


sCTFlow: 3D MRI-to-sCT with Conditional Rectified Flow

Pharuj Rajborirug^{1,2}

RAJBORI.P@CMKL.AC.TH, PHARUJ.RA@KMITL.AC.TH

Supan Tungjitkusolmun^{1,2}

SUPAN@CMKL.AC.TH, SUPAN.TU@KMITL.AC.TH

Sarun Gulyanon¹ 

SARUN@CMKL.AC.TH

¹ CMKL University, Bangkok 10520, Thailand

² King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

Editors: Under Review for MIDL 2026

Abstract

MRI and CT images are both crucial for radiotherapy planning, since MRI provides superior soft-tissue contrast for tumor delineation, while CT provides Hounsfield units (HU) required for dose calculation. MR-only radiotherapy offers important advantages, including reduced registration errors, elimination of additional radiation exposure, and streamlined clinical workflows. Generating synthetic CT (sCT) from MRI remains challenging due to the need for realistic HU reconstruction and the high computational demands of processing large 3D image volumes. We propose sCTFlow, a Conditional Rectified Flow (CRF) framework for 3D MRI-to-sCT translation. Unlike diffusion probabilistic models (DDPMs), sCTFlow learns a deterministic velocity field mapping noise to data, ensuring stability and requiring fewer sampling steps. Our architecture, a 3D Attention U-Net, conditions on MRI and organ segmentation via feature-wise linear modulation to predict velocity fields, which are subsequently converted into HU estimates. We evaluated our approach on the SynthRAD2023 dataset. sCTFlow achieves an MAE of 81.18 ± 19.48 HU, PSNR of 26.93, and SSIM of 0.830. We also investigated our method and found that the model captures HU distributions rather than relying on simple intensity transformations, indicating its capacity to model underlying CT characteristics. These findings demonstrate that sCTFlow has potential for reliable and clinically applicable MR-only radiotherapy workflows.

Keywords: synthetic CT, MRI-only radiotherapy, rectified flow, flow matching, medical image synthesis

1. Introduction

Radiotherapy conventionally relies on computed tomography (CT), which provides the Hounsfield units (HU) required for dose calculation, and magnetic resonance imaging (MRI) for tumor and organs-at-risk (OAR) delineation due to its superior soft-tissue contrast (Metcalfe et al., 2013). In standard clinical workflows, both modalities are acquired and subsequently registered, a process that introduces geometric misalignment, increases treatment latency, and adds cumulative exposure to radiation from CT scans (Huijben et al., 2024; McNair et al., 2022; Zhou et al., 2018). Recent clinical studies have demonstrated that MRI-only workflows, enabled by synthetic CT (sCT), can eliminate the CT-MRI registration step, shorten the treatment pathway, reduce patient radiation exposure and CT device utilization while maintaining accurate dose calculation (Bahloul et al., 2024; Thummerer et al., 2023; Bird et al., 2021).

Despite these advantages, MRI-to-sCT synthesis remains challenging for two key reasons. First, MRI intensities do not uniquely encode electron density, particularly in regions with ambiguous tissue boundaries (Hoesl et al., 2023; Edmund and Nyholm, 2017). As a result, a single MRI can correspond to multiple plausible CT appearances. This inherent ambiguity requires generative models that capture the underlying distribution of possible sCTs, rather than collapsing to a single deterministic prediction that averages over plausible anatomies (Nie et al., 2017; Isola et al., 2017). This motivates probabilistic modeling for MRI-to-sCT synthesis (Monteiro et al., 2020; Kohl et al., 2018). Second, high-resolution 3D medical volumes contain tens of millions of voxels, imposing substantial GPU memory and computational constraints that limit model capacity, spatial resolution, and training efficiency (Boulanger et al., 2021).

Early deep-learning sCT methods used supervised convolutional neural networks (CNNs) such as 2D and 3D U-Nets to directly regress HU values from paired MRI-CT data (Bird et al., 2021; Boulanger et al., 2021; Spadea et al., 2021). Although these models offer effective and stable training, their voxel-wise losses often produce overly smooth structures and poor bone depiction, motivating research into more expressive generative approaches. GAN-based models improve CT-like realism by enhancing edge sharpness and tissue contrast through adversarial training (Texier et al., 2023; Bird et al., 2021), but they remain susceptible to hallucinations, mode collapse, and training instability (Yazdani et al., 2025). Diffusion models have recently emerged as an alternative, synthesizing images via iterative denoising steps that produce anatomically realistic sCTs (Pan et al., 2024; Rombach et al., 2022). While they avoid adversarial instability and offer high fidelity, standard diffusion pipelines require many sampling steps, making them computationally expensive for high-resolution 3D volumes. Latent diffusion models (LDMs) reduce this burden by operating in a compressed latent space (Rombach et al., 2022), but diffusion-based sCT generation can still be slow for clinical deployment.

Flow-based generative models provide an emerging alternative. Instead of relying on stochastic denoising or adversarial training, flow matching regresses a time-dependent vector field that transports the input distribution to the target distribution (Lipman et al., 2022, 2024). Rectified flows simplify this further by learning a deterministic ordinary differential equation (ODE) that maps noise to data along a straightened trajectory (Yazdani et al., 2025). This formulation enables high-quality synthesis, captures the underlying distribution of plausible outputs, and requires only a few integration steps, while demonstrating strong performance and training stability in medical image synthesis tasks (Yazdani et al., 2025). An overview of our method is shown in Figure 1. Our approach adopts the conditional rectified flow (CRF) framework by conditioning the rectified flow on the input MRI, which guides the generation process to ensure anatomical consistency.

In this work, we make two contributions: (i) we introduce the first CRF framework for MRI-to-CT synthesis, enabling effective and stable sCT generation; and (ii) we demonstrate that our model captures the underlying distribution of CT appearances rather than relying on simple intensity transformations, indicating its ability to learn meaningful CT characteristics. Our findings show that our method learns clinically meaningful tissue-density relationships, producing anatomically coherent CT outputs rather than behaving as a black-box intensity mapping from MRI to CT.

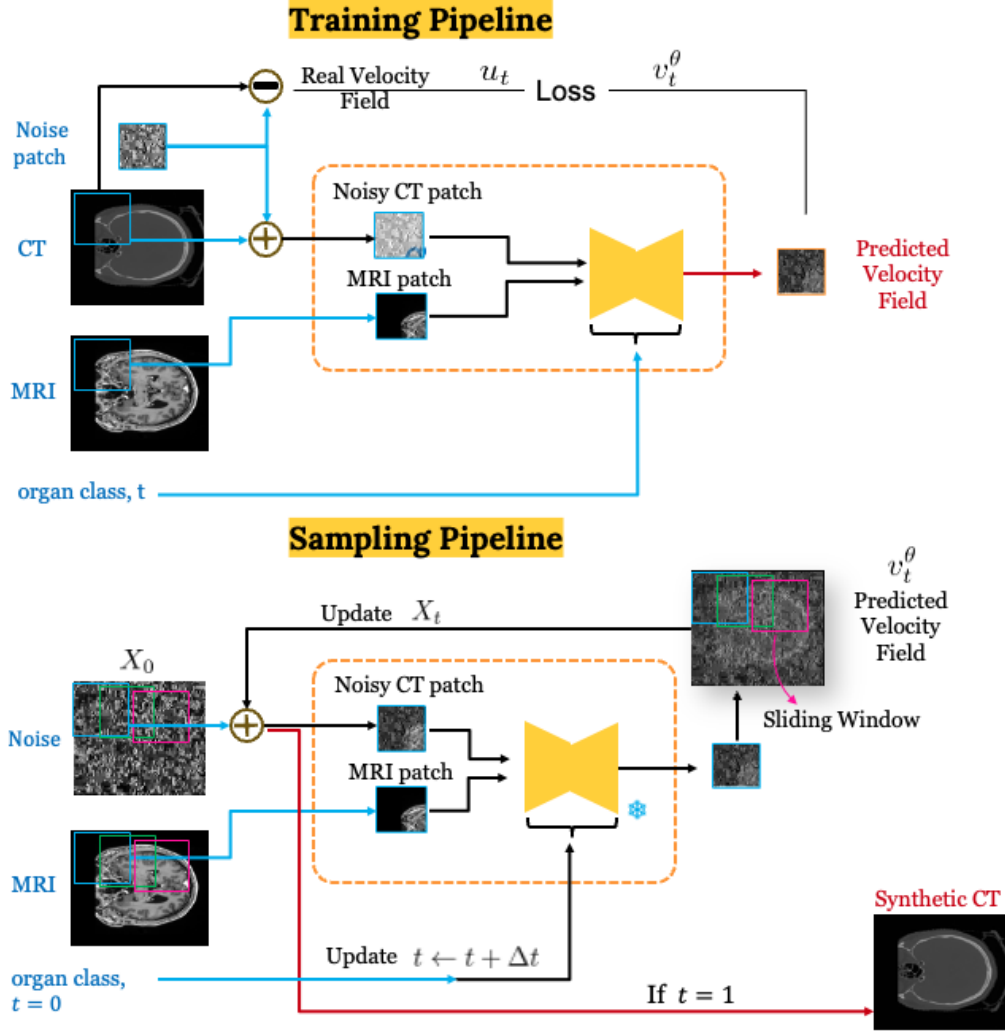


Figure 1: Overview of the proposed CRF framework for MRI-to-CT synthesis with *blue* labels representing inputs and *red* labels representing outputs. (Top) Training stage: A noisy CT patch and its paired MRI patch, along with time and organ-class conditioning, are provided to the model to predict the velocity field that transports the sample toward the ground-truth CT. (Bottom) Sampling stage: Starting from Gaussian noise, the model iteratively integrates the learned velocity field in a sliding-window manner with Gaussian blending, conditioned on the MRI and organ class, to generate the full sCT volume.

2. Related Work

2.1. Flow Matching and Rectified Flow Models

Flow-based models are an alternative approach that transports a simple distribution $X_0 \sim \pi_0$ (e.g., Gaussian noise) into a target data distribution $X_1 \sim \pi_1$ through the integration of an ordinary differential equation (ODE). Unlike diffusion models, which learn a stochastic reverse process, flow matching learns a deterministic time-dependent velocity field that enables few-step sampling with diffusion-level fidelity (Liu et al., 2022),

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t, \pi_0, \pi_1} \left\| v_t^\theta(X) - u_t(X_t | X_0, X_1) \right\|_2^2, \quad (1)$$

where $u_t(X_t | X_0, X_1)$ is the conditional vector field whose form depends on the chosen transport formulation.

A particularly effective variant of flow matching is rectified flow (Liu et al., 2022), simplifying X_t to a linear interpolation between data and noise,

$$X_t = tX_1 + (1 - t)X_0, \quad t \in [0, 1]. \quad (2)$$

Differentiating Equation (2) with respect to time gives the velocity field u_t in Equation (1), which reduce to a straight-line velocity between noise and data,

$$u_t(X_t | X_0, X_1) = X_1 - X_0, \quad (3)$$

enforcing a linear trajectory. This simplification yields a conceptually simple and intuitive transport process, improves inference stability, and enables high-quality sampling in only a few integration steps. Medical imaging applications such as MOTFM (Yazdani et al., 2025) demonstrate high fidelity and substantially faster inference compared to diffusion models.

Motivated by these strengths, we adopt the CRF framework for generating sCT, by conditioning the rectified flow model on MRI and organ label (Rombach et al., 2022; Yazdani et al., 2025).

2.2. Conditional Rectified Flow Formulation

To adapt rectified flows for MRI-to-CT synthesis, the framework is extended to a conditional setting by incorporating MRI volumes and auxiliary structural information like conditioning variables. This enables the model to learn anatomy-aware velocity fields that guide the transport from the noise distribution to the target CT space.

Let $X_0 \sim \pi_0$ be a Gaussian noise sample and $X_1 \sim \pi_1$ be a ground-truth CT paired with MRI I and class context c , the rectified flow aims to learn a velocity field v_θ such that:

$$\frac{dX_t}{dt} = v_t^\theta(X_t, I, c), \quad t \in [0, 1]. \quad (4)$$

The training minimizes the deviation from the straight-line velocity:

$$\begin{aligned} \mathcal{L}_{\text{CRF}} = & \lambda_{L_1} \mathbb{E}_{t, X_0, X_1} \left\| v_t^\theta(X_t, I, c) - (X_1 - X_0) \right\|_1 \\ & + \lambda_{L_2} \mathbb{E}_{t, X_0, X_1} \left\| v_t^\theta(X_t, I, c) - (X_1 - X_0) \right\|_2^2, \end{aligned} \quad (5)$$

where $X_t = (1 - t)X_0 + tX_1$ is the linear interpolation of X_0 and X_1 , and $\lambda_{L_1}, \lambda_{L_2} \in [0, 1]$ are weight constants. The L_1 term promotes sharp structural detail and stable gradients even under large prediction error. The L_2 term stabilizes optimization by penalizing large velocity errors and improving global HU accuracy.

By aligning the learned velocity field with the constant transport direction $X_1 - X_0$, the CRF framework produces high-quality sCT generation with far fewer integration steps than diffusion-based approaches.

3. Method

3.1. Conditional Rectified Flow Framework

Given an input MRI volume I and conditioning information c (e.g., organ class), the objective is to generate a synthetic CT \hat{X}_1 that matches the ground-truth CT X_1 . Our model learns a conditional velocity field v_t^θ that governs the rectified flow trajectory from a noise sample, $X_0 \sim \pi_0$, to the target CT distribution, $X_1 \sim \pi_1$.

During training, a time parameter t is uniformly sampled from $[0, 1]$ and the model minimizes Equation (5). At inference, the synthetic CT is obtained by integrating the learned velocity field along the flow path:

$$\hat{X}_1 = X_0 + \int_0^1 v_t^\theta(X_t, I, c) dt. \quad (6)$$

In practice, Equation (6) is approximated using a first-order (Euler) solver with K steps:

$$\hat{X}_1 = X_0 + \sum_{k=0}^{K-1} v_{t_k}^\theta(X_{t_k}, I, c) \Delta t, \quad \Delta t = \frac{1}{K}. \quad (7)$$

This deterministic ODE-based sampling enables high-quality CT synthesis with substantially fewer steps than diffusion-based methods, while MRI and organ conditioning preserve anatomical fidelity and consistency with the ground-truth CT.

3.2. Network Architecture

Our sCTFlow framework is inspired by MAISI, the 3D CT generation framework (Guo et al., 2025). While keeping the attention residual U-Net backbone (Rombach et al., 2022; Ni et al., 2019), we replace the noise scheduler with Equation (2), the model objective with Equation (5), and the sampling step with Equation (7). In Figure 1, MRI is used as an additional spatial feature by concatenating it with the noisy CT patch along the channel dimension (after spatial alignment), forming a two-channel input to the U-Net.

Training pipeline: As shown in Figure 1 (Top), sCTFlow receives CT–MRI patch pairs, the organ class, and the time step. Time and class embeddings are processed by a sinusoidal encoder and multilayer perceptron (MLP) and injected into each U-Net block as channel-wise biases. The CT patch is mixed with noise following Equation (2), and the model predicts a velocity field v_t^θ to minimize Equation (5). This conditioning mechanism enables the model to learn anatomy-aware velocity fields conditioned on both patient MRI and the type of organ structure. We train the model on small patches to avoid GPU memory limits.

Sampling pipeline: As shown in Figure 1 (Bottom), sCTFlow takes the sampling noise X_t , MRI conditioning patches, organ class, and time step t . Starting at $t = 0$, the model predicts the velocity field v_t^θ and updates $X_{t+\Delta t} = X_t + v_t^\theta \Delta t$. Following Equation (7), the sCTFlow sampling process iterates $K = 1/\Delta t$ times and returns synthetic CT output X_1 . To avoid GPU memory limits, the sampling process is performed patch-wise using a sliding-window approach from MONAI (Cardoso et al., 2022; Toraci et al., 2014; Wang et al., 2024). Overlapping regions between the patches are blended with Gaussian weighting to reduce seam artifacts and maintain anatomical continuity across the reconstructed volume. A larger overlap ratio produces smoother transitions but increases computational runtime.

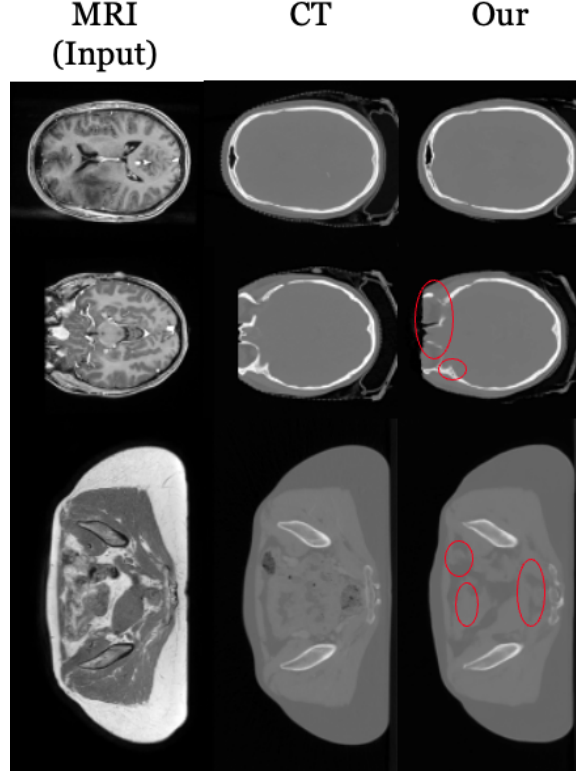


Figure 2: Qualitative sCT results. The model accurately reconstructs global anatomy and soft-tissue contrast. Red circles highlight remaining high-frequency errors, including slightly blurred cortical bone boundaries and missing or inaccurately reconstructed small air cavities (dark bubbles).

4. Experiments

sCT provides HU required for dose calculation in MRI-only radiotherapy, so the synthesis quality is evaluated using both image-based and dose-related metrics. Common image-based metrics include mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and struc-

tural similarity index (SSIM), which measure HU accuracy, signal fidelity, and structural preservation, respectively (Thummerer et al., 2023).

All experiments were conducted on the publicly available SynthRAD2023 Task 1 dataset, which contains paired 3D MRI-CT volumes for brain and pelvic regions (Thummerer et al., 2023). The official dataset split consists of 180 training, 30 validation, and 60 test cases. All volumes were resampled to 1 mm isotropic resolution, and CT intensities were clipped to the $[-1024, 3000]$ HU range. The data volumes vary in size from approximately 7×10^6 to 36×10^6 voxels.

Since ground-truth CT is not available for the official validation partition, we follow common practice during model development by randomly splitting the original training dataset into an 80:20 ratio for training and validation. The resulting validation subset is used for hyperparameter tuning and model comparison in our experiments.

Because our model operates in a patch-based manner, each volume is divided into $96 \times 96 \times 96$ voxel patches extracted from the full 3D images. Patch-based training substantially reduces memory requirements and enables effective processing of high-resolution volumetric data on hardware with limited GPU capacity. The model is optimized using AdamW with a learning rate of 10^{-4} . During inference, we employ sliding-window inference with $96 \times 96 \times 96$ patches and a 25% overlap. The rectified-flow ODE was solved using $K = 20$ Euler steps. The final model contains approximately 38M parameters and was trained for 200 epochs on an NVIDIA RTX 3090 GPU (24 GB), taking approximately 4.7 days of training time with sampling time of 1 to 2 minutes per image.

5. Results

We evaluated our approach on the SynthRAD2023 validation partition and compared image-based metrics against two of the top fifteen leaderboard methods: SMU-MedVision and Breizh-CT. SMU-MedVision method is a CNN-based approach that employs a hybrid 3D patch-based CNN-Transformer U-Net with multi-scale structure extraction and preservation (MSEP). Breizh-CT, in contrast, is a GAN-based method employing a pix2pix framework with a six-block residual (ResNet) generator and a PatchGAN discriminator (Huijben et al., 2024).

Our method achieves an MAE of 81.18 ± 19.48 HU, PSNR of 26.93, and SSIM of 0.830, as shown in Table 1. Qualitative results in Figure 2 show that our model successfully preserves global anatomical structures and produces coherent HU distributions across both brain and

Method	MAE ↓	PSNR ↑	SSIM ↑
SMU-MedVision	59.66 ± 13.21	29.42	0.885
Breizh-CT	94.43 ± 17.56	25.74	0.805
Ours	81.18 ± 19.48	26.93	0.830

Table 1: Comparison of image-based metrics against two of the top fifteen methods in SynthRAD2023 on the reconstructed volumes from the validation partition. ↑ and ↓ indicate whether higher or lower values are preferred.

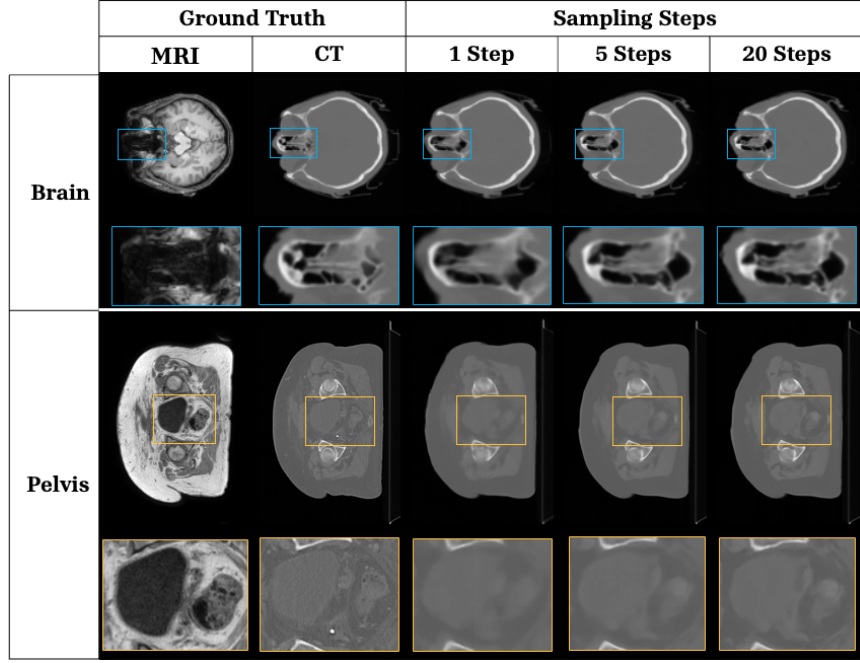


Figure 3: Effect of sampling time steps on sCT reconstruction quality. Increasing the number of steps yields sharper bone boundaries and improved fine-detail fidelity.

pelvic regions. Most residual errors occur in high-frequency areas, such as thin cortical bone interfaces or small air cavities, which may appear blurred or receive inaccurate HU estimates, indicating limitations in fine-detail reconstruction.

Figure 3 illustrates the impact of varying the number of ODE sampling steps during inference. Even with a very small number of steps, the model retains accurate global anatomical structure, and quantitative metrics (MAE, PSNR, SSIM) remain comparable across $K = 1, 5, 20$. However, visual inspection reveals that higher time-step counts yield sharper cortical bone boundaries and improved reconstruction of fine structures. Empirically, we found $K = 20$ to provide the best trade-off between computational efficiency and detail fidelity.

Although our image-based performance is slightly lower than that of the leading CNN-based approaches, our method outperforms the GAN-based model. These pixel-level metrics alone do not fully capture the strengths of our approach. Unlike methods that rely primarily on intensity mapping, our CRF model is designed to learn and utilize the underlying CT distribution, as demonstrated in the following robustness analysis.

5.1. Assessment of Learned CT Priors

To determine whether the proposed CRF model learns the underlying CT distribution rather than merely performing an intensity-based mapping from MRI to CT, we evaluate the model’s robustness when the MRI input is systematically degraded. If the model can

Table 2: Model’s performance on the validation dataset under different MRI distortions.

	Setting	MAE↓	PSNR↑	SSIM↑
No distortion	Baseline	67.70	29.36	0.827
Intensity discretization (quantization level)	7	69.82	29.07	0.821
	4	77.38	28.17	0.803
	2	245.67	20.34	0.592
Mixing random noise (noise ratio σ)	0.05	69.02	29.16	0.824
	0.15	93.34	27.30	0.792
	0.30	193.07	21.28	0.660
Low-pass filtering (retained frequency fraction)	0.3	72.57	28.68	0.815
	0.2	89.16	27.03	0.775
	0.1	191.69	24.49	0.685

reconstruct plausible CT volumes despite corrupted MRI inputs, this would indicate that it leverages a learned CT prior rather than relying solely on high-fidelity MRI appearance.

We apply three types of degradations: a) intensity discretization, which quantizes fine MRI values to a limited number of intensity levels; b) mixing random noise, which perturbs fine-grained intensity details; and c) hard cut-off low-pass filtering, which suppresses high-frequency anatomical structures. These perturbations remove fine structural cues and intensity information.

Table 2 and Figure 4 summarize the model’s performance under different MRI distortions. Quantizing the MRI to 7 or 4 intensity levels causes only mild degradation: MAE increases slightly and SSIM remains above 0.80. The sCT volumes still preserve global anatomy and realistic HU patterns. Even in the extreme case of 2 intensity levels, which removes most tissue contrast and sharply worsens MAE and SSIM, the sCTs retain coarse anatomical structure. This indicates that the model can infer plausible CT anatomy without relying heavily on detailed MRI intensities, suggesting the presence of a learned CT distribution that compensates when MRI information becomes insufficient.

Noise levels up to $\sigma = 0.15$ result in only moderate performance drops. Qualitative results show that the sCTs remain visually coherent with bone and soft-tissue morphology largely preserved. At higher noise ($\sigma = 0.3$), fine bone edges blur and small cavities become difficult to recover, but global anatomy remains intact. This further supports that the model does not rely heavily on high-quality MRI appearance.

Low-pass filtering produces a predictable loss of sharp boundaries. With moderate filtering (retained fraction = 0.3), performance remains near baseline and bone contours are still identifiable. More aggressive filtering removes high-frequency CT detail, reflected in declining PSNR and SSIM. Even with heavily suppressed MRI details, the model produces anatomically plausible CT volumes, indicating that it can infer missing high-frequency content from learned CT priors rather than relying solely on MRI-provided edges.

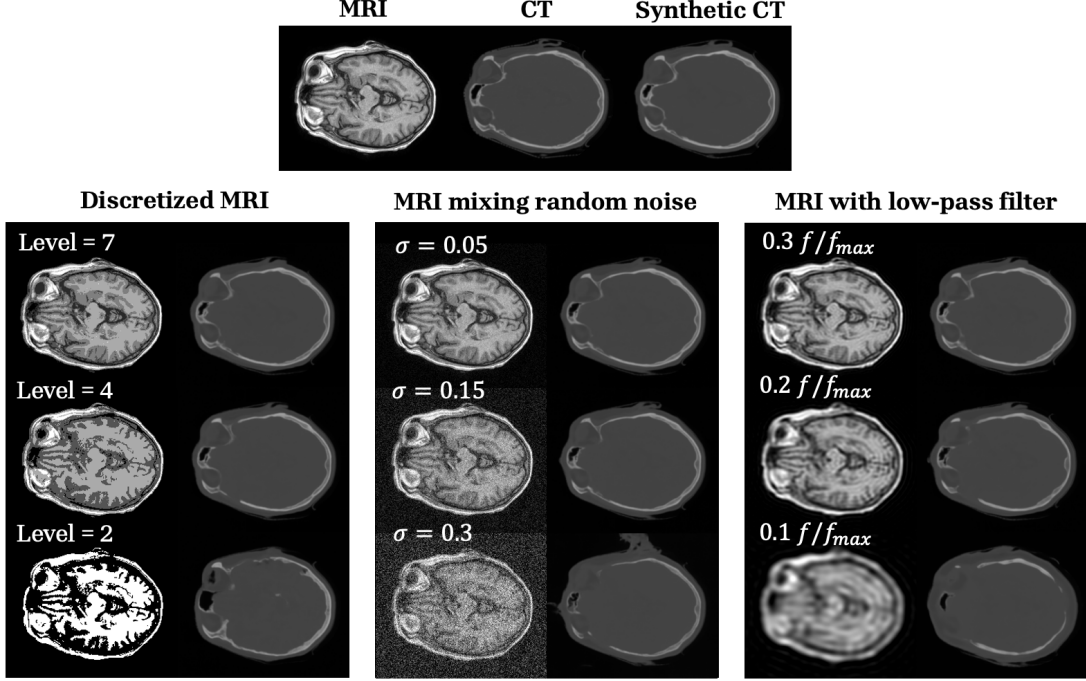


Figure 4: Effect of different MRI distortions on synthetic CT reconstruction: a) intensity discretization, b) mixing random noise, and c) low-pass filtering.

6. Conclusion

Developing a generative model for MR-only radiotherapy requires balancing anatomical fidelity, HU accuracy, and computational efficiency. In this work, we present sCTFlow, a Conditional Rectified Flow framework for 3D MRI-to-CT synthesis that addresses these challenges. sCTFlow leverages rectified flow modeling for volumetric medical image translation, enabling anatomy-aware CT generation with fewer sampling steps than the diffusion-based approaches. We demonstrate our results through extensive experiments on the SynthRAD2023 dataset. Our method produces accurate synthetic CTs that preserve structural detail, maintain realistic HU distributions, and remain robust under degraded MRI conditions. These results demonstrate the model’s ability to learn underlying CT characteristics and support its potential for reliable integration into MR-only radiotherapy workflows.

7. Acknowledgements

This work was supported by the APEX AI-HPC Project, CMKL University, Bangkok, Thailand, and the Center of Natural and Artificial Intelligence (CNAI).

References

Mohamed A Bahloul, Saima Jabeen, Sara Benoumhani, Habib Abdulmohsen Alsaleh, Zehor Belkhatir, and Areej Al-Wabil. Advancements in synthetic ct generation from mri: A

- review of techniques, and trends in radiation therapy planning. *Journal of Applied Clinical Medical Physics*, 25(11):e14499, 2024.
- David Bird, Michael G Nix, Hazel McCallum, Mark Teo, Alexandra Gilbert, Nathalie Casanova, Rachel Cooper, David L Buckley, David Sebag-Montefiore, Richard Speight, et al. Multicentre, deep learning, synthetic-ct generation for ano-rectal mr-only radiotherapy treatment planning. *Radiotherapy and Oncology*, 156:23–28, 2021.
- Marion Boulanger, Jean-Claude Nunes, Hilda Chourak, Axel Largent, Safaa Tahri, Oscar Acosta, R De Crevoisier, Caroline Lafond, and Anais Barateau. Deep learning methods to generate synthetic ct from mri in radiotherapy: a literature review. *Physica Medica*, 89:265–281, 2021.
- M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*, 2022.
- Jens M Edmund and Tufve Nyholm. A review of substitute ct generation for mri-only radiation therapy. *Radiation Oncology*, 12(1):28, 2017.
- Pengfei Guo, Can Zhao, Dong Yang, Ziyue Xu, Vishwesh Nath, Yucheng Tang, Benjamin Simon, Mason Belue, Stephanie Harmon, Baris Turkbey, et al. Maisi: Medical ai for synthetic imaging. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4430–4441. IEEE, 2025.
- Michaela Hoesl, Nuria E Corral, and Nilesh Mistry. White paper: Mr-based synthetic ct reimaged—an ai-based algorithm for continuous hounsfield units in the pelvis and brain—with syngo. via rt image suite (vb60), 2023.
- Evi MC Huijben, Maarten L Terpstra, Suraj Pai, Adrian Thummerer, Peter Koopmans, Many Afonso, Maureen Van Eijnatten, Oliver Gurney-Champion, Zeli Chen, Yiwen Zhang, et al. Generating synthetic computed tomography for radiotherapy: Synthrad2023 challenge report. *Medical image analysis*, 97:103276, 2024.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Led-sam, Klaus Maier-Hein, SM Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. *Advances in neural information processing systems*, 31, 2018.
- Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.
- Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky TQ Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024.

- Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.
- Helen A McNair, Kevin N Franks, and Marcel van Herk. On target 2: updated guidance for image-guided radiotherapy. *Clinical Oncology*, 34(3):187–188, 2022.
- Peter Metcalfe, GP Liney, Lois Holloway, A Walker, M Barton, GP Delaney, S Vinod, and W Tome. The potential for an enhanced role for mri in radiation-therapy treatment planning. *Technology in cancer research & treatment*, 12(5):429–446, 2013.
- Miguel Monteiro, Loïc Le Folgoc, Daniel Coelho de Castro, Nick Pawlowski, Bernardo Marques, Konstantinos Kamnitsas, Mark Van der Wilk, and Ben Glocker. Stochastic segmentation networks: Modelling spatially correlated aleatoric uncertainty. *Advances in neural information processing systems*, 33:12756–12767, 2020.
- Zhen-Liang Ni, Gui-Bin Bian, Xiao-Hu Zhou, Zeng-Guang Hou, Xiao-Liang Xie, Chen Wang, Yan-Jie Zhou, Rui-Qi Li, and Zhen Li. Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments. In *International Conference on Neural Information Processing*, pages 139–149. Springer, 2019.
- Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International conference on medical image computing and computer-assisted intervention*, pages 417–425. Springer, 2017.
- Shaoyan Pan, Elham Abouei, Jacob Wynne, Chih-Wei Chang, Tonghe Wang, Richard LJ Qiu, Yuheng Li, Junbo Peng, Justin Roper, Pretesh Patel, et al. Synthetic ct generation from mri using 3d transformer-based denoising diffusion model. *Medical Physics*, 51(4): 2538–2548, 2024.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- Maria Francesca Spadea, Matteo Maspero, Paolo Zaffino, and Joao Seco. Deep learning based synthetic-ct generation in radiotherapy and pet: a review. *Medical physics*, 48(11): 6537–6566, 2021.
- Blanche Texier, Cédric Hemon, Pauline Lekieffre, Emma Collot, Safaa Tahri, Hilda Chourak, Jason Dowling, Peter Greer, Igor Bessieres, Oscar Acosta, et al. Computed tomography synthesis from magnetic resonance imaging using cycle generative adversarial networks with multicenter learning. *Physics and Imaging in Radiation Oncology*, 28: 100511, 2023.
- Adrian Thummerer, Erik Van der Bijl, Arthur Galapon Jr, Joost JC Verhoeff, Johannes A Langendijk, Stefan Both, Cornelis (Nico) AT van den Berg, and Matteo Maspero. Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy. *Medical physics*, 50(7):4664–4674, 2023.

- Cristian Toraci, Gabriele Zaccaria, Stefano Ceriani, David Wilson, Marco Fato, and Michele Piana. Sliding window and compressive sensing for low-field dynamic magnetic resonance imaging. *arXiv preprint arXiv:1402.2453*, 2014.
- Haoshen Wang, Zhentao Liu, Kaicong Sun, Xiaodong Wang, Dinggang Shen, and Zhiming Cui. 3d meddiffusion: A 3d medical diffusion model for controllable and high-quality medical image generation. *arXiv preprint arXiv:2412.13059*, 2024.
- Milad Yazdani, Yasamin Medghalchi, Pooria Ashrafi, Ilker Hacihaliloglu, and Dena Shahriari. Flow matching for medical image synthesis: Bridging the gap between speed and quality. *arXiv preprint arXiv:2503.00266*, 2025.
- Li Zhou, Sen Bai, Yibao Zhang, Xin Ming, Ying Zhang, and Jun Deng. Imaging dose, cancer risk and cost analysis in image-guided radiotherapy of cancers. *Scientific reports*, 8(1):10076, 2018.