Diffusing Boundaries: CBCT-to-CT Translation with Extended Field of View

Quentin Spinat[1][0009−0008−7929−0528], Audrey Duran[1], Olivier Teboul[1], Nikos Paragios[1,2][0000−0002−9668−4763], and Nikos Komodakis[1,3,4,5]

[1] TheraPanacea, Paris, France,
[2] Universite Paris-Saclay, CentraleSuplec, 91190, Gif-sur-Yvette, France
[3] Computer Science Department, University of Crete
[4] IACM, Forth
[5] Archimedes, Athena RC
n.paragios@therapanacea.eu

Abstract. Cone-beam computed tomography (CBCT) is an essential imaging modality for adaptive radiotherapy, enabling the positioning and real-time verification of anatomical changes. However, CBCT images suffer from artifacts and lack the accurate Hounsfield unit (HU) calibration necessary for dose computation. Additionally, CBCT's limited field of view (FOV) further complicates its direct application for replanning. To address these limitations, we propose a novel framework leveraging diffusion models to synthesize a synthetic CT (sCT) from CBCT while inpainting the extended FOV using the original planning CT (pCT). Our method integrates with any CBCT-to-CT diffusion framework without degrading its performance, ensuring accurate HU values and comprehensive anatomical coverage for dose computation without requiring new CT acquisitions. Quantitative and qualitative evaluations demonstrate that our approach preserves the baseline CBCT-to-CT translation quality while effectively extending the FOV, offering a streamlined and effective solution for adaptive radiotherapy workflows.

Keywords: Medical Image Processing · Radiotherapy Workflow Optimization · Diffusion Models · Image Synthesis · Inpainting.

1 Introduction

Accurate dose computation in radiotherapy requires high-quality CT images with calibrated Hounsfield units (HU) reflecting electron density (ED). However, cone-beam CT (CBCT), used for patient setup and verification, suffers from artifacts, poor HU calibration, and a restricted field of view (FOV), making direct dose computation unreliable. As a result, additional CT scans are needed when significant anatomical changes occur, delaying treatment and increasing patient burden, particularly in online adaptive radiotherapy, where rapid imaging updates are essential.

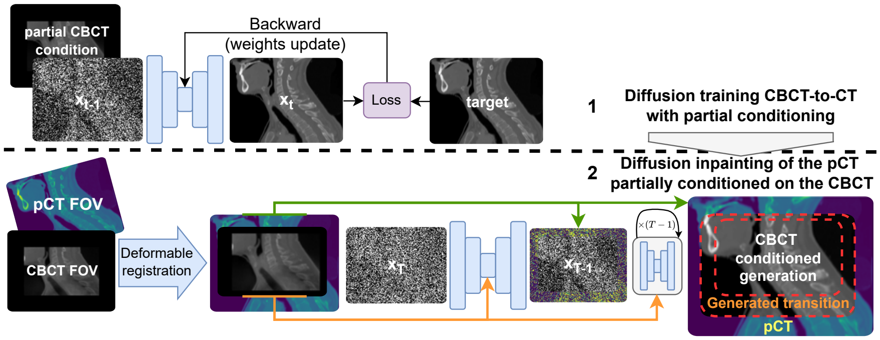

Fig. 1. sCT Generation with Extended FOV. A diffusion model is trained with partial conditioning on the CBCT. During inference, it extends the FOV by inpainting with a deformably registered pCT, ensuring a smooth and anatomically valid transition.

To overcome these limitations, we propose a novel diffusion-based framework for generating high-quality synthetic CTs (sCTs) from CBCT. Our approach addresses the FOV limitation through an inpainting-based diffusion process that seamlessly integrates information from the original planning CT (pCT). By conditioning sCT generation on CBCT input while incorporating pCT data, our method produces anatomically consistent volumes that leverage the strengths of both modalities—preserving the accurate HU representation of the pCT while capturing the most up-to-date anatomical details from the CBCT.

Our framework eliminates the need for new CT acquisitions, enabling dose computation on extended-FOV sCTs while maintaining the performance of existing CBCT-to-CT diffusion models. This advancement streamlines adaptive radiotherapy workflows, reducing delays and minimizing unnecessary radiation exposure. Quantitative results confirm that partial conditioning preserves diffusion model performance, while visual assessments demonstrate anatomically plausible inpainting, positioning our approach as a practical and flexible solution for modern adaptive radiotherapy.

2 Related work

2.1 Diffusion Models for Conditional Generation

Diffusion models [8, 21], have garnered significant attention for their exceptional generative capabilities. Unlike GANs, they operate within a probabilistic framework, ensuring stable training dynamics and producing high-quality outputs. In medical imaging, these models have demonstrated success in tasks such as image-to-image translation [26, 15], and volumetric synthesis [17], outperforming GANs-based methods.

2.2 Synthetic CT Generation in Radiotherapy

sCT generation from CBCT has become a key area of research in radiotherapy, aimed at enhancing dose computation accuracy and streamlining treatment planning. Traditional approaches, such as intensity mapping, often struggle to generalize across varying anatomical regions and acquisition settings, due to the artifacts and variability inherent in CBCT.

To overcome these limitations, generative adversarial networks (GANs) [7] and cycle-consistent GANs (CycleGAN) [25] have been proposed [2, 3, 11, 12, 14, 18, 23, 24]. While these models address domain translation, their inherent instability during training has hindered their widespread adoption in clinical settings. Recent advances in conditional diffusion models have brought a new wave of research to the CBCT-to-CT image translation task [6, 13, 16]. These models improved performance in handling CBCT artifacts and domain variability, thereby offering a more reliable solution for sCT generation in radiotherapy.

2.3 Field-of-View Expansion

Expanding the FOV is critical for dose computation in radiotherapy. Interpolation and registration-based methods [9, 20] have been used to reconstruct missing regions but often lack anatomical plausibility, particularly in cases with significant deformations. CNN-based methods [4, 5, 10] improve reconstruction by leveraging CBCT-domain priors. However, they still have difficulty incorporating pCT priors effectively. Recent advances in diffusion-based inpainting [22] offer a unified approach, seamlessly integrating generation and inpainting.

3 Method

3.1 Background

Diffusion models Diffusion models [8, 21] are a class of generative models that leverage a forward and reverse stochastic process. The forward process progressively adds Gaussian noise to the data x_0, resulting in increasingly noisy representations $x_1, x_2, \ldots, x_T$. This process is defined as a Markov chain:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

where β_t is the schedule controlling the noise added at each step.

The reverse process reconstructs the original data x_0 by denoising the corrupted representations. A neural network $\epsilon_\theta(x_t, t)$ predicts the noise added at each step, enabling the reconstruction of x_{t-1} from x_t:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)I) \tag{2}$$

where $\mu_\theta(x_t, t)$ and $\Sigma_\theta(x_t, t)$ are derived from the predicted noise $\epsilon_\theta(x_t, t)$.

The training objective minimizes the discrepancy between the predicted and true noise, using a simplified loss function:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{x_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right] \tag{3}$$

Diffusion models are known for their stability during training and ability to generate high-quality outputs. Moreover, the flexibility of the reverse process enables conditional generation by incorporating auxiliary information, making them suitable for structured tasks.

Latent Diffusion Models Latent diffusion models [19] (LDMs) operate in a compressed latent space, mapped from the original data space x via an encoder-decoder pair. The forward diffusion process is applied to the latent variables z, reducing computational demands while retaining generative capabilities. LDMs improve efficiency, allowing diffusion models to handle high-resolution inputs and complex tasks.

Latent Space with 3D VAE In this work, we employ a 3D VAE to encode volumetric data into a compressed latent space, enhancing the efficiency of diffusion by operating directly in 3D while preserving anatomical structure. Unlike 2D VAEs, a 3D VAE maintains spatial continuity along the z-axis, crucial for ensuring smooth transitions across slices and improving overall 3D coherence in the generated sCTs.

3.2 CBCT-to-CT Field of View Extension

Our approach leverages a latent diffusion framework as outlined above. To effectively integrate pCT information for FOV extension, we propose modifying the diffusion training with partial conditioning and adapting the inference process by incorporating a diffusion-based inpainting approach that leverages pCT data. Before inpainting, CBCT and pCT are aligned via deformable registration. Fig. 1 provides an overview. Importantly, these modifications seamlessly integrate into existing diffusion frameworks without performance loss.

Partial Conditioning A CBCT-conditioned diffusion model alone cannot ensure seamless FOV extension. It synthesizes sCT in CBCT-conditioned regions but depends on pCT for missing areas. Deformable registration helps but often causes boundary discontinuities due to misalignments or texture differences (Fig. 2). Similarly, diffusion-based inpainting [22] struggles with smooth transitions when trained on fully conditioned CT-CBCT pairs (Fig. 3), as it only learns to generate CT in CBCT-covered regions, failing in empty areas (Fig. 4).

To overcome this, we introduce partial conditioning, training the model with CBCT inputs that are only partially present while keeping the CT target fully available. This allows the model to generate sCT conditioned on CBCT while ensuring smooth and anatomically valid transitions into unconditioned regions (Fig. 3).

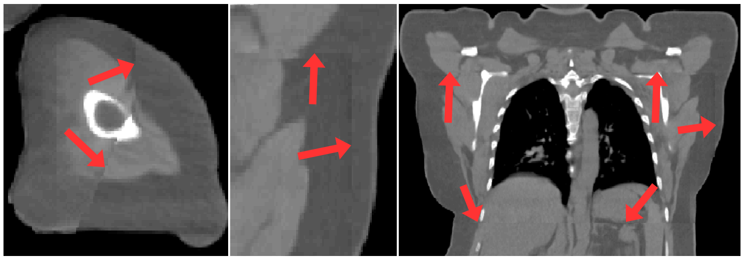

Fig. 2. Discontinuities arising from a naive inpainting, consisting of a deformable registration of the pCT on the generated sCT to fill in the missing parts. Those are not present in our approach as shown in 7

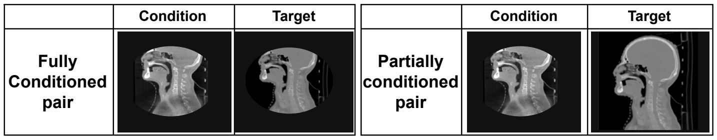

Fig. 3. Partial Conditioning training The model is trained with inputs where the CBCT can be only partially present, whereas the CT target is fully there. By learning to fill the gaps in the condition, the model is now able to generate anatomically consistent sCT between the CBCT-conditioned sCT and the pCT.

Field of View Extension To extend the FOV, we modify the inference process of the diffusion model using an inpainting-based approach that incorporates pCT information. To ensure a smooth and anatomically consistent transition between the CBCT-conditioned sCT and the pCT, we introduce two key ideas: (1) performing inpainting directly in the latent space of a 3D VAE to improve efficiency and 3D coherence, and (2) defining an inpainting region via a dilation margin, which extends the CBCT mask by a specified number of voxels (Fig. 5). In addition, we introduce a context region surrounding the inpainting area, called the context margin. Given these regions, the denoising process follows:

$$z_{t-1}^{\text{context}} = \mathcal{N}(\sqrt{\overline{\alpha}_t} z^{\text{pCT}}, (1-\overline{\alpha}_t)I) \tag{4a}$$

$$z_{t-1}^{\text{inpaint}} = \mathcal{N}(\mu_\theta(z_t,t), \Sigma(z_t,t)) \tag{4b}$$

$$z_{t-1} = m \odot z_{t-1}^{\text{context}} + (1-m) \odot z_{t-1}^{\text{inpaint}} \tag{4c}$$

where m is a binary mask that distinguishes context and inpainting regions, z_{t-1}^{context} is a noisy version of the known pCT region, z_{t-1}^{inpaint} is the predicted value for the unknown region, z^{pCT} is the latent encoding of the pCT, and $\overline{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ is the total noise variance.

The existence of a dilation margin allows the diffusion model to generate a seamless transition between the CBCT-conditioned sCT and the pCT, even when misaligned. Its size must be chosen to balance anatomical consistency and computational efficiency: a small margin risks discontinuities, while a large

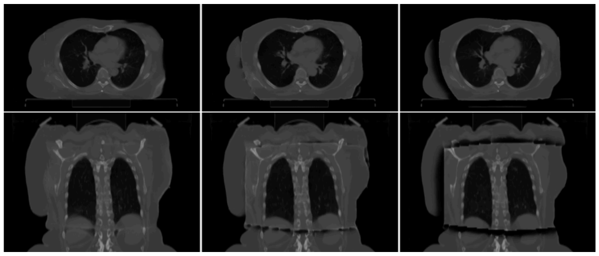

Fig. 4. FOV extension with a diffusion model trained without partial conditioning for different dilation margins. Artifacts are visible at the border of the CBCT conditioning, even when dilation margin is zero (left image). Omitting partial conditioning during training leads to poor transitions beyond the CBCT FOV.

margin reduces the influence of conditioning and increases computational cost (Fig. 6).

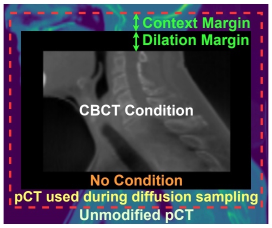

Fig. 5. Dilation and context margins. The inpainting area comprises the CBCT-conditioned region and a surrounding dilation margin, which defines the unconditioned region where a smooth transition is generated. Additionally, the context margin specifies the region where pCT information is incorporated during the diffusion process.

The context region plays a crucial role in anchoring the planning CT information. As shown in Eq. 4, we use it to incorporate a noised version of the planning CT directly into the denoising diffusion process. Given the significant difference in FOV between CBCT and pCT, inpainting the full pCT would be computationally prohibitive. To mitigate this, we restrict the diffusion process to the inpainting and context regions, leaving the remaining pCT unchanged. The context margin size is critical: a zero context margin can lead to discontinuity artifacts, while an excessively large context margin increases computational cost.

4 Experimental Results

4.1 Datasets and Hardware

Hardware Trainings and inferences were conducted on Nvidia RTX 3080Ti GPUs with 24GB. Most trainings used 4 GPUs with Distributed Data Parallel.

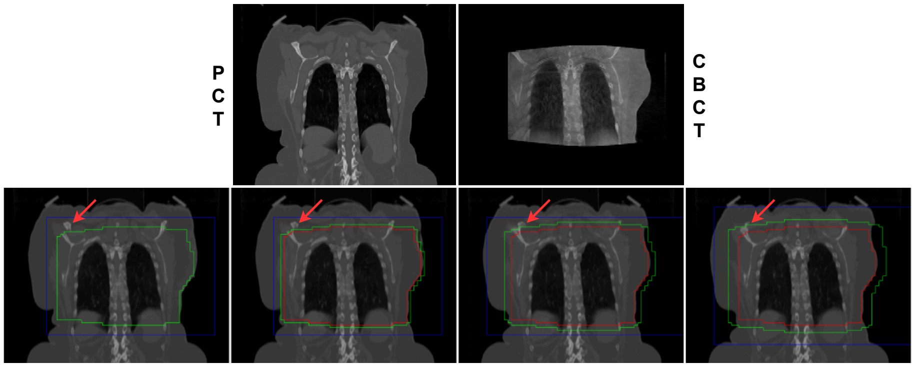

Fig. 6. Conditioned inpainting for different values of the sampling and context margins. Too small margins leads to discontinuities. CBCT borders in red, dilation margin borders in green, context margin borders in blue.

Models We employ a 3D latent diffusion architecture, with a 3D VAE (SDXL-inspired) and a 3D U-Net denoiser (depth 3, with attention at the bottleneck), using magnitude-preserving layers. The VAE (∼250M params) uses L1, KL-divergence, spatial-gradient, adversarial, and perceptual losses, on thousands of CT and CBCT scans. Volumes were cropped to $128 \times 128 \times 32$ patches, intensity clamped to [-1024, 3000] HU, and rescaled to [-1, 1]. The compression rate was [8,8,2] along [x,y,z]. The CBCT-to-CT U-Net diffusion model (∼300M params) was trained on 4,298 head-and-neck and 875 breast scan pairs, preprocessed similarly, with random cropping to $256 \times 256 \times 64$. All data are private, anonymized CBCT and CT scans from multiple centers.

Inference For memory reason, inference is done patchwise for the VAE encoding, decoding, and diffusion model sampling, using multidiffusion [1] for patch aggregation. Processing a typical $512 \times 512 \times 128$ volume required ∼5GB of memory and ∼1 minute.

4.2 Evaluation and Results

Evaluation Setup Quantitative evaluation of the extended FOV is challenging due to the absence of ground truth. CBCT images have a restricted FOV, and while the pCT provides an approximation, it does not reflect anatomical changes between acquisitions. Additionally, patient-specific deformations such as organ motion, weight loss, and tumor progression cause misalignments between pCT and CBCT, particularly in soft tissues, making quantitative metrics unreliable. Standard image similarity metrics like structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) fail to assess inpainting quality and transition smoothness adequately. A more nuanced approach is needed to evaluate anatomical consistency and clinical usability.

Qualitative Evaluation Given these challenges, we prioritize visual evaluation to assess anatomical plausibility, seamless blending at transition boundaries, and overall realism of the generated sCTs. Qualitative comparisons offer a more reliable means of determining whether the FOV extension produces clinically usable images. As shown in Fig. 7, our method achieves a seamless valid transition between sCT and pCT, outperforming a naïve pCT-to-sCT registration approach.

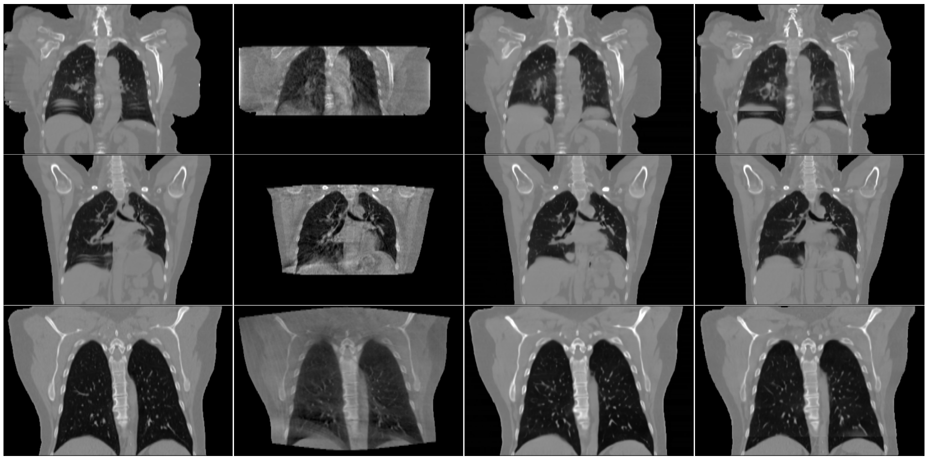

Fig. 7. Visual evaluation. Left to right: pCT, CBCT, Our method, naive method.

Quantitative Validation of Partial Conditioning While direct evaluation of FOV extension remains difficult, we verify that partial conditioning does not degrade diffusion model performance in CBCT-conditioned regions, proving that our method enables FOV extension without degrading the model's base CBCT-to-CT performance (Tab. 1). We stress that our objective is not to outperform specific CBCT-to-CT models, but to enable FOV extension without degrading performance within the original FOV This ensures that our method can be seamlessly integrated into existing CBCT-to-CT diffusion frameworks while preserving synthesis accuracy.

Expert Review Our method has been approved for clinical use by collaborating radiologists, following evaluation based on dose metrics and visual assessments.

5 Conclusion

We introduced a diffusion-based framework for sCT generation with extended FOV from CBCT, leveraging pCT for inpainting. Our method preserves CBCT-to-CT translation quality while seamlessly integrating FOV extension, addressing CBCT artifacts and limited coverage. By incorporating partial conditioning

Config.	**MAE ↓**	**SSIM ↑**	**PSNR↑**	**Axial ↓ FID**	**Sagittal ↓ FID**	**Organs ↑ Av. Dice**
Classical	**50,48**	0,946	**35,00**	**16,35**	**11,06**	0,916
Partial Cond.	50,73	**0,947**	34,78	16,46	11,13	**0,921**

Table 1. Performance with vs. without partial conditioning on CBCT-to-CT Breast translation task. Organ Dice is averaged over 9 organs. Metrics are computed within the CBCT mask, excluding FOV augmentation from metrics. Partial conditioning maintains the same performance within CBCT region, proving that our method enables FOV extension without degrading the model's base CBCT-to-CT performance.

and diffusion-based inpainting, we ensure a smooth transition between CBCT-conditioned and inpainted regions.

Given the lack of ground truth in the extended FOV, quantitative evaluation remains challenging. Instead, our qualitative analysis demonstrates that the method produces high-fidelity sCTs with realistic anatomical continuity and smooth transitions. Radiologist grading is planned for future work

Rather than fine-tuning an existing CBCT-to-CT model, we retrained a diffusion model from scratch with partial conditioning to demonstrate general applicability. However, fine-tuning a pretrained model with our approach could significantly reduce training time while maintaining performance. Future work will explore this strategy to enhance efficiency and accessibility. Additionally, we will focus on clinical validation in real-world adaptive radiotherapy settings, further advancing automated FOV extension and its impact on patient care.

Disclosure of Interests. All authors were employees of Therapanacea at the time of this study and received compensation as part of their employment. As such, they declare financial interest in the company.

References

1. Bar-Tal, O., Yariv, L., Lipman, Y., Dekel, T.: Multidiffusion: Fusing diffusion paths for controlled image generation. Proceedings of Machine Learning Research **202**, 1737–1752 (2023)
2. Brou Boni, K.N., Klein, J., Gulyban, A., Reynaert, N., Pasquier, D.: Improving generalization in mr-to-ct synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data. Medical physics **48**(6), 3003–3010 (2021)
3. Chen, J., Wei, J., Li, R.: Targan: target-aware generative adversarial networks for multi-modality medical image translation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 24–33. Springer (2021)
4. Fonseca, G.P., Baer-Beck, M., Fournie, E., Hofmann, C., Rinaldi, I., Ollers, M.C., van Elmpt, W.J., Verhaegen, F.: Evaluation of novel ai-based extended field-of-view ct reconstructions. Medical Physics **48**(7), 3583–3594 (2021)
5. Fournié, É., Baer-Beck, M., Stierstorfer, K.: Ct field of view extension using combined channels extension and deep learning methods. arXiv preprint arXiv:1908.09529 (2019)
6. Fu, L., Li, X., Cai, X., Miao, D., Yao, Y., Shen, Y.: Energy-guided diffusion model for cbct-to-ct synthesis. Computerized Medical Imaging and Graphics **113**, 102344 (2024)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
9. Hou, J., Guerrero, M., Chen, W., D'Souza, W.D.: Deformable planning ct to cone-beam ct image registration in head-and-neck cancer. Medical physics **38**(4), 2088–2094 (2011)
10. Huang, Y., Gao, L., Preuhs, A., Maier, A.: Field of view extension in computed tomography using deep learning prior. In: Bildverarbeitung für die Medizin 2020: Algorithmen–Systeme–Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2020 in Berlin. pp. 186–191. Springer (2020)
11. Kearney, V., Ziemer, B.P., Perry, A., Wang, T., Chan, J.W., Ma, L., Morin, O., Yom, S.S., Solberg, T.D.: Attention-aware discrimination for mr-to-ct image translation using cycle-consistent generative adversarial networks. Radiology: Artificial Intelligence **2**(2), e190027 (2020)
12. Lei, Y., Harms, J., Wang, T., Liu, Y., Shu, H.K., Jani, A.B., Curran, W.J., Mao, H., Liu, T., Yang, X.: Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks. Medical physics **46**(8), 3565–3581 (2019)
13. Li, Y., Shao, H.C., Liang, X., Chen, L., Li, R., Jiang, S., Wang, J., Zhang, Y.: Zero-shot medical image translation via frequency-guided diffusion models. IEEE transactions on medical imaging (2023)
14. Oulbacha, R., Kadoury, S.: Mri to ct synthesis of the lumbar spine from a pseudo-3d cycle gan. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI). pp. 1784–1787. IEEE (2020)
15. Pan, S., Chang, C.W., Peng, J., Zhang, J., Qiu, R.L., Wang, T., Roper, J., Liu, T., Mao, H., Yang, X.: Cycle-guided denoising diffusion probability model for 3d cross-modality mri synthesis. arXiv preprint arXiv:2305.00042 (2023)

16. Peng, J., Qiu, R.L., Wynne, J.F., Chang, C.W., Pan, S., Wang, T., Roper, J., Liu, T., Patel, P.R., Yu, D.S., et al.: Cbct-based synthetic ct image generation using conditional denoising diffusion probabilistic model. Medical physics **51**(3), 1847–1859 (2024)
17. Pinaya, W.H., Tudosiu, P.D., Dafflon, J., Da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J.: Brain imaging generation with latent diffusion models. In: MICCAI Workshop on Deep Generative Models. pp. 117–126. Springer (2022)
18. Qian, P., Xu, K., Wang, T., Zheng, Q., Yang, H., Baydoun, A., Zhu, J., Traughber, B., Muzic, R.F.: Estimating ct from mr abdominal images using novel generative adversarial networks. Journal of Grid Computing **18**, 211–226 (2020)
19. 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)
20. Ruchala, K.J., Olivera, G.H., Kapatoes, J.M., Reckwerdt, P.J., Mackie, T.R.: Methods for improving limited field-of-view radiotherapy reconstructions using imperfect a priori images. Medical physics **29**(11), 2590–2605 (2002)
21. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
22. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
23. Tie, X., Lam, S.K., Zhang, Y., Lee, K.H., Au, K.H., Cai, J.: Pseudo-ct generation from multi-parametric mri using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients. Medical physics **47**(4), 1750–1762 (2020)
24. Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J.: Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. pp. 174–182. Springer (2018)
25. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
26. 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)