SynthDiffuson at SynthRAD 2023 Task 1: Synthesizing Computed Tomography for Radiotherapy

Lujia Zhong[1,2], Zhiwei Deng[1,2], Shuo Huang[1,3], Wenhao Chi[1,2], Jianwei Zhang[1,2], and Yonggang Shi[1,2,3]⋆

[1] Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California (USC), Los Angeles, CA 90033, USA
[2] Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California (USC), Los Angeles, CA 90089, USA
[3] Alfred E. Mann Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California (USC), Los Angeles, CA 90089, USA

Abstract. This report presents a novel diffusion model based method to solve the Task 1 of SynthRAD 2023: MRI-to-sCT generation to facilitate MR-only RT. This task aims to convert MR to synthetic CT (sCT) for treatment planning and dose calculation in radio therapy (RT). Our proposed method utilizes the diffusion-based model, the Denoising Diffusion Implicit Models (DDIM), for the sCT generation, where, each slice of MRI images is converted to sCT images separately with the consistency constrain provided by the adjacent slices. Then, two U-Net models are employed as the refinement networks for noise and artifacts removal. Our method achieves high scores of mean average error (MAE), peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) on the validation set.

Keywords: Image synthesis · Synthetic CT · MRI · Denoising Diffusion Implicit Models.

1 Introduction

Radiotherapy (RT) plays an important role in the treatment of oncological patients. CT images are employed in RT to provide the tissue attenuation information for dose calculation and treatment planning. [1] However, obtaining the CT images introduces additional radiation to the patients.

The superb soft-tissue contrast of MRI can provide additional information to RT. [2] However, due to the lack of the tissue attenuation information, MRI cannot be directly used for dose calculation. Therefore, CT is still required, and MRI is generally registered to CT space. Developing MRI-only RT can not only reduce the radiation to patients, but also help to simplify and speed up the workflow.

Many methods have been proposed to convert MRI images to CT-equivalent images, obtaining synthetic CT (sCT) [3] that has tissue attenuation information for treatment planning and dose calculation. The organizers of SynthRAD 2023 [4] provide the first platform offering the public data evaluation metrics to compare the latest developments in sCT generation methods, and the Task 1 focuses on the MRI-to-CT generation. This facilities the comparison of different approaches.

In our work, we utilize the diffusion-based model for the sCT generation, then we employ two vanilla U-Net models as the refinement networks to remove the noise and artifact.

2   Method

The workflow of our method is shown in Fig. 1. Each MRI slice is padded and re-sampled to the size of 256×256 firstly, and then the diffusion model [5, 6] with consistency constraint is employed to generate the sCT images. After that, the generated sCT images are re-sampled to the same shape as the original MRI images. Finally, two vanilla U-Net [7] models are employed to remove the noise and artifacts.

The CT data are normalized to -1 to 1 using Eq. 1 (donated as CT_n) and the MRI data are normalized to 0 to 1 using the min-max normalization.

$$CT_n = (CT + 1024)/(3000 + 1024) \times 2 - 1 \tag{1}$$

Our method uses diffusion-based method to generate the sCT images from 2D MRI images. The Denoising Difusion Implicit Models (DDIM) [5, 8] is employed in our diffusion model sampling, which learns a Markov chain to reverse the non-Markov perturbation process using Eqs. 2, 3 and 4.

$$q\left(\mathbf{x}_1, \ldots, \mathbf{x}_T \mid \mathbf{x}_0\right) = \prod_{t=1}^{T} q\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{x}_0\right) \tag{2}$$

$$q_\sigma\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0\right) = \mathcal{N}\left(\mathbf{x}_{t-1} \mid \tilde{\mu}_t\left(\mathbf{x}_t, \mathbf{x}_0\right), \sigma_t^2 \mathbf{I}\right) \tag{3}$$

$$\tilde{\mu}_t\left(\mathbf{x}_t, \mathbf{x}_0\right) := \sqrt{\bar{\alpha}_{t-1}} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_{t-1} - \sigma_t^2} \cdot \frac{\mathbf{x}_t - \sqrt{\bar{\alpha}_t} \mathbf{x}_0}{\sqrt{1 - \bar{\alpha}_t}} \tag{4}$$

where x_t means the destroyed CT images after adding Gaussian noise for t times. q means the distribution of x_0 that $x_0 \sim q(x_0)$. $\bar{\alpha}_t \in (0, 1)$ is a hyperparameter chosen ahead of model training. T is the total number of iterations in the adding and removing of noise, $t \in 0, 1, \cdots, T$.

The input of the diffusion model contains 4 channels. The first 3 channels are 3 adjacent MRI slices in size of 256×256, padded and re-sized from original shape, which reduces the computational cost and also avoids input of pure background of MRI images. The second channel is the Gaussian noise x_T follows $x_T \sim \mathcal{N}(0, 1)$.

In each generation, we use 3 adjacent MR slices as input to generate the sCT of the middle one. The overall loss function consists of two losses: reconstruction loss and consistency loss. The reconstruction loss is chosen to be the $L1$ loss between the ground truth CT and generated sCT. To prompt the consistency of the generated sCT image with its adjacent slices, we also employ the $L1$ loss as a consistency constraint between the sCT and its adjacent two CT ground truth slices.

The generated sCT is then re-sampled to the same size as the MRI counterpart. After that, We use two supervised U-Net models to remove noises and artifacts in the generated sCT images. The $L1$ loss between sCT and ground truth CT images is used as refinement loss while training. Note that the first refinement network is to refine the sCT in the direction from top to bottom; the second refinement network refines the sCT in the direction from front to back.

3   Implementation Details

We trained in total 6 models for both pelvis and brain data: 2 diffusion models and 4 refinement networks. The diffusion model and refinement networks for the brain and pelvis are in the same structure with also the same training hyperparameters, except for the time embedding in the diffusion model. While training the diffusion models, we first train it for 140,000 iterations with a batch size of 16, a learning rate of 2e-4 for the first 100,000 iterations, and 5e-5 for the last 40,000 iterations. While training the refinement networks, a vanilla U-net is trained for 20,000 iterations with a batch size of 16 and a learning rate of 2e-4. While inference of diffusion models, DDIM sampling with 20 steps is adopted for consideration of hardware limitation.

References

1. Chernak, E.S., Rodriguez-Antunez, A., Jelden, G.L., Dhaliwal, R.S., Lavik, P.S.: The use of computed tomography for radiation therapy treatment planning. Radiology **117**(3), 613–614 (1975)
2. Lagendijk, J.J.W., Raaymakers, B.W., den Berg, C.A.T.V., Moerland, M.A., Philippens, M.E., van Vulpen, M.: Mr guidance in radiotherapy. Physics in Medicine & Biology **59**(21), R349 (Oct 2014). https://doi.org/10.1088/0031-9155/59/21/R349
3. Kida, S., Nakamoto, T., Nakano, M., Nawa, K., Haga, A., Kotoku, J., Yamashita, H., Nakagawa, K.: Cone beam computed tomography image quality improvement using a deep convolutional neural network. Cureus **10**(4) (2018)
4. Thummerer, A., Huijben, E., Terpstra, M., Gurney-Champion, O., Afonso, M., Pai, S., Koopmans, P., van Eijnatten, M., Perko, Z., Maspero, M.: SynthRAD2023 Challenge design: Synthesizing computed tomography for radiotherapy (Apr 2023). https://doi.org/10.5281/zenodo.7781049
5. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Cui, B., Yang, M.H.: Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796 (2022)

6. Zhang, L., Agrawala, M.: Adding conditional control to text-to-image diffusion models. arXiv preprint arXiv:2302.05543 (2023)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
8. Song, K.U.: Applying regularized schr\" odinger-bridge-based stochastic process in generative modeling. arXiv preprint arXiv:2208.07131 (2022)
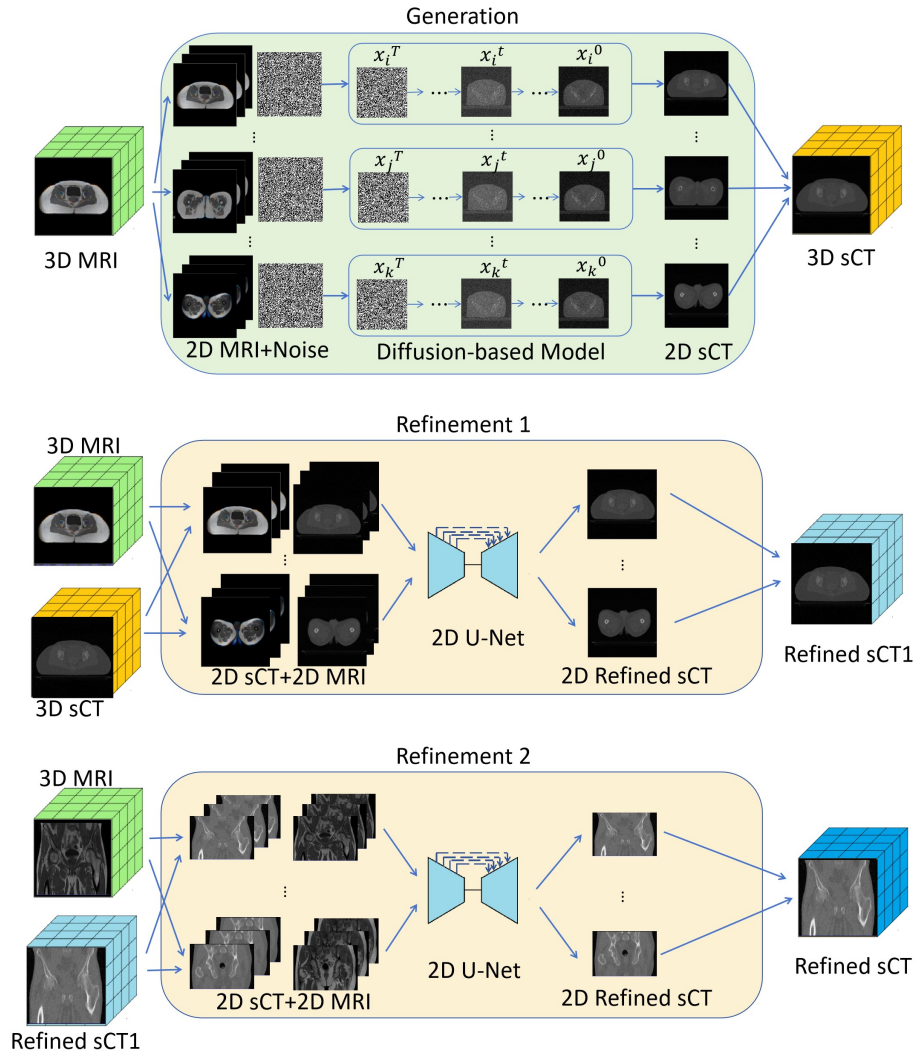
Fig. 1. Workflow of our method.