Learning Respiratory Dynamics in Fast Helical Free Breathing CT Imaging for Radiotherapy Planning

Rishi Upadhyay[1], Pascal Paysan[2], Supratik Bose[2], William Delery[1], Yunzheng Zhu[1], William Hsu[1], Daniel Low[1], Stefan Scheib[2], Achuta Kadambi[1], and Ricky R. Savjani[1]

[1] University of California, Los Angeles
[2] Varian Medical Systems

Abstract. Modeling the respiratory dynamics of breathing is crucial when using lung CT scans for radiotherapy. Especially in cases such as tumors in the lung, breathing motion can greatly affect targeted therapies, reducing their effectiveness and inducing negative side-effects. Recent work has introduced 5DCT data which pairs these lung CT scans with breathing traces which record breathing amplitude over time. In this work, we leverage this data and introduce deep learning based approaches to model human respiratory dynamics and warp lung volumes from arbitrary breathing traces. We explore two architectures, Variation Auto-Encoders (VAEs) and Latent Diffusion Models (LDMs), and show promising results, achieving 60.4% and 40.7% improvement respectively in masked RMSE over un-warped scans.

Keywords: Lung Radiotherapy · 5DCT · Respiratory Dynamics · Deep Generative Models

1 Introduction

When performing radiation therapy, accurately locating and targeting regions of interest is extremely important. To avoid irradiating healthy tissue, clinicians typically first determine the clinical target volume (CTV) and then add a margin for uncertainty and error, arriving at the planning target volume (PTV) which can be used to determine the optimal arrangements and shapes of radiation beams [3]. Despite these efforts, it is still common for healthy tissue to be irradiated due to body movement, variation in scans, or other factors [22]. This problem is particularly acute when dealing with tumors in the lungs, as regular breathing movements can move lung regions and these tumors significantly.

To mitigate these errors, a variety of techniques have been introduced [11]. One popular technique is respiratory gating, in which radiation is only applied during certain portions of a patient's breathing cycles [23]. These techniques rely on devices that measure the breathing cycles and then adaptively continue or pause radiation. This greatly reduces the amount of radiation error but also

increases the treatment time since radiation is not continuously applied [11]. Another popular technique are breath-hold methods which can either instruct the patient verbally to hold at a deep inhale [6] or can use specialized machines which artificially hold the breathing level at a pre-set level [28]. These techniques can significantly improve targeting accuracy, but rely on specialized instruction or equipment which can greatly limit their spread and usefulness. In this work, we focus on motion-based techniques which record breathing based motion and attempt to compensate for it. In specific, we build on top of 4DCT techniques which record breathing levels across free breathing scans and then use computational techniques to sort individual slices into full scans at specific breathing levels ("full exhale", "full inhale") or according to breathing traces [5]. 5DCT is a new paradigm which was first introduced in [16] to avoid sorting artifacts which can arise with 4DCT data. The term 5DCT comes from the fact that the data has five degrees of freedom: *x,y,z* dimensions of the CT scan, breathing amplitude, and breathing rate. 5DCT data has started to expand to real clinical settings, with a particularly relevant paper being [24] which leverages this data to learn a linear breathing motion model for patients. They first use a pre-trained registration network to learn deformations between various scans of the same patient. They then use these registrations to fit a breathing model of the following form:

$$\overrightarrow{X} = \overrightarrow{X_0} + \overrightarrow{\alpha} * A + \overrightarrow{\beta} * \dot{A}$$

where $\overrightarrow{X_0}$ is the initial location, A is breathing amplitude and $\dot{A}$ is the derivative of breathing amplitude over time. This model is fit individually for each voxel, allowing for voxel specific movement. Once this model is fit, it can be applied to warp any desired volume to any desired breathing amplitudes and rates. This model shows strong performance, but has long runtimes due to the need for registering all scans (25 in their dataset) and fitting the model per patient. Our techniques instead replace these steps by leveraging deep learning to directly learn a general model which can be applied across patients.

Another relevant line of work is that of probabilistic diffeomorphic registration [2,4,14]. These techniques learn probabilistic models that take in two images and predict a displacement that maps from the moving image to the fixed image. In order to ensure the resulting transform is diffeomorphic (topology preserving), these models learn to predict a velocity field which is then integrated to obtain the displacement map. We build on top of these but instead learn a generative transform: our models do not have access to the fixed image at test time.

1.1 Problem Statement

In this paper, we are interested in answering the question: *What would a moving scan M look like if the patient had been breathing according to a new breathing trace b_{new}? b_{new}* can be any number of traces, from traces collected in scanning sessions and evaluated offline to live traces collected at real time during treatment. More specifically, we can write our problem as follows: Given an input

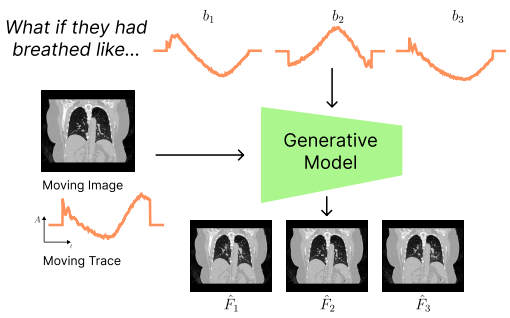

Fig. 1. Problem Overview: Given a CT scan and its trace *(left)*, we are interested in what this scan would looked like under a different trace. Our generative model takes in a new trace *(top row)* and outputs a predicted fixed image *(bottom row)*.

moving scan $M \in R^{512\times512\times448}$, the moving breathing trace $b_m \in R^{448}$, and the desired fixed breathing trace $b_f \in R^{448}$, we train our model $\mathcal{F}$ to output a dense stationary velocity field (DVF) $v \in R^{3\times512\times512\times448}$. This is then converted into a dense deformation field (DDF) $\phi : R^3 \rightarrow R^3$ using the *scaling and squaring* layers introduced in [4]. Once we have estimated ϕ, we can apply this to our moving image to obtain a predicted fixed image $\tilde{F} = \phi \circ M$, effectively imagining what M would look like under a new trace. A visual overview of our problem statement is shown in Fig. 1. To tackle this problem, we introduce two novel techniques based on different deep learning architectures: variational auto-encoders and latent diffusion models. We show promising results with both these architectures, achieving a 60.4% and 40.7% reduction in RMSE in the lung region over un-warped scans.

2 Variational Auto-Encoder Approach

2.1 Background and Formulation

Variational Auto-Encoders (VAEs) are a key development in generative modeling that combined deep learning and latent variable models [13]. They generate new samples by relying on an intermediate latent space on which they impose certain conditions, allowing for efficient inference. In particular, VAEs are generally composed of two networks: an encoder, which models the distribution $q_\phi(z|x)$, and a decoder, which models $p_\theta(x|z)$. The encoder takes in an input sample x and maps it to a pair of mean $\mu_{z|x}$ and standard deviation $\sigma_{z|x}$ from which a latent vector $z \sim \mathcal{N}(\mu_{z|x}, \sigma_{z|x})$ is sampled. The decoder then takes this latent vector as input and aims to recreate x. During training, [13] showed you can train these models by optimizing the variational lower bound:

$$\mathcal{L} = -\mathbb{E}_q \left[\log p(x|z)\right] + \mathrm{KL}\left[q_\phi(z|x)||p(z)\right]$$

In practice, this turns into two losses: a reconstruction loss which measures how close the predicted $\hat{x}$ is to x (the left hand term) and a KL divergence loss (right

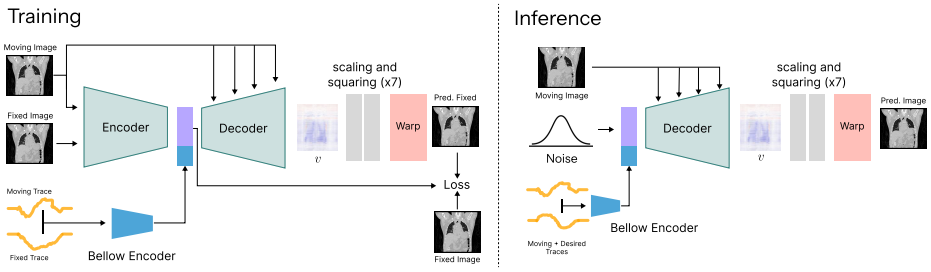

Fig. 2. Architecture of the 5DCT VAE. Training and Inference architectures vary for our conditional Variational Auto-Encoder. During training, the encoder is used to generate a pair μ, σ which we use to sample z. During inference, the encoder is entirely skipped, and we instead sample $z \sim \mathcal{N}(0, 1)$, meaning we do not rely on F at test time.

hand term) which encourages the distribution of latents z to match a prior $p(z)$, generally a simple distribution whose KL divergence is easily computable. In our case, this is the standard Normal distribution.

2.2 Model Architecture & Conditioning

Our VAE is built on top of the original architecture introduced in [4]. Their model is a U-Net modified to use 3D convolutions. Just as in their work, our encoder takes in both the moving and fixed image and accordingly we can write our encoder as $q_\phi(z|M, F)$. We also add a secondary network, the bellow trace encoder, $\mathcal{E}_b$. This encoder is a small network of convolutional and fully connected layers which encodes both of the 448 dimensional breathing traces, b_m and b_f into a bellow embedding. This vector is then concatenated with the sampled z to get the full latent $\tilde{z} = z \oplus \mathcal{E}_b(b_m, b_f)$. We also modify the decoder such that it is conditioned on the moving image through concatenating in intermediate layers. This makes our decoder $p_\theta(v|M, \tilde{z})$. Given a predicted deformation field $\hat{\phi} = \mathcal{S}(p_\theta(v|M, \tilde{z}))$ where $\mathcal{S}$ are the scaling and squaring layers introduced in Section 1.1, we can then write our loss function as:

$$\mathcal{L} = ||F - \hat{\phi} \circ M||_2^2 + \mathrm{KL}(\mathcal{N}(\mu_{z|M,F}, \sigma_{z|M,F})||\mathcal{N}(\mathbf{0}, \mathbf{I}))$$

An overview of our architecture is shown in Fig. 4. During inference, we forgo the encoder entirely and sample $z \sim \mathcal{N}(0, 1)$ but still condition the decoder on M and use the bellow encoder network to compute $\tilde{z}$. Therefore at test time, our network sees only the moving image and both breathing traces. Results of the VAE model are shown in Section. 4. We train our model for 150k steps with a batch size of 4. We use the Adam optimizer with learning rate 1e-4.

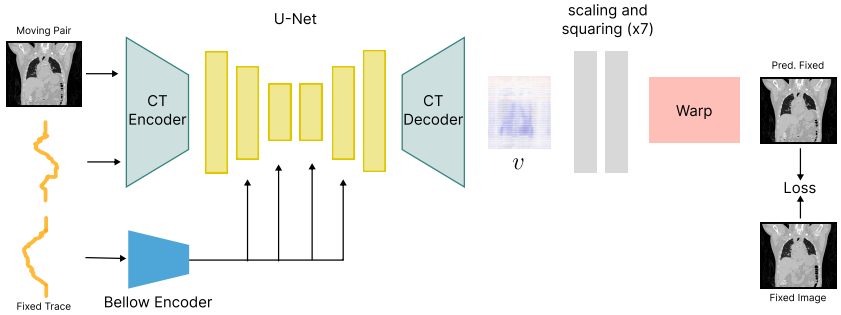

Fig. 3. Architecture of the 5DCT LDM. The LDM architecture is unchanged across training and inference, but is used differently. During training, a timestep is randomly sampled and we run one forward pass to get the output. During inference, we start with a high timestep, denoise, and then add noise again, decreasing the noise level through timesteps, meaning we run the forward pass multiple times per sample.

3 Latent Diffusion Approach

3.1 Background and Formulation

Diffusion Models have recently become state of the art for a variety of computer vision tasks including image generation [7], super-resolution [18], and many more [19]. They have also been applied to medical imaging domains including for X-Rays [15, 27], MRIs [12, 17], and CT-Scans [29, 30]. These models operate on the principle of learned denoising, where a forward process gradually adds Gaussian noise to data, and a neural network is trained to reverse this process [8]. Concretely, we begin with a forward process which gradually adds noise to a sample $\mathbf{x}$:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)I)$$

where α_t make up a pre-defined variance schedule. Traditionally, instead of sampling one timestep at a time, we sample $\epsilon \sim \mathcal{N}(0,1)$ and then compute: $q(\mathbf{x}_t|\mathbf{x}_0) = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + (1-\bar{\alpha}_t)\epsilon$ where $\bar{\alpha}_t = \Pi_t\alpha_t$ We can then parameterize the reverse process as:

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t), \Sigma(\mathbf{x}_t, t))$$

We generally use a pre-selected Σ and use the model to learn μ_θ. Thre are several options of how: One option is to have the model predict the noise added and then use the above equations to calculate μ_θ, and another is to directly have the model predict μ_θ. In this paper, we choose to directly have the model predict μ_θ. Once we have learned μ_θ, we can start from pure noise ($\mathbf{x}_T$) and iteratively take steps towards $\mathbf{x}_0$. Similar to the conditional VAEs introduced earlier, we can condition diffusion models on any number of input signals [10, 20]. Traditionally, this conditioning is done through cross-attention layers [25], which we also use in this paper to condition our model on the fixed bellows trace.

3.2 CT Auto-Encoder

A key property of latent diffusion models is that instead of operating in the high level image space (which would be a 117M dimensional space in our case), they operate in a lower dimensional latent space [20]. This significantly improves computational efficiency and makes the iterative process feasible. Many diffusion models that work with image data use auto-encoders pre-trained on natural images, but we opt to train our own auto-encoder specially focused on lung CT scans. We build on top of a standard 3D VAE architecture described in Section 2.1 and design the encoder such that the input is encoded into a $16 \times 16 \times 16$ latent. We additionally augment our model to add a second branch dedicated to encoding and decoding the bellows trace from the same latent. This allows us to easily encode both the moving scan and the moving bellow trace into a single latent space which we can then use for the diffusion process. We train this model for 100k steps using an effective batch size of 8.

3.3 Model Architecture & Conditioning

Our Latent Diffusion Model is built on top of the traditional U-NET architecture [21], with modifications. Although we work with 3D volumes, it is computationally prohibitive to use 3D convolutions throughout the whole network. Instead, we keep mostly 2D convolutions but augment them with psuedo-3D attention layers inspired by [29]. Specifically, where-ever we use self-attention blocks in the U-Net, we also perform our psuedo-3D attention as follows. If we have input $B \times H \times W \times D$, traditional attention reshapes this into $B \times (H * W) \times D$ vectors so that each spatial location can attend to each other. We instead rearrange and reshape to $B \times D \times (H * W)$ so that each slice can attend to each other across all spatial locations.

We condition our diffusion model on both the moving image, moving trace and on the fixed breathing trace. The moving image and trace are first encoded using our CT encoder and then concatenated on top of the noisy latent passed to the model. We condition on the fixed breathing trace using cross-attention similar to text-conditioning models [20]. This conditioning is done at the 3 smallest feature levels on both the down and up branches of the U-Net. Since the CT decoder was initially designed to reconstruct a CT scan instead of a velocity field, we add a few newly initialized 3D convolutional layers to the end and make it trainable while training the diffusion model, albeit with a significantly lower learning rate. On the contrary, we keep our CT encoder frozen so to that the input latent space remains consistent through the training process. We train this model for 150k steps with an effective batch size of 16. We use the AdamW optimizer with a cosine annealing learning rate scheduling starting at 1e-4 and decaying to 1e-6.

4 Experiments

We train and evaluate on an internal dataset collected at ******* Hospital. Our dataset consists of pairs of 3D Lung CT Scans with breathing traces: records of

Table 1. Quantitative Results on our test set. LDM performs 50 sampling steps with a DDIM scheduler. Masked MSE refers to the MSE only within the lung region. GPU time is computed on 1 A100 GPU.

Model	Masked MSE ↓	MSE ↓	MS-SSIM ↑	$\lvert\nabla J_\phi\rvert < 0$ ↓	GPU Sec.
No Deformation	0.02089	0.00262	0.9357	-	-
cVAE	**0.00827**	0.00253	0.9411	**0.0**	**0.68 ± 0.02**
LDM	0.01239	**0.00214**	**0.9433**	**0.0**	1.98 ± 0.03

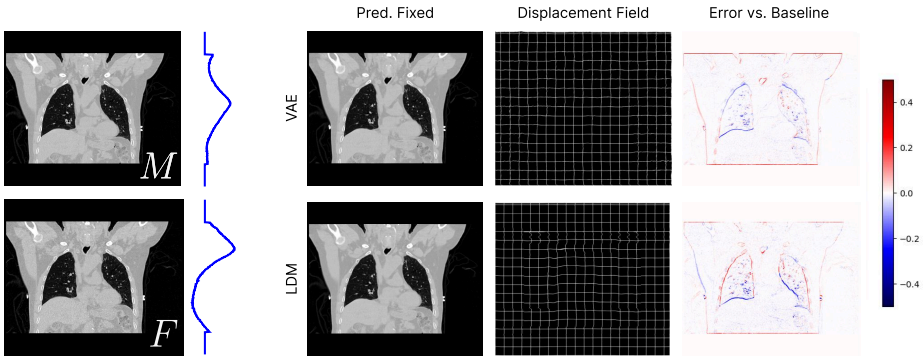

Fig. 4. Qualitative Example of our models. On the left, the moving M and fixed F images along with their corresponding breathing traces. Error vs. Baseline shows the error relative to no deformation (lower is better). Both the VAE and LDM lower errors near the lower lung, but result in increased artifacts on the boundaries of the scan.

normalized breathing amplitude over time. These traces represent a surrogate marker for patient breathing that can obtained during the course of treatment. The raw breathing traces are first padded, aligned, and then re-sampled such that each breathing amplitude value is paired with an axial slice in the 3D scan, providing information on the appearance of different slices across breathing levels along with information on breathing cycles. This dataset consists of 146 total patients, which we split into 126 training, 10 validation, and 10 testing. Each patient has 25 fast helical free breathing scans, resulting in 75,600 pairs of training scans and 6,000 validation and testing scans. CT scans were acquired using the protocol described in [16]. Briefly: they were acquired at 1mm resolution with 512x512 pixels in plane spanning a field of view that includes the thorax including the entire lung extent. We use them at full-resolution for all our techniques. CT scans were clipped to (-1000,400) hounsfield units before being normalized to (0,1). Breathing traces were recorded with a digital acquisition (DAQ) box time-synced to the X-Ray. All models were trained on a machine with 4 A100 GPUs with 80 GB memory each.

We report quantitative results on our test set in Table 1. As a baseline, we compute the error if we do no warping at all, i.e. treat the moving image as the fixed image directly. We compute the mean squared error (MSE) and mutli-scale

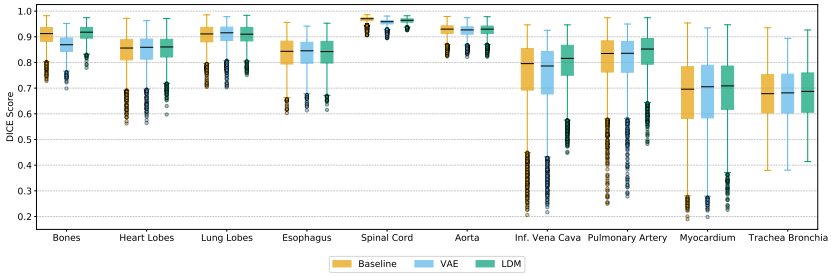

Fig. 5. Dice Scores across 10 buckets of segmentations in the lung regions. Some buckets, such as the Bones and Spinal Cord, generally do not deform with breathing motion and should stay constant while others, such as Lung Lobes, and Trachea Bronchia, change significantly in breathing.

structural similarity index measure (MS-SSIM), a metric which places stronger weight on sharp details. We additionally compute a masked MSE, which is computed only within a general lung mask: we run segmentation and then expand the mask by 10mm in every direction. This masked value better captures the MSE in the areas that we are most interested in. We also report the number of voxels for which the determinant of the gradient of the Jacobian of our displacement ∇J_ϕ is less than 0: the determinant of the Jacobian measures the relative change in volume, so a determinant less than zero implies our transformation results in "folding" or "inverting" of space, meaning it is no longer diffeomorphic and cannot be inverted 1-1 [1]. Both of our deep learning techniques entirely avoid inverted warps. Both the VAE and LDM models outperform the baseline across all metrics, with particularly strong performance gains on Masked MSE.

4.1 Segmentation Comparisons

We also compare our two approaches using DICE scores on various segmentations in the lung area. Segmentations for all scans in our dataset were run using TotalSegmentator [9, 26] and a specific subset were grouped into buckets and evaluated. This is done for computational efficiency but is done such that within each bucket there are no overlapping masks. The 10 buckets are Bones, Heart Lobes, Lung Lobes, Esophagus, Spinal Cord, Aorta, Inferior Vena Cava, Pulmonary Artery, Myocardium, Trachea Bronchia. Results are shown in Fig. 5.

5 Conclusion

In this paper, we have introduced novel deep learning based techniques for diffeomorphic image generation based on respiratory motion. Our models are trained once on a diverse dataset of lung CT scans paired with abdominal breathing traces which measure breathing amplitude over time. Although we only benchmark these against pre-collected datasets, our work can have real clinical applications: being able to dynamically predict motion based on real time breathing

traces or to simulate and visualize treatment deformation would open up new avenues for safe, effective radiotherapy treatment.

References

1. Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage **38**(1), 95–113 (2007)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9252–9260 (2018)
3. Burnet, N.G., Thomas, S.J., Burton, K.E., Jefferies, S.J.: Defining the tumour and target volumes for radiotherapy. Cancer Imaging **4**(2), 153 (2004)
4. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical image analysis **57**, 226–236 (2019)
5. Ford, E.C., Mageras, G., Yorke, E., Ling, C.: Respiration-correlated spiral ct: a method of measuring respiratory-induced anatomic motion for radiation treatment planning. Medical physics **30**(1), 88–97 (2003)
6. Hanley, J., Debois, M.M., Mah, D., Mageras, G.S., Raben, A., Rosenzweig, K., Mychalczak, B., Schwartz, L.H., Gloeggler, P.J., Lutz, W., et al.: Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation. International Journal of Radiation Oncology* Biology* Physics **45**(3), 603–611 (1999)
7. Hatamizadeh, A., Song, J., Liu, G., Kautz, J., Vahdat, A.: Diffit: Diffusion vision transformers for image generation. In: European Conference on Computer Vision. pp. 37–55. Springer (2024)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
10. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D.: Diffusion models for medical image analysis: A comprehensive survey. arXiv preprint arXiv:2211.07804 (2022)
11. Keall, P.J., Mageras, G.S., Balter, J.M., Emery, R.S., Forster, K.M., Jiang, S.B., Kapatoes, J.M., Low, D.A., Murphy, M.J., Murray, B.R., et al.: The management of respiratory motion in radiation oncology report of aapm task group 76 a. Medical physics **33**(10), 3874–3900 (2006)
12. Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baeßler, B., Foersch, S., et al.: Denoising diffusion probabilistic models for 3d medical image generation. Scientific Reports **13**(1), 7303 (2023)
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2022), https://arxiv.org/abs/1312.6114
14. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE transactions on medical imaging **38**(9), 2165–2176 (2019)

15. Liu, X., Qiao, Z., Liu, R., Li, H., Zhang, J., Zhen, X., Qian, Z., Zhang, B.: Diffux2ct: Diffusion learning to reconstruct ct images from biplanar x-rays (2024), https://arxiv.org/abs/2407.13545
16. Low, D.A., Parikh, P.J., Lu, W., Dempsey, J.F., Wahab, S.H., Hubenschmidt, J.P., Nystrom, M.M., Handoko, M., Bradley, J.D.: Novel breathing motion model for radiotherapy. International Journal of Radiation Oncology* Biology* Physics **63**(3), 921–929 (2005)
17. Mirza, M.U., Dalmaz, O., Bedel, H.A., Elmas, G., Korkmaz, Y., Gungor, A., Dar, S.U., Çukur, T.: Learning fourier-constrained diffusion bridges for mri reconstruction (2023), https://arxiv.org/abs/2308.01096
18. Moser, B.B., Shanbhag, A.S., Raue, F., Frolov, S., Palacio, S., Dengel, A.: Diffusion models, image super-resolution, and everything: A survey. IEEE Transactions on Neural Networks and Learning Systems (2024)
19. Po, R., Yifan, W., Golyanik, V., Aberman, K., Barron, J.T., Bermano, A., Chan, E., Dekel, T., Holynski, A., Kanazawa, A., et al.: State of the art on diffusion models for visual computing. In: Computer Graphics Forum. vol. 43, p. e15063. Wiley Online Library (2024)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
22. Stone, H.B., Coleman, C.N., Anscher, M.S., McBride, W.H.: Effects of radiation on normal tissue: consequences and mechanisms. The lancet oncology **4**(9), 529–536 (2003)
23. Tada, T., Minakuchi, K., Fujioka, T., Sakurai, M., Koda, M., Kawase, I., Nakajima, T., Nishioka, M., Tonai, T., Kozuka, T.: Lung cancer: intermittent irradiation synchronized with respiratory motion–results of a pilot study. Radiology **207**(3), 779–783 (1998). https://doi.org/10.1148/radiology.207.3.9609904, pMID: 9609904
24. Thomas, D., Lamb, J., White, B., Jani, S., Gaudio, S., Lee, P., Ruan, D., McNitt-Gray, M., Low, D.: A novel fast helical 4d-ct acquisition technique to generate low-noise sorting artifact–free images at user-selected breathing phases. International Journal of Radiation Oncology* Biology* Physics **89**(1), 191–198 (2014)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
27. Weber, T., Ingrisch, M., Bischl, B., Rügamer, D.: Cascaded latent diffusion models for high-resolution chest x-ray synthesis. In: Pacific-Asia conference on knowledge discovery and data mining. pp. 180–191. Springer (2023)
28. Wong, J.W., Sharpe, M.B., Jaffray, D.A., Kini, V.R., Robertson, J.M., Stromberg, J.S., Martinez, A.A.: The use of active breathing control (abc) to reduce margin for breathing motion. International Journal of Radiation Oncology*Biology*Physics **44**(4), 911–919 (1999)
29. Zhu, L., Codella, N., Chen, D., Jin, Z., Yuan, L., Yu, L.: Generative enhancement for 3d medical images. arXiv preprint arXiv:2403.12852 (2024)

30. Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., Yu, L.: Make-a-volume: Leveraging latent diffusion models for cross-modality 3d brain mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 592–601. Springer (2023)