Team KoalAI: Locally-enhanced 3D Pix2Pix GAN for Synthetic CT Generation

Bowen Xin[1], Aaron Nicolson[1], Hilda Chourak[1], Gregg Belous[1], and Jason Dowling[1]

Australian e-Health Research Centre, CSIRO, Herston, Queensland, 4006, Australia
{bowen.xin, aaron.nicolson, hilda.chourak, gregg.belous, jason.dowling}@csiro.au

Abstract. Synthetic CT generation from MRI and CBCT is an essential task to facilitate clinical workflow in radiation therapy. In this paper, we describe the participation of team KoalAI for both MRI-to-CT and CBCT-to-CT tasks in SynthRAD2023 challenge. The major contribution is the proposed Locally-enhanced 3D Pix2pix GAN, which improves the local details of 3D pix2pix GAN by incorporating an additional 2D local discriminator. This model outperformed other implemented comparison methods including diffusion models and unsupervised GAN in local validation. Model ensembling was used to further improve the generative performance.

Keywords: synthetic CT · 3D GAN

1 Introduction

Radiation Therapy (RT) is a widely used medical treatment for cancer. It involves delivering targeted radiation to cancerous cells and is often used in conjunction with surgery or chemotherapy. This type of treatment is non-invasive and is commonly employed to treat various types of cancer, including prostate, breast, lung, and head and neck cancers. The standard RT workflow relies on three imaging modalities: computed tomography (CT) for dose calculation based on electron density information, magnetic resonance imaging (MRI) for better soft tissue contrast, enabling more accurate target delineation [4] and minimising the risk of toxicity in healthy tissue [2], and cone-beam CT (CBCT) to position the patient under the linear accelerator (LINAC) by registration with the planning CT.

Traditionally, both CT and MRI images are co-registered to define the treatment plan. However, this step introduces uncertainties, with reported calculations of up to 2 mm for prostate cancer patients [8]. In order to enhance efficiency and precision in the clinical workflow, MRI-only RT has gained popularity, rendering the use of CT obsolete. The interest in MRI for RT has led to the development of MRI-LINAC machines, which combine an on-board MRI scanner and a linear accelerator. The advantage of this device is its ability to accommodate image guided adaptive RT (IGART), which considers daily anatomical changes and recalculates the dose distribution prior to each session. However, a major limitation of MRI for treatment planning is its lack of information on electron density, which is essential for accurate dose calculation. To address this, several methods have been proposed in the literature for generating synthetic CT scans (sCT) from MRI, including bulk-density, atlas-based, and voxel-based techniques[5], and more recently machine-learning based methods [6]. Among these, deep learning methods have shown promise in terms of both robustness and accuracy [1].

CBCT, on the other hand, enables patient positioning under the LINAC as well as real-time monitoring of the tumour during treatment. However, these imaging modality suffer from limited soft tissue contrast, artefacts and noise, and expose patients to additional radiation. sCT generation from CBCT would allow for accurate dose calculation and improved image-guided adaptive RT (IGART). As for sCT generated from MRI, deep learning based methods offer precise Hounsfield Units (HU) prediction from CBCT [10].

Deep learning-based sCT generation methods employ convolutional neural networks (CNNs) as model architectures, but recently transformers have demonstrated great potential in image synthesis [12]. Hybrid networks, combining CNNs and transformers [3, 13], or diffusion model [7, 9], have been proposed to extract both local texture and global information. The primary advantage of transformers lies in their ability to better understand contextual information compared to CNNs. However, they do tend to come with a higher computational cost and require larger amounts of data.

In this paper, we proposed a locally-enhanced 3D pix2pix GAN for synthetic CT generation. We also implemented other generative models including diffusion models and unsupervised GAN network. The experimental results show the proposed model achieved the best results to compared to other implemented methods.

2 Methodology

2.1 Data description

Multi-center datasets were provided for both task1 (MRI-to-CT) and task2 (CBCT-to-CT) [11]. The data were collected from three centers (UMC Utrecht, UMC Groningen, and Radboud Nijmegen). As shown in Fig 1, each task contains a brain dataset and a pelvis dataset. A total of 270 samples per dataset was separated into 180 training samples, 30 validation samples and 60 testing samples. For training data, masks were provided in addition to MRI-CT or CBCT-CT pairs. For validation and testing data, target CT was not provided to participants. During the algorithm development, we further randomly divided the training data into 80% local train set and 20% local validation set for the model selection purpose.

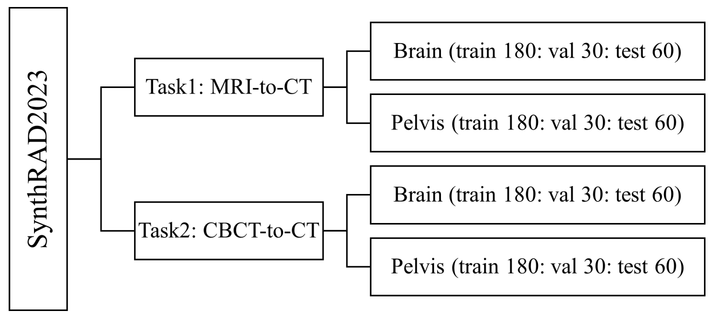

Fig. 1. Data description.

2.2 Preprocessing

Task1 Both brain and pelvis MRI were preprocessed with 1) histogram matching with a random MRI sample, 2) N4 bias field correction, 3) smoothing with gradient anisotropic diffusion filter, and 4) masking using provided body mask. Additionally, arms in pelvis MRI were removed using a 2d connected-component-based algorithm. All MRI was scaled to [-1, 1] using min-max normalisation, while CT was scaled to [-1, 1] using intensity range [-1024, 3000].

Task2 Both brain and pelvis CBCT were preprocessed with 1) lower-bound intensity scaling (from 0 to -1024), 2) masking using the provided body mask, 3) intensity clipping to [-1000, 3000]. Additionally, to remove bright spots around the body region in pelvis CBCT, a thresholding-based algorithm was applied in the adjacent area of the body, followed by a 2d connected-component-based algorithm to remove noise. All CBCT and CT were scaled to [-1, 1] using an intensity range of [-1024, 3000].

2.3 Locally-enhanced pix2pix GAN

The proposed locally-enhanced pix2pix GAN is described in Fig. 2. In terms of model architecture, the model consists of a 3D Generator and a Locally-enhanced 3D Discriminator. The 3D generator takes a 3D patch as input and produces a translated 3D image. The generator can be implemented using Resnet, Unet, etc. The locally-enhanced 3D discriminator takes both translated 3D images and target 3D images as input and produces binary decisions. Specifically, the input 3D patch will be separated into two pathways with an image iterator. The first pathway feeds the original 3D patch to a 3D patch discriminator. The second pathway randomly samples 2D images from the 3D patch and feeds them to 2D patch discriminator. In terms of the loss function, MSE loss was used for the adversarial training, while l1 loss was used for pixel-wise supervision. An ensemble of models consisting of different 3D generators is used to enhance generative performance.

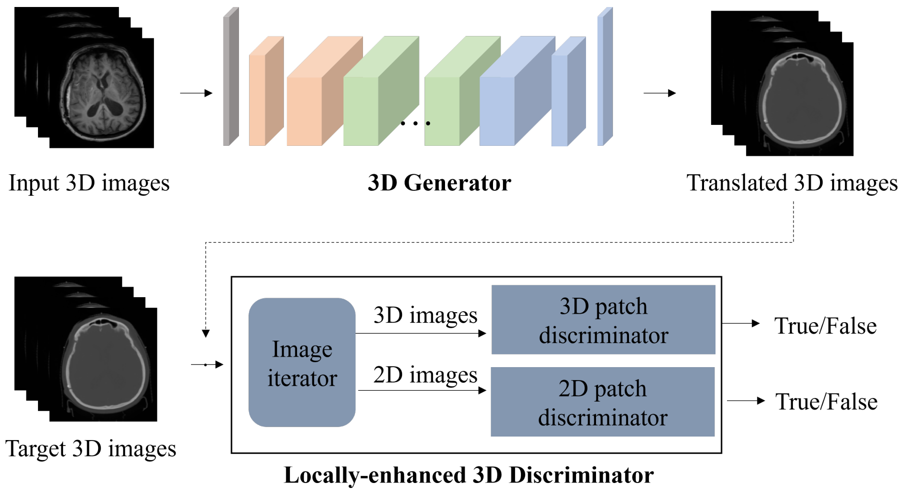

Fig. 2. Model architecture of Locally-enhanced 3D Pix2Pix GAN.

3 Results

3.1 Metrics

The quality of synthetic CT was evaluated using both image similarity metrics and dose evaluation metrics. Image similarity metrics included mean absolute error (MAE), peak-signal-to-noise (PSNR), and structural similarity index (SSIM) between synthetic CT and CT. These image metrics used in local validation were implemented by the code provided at https://github.com/SynthRAD2023/metrics. Dosimetric evaluation included 1) relative dose difference, 2) dose-volume histogram, and 3) gamma index. Dose calculation is performed with matRad.

3.2 Implementation

The Adam optimizer was used with a learning rate of 0.0002. The weight for adversarial loss was set to 1, and the weight for l1 loss was set to 100. The batch size was set to 3. The training stopped when the validation MAE converged. Various 3D patch sizes were experimented including (256, 256, 56), (256, 56, 256), (56, 256, 256), (128, 128, 128). We implemented data augmentation, including affine transformation, elastic deformation, random intensity shift, random contrast adjustment, and random histogram shift.

3.3 Task1-specific implementation and results

As shown in Table 1, the model for task1 pelvis was implemented with a single Locally-enhanced 3D pix2pix network, while the model for task1 brain was implemented with an ensemble of three networks with different patch sizes (256, 256, 56), (256, 56, 256), and (56, 256, 256). The best MAE on the validation leaderboard was 69.41 for task1 pelvis and 74.28 for task1 brain.

Table 1. Task1 implementation and validation leaderboard results

Data	Generator	Discriminator	Patch size	Val-MAE
Task1_pelvis	Resnet	LE Discriminator	(256, 256, 56)	69.41
	Resnet	LE Discriminator	(256, 256, 56)	
Task1_brain	Resnet	Patch Discriminator	(256, 56, 256)	74.28
	Resnet	LE Discriminator	(56, 256, 256)	

3.4 Task2-specific implementation and results

As shown in Table 2, the model for task2 pelvis was implemented with an ensemble of 3 networks with different generator architecture and patch, while the model for task2 brain was implemented with an ensemble of three networks with different patch sizes (256, 256, 56), (256, 56, 256), and (56, 256, 256). The best MAE on the validation leaderboard was 62.48 for task2 pelvis and 74.28 for task1 brain.

Table 2. Task1 implementation and validation leaderboard results

Data	Generator	Discriminator	Patch size	Val-MAE
	Dynet	Patch Discriminator	(128, 128, 128)	
Task2_pelvis	Resnet	LE Discriminator	(256, 256, 56)	62.48
	Unet	Patch Discriminator	(448, 448, 64)	
	Resnet	LE Discriminator	(256, 256, 56)	
Task2_brain	Resnet	LE Discriminator	(256, 56, 256)	52.89
	Resnet	LE Discriminator	(56, 256, 256)	

References

1. Boulanger, M., Nunes, J.C., Chourak, H., Largent, A., Tahri, S., Acosta, O., Crevoisier, R.D., Lafond, C., Barateau, A.: Deep learning methods to generate synthetic ct from mri in radiotherapy: A literature review (9 2021). https://doi.org/10.1016/j.ejmp.2021.07.027
2. Bruynzeel, A.M., Tetar, S.U., Oei, S.S., Senan, S., Haasbeek, C.J., Spoelstra, F.O., Piet, A.H., Meijnen, P., van der Jagt, M.A.B., Fraikin, T., Slotman, B.J., van Moorselaar, R.J., Lagerwaard, F.J.: A prospective single-arm phase 2 study of

stereotactic magnetic resonance guided adaptive radiation therapy for prostate cancer: Early toxicity results. International Journal of Radiation Oncology Biology Physics **105**, 1086–1094 (12 2019). https://doi.org/10.1016/j.ijrobp.2019.08.007
3. Dalmaz, O., Yurt, M., Çukur, T.: Resvit: Residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**, 2598–2614 (2022). https://doi.org/10.1109/TMI.2022.3167808
4. Hentschel, B., Oehler, W., Straußß, D., Ulrich, A., Malich, A.: Definition of the ctv prostate in ct and mri by using ct-mri image fusion in imrt planning for prostate cancer. Strahlentherapie und Onkologie **187**, 183–190 (3 2011). https://doi.org/10.1007/s00066-010-2179-1
5. Johnstone, E., Wyatt, J.J., Henry, A.M., Short, S.C., Sebag-Montefiore, D., Murray, L., Kelly, C.G., McCallum, H.M., Speight, R.: Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy (2018). https://doi.org/10.1016/j.ijrobp.2017.08.043
6. Largent, A., Barateau, A., Nunes, J.C., Lafond, C., Greer, P.B., Dowling, J.A., Saint-Jalmes, H., Acosta, O., de Crevoisier, R.: Pseudo-ct generation for mri-only radiation therapy treatment planning: Comparison among patch-based, atlas-based, and bulk density methods. International Journal of Radiation Oncology Biology Physics **103**, 479–490 (2019). https://doi.org/10.1016/j.ijrobp.2018.10.002
7. Li, X., Shang, K., Wang, G., Butala, M.D.: Ddmm-synth: A denoising diffusion model for cross-modal medical image synthesis with sparse-view measurement embedding (2023)
8. Nyholm, T., Nyberg, M., Karlsson, M.G., Karlsson, M.: Systematisation of spatial uncertainties for comparison between a mr and a ct-based radiotherapy workflow for prostate treatments. Radiation Oncology **4** (11 2009). https://doi.org/10.1186/1748-717X-4-54
9. Pan, S., Wang, T., Qiu, R.L., Axente, M., Chang, C.W., Peng, J., Patel, A.B., Shelton, J., Patel, S.A., Roper, J., Yang, X.: 2d medical image synthesis using transformer-based denoising diffusion probabilistic model. Physics in Medicine and Biology **68** (5 2023). https://doi.org/10.1088/1361-6560/acca5c
10. Spadea, M.F., Maspero, M., Zaffino, P., Seco, J.: Deep learning based synthetic-ct generation in radiotherapy and pet: A review. Medical Physics (2021). https://doi.org/10.1002/mp.15150
11. Thummerer, A., van der Bijl, E., Galapon Jr, A., Verhoeff, J.J.C., Langendijk, J.A., Both, S., van den Berg, C.N.A.T., Maspero, M.: SynthRAD2023 Grand Challenge dataset: Generating synthetic CT for radiotherapy. Medical Physics **50**(7), 4664–4674 (2023). https://doi.org/10.1002/mp.16529, https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.16529, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.16529
12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
13. Zhao, B., Cheng, T., Zhang, X., Wang, J., Zhu, H., Zhao, R., Li, D., Zhang, Z., Yu, G.: Ct synthesis from mr in the pelvic area using residual transformer conditional gan. Computerized Medical Imaging and Graphics **103**, 102150 (2023)