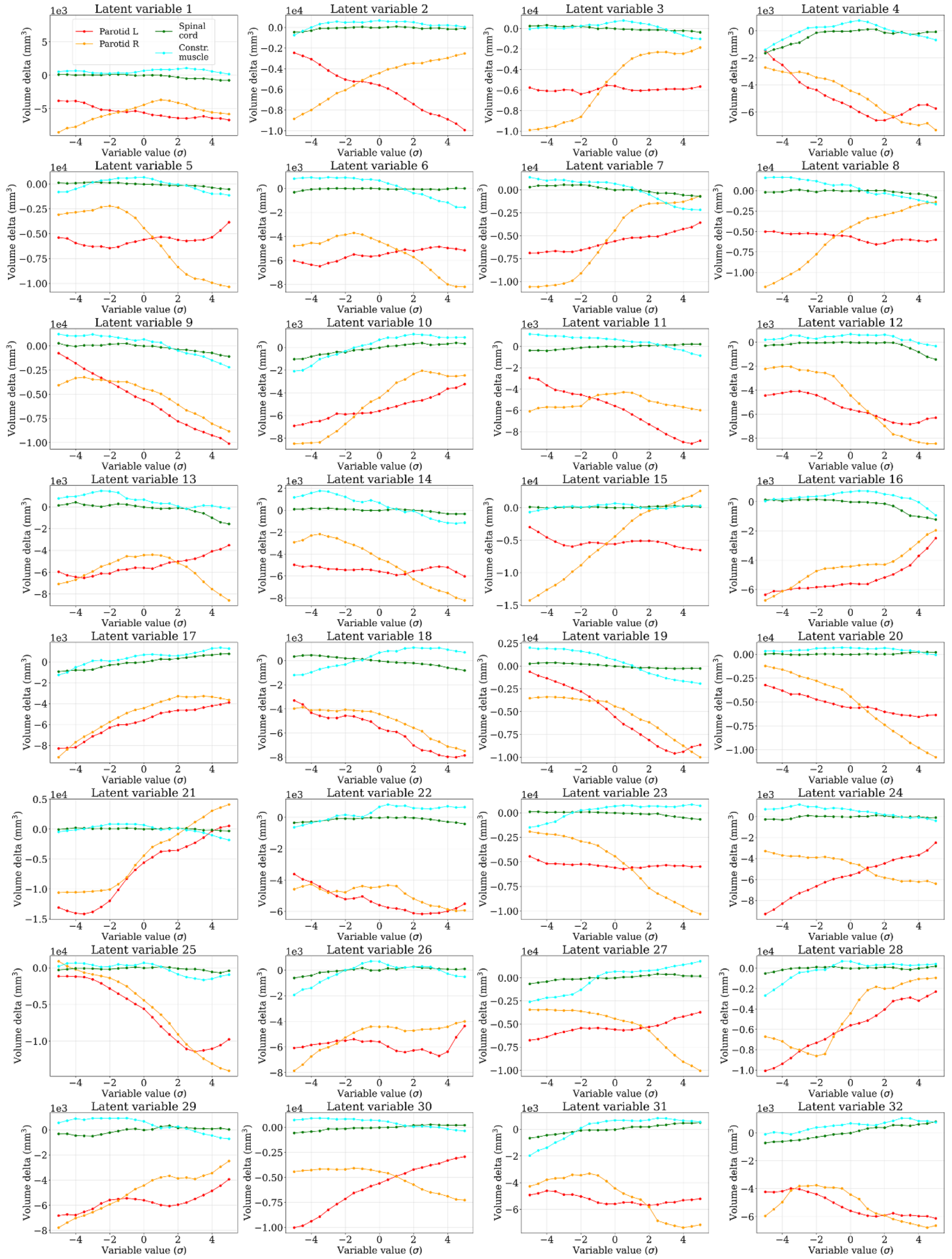


Figure 9. Latent space variations. The figure displays the organ (left parotid in red, right parotid in orange, spinal cord in green and constrictor muscle in blue) volume change that individual latent variables cause. The latent variables were varied from -5σ to 5σ while the remaining variables were set to 0.

the framework, is that the latent variables are not encouraged to generate non-correlated deformations and therefore, it is difficult to relate specific latent variables to specific induced anatomical changes.

5. Conclusion

This work presented a probabilistic deep learning model for generating future anatomical changes in H&N RT patients. The model was trained on a training set coming from 83 PT H&N patients and was assessed on test set coming from 9 patients. On the test set the model achieved a DICE score of 0.83 and an NCC score of

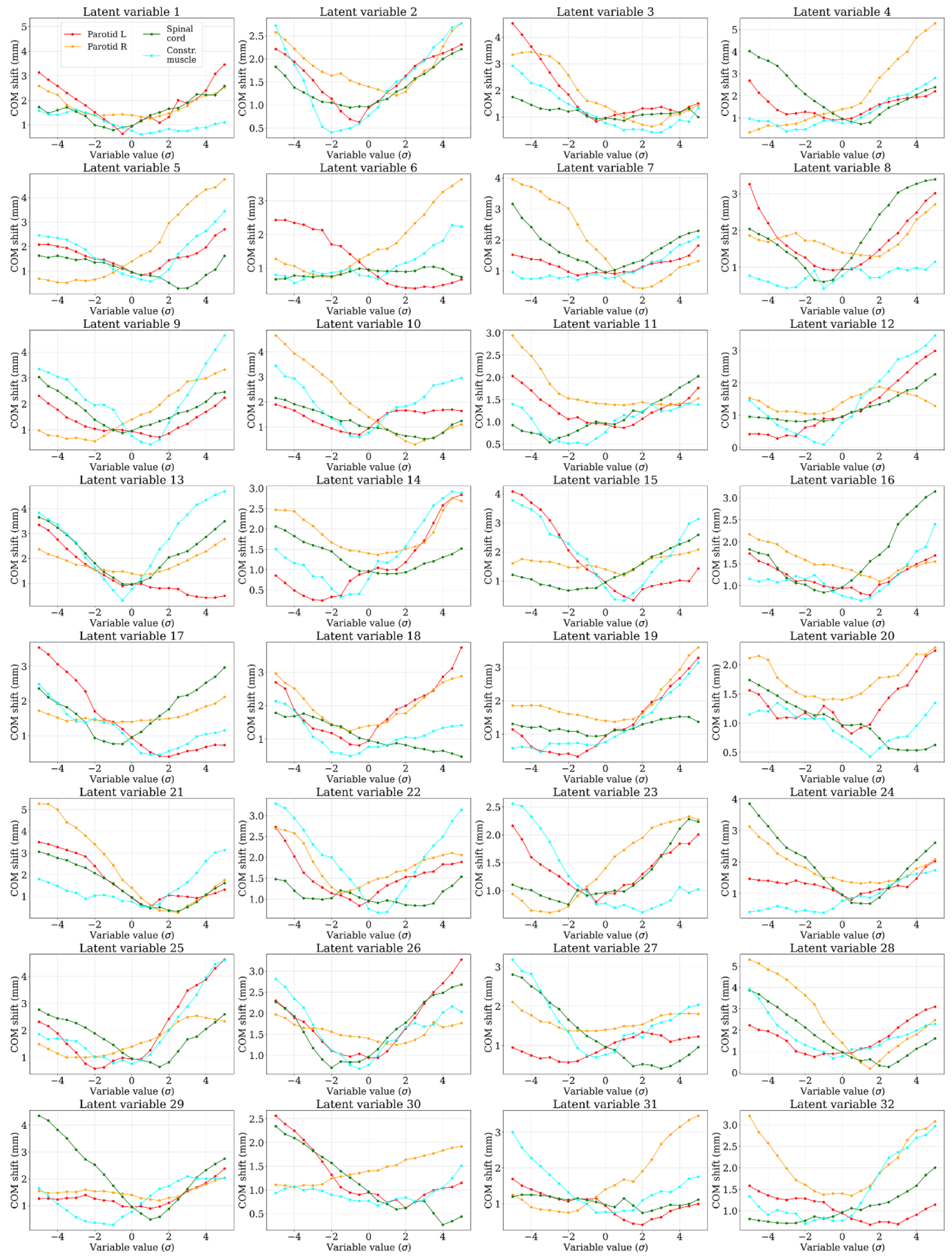


Figure 10. Latent space variations. The figure displays the organ (left parotid in red, right parotid in orange, spinal cord in green and constrictor muscle in blue) COM shift that individual latent variables cause. The latent variables were varied from -5σ to 5σ while the remaining variables were set to 0.

0.60 using 32 latent variables. The model produces volumetric changes and COM shift distributions that are broad enough to capture the real, observed ones, with the predicted means being close to the real ones. DAM_{HN} was compared to the state of the art DDPM for H&N anatomical changes presented by Smolders *et al* (2024). For both parotid glands, DAM_{HN} achieved similar WDs to the ones obtained by the DDPM model between the true and generated volume loss distributions (0.12 versus 0.60 and 0.38 versus 0.28) and between the COM shift distributions (0.37 versus 0.31 and 0.19 versus 0.22). The latent space analysis showed that the model learns a smooth latent space, that displays some correlation between the latent variables (which was not discouraged in the model framework). Although this work focused on data coming

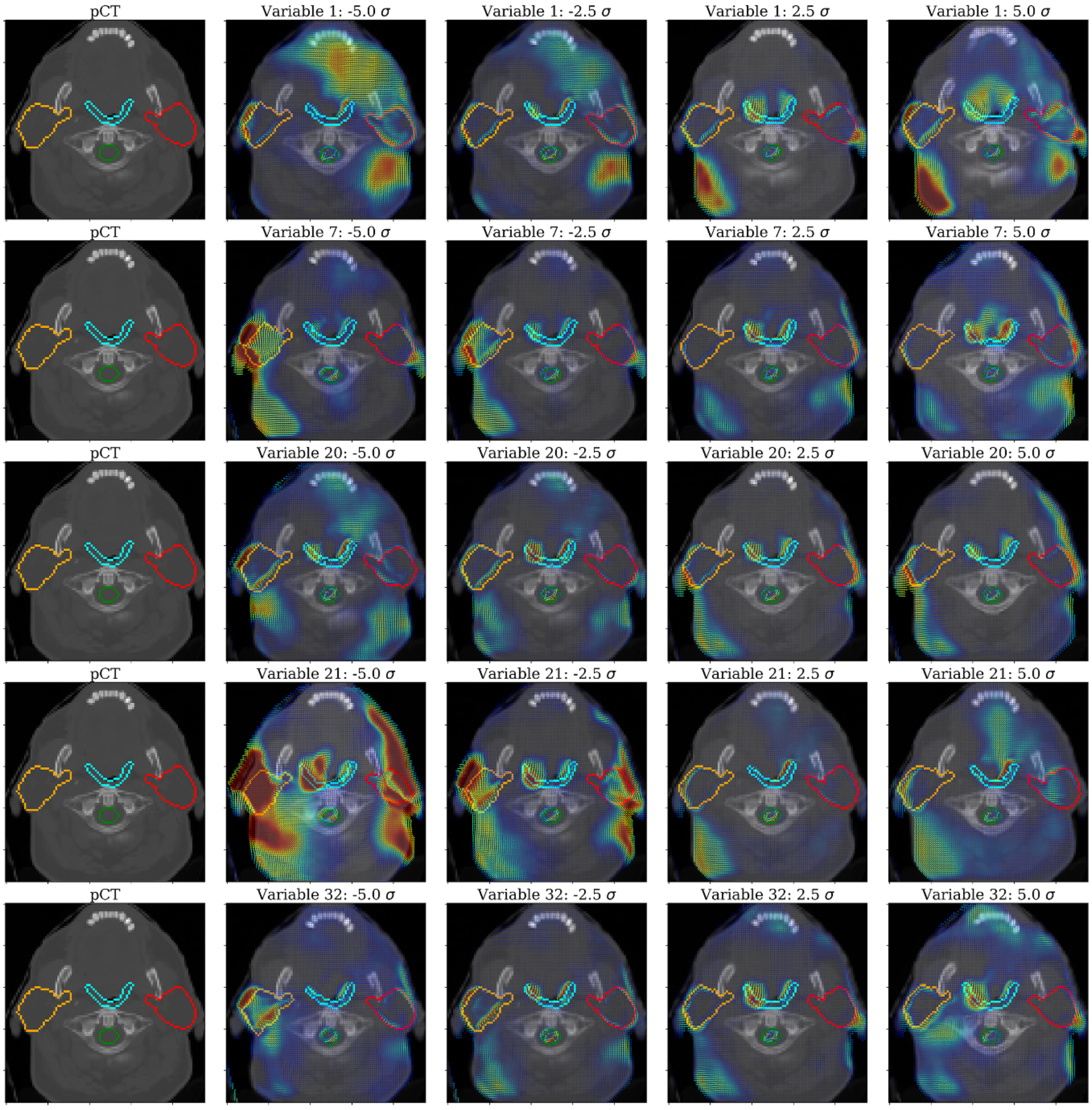


Figure 11. Latent space visualization. The figures display in the first column for a given patient, the pCT and associated organs (left parotid in red, right parotid in orange, spinal cord in green and constrictor muscle in blue). In the following columns the figure displays in the title the chosen latent variable number (and the value it was set to), the generated image and its organs. Overlaid is plotted the deformation vector field that the model learns, where the color represents the magnitude of the field.

from a PT patient cohort, the methodology is valid for a wider range of problems in adaptive RT, including adaptive photon RT.

There are several limitations of the current methodology and model framework and points for future improvement and studies. First, the dataset contained a different number of rCTs for each patient and is therefore biased towards patients with larger anatomical changes (as those patients are more likely to be re-imaged). This bias was not accounted for in this model and likely leads to the model predicting larger than observed anatomical changes for patients with small ones. However, given that a dataset with larger anatomical changes is more difficult to encode in the latent space, a dataset that contains rCTs from patients with small anatomical changes should not significantly decrease the overall population accuracy of $\mathrm{DAM}_{\mathrm{HN}}$. Second, a limitation of the model is that, despite allowing the parameters of the prior distribution to vary on an individual patient level, the model is intrinsically a population based one, as it optimizes the log likelihood of observing the full dataset. This, coupled to a limited number of repeat CTs in the dataset, leads to degraded accuracy for some patients. Third, if the large number of structures present in the H&N area would be included in the dataset, it is expected that the model would require a change in architecture (specifically an increase in the number and size of layers and latent space dimensionality) to correctly model those datasets. More generally, the necessary minimal architecture and the optimal weights of the different loss terms should be further investigated. Moreover, the inclusion of the regularization term that penalizes large gradients in the deformation could be detrimental for anatomical regions where such changes do occur (e.g. tongue position). Fourth, the comparison between $\mathrm{DAM_{HN}}$ and the model of Smolders *et al* (2024), is ultimately difficult due to the different datasets that the models were trained and evaluated on. Thus, both models should be trained and evaluated on the same sets that contain more structures than they presently do (e.g. additional useful structures could be the submandibular glands and the oral cavity). Fifth, the structure of the dataset could be changed from pCT–$\mathrm{rCT_1}$, pCT–$\mathrm{rCT_2}$, and so on to pCT–$\mathrm{rCT_1}$, $\mathrm{rCT_1}$-$\mathrm{rCT_2}$ and so on. In doing so, a model that predicts changes on the time scales on which patients are re-imaged (weekly or daily depending on the workflow) could be obtained. Such a model would be applicable to an adaptive PL approach or for plan quality assurance. Moreover, a time variable could be included in the architecture to encode information in addition to the rCTs. Sixth, as anatomy change predictions has applications in dose change predictions, an analysis of the effect on dose characteristics (including a robustness analysis) of delivering treatment plans to the generated images is a natural next step for this model. Next to such a study, work on additional standards (beyond volume changes and COM shifts) for assessing the degree to which generated CT images are realistic should be established.

Overall, $\mathrm{DAM_{HN}}$ was capable of quickly generating hundreds of realistic images of inter-fractional anatomies. As already mentioned, such a model has a number of applications in the RT workflow, such as improving robust optimization, as a component in plan quality assurance in OAPT or in expanding the PL approach.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

The authors wish to acknowledge that the manuscript is partly funded by Varian, a Siemens Healthineers Company.

Conflict of 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.

Credit statement

Tiberiu Burlacu: Conceptualization, methodology, software, validation, formal analysis, data curation, investigation, writing—original draft, writing—review & editing, visualization.
Zoltán Perkó: Conceptualization, methodology, validation, resources, writing—review & editing, supervision, project administration, funding acquisition.
Danny Lathouwers: resources, writing—review & editing, supervision.
Mischa Hoogeman: resources, writing—review & editing.

ORCID iDs

Tiberiu Burlacu https://orcid.org/0000-0001-5542-6971
Mischa Hoogeman https://orcid.org/0000-0002-4264-9903
Danny Lathouwers https://orcid.org/0000-0003-3810-1926
Zoltán Perkó https://orcid.org/0000-0002-0975-4226

References

Barker J L *et al* 2004 Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system *Int. J. Radiat. Oncol. Biol. Phys.* **59** 960–70

Beare R, Lowekamp B and Yaniv Z 2018 Image segmentation, registration and characterization in R with SimpleITK *J. Stat. Softw.* **86** 1–35

Bhide S A, Davies M, Burke K, McNair H A, Hansen V, Barbachano Y, El-Hariry I A, Newbold K, Harrington K J and Nutting C M 2010 Weekly volume and dosimetric changes during chemoradiotherapy with intensity-modulated radiation therapy for head and neck cancer: a prospective observational study *Int. J. Radiat. Oncol. Biol. Phys.* **76** 1360–8

Chen Z, Dominello M M, Joiner M C and Burmeister J W 2023 Proton versus photon radiation therapy: a clinical review *Front. Oncol.* **13** 1133909

Cubillos-Mesías M, Troost E G C, Lohaus F, Agolli L, Rehm M, Richter C and Stützer K 2019 Including anatomical variations in robust optimization for head and neck proton therapy can reduce the need of adaptation *Radiother. Oncol.* **131** 127–34

Deiter N, Chu F, Lenards N, Hunzeker A, Lang K and Mundy D 2020 Evaluation of replanning in intensity-modulated proton therapy for oropharyngeal cancer: factors influencing plan robustness *Med. Dosim.* **45** 384–92

dos Santos W P, Perez Gomes J P, Nussi A D, Altemani J M, Botti Rodrigues dos Santos M T, Hasseus B, Giglio D, Braz-Silva P H and Ferreira Costa A L 2020 Morphology, volume and density characteristics of the parotid glands before and after chemoradiation therapy in patients with head and neck tumors *Int. J. Dent.* **2020** 8176260

Ericson S 1970 The normal variation of the parotid size *Acta Oto-Laryngol.* **70** 294–300

Fiorentino A, Caivano R, Metallo V, Chiumento C, Cozzolino M, Califano G, Clemente S, Pedicini P and Fusco V 2012 Parotid gland volumetric changes during intensity-modulated radiotherapy in head and neck cancer *Br. J. Radiol.* **85** 1415–9

Ghojogh B, Ghodsi A, Karray F and Crowley M 2022 Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: tutorial and survey (arXiv:2101.00734)

Jaderberg M, Simonyan K, Zisserman A and Kavukcuoglu K 2016 Spatial transformer networks (arXiv:1506.02025)

Kingma D P and Welling M 2019 An introduction to variational autoencoders *Found. Trends® Mach. Learn.* **12** 307–92

Krebs J, Delingette H, Mailhé B, Ayache N and Mansi T 2019 Learning a probabilistic model for diffeomorphic registration *IEEE Trans. Med. Imaging* **38** 2165–76

Liu W, Frank S J, Li X, Li Y, Park P C, Dong L, Ronald Zhu X and Mohan R 2013 Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers *Med. Phys.* **40** 051711

Medbery R, Yousem D M, Needham M F and Kligerman M M 2000 Variation in parotid gland size, configuration and anatomic relations *Radiother. Oncol.* **54** 87–89

Oud M, Breedveld S, Giżyńska M, Kroesen M, Hutschemaekers S, Habraken S, Petit S, Perkó Z, Heijmen B and Hoogeman M 2022 An online adaptive plan library approach for intensity modulated proton therapy for head and neck cancer *Radiother. Oncol.* **176** 68–75

Oud M, Breedveld S, Rojo-Santiago J, Giżyńska M K, Kroesen M, Habraken S, Perkó Z, Heijmen B and Hoogeman M 2024 A fast and robust constraint-based online re-optimization approach for automated online adaptive intensity modulated proton therapy in head and neck cancer *Phys. Med. Biol.* **69** 075007

Pastor-Serrano O, Habraken S, Hoogeman M, Lathouwers D, Schaart D, Nomura Y, Xing L and Perkó Z 2023 A probabilistic deep learning model of inter-fraction anatomical variations in radiotherapy *Phys. Med. Biol.* **68** 085018

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L and Lerer A 2017 Automatic differentiation in PyTorch *31st Conf. on Neural Information Processing Systems (NIPS 2017)* (*Long Beach, CA, USA*) https://openreview.net/forum?id=BJJsrmfCZ

Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30

Smolders A J, Rivetti L, Vatterodt N, Korreman S S, Lomax A J, Sharma M, Studen A, Weber D C, Jeraj R and Albertini F 2024 DiffuseRT: predicting likely anatomical deformations of patients undergoing radiotherapy *Phys. Med. Biol.* **69** 155016

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

van de Schoot A J A J, de Boer P, Crama K F, Visser J, Stalpers L J A, Rasch C R N and Bel A 2016 Dosimetric advantages of proton therapy compared with photon therapy using an adaptive strategy in cervical cancer *Acta Oncol.* **55** 892–9

Van de Water S, Albertini F, Weber D C, Heijmen B J M, Hoogeman M S and Lomax A J 2018 Anatomical robust optimization to account for nasal cavity filling variation during intensity-modulated proton therapy: a comparison with conventional and adaptive planning strategies *Phys. Med. Biol.* **63** 025020

van Kranen S, van Beek S, Rasch C, van Herk M and Sonke J-J 2009 Setup uncertainties of anatomical sub-regions in head-and-neck cancer patients after offline CBCT guidance *Int. J. Radiat. Oncol. Biol. Phys.* **73** 1566–73

Vásquez Osorio E M, Hoogeman M S, Al-Mamgani A, Teguh D N, Levendag P C and Heijmen B J M 2008 Local anatomic changes in parotid and submandibular glands during radiotherapy for oropharynx cancer and correlation with dose, studied in detail with nonrigid registration *Int. J. Radiat. Oncol. Biol. Phys.* **70** 875–82

Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72