



Published in final edited form as:

Med Phys. 2024 April ; 51(4): 2538–2548. doi:10.1002/mp.16847.

Synthetic CT generation from MRI using 3D Transformer-based Denoising Diffusion Model

Shaoyan Pan^{1,2}, Elham Abouei¹, Jacob Wynne¹, Chih-Wei Chang¹, Tonghe Wang³, Richard L.J. Qiu¹, Yuheng Li¹, Junbo Peng¹, Justin Roper¹, Pretesh Patel¹, David S. Yu¹, Hui Mao⁴, Xiaofeng Yang^{1,2}

¹Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA

²Department of Biomedical Informatics, Emory University, Atlanta, GA 30322, USA

³Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

⁴Department of Radiology and Imaging Sciences and Winship Cancer Institute, Atlanta, GA 30308

Abstract

Background and Purpose: Magnetic resonance imaging (MRI)-based synthetic computed tomography (sCT) simplifies radiation therapy treatment planning by eliminating the need for CT simulation and error-prone image registration, ultimately reducing patient radiation dose and setup uncertainty. In this work, we propose a MRI-to-CT transformer-based improved denoising diffusion probabilistic model (MC-IDDPM) to translate MRI into high-quality sCT to facilitate radiation treatment planning.

Methods: MC-IDDPM implements diffusion processes with a shifted-window transformer network to generate sCT from MRI. The proposed model consists of two processes: a forward process, which involves adding Gaussian noise to real CT scans to create noisy images, and a reverse process, in which a shifted-window transformer V-net (Swin-Vnet) denoises the noisy CT scans conditioned on the MRI from the same patient to produce noise-free CT scans. With an optimally trained Swin-Vnet, the reverse diffusion process was used to generate noise-free sCT scans matching MRI anatomy. We evaluated the proposed method by generating sCT from MRI on an institutional brain dataset and an institutional prostate dataset. Quantitative evaluations were conducted using several metrics, including Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), Multi-scale Structure Similarity Index (SSIM), and Normalized Cross Correlation (NCC). Dosimetry analyses were also performed, including comparisons of mean dose and target dose coverages for 95% and 99%.

Results: MC-IDDPM generated brain sCTs with state-of-the-art quantitative results with MAE 48.825 ± 21.491 HU, PSNR 26.491 ± 2.814 dB, SSIM 0.947 ± 0.032 , and NCC 0.976 ± 0.019 . For the

Corresponding author: Xiaofeng Yang, PhD, DABR, xiaofeng.yang@emory.edu, 1365 Clifton Road, NE Suite A1320, Atlanta, GA 30322.

Conflict of interest: The authors have no conflict of interests to disclose.

prostate dataset: MAE 55.124 ± 9.414 HU, PSNR 28.708 ± 2.112 dB, SSIM 0.878 ± 0.040 , and NCC 0.940 ± 0.039 . MC-IDDPM demonstrates a statistically significant improvement (with p -values < 0.05) in most metrics when compared to competing networks, for both brain and prostate synthetic CT. Dosimetry analyses indicated that the target dose coverage differences by using CT and sCT were within $\pm 0.34\%$.

Conclusions: We have developed and validated a novel approach for generating CT images from routine MRIs using a transformer-based improved DDPM. This model effectively captures the complex relationship between CT and MRI images, allowing for robust and high-quality synthetic CT images to be generated in a matter of minutes. This approach has the potential to greatly simplify the treatment planning process for radiation therapy by eliminating the need for additional CT scans, reducing the amount of time patients spend in treatment planning, and enhancing the accuracy of treatment delivery.

1. Introduction

Magnetic resonance imaging (MRI) and computed tomography (CT) are vital tools in medical diagnostics and radiation therapy, each offering unique advantages for treatment planning. While MRI delivers exceptional soft tissue contrast, enhancing the delineation of organ risks and offering superior contrast near many tumor targets, particularly in the pelvis and brain areas^{1–3}, CT offers precise geometrical accuracy vital for dose calculations in treatment planning^{4–6}. The potential to generate sCT images from MRI has thus emerged as a promising avenue to combine the strengths of both modalities, improving radiotherapy treatment planning by not only bypassing the uncertainties of MRI and CT co-registration but also reducing radiation exposure, lowering costs, and enhancing patient comfort. This initiative to utilize MRI-based sCT as a stand-in for CT is currently a significant focus of attention in the radiotherapy community.

Artificial intelligence and machine learning techniques have advanced rapidly in recent years. Deep learning-based image synthesis methods rely on complex, nonlinear, and trainable mapping methods, such as convolutional neural networks (CNNs)⁷ and generative adversarial networks (GANs)⁸. GAN-based approaches have been particularly successful in generating realistic CT images from MRI. Lei *et al.*⁹ developed sCTs from MRIs based on dense cycle GAN model to effectively capture the relationship between the CT and MRIs. Their method generated robust, high-quality synthetic CT (sCT) in minutes on brain and pelvic datasets. Zhao *et al.*¹⁰ proposed an approach using a hybrid CNN and transformer architecture as a generator in the GAN framework. Their results show that the proposed method can generate more accurate CT images from pelvic MRI and is more robust against local mismatch between MR and CT images. Wolterink *et al.*¹¹ can use GAN to generate sCT volumes with unpaired brain MR and CT images. Many other GAN-based studies^{12–15} have successfully demonstrated the capacity to create visually appealing sCTs from MRIs. However, these methods typically experience unstable training and output homogeneity issues, a consequence of the adversarial strategy employed in GAN training, as documented in studies across natural and medical imaging spheres^{16–18}. Accordingly, they demand substantial time to fine-tune the network and adjust training hyper-parameters, seeking a harmonious balance between their discriminators and generators¹⁷. This painstaking tuning

process recurs with each introduction of new training MRI-CT paired data to an existing model, a factor that largely impinges on its practical applicability in real-world clinical settings.

As an alternative to GAN, diffusion and score matching models are generative approaches inspired by non-equilibrium thermodynamics in physics. They define a Markov chain of diffusion steps to gradually add random noise to data and then learn to reverse the diffusion process to generate samples from the noise^{19–23}. Diffusion models utilize a neural network (typically a U-shape CNN) to learn denoising. As opposed to GANs, diffusion models are not reliant upon adversarial training methods²⁴. This improves training stability and results in more authentic output images with higher quality and greater semantic diversity without the need of overwhelmingly hyper-parameter fine-tuning. Several diffusion-based generative models^{25–28} have been proposed for medical image synthesis, including generating sCT from MRI, and demonstrate state-of-the-art image quality superior to CNN-based and GAN-based methods. However, a common drawback plaguing most of these diffusion models is their low efficiency; the substantial amount of time they necessitate to synthesize 2D images (e.g., about 1000 times slower than GAN-based methods) even significantly escalates when tasked with generating 3D volumes. This limitation not only extends the production time but also constrains their practical applicability in real-time medical settings.

In contrast to the diffusion-based methods depicted in prior studies including the one denoted as²⁷, our research introduces a pioneering approach leveraged on the 3D MRI-to-CT improved denoising diffusion probabilistic model (MC-IDDPM). The novelty of our endeavor hinges on two pivotal advancements: Firstly, while conventional diffusion-based models using in MRI-to-CT synthesis²⁷ principally focus on predicting the mean of the reverse Gaussian process, the proposed MC-IDDPM, motivated by improved DDPM²², forecasts both the mean and the variance, thereby substantially reducing the necessary timesteps by about 20 times for MRI-CT synthesis. Accordingly, the proposed MC-IDDPM improves efficiency and makes the diffusion-based 3D volume generation more practical in terms of generation time. Secondly, we integrate the Swin-Vnet, a shifted-window Vision Transformer-based network, in lieu of the frequently adopted Unets to study the reverse process. This innovation manifests a superior efficacy in medical image processing, amplifying the accuracy in predicting both the mean and variance in the reverse process and thus significantly enhancing the final synthesis accuracy. Our work is the inaugural effort to apply conditional IDDPM to the translation of 3D medical volumes, thereby introducing a diffusion model solution for 3D image translation. This marks a departure from earlier works which either employed 2D diffusion models to generate volumes slice-by-slice losing the spatial information among the axial slices or reserved the use of IDDPM exclusively for non-conditional image creation²⁶. Consequently, our MC-IDDPM exhibits a remarkable competency in generating high-quality sCT images with reduced artifacts from MRI images, standing as a testimony to its pioneering architecture and process enhancements. These innovations mark a significant stride in MRI-based sCT technology, paving the way for more robust, accurate, and efficient radiation therapy planning, thereby unveiling a new horizon in the medical imaging landscape. In addition, the advantages and disadvantages are summarized in the Appendix.

2. Method

The proposed MC-IDDPM translates a patient's MRI into a sCT image. As shown in the Fig. 1, a diffusion-based process converts a three-dimensional isotropic Gaussian noise $T \sim \mathcal{N}(0, 1)$ into a sCT image conditioning on the corresponding MR. The diffusion process relies on a significant assumption: by adding a small amount of noise ϵ to a real CT scan X over n timesteps, we can transform X into a purely Gaussian noise sample T , with n being sufficiently large. Consequently, the noise-free image X can be retrieved by eliminating the added noise from the noise sample T , which can also be generalized to any noise sample. Consequently, the diffusion process consists of two processes: a forward process that adds Gaussian noise to the CT image, and a reverse process learned by the proposed neural network, Swin-Vnet, that removes the Gaussian noise from the noisy CT images. The architecture of Swin-Vnet is described in Appendix. A, and the visualizations are shown in Fig. A1 and A2. Furthermore, the mathematical formulation of the diffusion process is introduced here.

2.A Diffusion process

Following ²⁰, we present a three-stage formulation for the proposed diffusion model: first, a forward diffusion process is executed, wherein small amounts of Gaussian noise T are gradually applied to a given CT image X_0 over N timesteps, to gradually transform the CT scan into pure Gaussian noise X_N . Next, the proposed MC-IDDPM network is trained to learn a reverse diffusion process, conditioned on the MRI Z , which effectively removes the small amounts of noise overlayed at each timestep, and subsequently denoises X_t back to the original image X_0 . With an optimal MC-IDDPM denoiser, we can recursively remove Gaussian noise to obtain the noise-free CT image paired to the input MRI.

2.A.1 Forward diffusion—In the forward diffusion process, the noisy image at timestep n is defined to depend solely on the noisy image at timestep $n-1$. Specifically, we define the noisy image generation process q as a Markov process where the transition probability from image X_{n-1} to n follows a Gaussian distribution $\mathcal{N}$:

$$q(X_n|X_{n-1}) = \mathcal{N}(X_n; \sqrt{1 - \beta_n}X_{n-1}, \beta_n I) \quad [6]$$

where β_n is a pre-determined variance at timestep n . Practically, we are able to efficiently represent noisy images at any arbitrary timestep n using reparameterization ²⁰:

$$X_n = \sqrt{\prod_{i=1}^n (1 - \beta_i)} X_0 + \sqrt{1 - \prod_{i=1}^n (1 - \beta_i)} \epsilon_n \quad [7]$$

where $\epsilon_n \sim \mathcal{N}(0, I)$ is noise sampled from a normal distribution. The selection of the maximum timestep N is 4000 so the β_n is $6e^{-6}t$ for all experiments.

2.A.2 Reverse diffusion—In the reverse diffusion process, we calculate the inverse Gaussian distribution $p(X_{t-1}|X_t) = \mathcal{N}(X_{t-1}; \mu, \Sigma)$ so we can recursively move in a reverse direction from X_T to X_0 . Following Ho et al. and Song et al.'s works²³, the inverse probability can be calculated in a closed form only when the clean image X_0 is already known, which is impractical. Accordingly, we propose a neural network θ to approximate the incompatible $p(X_{n-1}|X_n)$:

$$p_\theta(X_{n-1}|X_n, Z) = \mathcal{N}(X_{n-1}; \mu_\theta(X_n, n|Z), \Sigma_\theta(X_n, n|Z)) \quad [8]$$

where Z is the conditioning MRI, μ_θ is an estimated mean matrix and Σ_θ is an estimated variance matrix of the inverse Gaussian distribution. Using reparameterization, the less noisy image X_{n-1} can be calculated as:

$$X_{n-1} = \mu_\theta(X_n, n|Z) + \sigma_\theta(X_n, n|Z)\epsilon \quad [9]$$

where $\epsilon \sim \mathcal{N}(0, I)$ is a noise sampled from a normal distribution, σ_θ is the standard deviation of the inverse distribution. In conclusion, we apply a transformer-based network (MC-IDDPN network) taking inputs of a noisy image X_n , timestep n , and MR scan Z from the same patient, to estimate the mean and standard deviation of the inverse distribution. We thus generate the less noisy image X_{n-1} without foreknowledge of the clean image X_0 .

Accordingly, the proposed network is optimized to estimate the mean and variance. For the mean μ_θ , we firstly optimize the network to predict the noise ϵ_θ ²⁰:

$$\operatorname{argmin}_{\epsilon_\theta} L = \operatorname{argmin}_{\epsilon_\theta} MAE(\epsilon_n, \epsilon_\theta(X_n, n|Z)) \quad [10]$$

where MAE is the mean absolute error. Then the mean μ_θ can be calculated using the estimated noise ϵ_θ :

$$\mu_\theta(X_n, n|Z) = \frac{1}{\sqrt{1-\beta_n}} \left(X_n - \frac{\beta_n}{\sqrt{1-\sum_{i=1}^n 1-\beta_i}} \epsilon_\theta(X_n, n|Z) \right) \quad [11]$$

For estimating the variance, we employ Prafulla et al.'s work²¹ to first train the network to estimate an interpolation variance coefficient k_n which can be used for the variance matrix:

$$\Sigma_\theta(X_n, n|Z) = \exp \left(k_n(X_n, n|Z) \log \beta_n + (1 - k_n(X_n, n|Z)) \log \left(\frac{1 - \prod_{i=1}^{t-1} (1 - \beta_i)}{1 - \prod_{i=1}^t (1 - \beta_i)} \beta_t \right) \right) \quad [12]$$

where $k_\theta(X_n, n|Z)$ is the only unknown parameter. To obtain the coefficient k_θ , the network is optimized by a variational lower bound loss L_{VLB} ²²:

$$\underset{k_\theta}{\operatorname{argmin}} L_{var} = \underset{k_\theta}{\operatorname{argmin}} L_{VLB}(\Sigma_n, \Sigma_\theta(X_n, n|Z)) \quad [13]$$

More details of the implementation are shown in the Appendix. B. The overall optimization function is presented as:

$$L = L_{mean} + \gamma L_{var} \quad [14]$$

where γ is a weighting parameter which is empirically selected as the ratio between the inference diffusion step to the training diffusion step, which is 0.0125. More implementation details are shown in Append. B.

2.A.3 Generated sCT volume—Once the network is optimized, we can substitute the estimated μ_θ and Σ_θ into Eq. 9 to denoise X_{n-1} from X_n . In the generation stage, we evenly spaced steps between 1 and 4000 by 50 timesteps and denote the new number set as the inference timestep set $s \in [S_1, S_2, \dots, S_{50}]$. With the resampled timestep, we recursively denoise X_s until we obtain the final noise-free CT image X_0 , a synthetic image anatomically matching the input MRI. Notice that the noise ϵ introduces randomness into the generation process, we generate the final sCT scan in a manner of Monte Carlo-based (MC-based) generation: we run the generation process 5 times for each patient and took the averaged result as the final sCT scans.

3. Data Acquisition and Preprocessing

We utilized a dataset for our MRI-to-sCT translation, which consists of MR-CT scan pairs from 36 patients with brain imaging and a dataset with prostate imaging collected from 28 patients. The collection details, along with the code-implementation details, and shown in the Appendix. C. We centered, removed the air background, and central-cropped the slices to highlight the Cerebrospinal fluid region. The final MR-CT scan pairs have a voxel size of 192x192x96. For training and inference, we used the same patch-based input approach and sliding window prediction technique as with the institutional prostate dataset. The patch size chosen for the brain dataset was 64x64x4. Data augmentation and normalization techniques used for the institutional prostate dataset were also applied to this brain dataset. We trained our model using the first 75% patients in the dataset, while the remaining 20% patients were used for testing and 5% patients were reserved for validation. Prior to training and inference, the voxel intensities of CT scans were cut to $[-1024, 1650]$, and jointly normalized to the interval $[-1, 1]$. The voxel intensities of MRIs were independently normalized to $[-1, 1]$. All networks were trained using an AdamW optimizer with an initial learning rate of 3e-5 and weight decay of 1e-5 across 500 epochs. In this experiment, the number of training diffusion timesteps is 1000, and the number of generating timesteps is 50.

4. Implementation detail and performance evaluation

4.A sCT evaluation

To measure the quality of the sCT from the proposed MC-IDDPM, we compare the sCT and the real CT. We evaluate the mean absolute error (MAE), peak signal-to-noise ratio (PSNR), multi-scale structure similarity index (MS-SSIM) with evaluation scale of 5 and normalized cross correlation (NCC) indices to quantify the absolute difference, peak signal similarity, image overall visual similarity, and image correlation, respectively. Greater PSNR, MS-SSIM and NCC indicate better quality of the sCT. Final performance was reported after evaluating all sCT scans. We compared MC-IDDPM's performance to state-of-the-art methods, including the MRI-to-CT pixel-to-pixel generative adversarial network (MR-GAN)⁹, MRI-to-CT cycle-consistent generative adversarial network (MR-CGAN)⁹, 2D improved DDPM (2D-IDDPM)²², 3D-DDIM¹⁹, and 3D DDPM (3D-DDPM)²¹. Except for the 2D-IDDPM, which is based on a 2D network, all the other competing methods utilize 3D networks. For mitigating the selection bias, the comparisons are implemented in a manner of 4-fold cross-validation. Pair-wise comparisons between MC-IDDPM and competing networks were made using Mann Whitney U test with $\alpha = 0.05$.

In order to comprehensively assess the performance of our proposed method, we conducted three ablation studies using a dataset of institutional prostate images with the last-fold testing data. The first study aimed to investigate the impact of using different state-of-the-art deep learning neural networks on the quality of the generated sCT images. Specifically, we compared the performance of three different networks - V-net²⁹, Swin VIT (proposed), and token-based multi-layer linear Mixer Vnet (MLP-Vnet)³⁰ - in terms of various evaluation metrics for the diffusion model. More specifically, we used the convolutional block (shown in Fig. A1) to replace all the Swin-attention blocks to build the Vnet; and used the token-based MLP-Mixer block³⁰ to replace all the W-SA and SW-SA modules to build the MLP-Vnet. The second study focused on the effect of the training diffusion timesteps on the quality of the generated sCT images. We tested the performance of MC-IDDPM using maximum timestep values of $N = 250, 500, 1000$ (proposed), and 2000. In the third study, we evaluated the relationship between the number of axial slices in the input scan and the resulting number of generated diffusions timesteps. We reported the MAE and SSIM for different settings, including slices = 4 (proposed), 8, 12, and generating timestep values of $S = 50$ (proposed), 100, 256. Finally, we reported the generating efficiency of our proposed method by reporting the generating time for each patient from one run for each setting.

Additionally, we executed an in-depth dosimetric analysis that compares the sCT and planning CT volumes, with a particular focus on a patient from the brain dataset, with four Planning Target Volumes (PTVs). Specifically, we examined the mean dose and target coverage of D95% and D99% in PTV for both the CT and sCT volumes. Comprehensive discussions on these findings will be further elaborated in the discussion section.

5. Results

The sCTs generated using the brain and prostate datasets are displayed in Fig. 2 and 3. More visualizations are shown in the Appendix. E. To quantitatively evaluate the performance

of the MC-IDDPM sCT synthesis from MRI, we present the quantitative and statistical comparison between MC-IDDPM and other state-of-the-art methods.

5.A: Quantitative result

In Table. 1, compared to all other methods, MC-IDDPM has the lowest MAE (48.825 ± 21.491 HU), indicating that it has the smallest difference between the sCT and CT images. Additionally, MC-IDDPM has the highest PSNR (26.491 ± 2.814 dB), highest SSIM (0.947 ± 0.032) and NCC (0.976 ± 0.019) among all techniques, indicating that it preserves the image structure and details better while providing the highest peak signal similarity, image visual similarity and overall image correlation. On the other hand, MC-IDDPM achieves statistically significant improvements over all metrics when compared with MC-GAN, MC-CGAN, 3D-DDIM, and 3D-DDPM. MC-IDDPM also demonstrates significant improvements ($p < 0.05$) when compared to 2D-IDDPM in terms of MAE, PSNR and NCC, but not SSIM.

In Table. 1, compared to all other methods, MC-IDDPM has the lowest MAE (55.124 ± 9.414 HU), indicating that it produces prostate sCT images that have the smallest absolute difference relative to the CT images. Additionally, it has the highest PSNR (28.708 ± 2.112 dB), SSIM (0.878 ± 0.040) and NCC (0.940 ± 0.039), indicating that the generated sCT images are closer to the peak signal ratio and visual appearance to ground truth, and have a higher correlation with the reference images. Similar to the brain sCT image evaluation, 2D-IDDPM outperforms GAN-based methods in terms of quantitative results, while 3D-DDPM shows worse results. MC-IDDPM demonstrates significant improvements over all metrics ($p < 0.05$) when compared to GAN-based methods. Compared to 2D-IDDPM and 3D-DDIM, MC-IDDPM shows significant improvement ($p < 0.05$) in MAE. Furthermore, compared to the 3D-DDPM, the MC-IDDPM shows a statistically significant improvement ($p\text{-value} < 0.05$) across all metrics.

5.B: Hyperparameters study of the diffusion process and network settings

The quantitative results for the hyperparameters study in detail are presented in the Appendix. D. The Swin-Vnet used in MC-IDDPM quantitatively outperforms Vnet and MLP-Vnet models (Table. D1) with MAE (59.953 ± 12.462 HU), PSNR (26.92 ± 2.429 dB), and NCC (0.948 ± 0.018). Swin-Vnet also demonstrates a statistically significant improvement ($p < 0.05$) over Vnet in terms of MAE, while there was no difference in terms of other metrics. Additionally, there was no difference between Swin-Vnet and MLP-Vnet ($p > 0.05$). Furthermore, Swin-Vnet has a lower generation time than Vnet, but a higher generation time than the MLP-Vnet. The results may indicate the effectiveness of the global information technique (Swin-attention in the Swin-Vnet and the MLP-Mixer in the MLP-Vnet) in reducing the MAE in the DDPM-based sCT synthesis.

In addition, we evaluated the impact of training timesteps on MC-DDPM performance (Table. D2), with 1000 timesteps producing second-best results for MAE (59.953 ± 12.462 HU), PSNR (26.92 ± 2.429 dB), and NCC (0.948 ± 0.018), and the best result for SSIM (0.844 ± 0.043). MC-DDPM with 1000 training timesteps demonstrated a significant

improvement ($p < 0.05$) over MC-DDPM with 250 and 500 timesteps, but no improvement over MC-DDPM with 1000 and 2000 training timesteps.

Finally, we examined the impact of input size on MC-DDPM performance (Table. D3). When taking input of the Gaussian noise and MR scan with 4 axial slices (size of $128 \times 128 \times 4$), MC-DDPM performed well using generation timesteps of 50, 100, and 200. When taking input with 8 slices (input size of $128 \times 128 \times 8$), MC-DDPM required 100 or 200 generation timesteps to achieve performance close to the proposed setting combination (axial slices of 4 and generation timestep of 50). With 16 slices (input size of $128 \times 128 \times 16$), MC-DDPM demonstrated increasing performance with increasing generation timestep. However, for inputs with 16 axial slices, MC-DDPM requires more generation timesteps for optimal performance, which could reduce efficiency due to the approximately proportional relationship between generation time and generation timesteps.

5.C: Dosimetric analysis

On the other hand, our dosimetric analysis (quantitative table shown in Appendix. F) reveals a remarkable consistency between sCT and planning CT, particularly when evaluating mean doses across four PTVs. Table. F1 indicates that the dose differences are within $\pm 0.34\%$. Fig. 4 shows good agreements of dose calculation by using CT and pCT for different targets. Notably, the Synthetic CT not only excels in quantitative image-based metric performance but also demonstrates robust dose-based performance.

6. Discussion

Our method introduces MC-DDPM, a novel algorithm for converting MRI to synthetic CT (sCT) images, crucial in radiation therapy planning. MC-DDPM combines diffusion models and Swin-transformer neural networks, potentially eliminating the need for actual CT scans. This advancement could reduce radiation exposure, costs, and improve patient comfort. MC-DDPM comprises two phases: forward and reverse diffusion. In the forward diffusion, real CTs are progressively noised through Gaussian noise addition, creating increasingly noisy CT sequences. These are inputs for the reverse phase, where a Swin-Vnet network denoises them iteratively. This reverse diffusion interprets denoising as predicting a Gaussian Markov process's mean and variance. The Swin-Vnet, trained on noisy CT scans conditioned on MR scans, estimates noise and predicts variance interpolation coefficients. By effectively transforming Gaussian noise into sCT scans corresponding to MR scans, our Swin-Vnet optimized process showcases a pioneering approach in sCT synthesis and MRI-only radiotherapy. This is the first use of a 3D Swin-transformer-based network in a diffusion model to elevate image synthesis quality.

In the institutional brain and prostate dataset, the MC-DDPM achieves state-of-the-art results: 1) In the brain dataset, by average among all the testing patients, the methods can generate sCT achieving MAE 48.825 ± 21.491 HU, PSNR 26.491 ± 2.814 dB, SSIM 0.947 ± 0.032 , and NCC 0.976 ± 0.019 . MC-DDPM demonstrates statistical improvement ($p < 0.05$) over the competing methods. 2) In the prostate dataset, the MC-DDPM can generate sCT achieving MAE 55.124 ± 9.414 HU, PSNR 28.708 ± 2.112 dB, SSIM 0.878 ± 0.040 , and NCC 0.940 ± 0.039 . MC-DDPM can obtain statistical improvements over all metrics

compared to the MC-GAN, MC-CGAN, and 3D-DDPM, and mainly improvement over the MAE compared to the 2D-DDPM and 3D-DDIM. MC-DDPM therefore demonstrates utility in generating sCT images from MRIs, effectively streamlining the treatment planning workflow, reducing inefficiency, improving patient experience and reducing costs. The dosimetric result also robust dose-based performance. This strength strongly suggests that sCT holds the potential to replace Truth CT in radiotherapy planning. Such a transition could revolutionize the treatment process by obviating the need for additional CT scans, thereby reducing patient exposure to ionizing radiation.

However, one of the limitations is that it took approximately 760 seconds for MC-DDPM using a sampling timestep of 50 and MC-based generation of 5 runs to generate a 192x256x32 resolution sCT on the workstation described in the Appendix. C.3. MC-GAN and MC-CGAN were able to generate the same resolution image in only 22 seconds. There are two reasons for this low efficiency. First, a high quality sCT requires a relatively large number of generation iterations. Second, the stochasticity of generation requires MC-based generation which contains multiple runs for a single image. These limitations make 3D MRI to sCT synthesis prohibitively slow, and it can be inferred that the same low efficiency will be observed for other 3D images. Nevertheless, this limitation does not detract from the value of MC-DDPM, but instead highlights the need for further hardware and software optimization. Song *et al.*³¹, Zhang *et al.*³² Pan *et al.*³³, and Kong *et al.*³⁴ have proposed algorithms for improving efficiency, such as designing an exponential forward process to generate noisy images with fewer timesteps or deploying a pre-trained network to accelerate the image synthesis process. These methods can enhance the efficiency of MC-DDPM. In addition, 3D diffusion-based methods for ultrasound and cone beam CT²⁸ have not yet been explored and evaluating the synthetic medical images on these applications is another promising area of future inquiry.

We further intend to explore efficiency improvements for the diffusion framework for 3D synthesis, investigating more sophisticated network architectures to elevate the quality of image synthesis, and designing a deterministic diffusion process to eliminate the need for multiple runs of Monte Carlo-based generation. Moreover, we plan to extend the application of MC-DDPM to a wider range of medical image modalities and undertake a more comprehensive study to validate the effectiveness of our approach.

7. Conclusion

This work presents a 3D Magnetic Resonance image to Computed Tomography image (MRI-to-CT) denoising diffusion probabilistic model (MC-DDPM) for generating sCT scans from MR scans. The proposed method utilizes a 3D Shifted-window (Swin) transformer network to learn a diffusion process to convert a pure Gaussian noise into a realistic CT scan from a given MR scan. The method can achieve superior image quality compared to several competing state-of-the-art synthesis algorithms (GANs and conventional DDPMs). MC-DDPM generates high quality sCT images using only MRI inputs, eliminating the need for CT scans in radiation therapy planning, therefore improving quality of patient care.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This research was supported in part by National Institutes of Health R01CA215718, R56EB033332, R01EB032680 and P30CA008748.

Reference

1. Chandarana H, Wang H, Tijssen RHN, Das JJ. Emerging role of MRI in radiation therapy. *Journal of Magnetic Resonance Imaging*. 2018;48(6):1468–1478. [PubMed: 30194794]
2. Khoo V, Padhani A, Tanner S, Finnigan D, Leach M, Dearnaley D. Comparison of MRI with CT for the radiotherapy planning of prostate cancer: A feasibility study. *The British journal of radiology*. 1999;72:590–597. [PubMed: 10560342]
3. Chowdhury N, Toth R, Chappelw J, et al. Concurrent segmentation of the prostate on MRI and CT via linked statistical shape models for radiotherapy planning. *Medical Physics*. 2012;39(4):2214–2228. [PubMed: 22482643]
4. Decazes P, Hinault P, Veresezan O, Thureau S, Gouel P, Vera P. Trimodality PET/CT/MRI and Radiotherapy: A Mini-Review. *Front Oncol*. 2021;10:614008. [PubMed: 33614497]
5. Boulanger M, Nunes J-C, Chourak H, et al. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. *Physica Medica*. 2021;89:265–281. [PubMed: 34474325]
6. Pereira GC, Traughber M, Muzic RF. The Role of Imaging in Radiation Therapy Planning: Past, Present, and Future. *BioMed Research International*. 2014;2014:231090. [PubMed: 24812609]
7. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016, 2016.
8. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Communications of the ACM*. 2020;63(11):139–144.
9. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Medical physics*. 2019.
10. Zhao S, Geng C, Guo C, Tian F, Tang X. SARU: A self-attention ResUNet to generate synthetic CT images for MR-only BNCT treatment planning. *Medical Physics*. 2023;50(1):117–127.
11. Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Išgum I. Deep MR to CT synthesis using unpaired data. Paper presented at: Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017, Proceedings 22017.
12. Andreasen D, Van Leemput K, Edmund JM. A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis. *Medical Physics*. 2016;43(8Part1):4742–4752. [PubMed: 27487892]
13. Li W, Li Y, Qin W, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. *Quantitative Imaging in Medicine and Surgery*. 2020;10(6):1223–1236. [PubMed: 32550132]
14. Yang H, Sun J, Carass A, et al. Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN. Paper presented at: 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 42018.
15. Zhao B, Cheng T, Zhang X, et al. CT synthesis from MR in the pelvic area using Residual Transformer Conditional GAN. *Computerized Medical Imaging and Graphics*. 2023;103:102150. [PubMed: 36493595]
16. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training gans. *Advances in neural information processing systems*. 2016;29.

17. Kodali N, Abernethy J, Hays J, Kira Z. On convergence and stability of gans. arXiv preprint arXiv:170507215. 2017.
18. Shokraei Fard A, Reutens DC, Vegh V. From CNNs to GANs for cross-modality medical image estimation. *Computers in Biology and Medicine*. 2022;146:105556. [PubMed: 35504221]
19. Song J, Meng C, Ermon S. Denoising diffusion implicit models. arXiv preprint arXiv:201002502. 2020.
20. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*. 2020;33:6840–6851.
21. Dhariwal P, Nichol A. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*. 2021;34:8780–8794.
22. Nichol AQ, Dhariwal P. Improved Denoising Diffusion Probabilistic Models. *Proceedings of the 38th International Conference on Machine Learning; 2021; Proceedings of Machine Learning Research*.
23. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:201113456. 2020.
24. Gui J, Sun Z, Wen Y, Tao D, Ye J. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*. 2021.
25. Wolleb J, Bieder F, Sandkühler R, Cattin PC. Diffusion Models for Medical Anomaly Detection. arXiv preprint arXiv:220304306. 2022.
26. Pan S, Wang T, Qiu RL, et al. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. *Physics in Medicine and Biology*. 2023.
27. Lyu Q, Wang G. Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models. arXiv preprint arXiv:220912104. 2022.
28. Peng J, Qiu RLJ, Wynne JF, et al. CBCT-Based synthetic CT image generation using conditional denoising diffusion probabilistic model. *Medical Physics*.n/a(n/a).
29. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv*. 2018;abs/1804.03999.
30. Pan S, Chang C-W, Wang T, et al. Abdomen CT multi-organ segmentation using token-based MLP-Mixer. *Medical Physics*.n/a(n/a).
31. Song Y, Dhariwal P, Chen M, Sutskever I. Consistency models. arXiv preprint arXiv:230301469. 2023.
32. Zhang Q, Chen Y. Fast Sampling of Diffusion Models with Exponential Integrator. arXiv preprint arXiv:220413902. 2022.
33. Pan S, Abouei E, Peng J, et al. Full-dose PET Synthesis from Low-dose PET Using High-efficiency Diffusion Denoising Probabilistic Model. arXiv preprint arXiv:230813072. 2023.
34. Kong Z, Ping W. On fast sampling of diffusion probabilistic models. arXiv preprint arXiv:210600132. 2021.

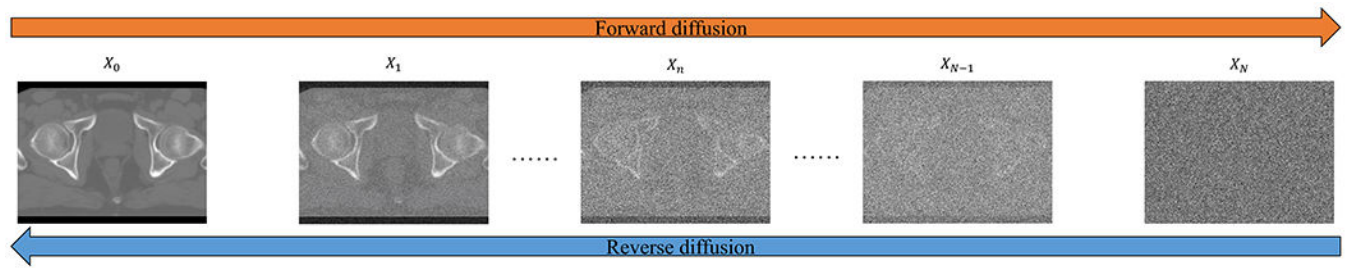


Figure 1:

The proposed diffusion process of the MC-IDDPM's synthesis: First, the 3D CT image is transformed into pure Gaussian noise through forward diffusion, which involves iterative addition of a small amount of noise. Second, to obtain the noise-free image, a neural network is used to repeatedly denoise the Gaussian noise through a reverse process.

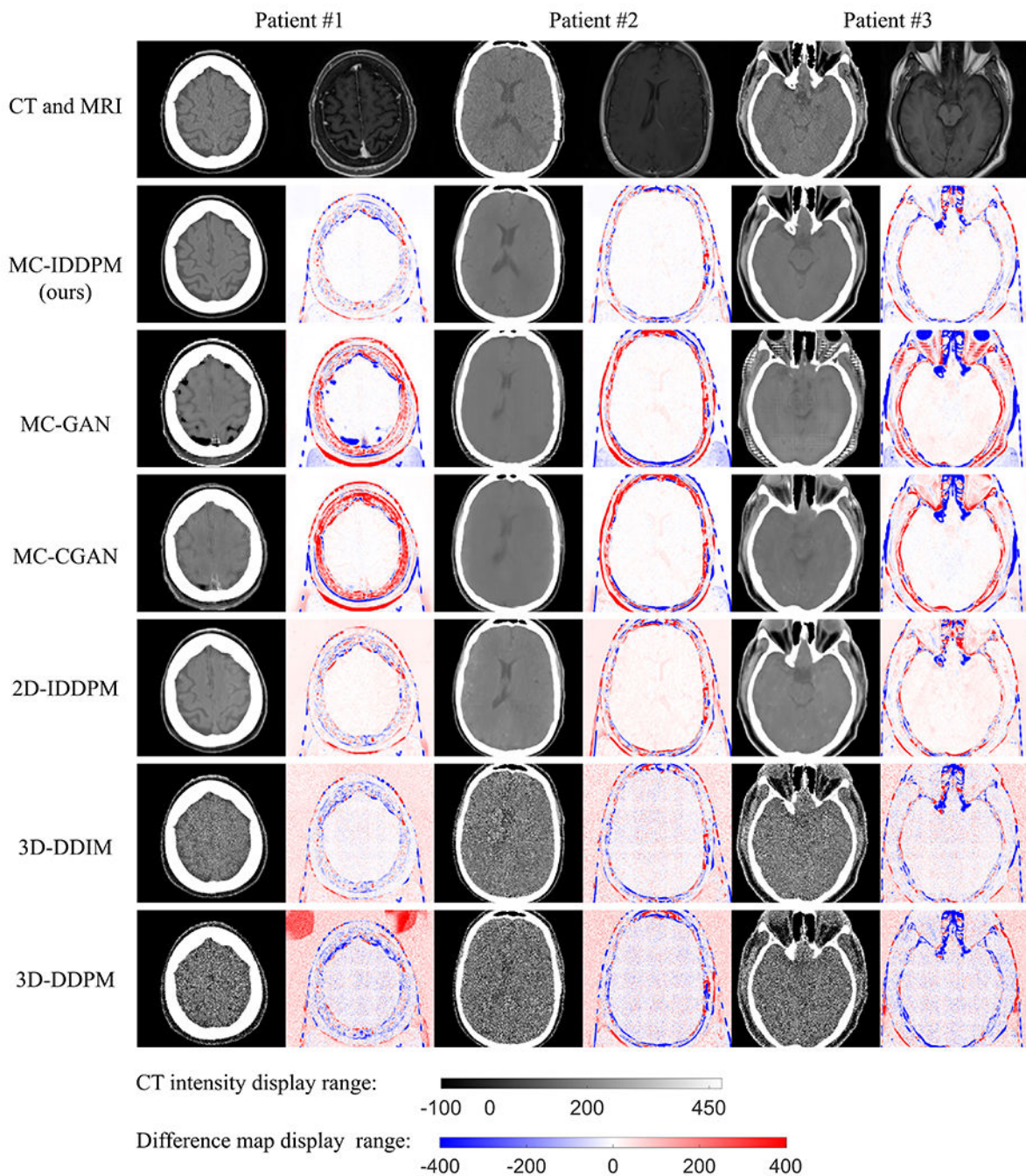
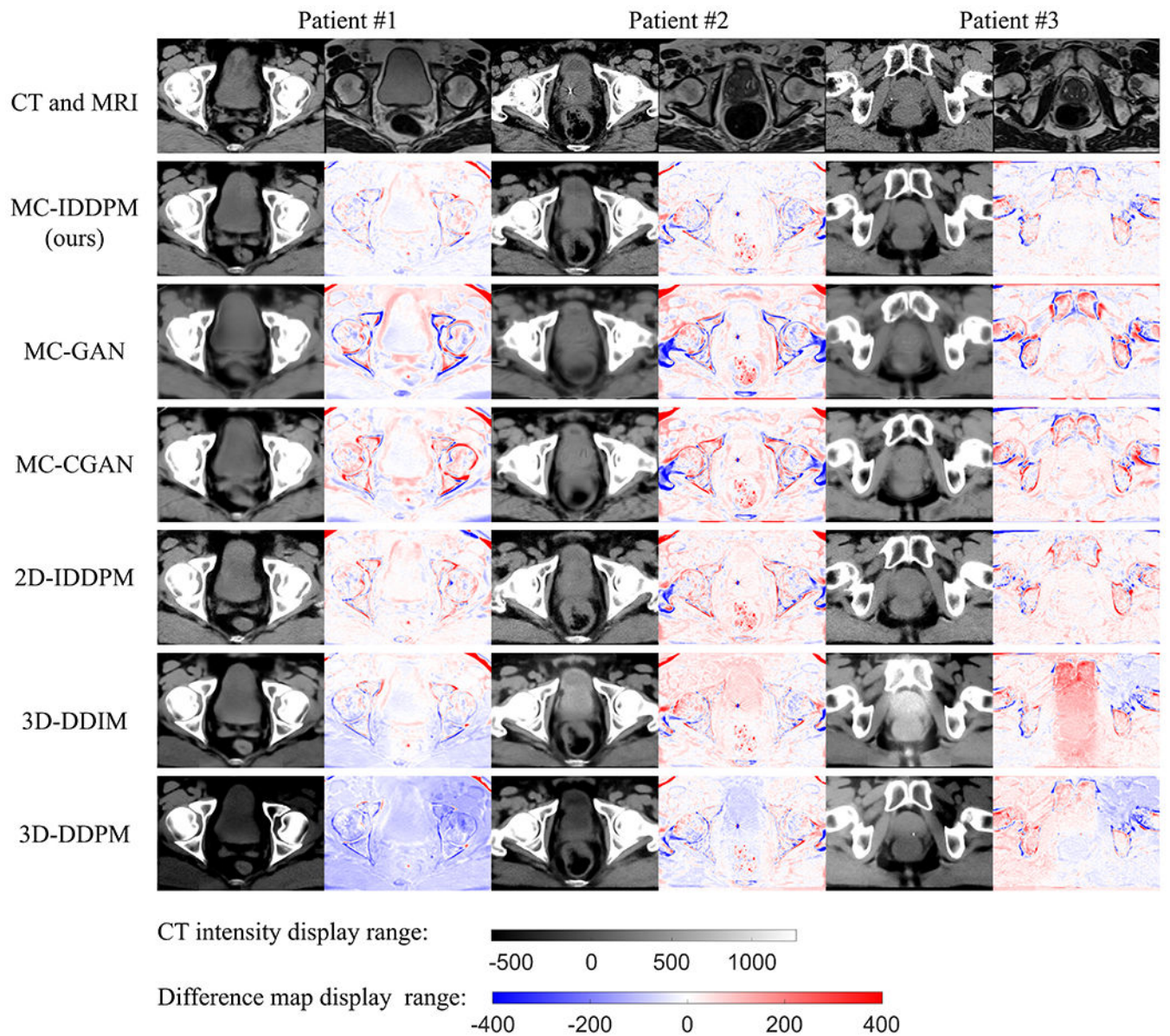


Figure 2:

Synthetic CT images generated from the brain dataset including three slices from three patient subjects. The first row displays three CT images (left) and paired input MRIs (right). In the second to the sixth row, the sCT image outputs of MC-IDDPM (row 2) and competing networks (row 3-7) are presented in the first, third and the fifth columns. The difference map between the sCTs with their corresponding CT images are presented in the second, fourth, and the sixth column.

**Figure 3:**

Synthetic CT images generated from the prostate dataset including three slices from three patient subjects. The first row displays three groups of the CT images and matched input MR scans. In the second to the sixth row, the sCT scans and the corresponding difference map of MC-IDDPM (row 2) and competing networks (row 3-7) are presented in every two columns, respectively.

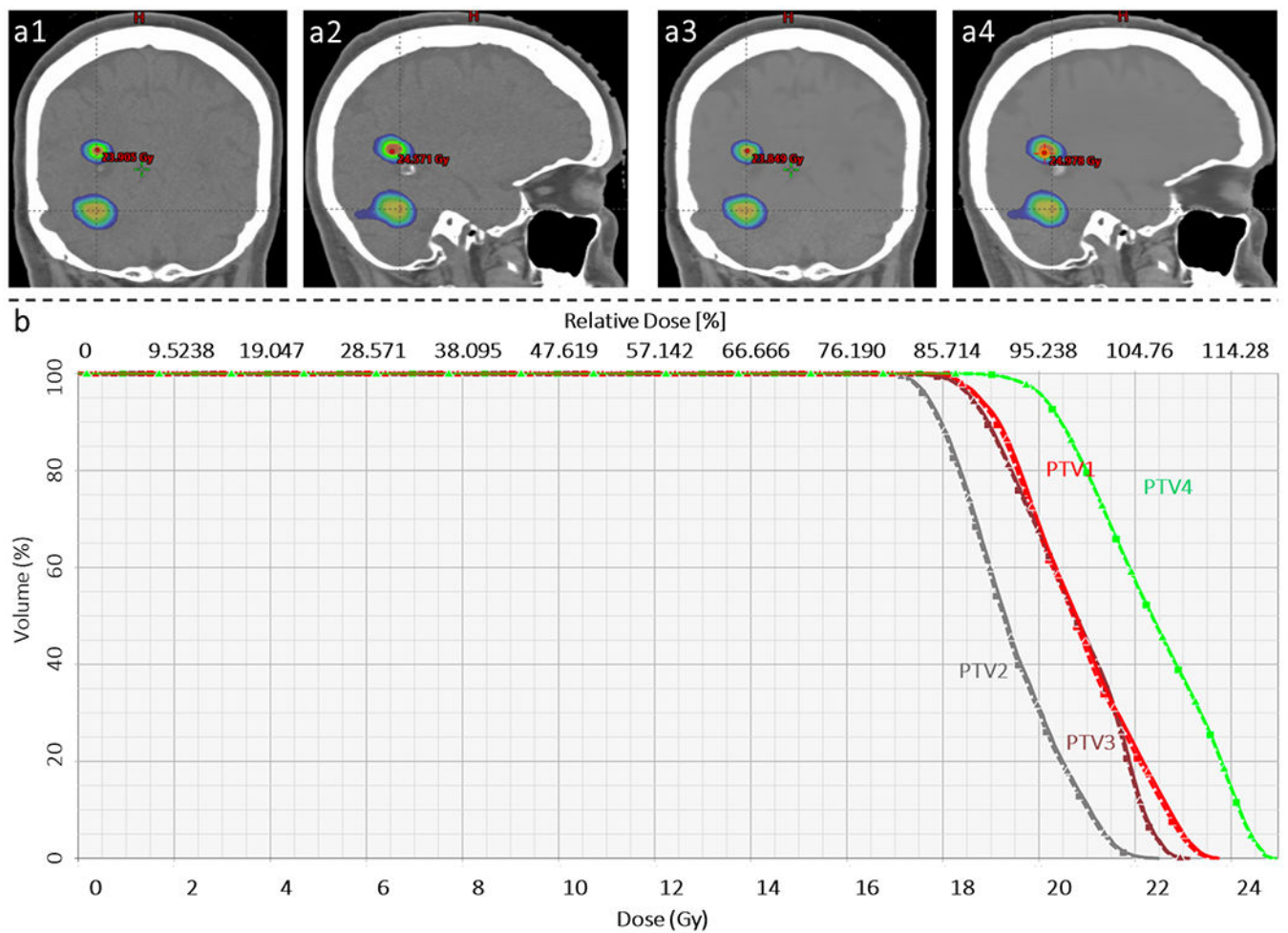


Figure 4:

The dosimetric analysis results using one example patient. a) Visualization of the dose distribution of a clinical radiotherapy plan calculated on planning CT (a1-a2) and sCT (a3-a4). b) Dose-Volume Histograms (DVHs) of the plan calculated on sCT (represented by dashed lines) and planning CT (represented by solid lines) for four Planning Target Volumes (PTVs). PTV1 is shown in red, PTV2 in gray, PTV3 in deep red, and PTV4 in green.

Table 1.

Quantitative analysis of sCT images from MC-IDDPM vs. MC-GAN, MC-CGAN, 2D-IDDPM, 3D-DDIM, and 3D-DDPM using the institutional brain dataset. The table highlights the best-performing network(s), indicated in bold, and the second-best network(s), underlined, based on the mean evaluation results. P-values are provided to compare the results of MC-IDDPM with those of the other competing methods. The reported values in the table are rounded to three decimal places.

	MAE (HU)	PSNR (dB)	SSIM	NCC
MC-IDDPM	48.825±21.491	26.491±2.814	0.947±0.032	0.976±0.019
<i>p-value</i>	N/A	N/A	N/A	N/A
MC-GAN	85.394±19.624	21.943±1.724	0.884±0.042	0.944±0.023
<i>p-value</i>	<0.010	<0.010	<0.010	<0.010
MC-CGAN	80.758±17.971	22.528±1.768	0.905±0.034	0.951±0.021
<i>p-value</i>	<0.010	<0.010	<0.010	<0.010
2D-IDDPM	<u>67.228±63.807</u>	24.964±3.404	<u>0.937±0.093</u>	0.956±0.090
<i>p-value</i>	<0.010	0.023	0.090	0.018
3D-DDIM	73.910±19.864	<u>25.424±2.475</u>	<u>0.931±0.032</u>	<u>0.971±0.021</u>
<i>p-value</i>	<0.010	0.028	<0.010	0.026
3D-DDPM	106.153±15.572	22.709±1.566	0.880±0.026	0.952±0.018
<i>p-value</i>	<0.010	<0.010	<0.010	<0.010

Table 2.

Quantitative analysis of sCTs from MC-IDDPM vs. MC-GAN, MC-CGAN, 2D-IDDPM, 3D-DDIM, and 3D-DDPM using the institutional prostate dataset. The table highlights the best-performing network(s), indicated in bold, and the second-best network(s), underlined, based on the mean evaluation results. P-values are shown below each competing method.

	MAE (HU)	PSNR (dB)	SSIM	NCC
MC-IDDPM	55.124±9.414	28.708±2.112	0.878±0.040	0.940±0.039
<i>p-value</i>	N/A	N/A	N/A	N/A
MC-GAN	80.366±28.880	24.712±2.970	0.800±0.050	0.846±0.068
<i>p-value</i>	<0.010	<0.010	<0.010	<0.010
MC-CGAN	68.278±19.948	26.023±2.781	0.852±0.043	0.884±0.061
<i>p-value</i>	<0.010	<0.010	0.024	<0.010
2D-IDDPM	<u>64.197±10.183</u>	27.786±2.069	<u>0.863±0.038</u>	0.930±0.040
<i>p-value</i>	<0.010	0.070	0.108	0.182
3D-DDIM	64.426±9.869	<u>28.124±1.953</u>	0.861±0.040	<u>0.932±0.046</u>
<i>p-value</i>	<0.010	0.176	0.093	0.381
3D-DDPM	73.696±16.794	26.754±2.047	0.839±0.031	0.912±0.051
<i>p-value</i>	<0.010	<0.010	<0.010	0.020