



OPEN ACCESS

EDITED BY

Abdul K. Parchur,
University of Maryland Medical Center,
United States

REVIEWED BY

Jaidip Jagtap,
Mayo Clinic, United States
Varekan Keishing,
Mayo Clinic, United States

*CORRESPONDENCE

Yue Yan

✉ yue.yan@dartmouth.edu

RECEIVED 24 July 2025

ACCEPTED 08 September 2025

PUBLISHED 29 September 2025

CITATION

Asher GL, Wang S, Zaki BI, Russo GA, Gill GS, Thomas CR, Prioleau TO, Li Y, Zhang R, Yan Y and Hunt B (2025) Dosimetric evaluations using cycle-consistent generative adversarial network synthetic CT for MR-guided adaptive radiation therapy.
Front. Oncol. 15:1672778.
doi: 10.3389/fonc.2025.1672778

COPYRIGHT

© 2025 Asher, Wang, Zaki, Russo, Gill, Thomas, Prioleau, Li, Zhang, Yan and Hunt. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Dosimetric evaluations using cycle-consistent generative adversarial network synthetic CT for MR-guided adaptive radiation therapy

Gabriel L. Asher¹, Shiru Wang², Bassem I. Zaki^{3,4}, Gregory A. Russo^{3,4}, Gobind S. Gill⁴, Charles R. Thomas^{3,4}, Temiloluwa O. Prioleau¹, Yuting Li⁵, Rongxiao Zhang^{2,6}, Yue Yan^{2,3,4*} and Brady Hunt^{2,3}

¹Department of Computer Science, Dartmouth College, Hanover, NH, United States, ²Thayer School of Engineering, Dartmouth College, Hanover, NH, United States, ³Geisel School of Medicine, Dartmouth College, Hanover, NH, United States, ⁴Department of Radiation Oncology & Applied Sciences, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States, ⁵Department of Radiation Physics, University of Texas MD Anderson Cancer Center, Houston, TX, United States, ⁶Department of Radiation Oncology, University of Missouri, Columbia, MO, United States

Background: Magnetic resonance (MR) guided radiation therapy combines high-resolution image capabilities of MRI with the precise targeting of radiation therapy. However, MRI does not provide the essential electron density information for accurate dose calculation, which limit the application of MRI. In this presented work, we evaluated the potential for Deep Learning (DL) based synthetic CT (sCT) generation using 3D MRI setup scans acquired during real-time adaptive MRI-guided radiation therapy.

Methods: We trained and evaluated a Cycle-consistent Generative Adversarial Network (Cycle-GAN) using paired MRI and deformably registered CT scan slices (dCT) in the context of real-time adaptive MRI-guided radiation therapy. Synthetic CT (sCT) volumes are output from the MR to CT generator of the Cycle-GAN network. A retrospective study was conducted to train and evaluate the DL model using data from patients undergoing treatment for kidney, pancreas, liver, lung, bone, and prostate tumors. Data was partitioned by patients using a stratified k-fold approach to ensure balanced representation of treatment sites in the training and testing sets. Synthetic CT images were evaluated using mean absolute error in Hounsfield Units (HU) relative to dCT, and four image quality metrics (mean absolute error, structural similarity index measure, peak signal-to-noise ratio, and normalized cross correlation) using the deformed CT scans as a reference standard. Synthetic CT volumes were also imported into a clinical treatment planning system and dosimetric calculations re-evaluated for each treatment plan (absolute difference in delivered dose to 3cm radius of PTV).

Results: We trained the model using 8405 frames from 57 patients and evaluated it using a test set of 357 sCT frames from 17 patients. Quantitatively, sCTs were comparable to electron density of dCTs, while improving structural similarity with on-table MRI scans. The MAE between sCT and dCT was 49.2 ± 13.2 HU, sCT NCC outperformed dCT by 0.06, and SSIM and PSNR were 0.97 ± 0.01 and 19.9 ± 1.6

respectively. Furthermore, dosimetric evaluations revealed minimal differences between sCTs and dCTs. Qualitatively, superior reconstruction of air-bubbles in sCT compared to dCT reveal higher alignment between sCT than dCT with the associated MR.

Conclusions: Accuracy of deep learning based synthetic CT generation using setup scans on MR-Linacs was adequate for dose calculation/optimization. This can enable MR-only treatment planning workflows on MR-Linacs, thereby increasing the efficiency of simulation and adaptive planning for MRgRT.

KEYWORDS

deep learning, cycle-consistent generative adversarial network, deformable registration, MRI-guided radiation therapy, synthetic CT

1 Introduction

MR-guided radiation therapy (MRgRT) is a relatively new approach to radiation therapy (RT) which combines the high-resolution imaging capabilities of magnetic resonance imaging (MRI) with the precise targeting of radiation therapy. By using real-time MRI linear accelerator systems (MRI-LINAC) during treatment, MRgRT allows for more accurate targeting of the tumor, which can lead to improved outcomes and reduced side effects for patients (1). Several studies have demonstrated the benefits of MRgRT, including improved target coverage, reduced toxicity, and improved overall survival (2–4). Additionally, MRgRT has been shown to be effective in treating a variety of cancer types, including brain, prostate, and breast cancer (1, 5–7). Overall, MRgRT is an innovative approach to radiation therapy that has the potential to improve the radiation therapy workflow and patient outcomes.

However, a critical weakness of MRgRT is a reliance on electron density maps which are derived from computed tomography (CT) images for dose planning. Thus, to successfully carry out MRgRT, MR images must be co-registered with CT images. Co-registration of CT and MRI images is a critical step in MRgRT, however, it can also be a source of errors which can impact the accuracy of the treatment. The co-registration process aligns the CT and MRI images, allowing for precise targeting of the tumor, but it is prone to errors due to the inherent differences in the imaging modalities. CT images have limited soft tissue contrast and do not provide functional information about the tumor. On the other hand, MRI images have a better soft-tissue contrast, and provide functional information about the tumor, however, they are sensitive to motion artifacts and the presence of metallic objects and contrast agents (1, 8). The co-registration of CT and MRI images is conventionally performed manually, which can also be a source of errors if not done carefully. For example, this co-registration process has a systematic uncertainty of approximately 2–5 mm (9). The errors that can occur during the

co-registration process can persist at multiple levels of the treatment workflow and bring systematic errors (10).

Current MR to CT registration techniques involve generating a deformed CT (dCT) from a baseline MR image and an existing CT image of the patient. Many of the conventional CT image registration techniques are time consuming and costly (11). Computational time for atlas-based methods rises linearly with dataset size and bulk segmentation requires longer acquisition time compared to conventional MR sequences (12, 13). Thus, deep learning (DL) based synthetic CT registration has been touted as a promising alternative to previous registration techniques. These deep-learning techniques offer several benefits over other methods. These methods consistently achieve state-of-the-art results in terms of registration performance according to Dice score evaluation. Additionally, these methods offer considerable speedups over traditional registration methods (12, 14, 15). These advantages make deep-learning based CT registration an optimal choice for real-time dose-calculations with novel MR-LINAC systems.

However, despite these advantages, there are several significant hurdles for training generalizable deep-learning synthetic CT generation models. First and foremost, datasets are often small. Training machine-learning (ML) models require ground truth labels, which means that for a MR to CT registration task, paired CT/MR datasets are needed. Thus, datasets used in training synthetic CT models tend to be small by machine learning standards and can be afflicted by batch errors due to limited diversity in acquisition settings such as MR and CT machines. Additionally, many datasets suffer from data leakage, which inflates model performance (16). Finally, large differences between intensity values from different MRI manufacturers means that trained models may struggle to generalize well.

Many studies evaluate the performance of various deep-learning architectures on synthetic CT registration from MR images. Generative adversarial networks (GAN) are among the more popular techniques for mapping MR images to sCTs (17).

However, GANs require strongly paired ground truths to train properly. This poses an issue in the context of dataset generation since MR images and CT images cannot be captured simultaneously. Thus, CycleGANs have shown themselves to be a promising method for synthetic CT generation. Due to a cycle-consistency loss, CycleGANs can train on paired or unpaired data, greatly increasing dataset size. Results from related CycleGAN studies have shown promise, with mean absolute error between ground-truth CT volume and sCT volumes between 30–150 Hounsfield Units (HU) (18–22). However, these studies demand further investigation for the following reasons. Firstly, most prior models were only trained with one or two sites in mind. Lei et al. study the brain and pelvis (21), Farjam et al. the pancreas (23), Wolterink et al. the brain (20), and Yang et al. study the brain and abdomen (with different models for each) (18). Furthermore, due to the limited size of datasets, many prior studies, such as Kang et al, Farjam et al, and Lei et al, do not use a held out validation data set in addition to their final test set (21–23). The lack of this validation set means that these models implicitly overfit on the test data since hyperparameters can be directly tuned on test data. Finally, although some studies evaluated treatment planning dose volumetrics on synthetic CTs (22), many of these studies did not (18–21, 23).

Thus, in this research study, we aim to further evaluate the performance of CycleGAN, a deep-learning algorithm, on synthetic CT generation. We employ a novel, large, paired image dataset with 6 different sites. Furthermore, we employ a rigorous data splitting regime to ensure minimal data leakage and the most generalizable results possible. Finally, we use the model trained from this dataset to investigate two tasks. Firstly, we assess the performance of CycleGAN in generating synthetic CT images that accurately correspond to their ground truths. Secondly, we evaluate our synthetic CT images with treatment planning software to determine whether they have adequate dosimetric outcomes to enable MR-only planning on MR-LINACs.

2 Materials and methods

2.1 Study dataset

This study analyzed patients undergoing stereotactic body radiation therapy using the ViewRay MRIdian MR-LINAC at the Dartmouth-Hitchcock Medical Center (DHMC). Patients underwent radiation therapy between March 2021 and June 2022. All DICOM and treatment delivery records were retrospectively accessed and anonymized before inclusion in the study in accordance with a protocol approved by the DHMC institutional review board. Simulation CT/MR scans were acquired on the same day at the outset of RT treatment planning, typically one hour between scans with CT scanning first. CT scans were acquired using a simulation scanner (Siemens EDGE) using routine clinical settings. MRI scans were obtained using a MR-Linac (ViewRay

MRIdian) using built-in clinical protocols. Ground-truth deformed CTs (dCT) were generated using the ViewRay treatment planning system registration pipeline.

2.2 Image pre-processing and data partitioning

After dCT generation, all dCT and MR images were extracted in DICOM format with 144 slices and a 3mm axial resolution. The in-plane dimensions were 310 x 360 pixels with a 1.5mm resolution. DICOM volumes were converted to tiff images using the python open-source pydicom and tifffile packages. After conversion to tiff format, all images were padded to 440x440 pixels. Next, we conducted a stratified k-fold data splitting scheme using treatment site specified in the DICOM data as the category for stratification including: adrenal, pancreas, liver, lung, bone, prostate, and other. This split yielded 58 training patients, 11 validation patients, and 17 testing patients, ensuring that train and test splits contained at least one patient from each treatment site category. Prior to use in model training, all 8405 CT and MR images were normalized. CT images were normalized with the following linear formula:

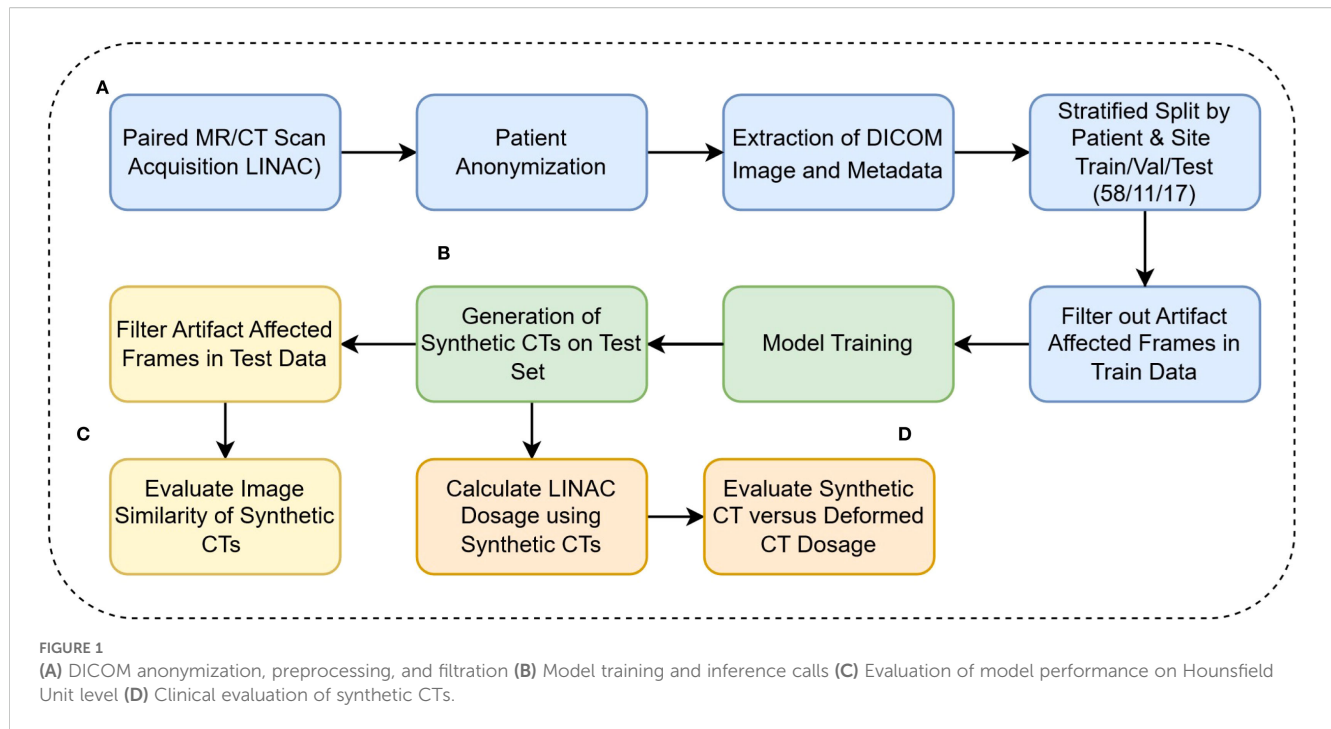
$$Y = \frac{X - X_0}{\alpha} \quad (1)$$

where X is the input voxel in HU, $X_0 = -1024$, and α = the 99.99 percentile of HU values in the training set range. To ensure that all normalized values were positive, in Equation 1, X_0 was set to -1024, considering that the minimum HU value in our training set was also -1024. MR images were also normalized with Eq.1, with X_0 set at 0, since MR intensity values are all already nonnegative. Our pre-processing pipeline is described in Figure 1.

2.3 Model and loss formulation

Our study utilized the CycleGAN (Cycle-Consistent Generative Adversarial Network) model architecture (24). Training workflow of the CycleGAN model has been shown in Figure 2. Figure 3 details the architecture of CycleGAN. This architecture is a derivative of the generative adversarial network (GAN) (25), a popular deep-learning architecture which leverages two competing networks: a generator and a discriminator. In GANs, a generator generates a synthetic image from an input image, and the discriminator predicts whether this synthetic image is real or fake. The model stops learning when the generator produces indistinguishable images from the ground truth.

CycleGAN adopts a similar architecture as a GAN with two key differences. Firstly, instead of having one generator and one discriminator, CycleGAN has two generators (G_{MR}, G_{CT}) and two discriminators (D_{MR}, D_{CT}). The generator G_{CT} takes a CT image and generates an MR image, while the generator G_{MR} takes an MR image



and generates a CT image. The loss functions used to train the generators and discriminators in a Cycle-GAN typically include three components: the adversarial loss, the cycle-consistency loss, and the identity loss. Adversarial loss is used to ensure that the generated images are realistic and can fool the discriminators. The adversarial loss, L_{adv} is calculated using the mean squared error between the discriminator's output and the ground-truth label. The loss formulation of our discriminators is a summation of the adversarial loss with sCT as input and adversarial loss with dCT as input. Cycle-consistency loss is used to ensure that the generated images preserve the content of the original images and is calculated as the mean absolute error (MAE) between the original image and the translated image that has been translated back to the original domain. The cycle loss, L_{cyc} for the generator G_{CT} is calculated in Equation 2 as:

$$L_{cyc} = \|G_{MR}(G_{CT}(CT_{gt})) - CT_{gt}\| \cdot \lambda \quad (2)$$

, where λ is a weight, and CT_{gt} is a ground truth CT. The identity loss is added to ensure that the images from the same domain should not change after passing through the generator. The identity loss, as shown in Equation 3, L_{idt} for G_{CT} is:

$$L_{idt} = \|G_{CT}(CT_{gt}) - CT_{gt}\| \quad (3)$$

The loss formulations for G_{MR} are formulated similarly as G_{CT} . Finally, the total loss as shown in Equation 4, L_{total} is calculated as follows:

$$L_{total} = L_{advMR} + L_{advCT} + L_{cycleMR} + L_{cycleCT} + L_{idtMR} + L_{idtCT} \quad (4)$$

We trained our model on 1 Nvidia RTX 2080 Ti with 12GB of GPU memory. Additionally, we used the following model

parameters: random crop to 256x256, batch size of 1, 100 epochs. Our model training code was from: <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix> (24, 26).

2.4 Evaluations

After model training, we used the lowest loss model (measured on the validation set) to generate a set of 2464 synthetic CT images from a held-out test set. However, whilst evaluating image similarity, we found a large amount of our test set dCTs were corrupted by artifacts in the beginning and end of each series in the axial plane. Thus, we only conducted our evaluations on the dosimetric relevant images within a 3cm radius of the PTV.

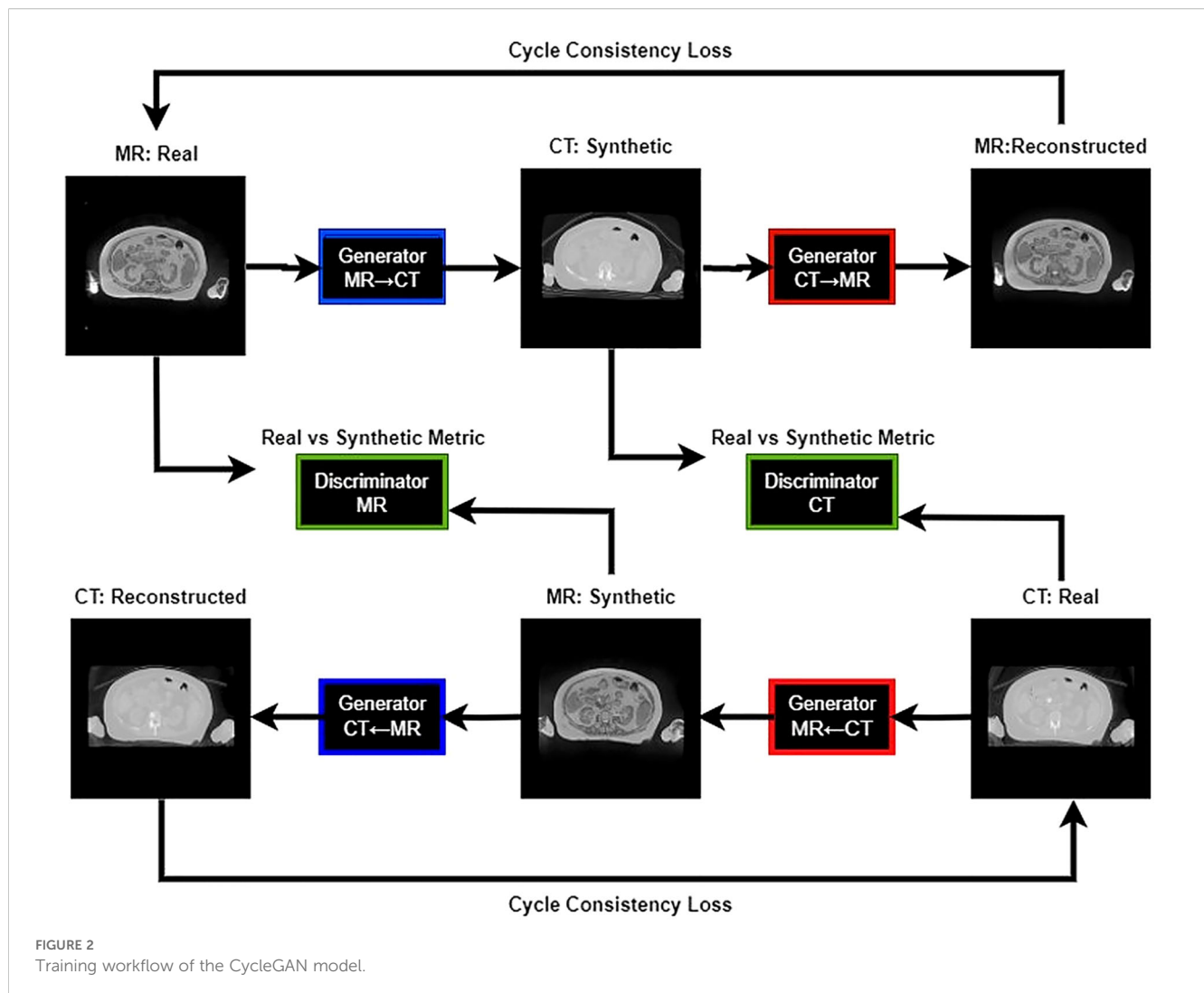
2.4.1 Synthetic image quality assessments

We first evaluated our test set through image quality and similarity metrics. We quantitatively measured the similarity between our sCT and dCT through mean average error (MAE), peak to signal noise ratio (PSNR), and structural similarity (SSIM). The formulas for these metrics are shown in Equations 5–7):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{defCT} - y_{synCT}| \quad (5)$$

$$PSNR = 10 \cdot \log_{10} \frac{\max(y_{defCT})^2}{\frac{1}{n} (\sum_{i=1}^n (y_{defCT} - y_{synCT})^2)} \quad (6)$$

$$SSIM(x, y) = \frac{(2\mu_{ix}\mu_{iy} + C_1)(2\sigma_{ixiy} + C_2)}{(\mu_{ix}^2 + \mu_{iy}^2 + C_1)(\sigma_{ix}^2 + \sigma_{iy}^2 + C_2)} \quad (7)$$



For SSIM, μ_{ix} , u_{iy} are the average HU values for the i th axial slice of our dCT and sCT series respectively. $\sigma_{ix}^2, \sigma_{iy}^2$ represent the variance of the aforementioned dCT and sCT slices. Finally, C_1, C_2 are constants applied as suggested by Wang et al (27). Additionally, we evaluated the similarity of sCT and dCT images to their corresponding MR inputs using Normalized Cross-Correlation (NCC). Normalized cross correlation (NCC) is a similarity measure that ranges from -1 to 1, used to determine the degree of similarity between two image regions, with 1 being most similar. NCC was calculated using the xcdskd package on python.

2.4.2 Dosimetric assessments

The second evaluation criteria of our sCTs consists of a comparison between the RT dose calculation in the sCT versus the dCT. In order to calculate these differences, our tiff images were reconverted to DICOM format. This was done by replacing the dCT DICOM “Pixel Data” tag with our generated sCT image. Next, we fed our DICOM sCTs into the ViewRay treatment planning software with the same parameters used on the dCTs to create dose volume histograms (DVHs). These DVHs were used to calculate the absolute difference of dose delivered to the PTV at

95%, 90%, and 85% of the volume. Additionally, we also calculate the difference in dose delivered above 33Gy to all 3cm OAR sites.

3 Results

3.1 Image quantitative comparisons

Table 1 reports the mean absolute error (MAE) of our cycleGAN model. In total, 357 synthetic CT frames were analyzed from 17 patients. Of the treatment sites analyzed, the most common sites were liver, pancreas, and lung, with 84, 105, and 63 frames respectively. MAE was calculated by comparing synthetic CTs versus deformed CTs. Median MAE values across sites ranged between 44.7.4 HU to 55.6 HU with an overall median MAE of 49.2 HU. Pancreas and lung scans had the highest MAE standard deviations with 20.1 and 9.8 HU respectively. Kidney showed the least difference in MAE between scans with a standard deviation of 1.2 HU.

Table 2 reports SSIM and PSNR metrics, as well as the NCC of our sCTs and dCTs versus our MR image ground truths. Our sCTs

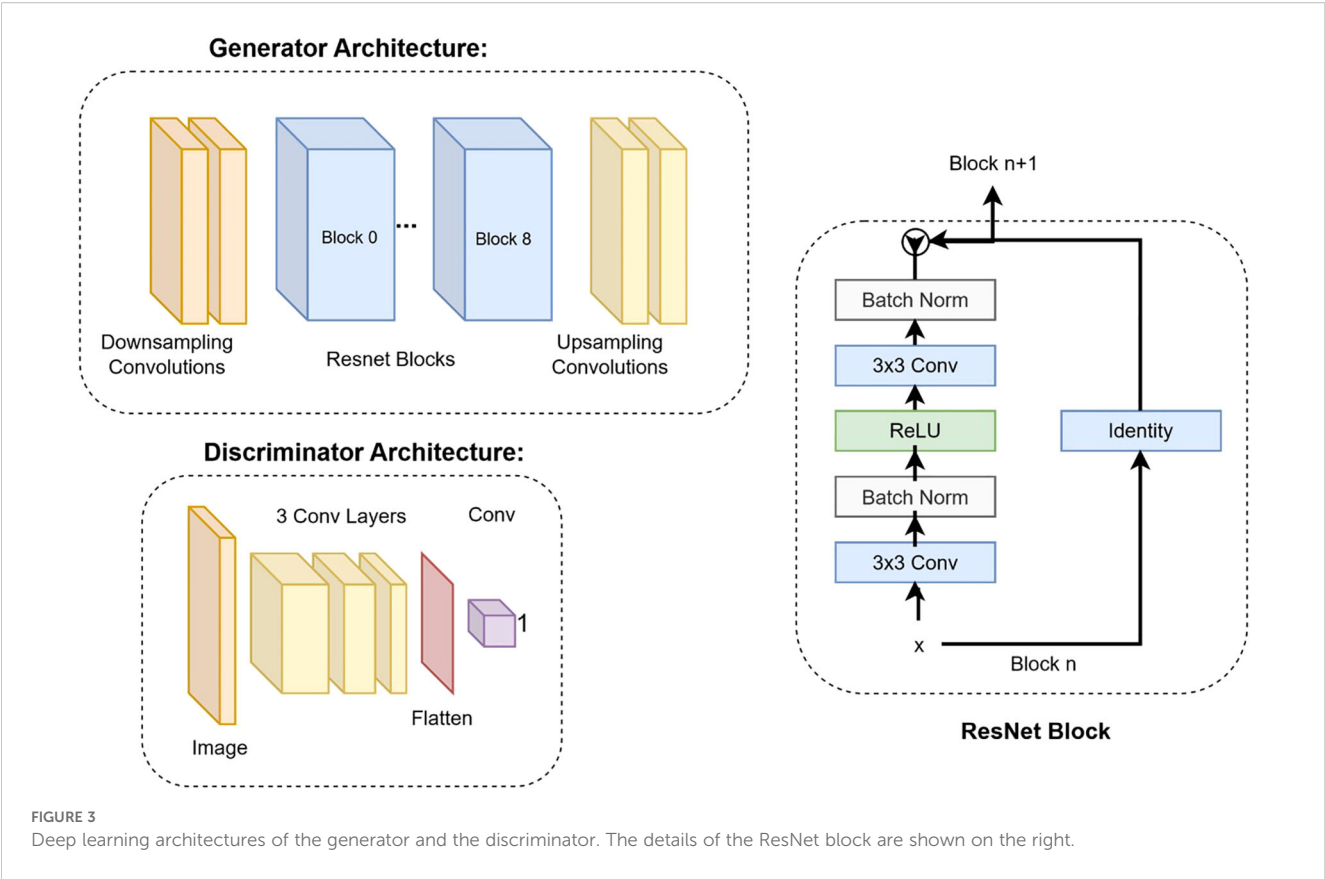


TABLE 1 The mean absolute error (MAE) of our cycle-GAN model per site.

Value description	Other	Kidney	Pancreas	Liver	Lung	Bone	Prostate	All
MAE min (HU)	44.4	42.6	37.7	35.6	46.9	52.3	48.3	35.6
MAE median (HU)	49.4	44.7	47.6	44.4	55.6	55.2	51.4	49.2
MAE max (HU)	57.7	47	100.4	67.5	87.8	58.3	53.9	100.4
std (HU)	4.3	1.2	20.1	7.7	9.8	1.6	1.8	13.2
Number of frames	42	21	105	84	63	21	21	357
Number of patients	2	1	5	4	3	1	1	17

TABLE 2 Image quality metrics per site. NCC refers to normalized-cross-correlation, SSIM refers to structural similarity, and PSNR refers to peak signal to noise ratio.

Value description	Other	Kidney	Pancreas	Liver	Lung	Bone	Prostate	All
Deformed NCC	0.769	0.825	0.827	0.839	0.826	0.803	0.707	0.818
Synthetic NCC	0.763	0.826	0.83	0.847	0.83	0.812	0.718	0.824
Median SSIM	0.974	0.978	0.969	0.974	0.967	0.965	0.976	0.971
SSIM std	0.002	0.002	0.016	0.008	0.015	0.003	0.002	0.013
Median PSNR	20.3	20.8	19.6	19.9	19.3	19	20.4	19.9
PSNR std	0.98	0.76	2.05	1.54	1.51	0.17	0.33	1.64
Number of frames	42	21	105	84	63	21	21	357
Number of patients	2	1	5	4	3	1	1	17

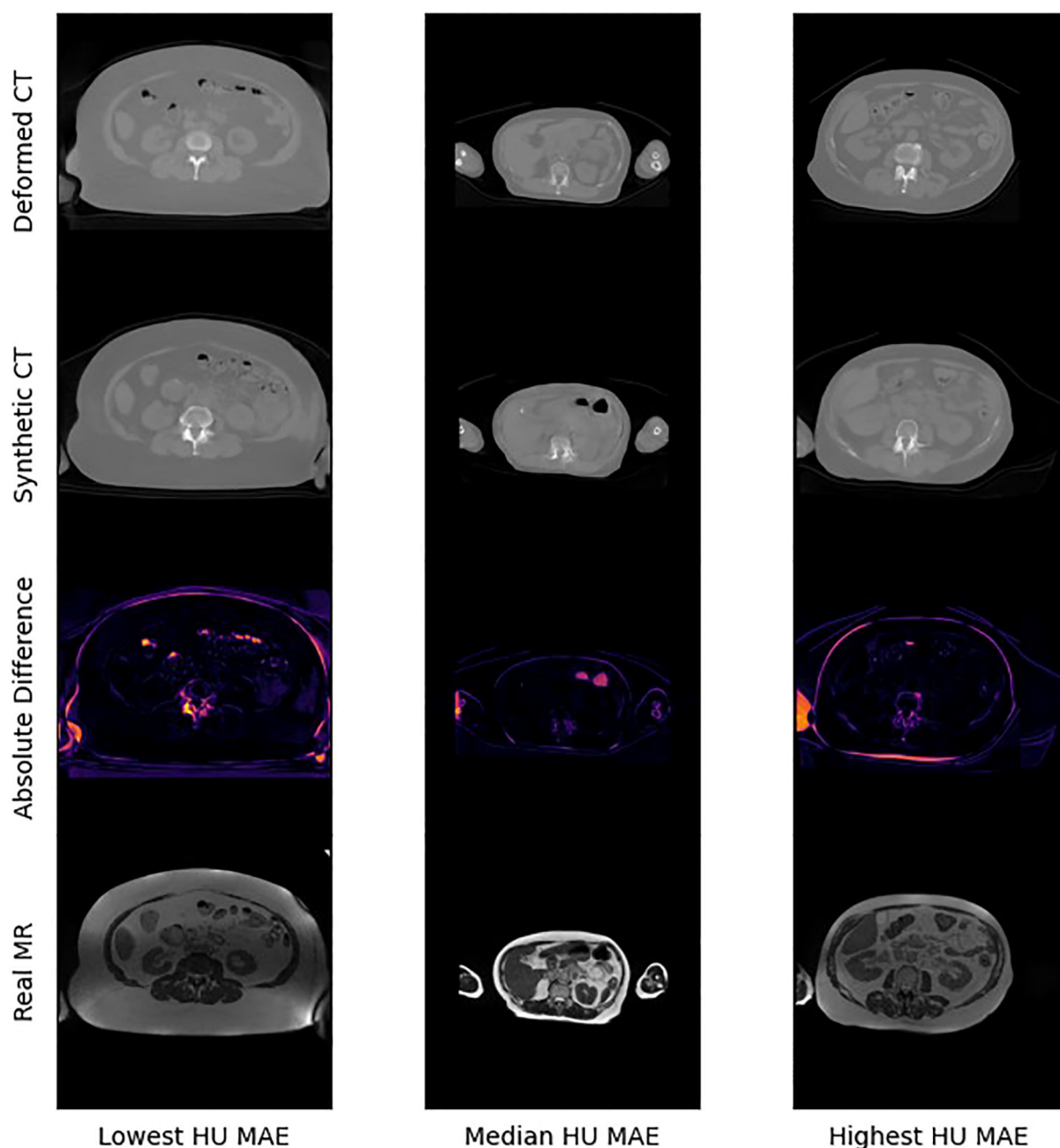


FIGURE 4
The lowest, median, and highest sCT mean average errors (HU) in the test set.

demonstrate a higher (better) NCC value in comparison to deformed CTs in all but one treatment site (Other). Although differences were generally minimal, sCT NCC scores on bone, prostate, and liver showed the largest improvements compared to our deformable registration ground truths. Additionally, we report high SSIM values across all sites, with a median SSIM value of 0.971. Thus, the structural similarity between our sCTs and dCTs is near perfect. Additionally, we report a median PSNR of 19.9 across all sites. This high PSNR value indicates that the sCT is a good representation of the dCT.

Figure 4 presents selected images from the scans with the lowest, median, and highest MAE (HU) from our test set. Figure 5 shows zoomed in panel of spinal region of interest. We compare sCT spinal reconstruction on lowest, median, and highest HU MAE patients. As is evident in all three examples, our model struggled to properly predict skin, bone, and limbs outside of the torso region (arms). Difficulty predicting bone is consistent with prior works (20, 21). Additionally, we view differences between the sCT and dCT with regards to air bubbles. For example, in the median HU images, we observe that our synthetic CT images

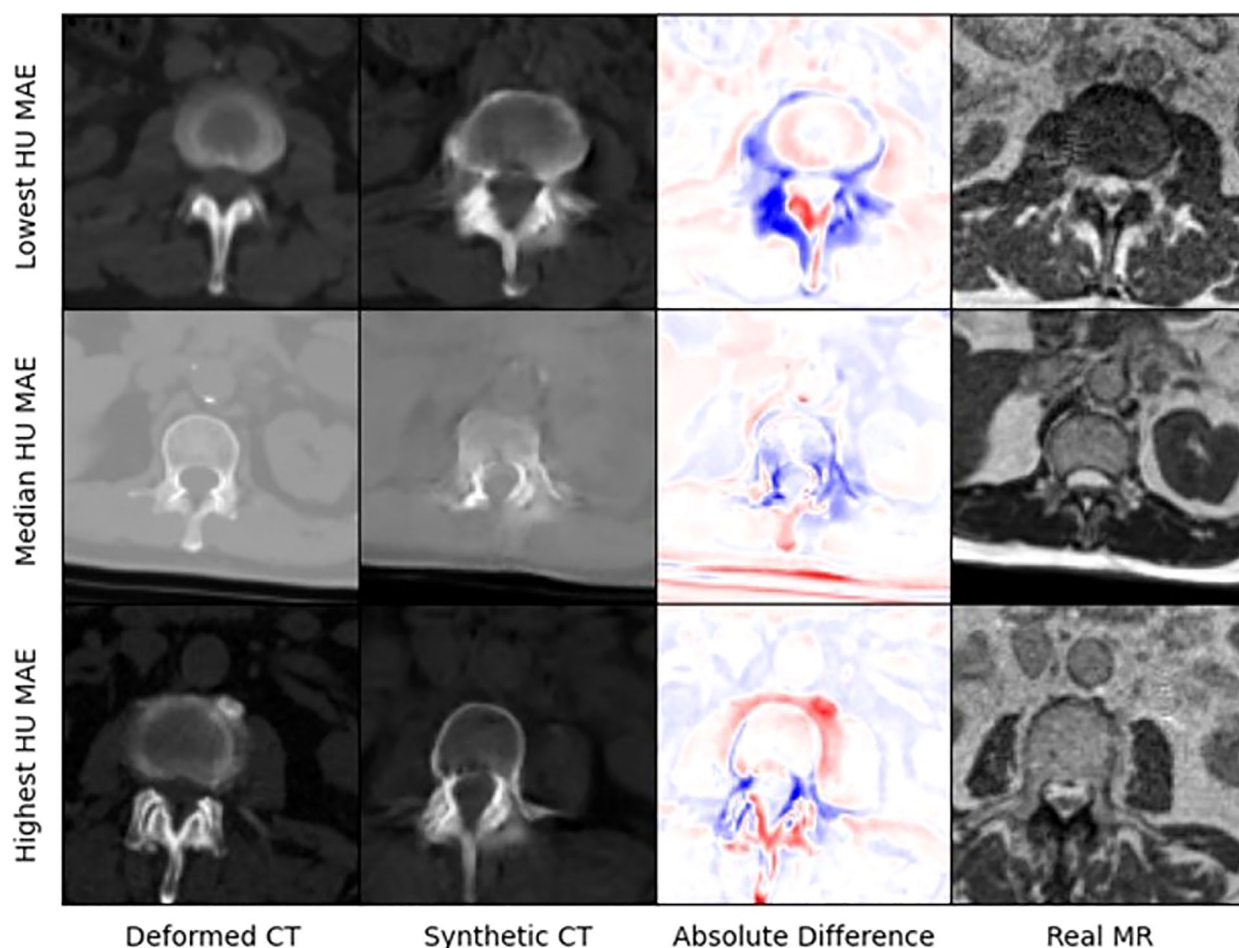


FIGURE 5

Zoomed in panel of spinal region of interest. We compare sCT spinal reconstruction on lowest, median, and highest HU MAE patients.

correctly include air bubbles present in MR, whereas the deformed CTs do not include said bubbles. However, from a qualitative perspective, our sCT reconstructions are minimally different from their deformed CT counterparts.

3.2 Dose comparisons

Table 3 reports the absolute difference of RT dose delivered to PTV at 85%, 90%, and 95% of the volume based on RT dose recalculations using our sCTs. Averaged sCT vs CT dose volumetrics per site is shown in Figure 6. Our results indicate minimal differences in dose delivered using sCTs. Median difference in dose delivered to PTV across all sites is 45Gy, 47Gy, and 46Gy for 85%, 90%, and 95% of volume respectively. Additionally, we found that dose above 33Gy delivered to OARs within 3cm of the PTV also showed minimum differences between sCT and dCT RT dose calculations. The median difference in dose delivered to 3cm OARs minimally increased by 0.01 Gy after switching from sCT to dCT. Additionally, 4 patients had no change in dose delivered to these

sites, and 6 patients had less dose above 33Gy delivered to these sites.

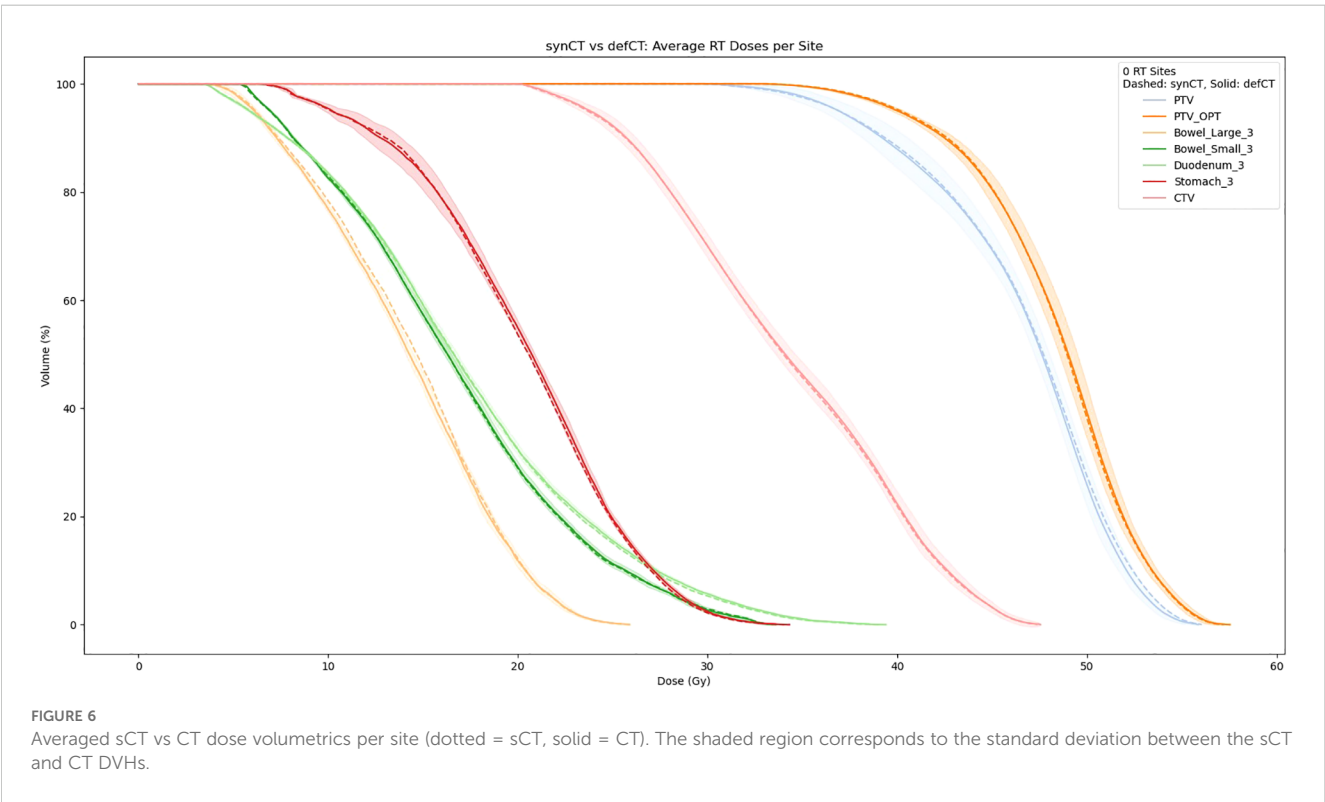
4 Discussion

This report presents strong results supporting an MRI-only RT workflow. From a synthetic image quality perspective, our CycleGAN implementation reports comparable or superior MAE compared to prior studies conducted using a CycleGAN architecture (18, 28). Additionally, our results indicate that CycleGAN architecture generalizes well to several treatment sites with minimal additional training data. For example, despite only having 1 kidney, bone, and 3 prostate series each in our training data, we still report a mean MAE of 50.4 HU across these sites. Also, our overall median MAE of 49.2 HU is a strong result in comparison to prior work. This improved synthetic image quality may be a result of a larger dataset. Of the prior CycleGAN studies performed, this study analyzed 86 patient scans, compared with 24, 45, and 38, patients in works from Wolterink et al. (20), Yang et al

TABLE 3 DVH metrics per patient. Reports absolute difference between sCT and CT of dose delivered to 85%, 90%, and 95%.

Patient	Site	Gy diff@85%	Gy diff@90%	Gy diff@95%	Diff OAR dose > 33gy
PT0049	OTHER	0.19	0.21	0.23	0
PT0059	KIDNEY	0.5	0.49	0.47	-0.16
PT0066	PANCREAS	0.23	0.14	0.09	0.07
PT0067	LIVER	0	0.06	0.01	0
PT0069	LIVER	0.48	0.56	0.47	0
PT0070	PANCREAS	1.56	1.16	1.21	-0.26
PT0074	LIVER	0.04	0.04	0.05	0
PT0075	LIVER	0.86	0.82	0.28	-0.18
PT0076	PANCREAS	0.7	0.7	0.45	0.65
PT0077	LUNG	0.16	0.04	0.42	0.98
PT0078	PANCREAS	0.12	0.07	0.01	-0.19
PT0080	PANCREAS	0.19	0.12	0.21	0.08
PT0081	BONE	0.57	0.57	0.41	NAN
PT0082	PROSTATE	0.16	0.16	0.17	0.25
PT0083	LUNG	1.23	1.64	1.76	0.11
PT0084	LUNG	0.25	0.68	1.13	-1.11
PT0085	OTHER	NAN	NAN	NAN	-0.03
ALL	ALL	0.45	0.47	0.46	0.01

This table also reports the difference in dose delivered above 33Gy in 3cm OAR sites, where a negative value indicates that less dose above 33Gy was delivered with the sCT.



(18), and Brou et al (28). Thus, given our superior reconstruction results, we demonstrate that dataset size is paramount in creating strong generative synthetic CT models.

Our treatment dose comparison also supports an MRI-only RT workflow. Our 0.45 Gy, 0.47 Gy, and 0.46 Gy average difference in dose delivered to PTV for 85%, 90%, and 95% of the volume indicates that sCTs have minimal effects on dose delivered. Additionally, NCC comparisons in Table 4 indicate that synthetic CTs capture an equivalent to better representation of the MR scans taken during treatment. Thus, we can extrapolate that synthetic CT scans may present a more precise image to calculate RT dose with. Exemplary of this claim are the air bubbles in the median MAE frame of Figure 3. From a visual comparison to the real MR, we can see that the synthetic CT better models air bubbles. Thus, the synthetic CT is likely a better representation to use when performing dose calculations. Another observed benefit of synthetic CTs is the elimination of artifacts compared to deformable CTs. Given that sCT generation is wholly dependent on the MR image fed into the model, if this MR image has no artifacts the resulting sCT will be artifact-free as well.

The key strength of using a deep learning model in synthetic CT generation is speed. Conventional image registration techniques rely on an iterative image update process. However, this process is slow and requires lengthy computation, which is a bottleneck when performing real-time dose calculation. On the other hand, our deep-learning model is fixed after training and requires a single forward pass to generate an output. Therefore, our DL model is much more suited to real-time RT strategies.

Some limitations of our approach include limited frame to frame cohesion on the axial plane. We observed that although primary image structures and features remained fixed, there were some frame-to-frame shifts in axial position along the extremities of each scan. These shifts may have occurred because the CycleGAN architecture trains on a single image at a time as opposed to a whole volume. Therefore, the model has more difficulty learning frame to frame continuance. Another observed limitation to our method is that our model has difficulty predicting HU values for high-intensity regions. For example, in Figure 4, we see that for all 3 image examples our

model had difficulty reconstructing and predicting spine intensity values. We believe that this is due to the intrinsic distribution of HU values. Bone HU values are typically above 700, whereas all other tissue HU values fall between -100 and 300 HU. Thus, distribution imbalances may have led to difficulties predicting HU values for high-intensity regions. Finally, we believe that some distribution-matching losses may have caused anomalously shaped structures in certain already hard-to-predict regions such as the spine. For example, the spinal structure of all three sCTs in Figure 4 differs from their dCT ground truths. Given their relative similarity to each other, we believe that these differences may correspond to hallucinated features caused by distribution-matching losses, a phenomenon previously observed in Cycle-GAN based image translation (29).

Future work will involve two key advancements. Firstly, larger and higher quality datasets must be created. Our study shows that despite similar architectures; dataset scale improved our results in comparison to prior studies. Additionally, we believe that on top of scaling dataset, some incorporation of newer DL architectures could improve results. For example, although some work has already been done on MRI to CT conversion using diffusion models (30), given the remarkable performance of diffusion models in other image-processing domains (31), these models are promising for a synthetic CT generation task. As our dataset and compute resources grow, we plan to evaluate volumetric (3D) CycleGAN variants to improve through-plane consistency and integrate them into our framework.

5 Conclusions

In this study, we developed a deep learning algorithm based on CycleGAN (24) to derive sCT from MRI. We have demonstrated that the sCT provides comparable dose accuracy as the clinical CT. Accuracy of deep learning based synthetic CT generation using setup scans on MR-Linacs was adequate for dose calculation/optimization. This can enable MR-only treatment planning workflows on MR-Linacs, thereby increasing the efficiency of simulation and adaptive planning for MRgRT.

TABLE 4 Quantitative evaluation of synthetic versus deformed images across multiple organs.

Value Description	LIVER	PANCREAS	KIDNEY	PROSTATE	LUNG	OTHER	BONE	ALL
Deformed NCC	0.827	0.825	0.817	0.685	0.827	0.759	0.808	0.815
Synthetic NCC	0.838	0.831	0.815	0.705	0.830	0.778	0.824	0.824
SSIM	0.967	0.966	0.974	0.953	0.967	0.969	0.956	0.966
SSIMSTD	0.028	0.052	0.038	0.079	0.010	0.017	0.009	0.040
PSNR	19.406	19.232	20.614	17.978	19.395	19.015	18.496	19.150
PSNR STD	2.050	2.649	1.685	3.367	1.539	1.418	0.923	2.235
Number of Frames	235	340	72	76	228	152	72	1175
Number of Patients	4	5	1	1	3	2	1	17

NCC stands for normalized cross-correlation. SSIM stands for structural similarity. PSNR stands for peak to signal noise ratio.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Dartmouth-Hitchcock Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because The study is a retrospective study based on previous patients' data.

Author contributions

GA: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. SW: Methodology, Writing – review & editing. BZ: Data curation, Writing – review & editing. GR: Data curation, Writing – review & editing. GG: Data curation, Writing – review & editing. CT: Data curation, Resources, Writing – review & editing. TP: Methodology, Writing – review & editing. YL: Funding acquisition, Writing – review & editing, Validation. RZ: Supervision, Writing – review & editing. YY: Funding acquisition, Supervision, Writing – review & editing. BH: Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported

in part by funding from the Dartmouth-Hitchcock Medical Center Department of Medicine through the Scholarship Enrichment in Academic Medicine program.

Conflict of interest

The authors declare that the research was developed in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Cao Y, Tseng CL, Balter JM, Teng F, Parmar HA, Sahgal A, et al. MR guided radiation therapy: transformative technology and its role in the central nervous system. *Neuro Oncol.* (2017) 19:ii16–29.
2. Chin S, Eccles CL, McWilliam A, Chuter R, Walker E, Whitehurst P, et al. Magnetic resonance-guided radiation therapy: a review. *J Med Imaging Radiat Oncol.* (2020) 64:163–77.
3. Wang W, Dumoulin CL, Viswanathan AN, Tse ZTH, Mehrtash A, Loew W, et al. Real-time active MR-tracking of metallic stylets in MR-guided radiation therapy. *Magn Reson Med.* (2015) 73:1803–11.
4. Wooten HO, Rodriguez VL, Green OL, Kashani R, Santanam L, Tanderup K, et al. Benchmark IMRT evaluation of a Co-60 MRI-guided radiation therapy system. *Radiother Oncol.* (2015) 114:402–5.
5. Tenhunen M, Korhonen J, Kapanen M, Seppälä T, Koivula L, Collan J, et al. MRI-only based radiation therapy of prostate cancer: workflow and early clinical experience. *Acta Oncol.* (2018) 57:902–7.
6. Bohoudi O, Bruynzeel AME, Senan S, Cuijpers JP, Slotman BJ, Lagerwaard FJ, et al. Fast and robust online adaptive planning in stereotactic MR-guided adaptive radiation therapy (SMART) for pancreatic cancer. *Radiother Oncol.* (2017) 125:439–44.
7. Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of image-guided radiotherapy will be MR guided. *Br J Radiol.* (2017) 90:20160667.
8. Hargreaves BA, Worters PW, Pauly KB, Pauly JM, Koch KM, Gold GE. Metal-induced artifacts in MRI. *AJR Am J Roentgenol.* (2011) 197:547–55.
9. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol.* (2017) 12:28.
10. Owangi AM, Greer PB, GlideHurst CK. MRI-only treatment planning: benefits and challenges. *Phys Med Biol.* (2018) 63:05TR01.
11. Liu Y, Lei Y, Wang T, Fu Y, Tang X, Curran WJ, et al. MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method. *Br J Radiol.* (2019) 92:20190067.
12. Arabi H, Dowling JA, Burgos N, Han X, Greer PB, Koutsouvelis N, et al. Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region. *Med Phys.* (2018) 45:5218–5233.
13. Rank CM, Tremmel C, Hünemohr N, Nagel AM, Jäkel O, Grelich S. MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. *Radiat Oncol.* (2013) 8:1–13.
14. De Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal.* (2019) 52:128–43.
15. Sentker T, Madesta F, Werner R. Gdl-fire: Deep learning-based fast 4d ct image registration. In: *International conference on medical image computing and computer-assisted intervention*. Springer International Publishing, Cham (2018). p. 765–73.
16. Yagis E, Atnafu SW, de Herrera AGS, Marzi C, Giannelli M, Tessa C, et al. Deep learning in brain MRI: effect of data leakage due to slice-level split using 2D convolutional neural networks. (2021).

17. Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical image synthesis with context-aware generative adversarial networks, in: *Medical Image Computing and Computer Assisted Intervention– MICCAI 2017: 20th International Conference*, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part III 20. *Lecture Notes in Computer Science(LNIP)*, Springer, Cham, Switzerland: Springer International Publishing. (2017) 10435, 417–25.
18. Yang H, Zhang Y, Chen Y, Zhang Y, Zhang H, Cai W, et al. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE Trans Med Imaging*. (2020) 39:4249–61.
19. Olberg S, Choi BS, Park I, Liang X, Kim JS, Deng J, et al. Ensemble learning and personalized training for the improvement of unsupervised deep learning-based synthetic CT reconstruction. *Med Phys*. (2023) 50:1436–49.
20. Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Išgum I. Deep MR to CT synthesis using unpaired data, in: *Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017*, Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017 Proceedings 2. *Lecture Notes in Computer Science(LNIP)*, Springer, Cham, Switzerland: Springer International Publishing. (2017) 10557, 14–23.
21. Lei Y, Harms J, Wang T, Liu Y, Shu HK, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. (2019) 46:3565–81.
22. Kang SK, An HJ, Jin H, Kim JI, Chie EK, Park JM, et al. Synthetic CT generation from weakly paired MR images using cycle-consistent GAN for MR-guided radiotherapy. *BioMed Eng Lett*. (2021) 11:263–71.
23. Farjam R, Nagar H, Kathy Zhou X, Ouellette D, Chiara Formenti S, DeWyngaert JK. Deep learning-based synthetic CT generation for MR-only radiotherapy of prostate cancer patients with 0.35 T MRI linear accelerator. *J Appl Clin Med Phys*. (2021) 22:93–104.
24. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE international conference on computer vision (ICCV 2017)*. Los Alamitos, California, USA: IEEE. (2017) pp. 2223–32.
25. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Commun ACM*. (2020) 63:139–44.
26. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017)*. Los Alamitos, California, USA: IEEE. (2017). pp. 1125–34.
27. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: *The thirty-seventh asilomar conference on signals, systems & Computers, 2003*, vol. 2. Los Alamitos, California, USA: IEEE (2003). p. 1398–402.
28. Brou Boni KN, Klein J, Gulyban A, Reynaert N, Pasquier D. Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data. *Med Phys*. (2021) 48:3003–10.
29. Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation, in: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain*, September 16–20, 2018 Proceedings, Part I. *Lecture Notes in Computer Science(LNIP)*, Springer, Cham, Switzerland: Springer International Publishing (2018) 11070, 529–36.
30. Lyu Q, Wang G. Conversion between CT and MRI images using diffusion and score-matching models. *arXiv preprint arXiv:2209.12104*. (2022).
31. Yang L, Zhang Z, Song Y, Hong S, Xu R, Zhao Y, et al. Diffusion models: A comprehensive survey of methods and applications. *ACM. Comput Surv*. (2023) 56:1–39.