Unsupervised pseudo CT generation using heterogenous multicentric CT/ MR images and CycleGAN

Jabbarpour, Amir; Mahdavi, Seied Rabi; Sadr, Alireza Vafaei; Esmaili, Golbarg; Shiri, Isaac; Zaidi, Habib

Published in:
Computers in biology and medicine

DOI:
10.1016/j.compbiomed.2022.105277

Publication date:
2022

Computers in Biology and Medicine 143 (2022) 105277

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Unsupervised pseudo CT generation using heterogenous multicentric CT/MR images and CycleGAN: Dosimetric assessment for 3D conformal radiotherapy

Amir Jabbarpour [a], Seied Rabi Mahdavi [a,b,*], Alireza Vafaei Sadr [c,i], Golbarg Esmaili [d], Isaac Shiri [e,f,g,h], Habib Zaidi [e,f,g,h]

[a] Medical Physics Department, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
[b] Radiation Biology Research Center, Iran University of Medical Sciences, Tehran, Iran
[c] Institute of Pathology, RWTH Aachen University Hospital, Aachen, Germany
[d] Pars Hospital, Tehran, Iran
[e] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211, Geneva 4, Switzerland
[f] Geneva University Neurocenter, Geneva University, Geneva, Switzerland
[g] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
[h] Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark
[i] Department of Theoretical Physics and Center for Astroparticle Physics, Geneva University, Geneva, Switzerland

ARTICLE INFO

Keywords:
MRI-Only radiotherapy
Brain tumors
Unsupervised deep learning
CycleGAN

ABSTRACT

Purpose: Absorbed dose calculation in magnetic resonance-guided radiation therapy (MRgRT) is commonly based on pseudo CT (pCT) images. This study investigated the feasibility of unsupervised pCT generation from MRI using a cycle generative adversarial network (CycleGAN) and a heterogenous multicentric dataset. A dosimetric analysis in three-dimensional conformal radiotherapy (3DCRT) planning was also performed.

Material and methods: Overall, 87 T1-weighted and 102 T2-weighted MR images alongside with their corresponding computed tomography (CT) images of brain cancer patients from multiple centers were used. Initially, images underwent a number of preprocessing steps, including rigid registration, novel CT Masker, N4 bias field correction, resampling, resizing, and rescaling. To overcome the gradient vanishing problem, residual blocks and mean squared error (MSE) loss function were utilized in the generator and in both networks (generator and discriminator), respectively. The CycleGAN was trained and validated using 70 T1 and 80 T2 randomly selected patients in an unsupervised manner. The remaining patients were used as a holdout test set to report final evaluation metrics. The generated pCTs were validated in the context of 3DCRT.

Results: The CycleGAN model using masked T2 images achieved better performance with a mean absolute error (MAE) of 61.87 ± 22.58 HU, peak signal to noise ratio (PSNR) of 27.05 ± 2.25 (dB), and structural similarity index metric (SSIM) of 0.84 ± 0.05 on the test dataset. T1-weighted MR images used for dosimetric assessment revealed a gamma index of 3%, 3 mm, 2%, 2 mm and 1%, 1 mm with acceptance criteria of 98.96% ± 1.1%, 95% ± 3.68%, 90.1% ± 6.05%, respectively. The DVH differences between CTs and pCTs were within 2%.

Conclusions: A promising pCT generation model capable of handling heterogenous multicenteric datasets was proposed. All MR sequences performed competitively with no significant difference in pCT generation. The proposed CT Masker proved promising in improving the model accuracy and robustness. There was no significant difference between using T1-weighted and T2-weighted MR images for pCT generation.

1. Introduction

Computed Tomography (CT) is one of the principal requirements in radiation therapy workflow, which provides three-fold benefits in radiation treatment planning. First and foremost, CT reflects information about electron density which is correlated with attenuation coefficients that yield CT numbers in Hounsfield units (HUs), which is required for accurate dosimetric calculations by treatment planning

* Corresponding author. Department of Medical Physics, Iran University of Medical Sciences Hemmat Highway, Tehran, Iran.
E-mail address: mahdavi.r@iums.ac.ir (S.R. Mahdavi).

https://doi.org/10.1016/j.compbiomed.2022.105277
Received 10 October 2021; Received in revised form 9 January 2022; Accepted 27 January 2022
Available online 31 January 2022

systems (TPSs). Next, subtle comparison of digitally reconstructed radiographs (DRRs) by means of CT with either on-board linear accelerator (Linac) cone-beam CT or megavoltage CT (MVCT) benefits image-guided radiation therapy (IGRT) and patient positioning. Finally, CT preserves the presence of bone signal, in contrast to magnetic resonance imaging (MRI), which is mandatory for dose calculation. MR images provide better soft-tissue contrast compared to CT images, which has prioritized its choice for contouring to enable accurate delineation of regions/volumes of interest and organs at risk (OARs). Accurate contouring is essential when treating a patients presenting with brain tumors for precise delivery of radiation beams to small planning target volume (PTVs) and sparing of OARs [1,2]. In addition, MRI does not deliver extra non-therapeutic radiation dose to patients compared to CT [2,3].

In clinical practice, MR images are fused to CT images to take advantage of both imaging modalities. Image registration algorithms could result in 2–5 mm systematic error while transferring structures from MRI to CT in various treatment sites. These errors originate from variations in organs' anatomical position between different imaging sessions and discrepancies in patient positioning during scanning with the different imaging modalities, such as neck flexion. Although the former is not frequent in brain scanning, since the brain is subject to very low anatomical variability, the latter could still be an outstanding source of error during the fusion process in the brain region. The dosimetric errors could be significant when irradiating tiny and sensitive targets, as is the case when treating brain tumors [2,4–6]. Considering these circumstances, the errors might end up with a loss of tumor control probability and a reduction in patients' quality of life. To overcome these challenges in radiotherapy workflow, pseudo CT (pCT) or synthetic CT (synCT) generation from MR images using deep learning algorithms could be a potential and practical approach in clinical setting. In addition, pCT generation facilitates MRI-only radiotherapy with MR Linacs. Nevertheless, the lack of fiducial markers on pCT may seem an impediment to instituting MRI-only radiotherapy, where pCT could be registered to the initial MRI performed with MR-compatible markers to pave the way for treatment setup. To sum up, pCT generation aiming at eliminating excessive CT scans in IGRT, enables to decrease patient radiation dose, scanning time, and cost without compromising treatment accuracy and potentially improving treatment outcome.

Different approaches have been proposed to tackle pCT generation, which can be classified into three main groups. This includes segmentation-based methods which use bulk density assignment to each contoured region of MR images [7], atlas-based methods which utilize rigid and deformable registration for mapping MRI to CT [2,8,9], and finally learning-based methods that emerged through statistical modeling [10] and evolved to regression and machine learning models [3,11]. With the advent of artificial intelligence (AI)-based algorithms, recent research focused on the development of deep learning algorithms, particularly convolutional neural networks (CNNs) to address challenges in medical image analysis [12–14]. Recently, increasing interest focused on variants of generative adversarial networks (GANs) including deep convolutional generative adversarial networks (DCGANs), conditional generative adversarial networks (CGANs) and cycle consistent generative adversarial networks (CycleGANs). These learning-based algorithms could be trained using paired and unpaired data. The training of most CNN models for image-to-image translation relies on minimization of voxel-wise differences between pCT and ground truth image, which requires paired datasets. Using paired datasets, tiny voxel-wise misalignment of input and target images may result in blurring of synthesized images. To address this issue, Nie et al. [15] combined voxel-wise loss and an image-wise adversarial loss in GAN. However, there is still a voxel-wise loss term, which requires a paired dataset for training.

GAN models are prone to overgeneralization and mode collapse where some generated modes should not exist and some modes are not well represented in the generated images even though the dataset supports these modes [16]. In practice, preparing clean and cured paired datasets is time-consuming, troublesome, and may not be a good choice for real-world problems due to the scarcity of cases that have undergone both imaging modalities. To tackle dataset preparation, the CycleGAN introduced by Zhu et al. [17] not only does not require pixel-wise alignment of images but is also able to use images of patients not necessarily scanned on both modalities. A number of studies reported on investigations in the pelvic region for proton [18] and photon [19] radiotherapy, as well as the head and neck region [20–22]. In this study, a CycleGAN model providing state-of-the-art performance for image generation using heterogeneous datasets was elaborated. To the best of our knowledge, this is the first work evaluating MRI-only radiotherapy workflow in 3D conformal radiotherapy (3DCRT).

2. Materials and methods

2.1. Image acquisition

T1-weighted and T2-weighted MR images of 87 and 102 patients with brain cancer who had previously undergone radiation therapy were anonymously collected from three different radiation therapy centers. MRI and CT were acquired on 18 and 3 different scanners, respectively. The included scanner models and percentage of included studies from each one are illustrated in Fig. 1. Most of the included patients have undergone CT and MRI on Somatom Scope and Avanto scanners, respectively. T1-weighted and T2-weighted MR images acquired with different protocols, such as blade, gadolinium contrast-enhanced, dark fluid, flair, etc. are summarized in Table 1.

Detailed information about MRI and CT image acquisition parameters can be found in Tables 2 and 3, respectively. The patients were a mixed bag of primary tumors (astrocytomas, meningiomas, gliomas, etc.) and secondary cancer (breast, lung, colon, etc.). MRI and CT transverse slices from 60 T1 (3315 slices) and 65 T2 (3655 slices) volumes were used as training set, whereas 10 T1 (411 slices) and 15 T2 (603 slices) volumes served as the validation set. Seventeen T1 (670 slices) and 22 T2 (882 slices) volumes were used as holdout test set. CT and MR image volumes had initial voxel sizes ranging from $0.48 \times 0.48 \times 1\ mm^3$ to $0.68 \times 0.68 \times 6\ mm^3$ and from $0.48 \times 0.48 \times 1\ mm^3$ to $0.68 \times 0.68 \times 8\ mm^3$, respectively.

2.2. Image pre-processing

Fig. 2 illustrates schematically the preprocessing steps. Initially, the Elastix open-source package was employed to register MR to CT images [23,24]. N4 bias correction was implemented using SimpleITK to compensate for the intrinsic inhomogeneities in the magnetic fields of an MRI machine and inhomogeneities of the fields created by the introduction of a patient. Since previously proposed adaptive thresholding approach showed deficiency in some CT slices in the presence of other objects adjacent to the head, a novel masker algorithm was developed to create binary head masks, excluding every pixel outer to the patient head, henceforth referred to as CT Masker.

The proposed masker algorithm consists of adaptive thresholding to binarize the input image followed by a Sobel filter to make the edges of the body clear. Subsequently, the external body contours are defined. Points of body contours were used to assign inner pixel labels of one and zero for outer pixels. After masking structures, CT and MR images were resampled and resized to $0.5 \times 0.5 \times 2\ mm^3$ voxels and 256×256 matrix size, respectively. Data augmentation increased T1 and T2 training set to 5216 and 5961 slices with cropped, rotated (up to 180°), sheared, and horizontally/vertically flipped images. Lastly, all images were rescaled to -1 and 1 prior to feeding the CycleGAN. CycleGAN trained using T1 and T2 images with and without CT masker are referred to as T1, T1 Masked, T2, and T2 Masked models, respectively.

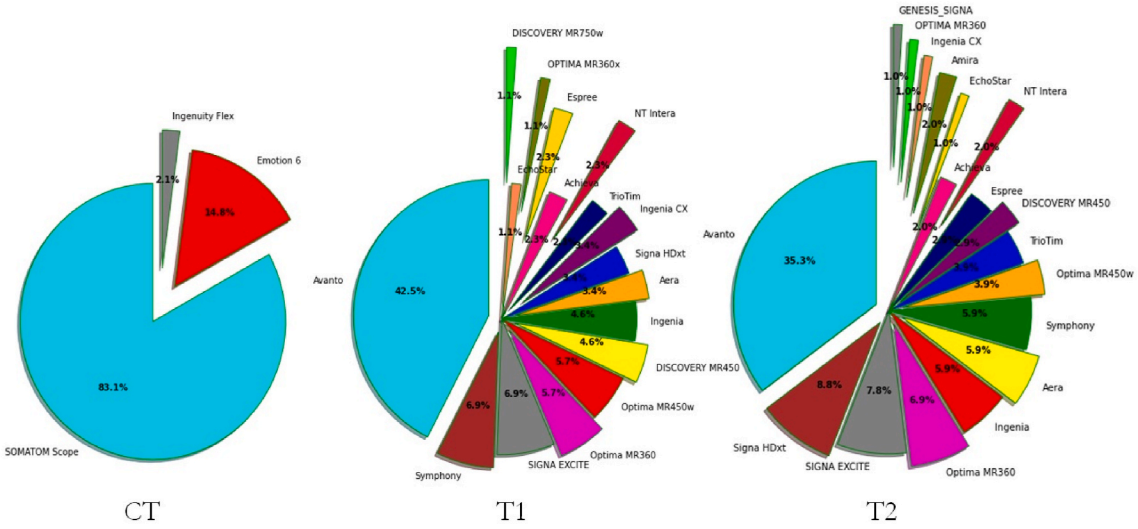

Fig. 1. Percentage of included patients from different scanners.

Table 1
List of MRI sequences used for data acquisition of the datasets used in this study. Multiplanar reformation or reconstruction (MPR), Gadolinium-based MRI contrast agents (GD), spin echo (SE), Multi Echo Multi Planar (MEMP), fluid-attenuated inversion recovery (FLAIR), Periodically rotated overlapping parallel lines with enhanced reconstruction (PROPELLER), Pre contrast (PRE), Gradient Echo (GE), Turbo inversion recovery magnitude (TIRM), Fast or turbo spin echo (FSE/TSE), fat suppression (FS), BLADE: proprietary name for PROPELLER in MRI systems from Siemens Healthcare.

MRI Sequences	T1-weighted	T2-weighted
MPR GD	12	0
TRIM	0	3
TRIM DARK FLUID	1	17
SE GD	14	0
BLADE	0	7
SE	27	0
MPR P2 ISO	10	0
TSE	6	44
MEMP	2	0
FLAIR	4	13
PROPELLER	1	6
PRE	1	0
GE	1	1
TRIM FLAIR	0	3
FL2D	1	0
BLADE DARK FLUID	0	1
FIL	1	1
FIL GD	1	0
FLAIR FS	0	2
TFE	1	0
TSE GD	2	0
SPACE P4 ISO	0	3
TSE BLADE	0	1
FSE GD	1	0
FS	1	0
Sum	87	102
	189	

2.3. Deep neural network architecture

Fig. 3 shows the architecture of CycleGAN and workflow adopted in this study. The CycleGAN is composed of two cycles, called forward and backward cycles. Each cycle contains two networks, including a generator and discriminator. The generator in the forward cycle ($G_{MR \to CT}$) takes MR images as input and translates to target domain (CT). The discriminator attempts to differentiate real and fake images synthesized by the generator and assigns 1 and 0 to each of them, respectively.

Table 2
Image acquisition parameters for the included MR images for both T1-and T2-weighted sequences.

MRI acquisition parameter	T1	T2
TR (min,max,avg)	[9, 6000, 891]	[665, 9075, 5970]
TE (min,max,avg)	[2, 100, 10.60]	[23, 383, 106]
FA (min,max,avg)	[8, 160, 75]	[18, 180, 133.87]
AT (min,max,avg)	[11618, 234236, 137519]	[2123, 233044, 132173]
Matrix size (min,max,avg)	[256, 551, 439]	[192, 1024, 445.65]
Magnetic Field Strength (min, max,avg)	[1.5, 3.0, 1.54]	[1.5, 3.0, 1.54]
Slice Thickness (min,max,avg)	[1, 8, 4.24]	[1, 8, 5.08]

Table 3
Image acquisition parameters for the included CT scans.

CT acquisition parameter	
kVp (min,max,avg)	[80, 130, 109]
Tube current (min,max,avg)	[23, 341, 147]
Matrix size (min,max,avg)	[512, 512, 512]
Slice thickness (min,max,avg)	[1.0, 6.0, 2.20]

Translating back the synthesized image (pCT) to the initial domain (MRI), backward cycle aims to verify MRI to CT mapping by calculating the loss between the initial MR image and the secondary MR image. In each cycle, the generator and discriminator compete to fool each other.

A residual block was used in the generator to aid the training of deep CNNs through deploying skip connections which alleviates gradient vanishing/exploding problem in deep neural networks [25]. Each residual block started with a reflection padding layer followed by a convolutional layer. Instance normalization was applied after each convolutional layer. Thereafter, an activation function, reflection padding, convolutional layer, skip connection, and instance normalization, was exploited for the residual block. The right upper panel of Fig. 3 visualizes the residual blocks.

In the generator, after reflection padding, a convolutional layer was applied to the input image followed by instance normalization and ReLU activation function. A downsampling block, consisting of a convolutional layer with doubled initial number of filters, instance normalization and an activation function was applied. The residual part was followed by an up-sampling block including transpose convolution, instance normalization, and activation function. Ultimately, the reflection padding, a convolutional layer with one filter, and an activation

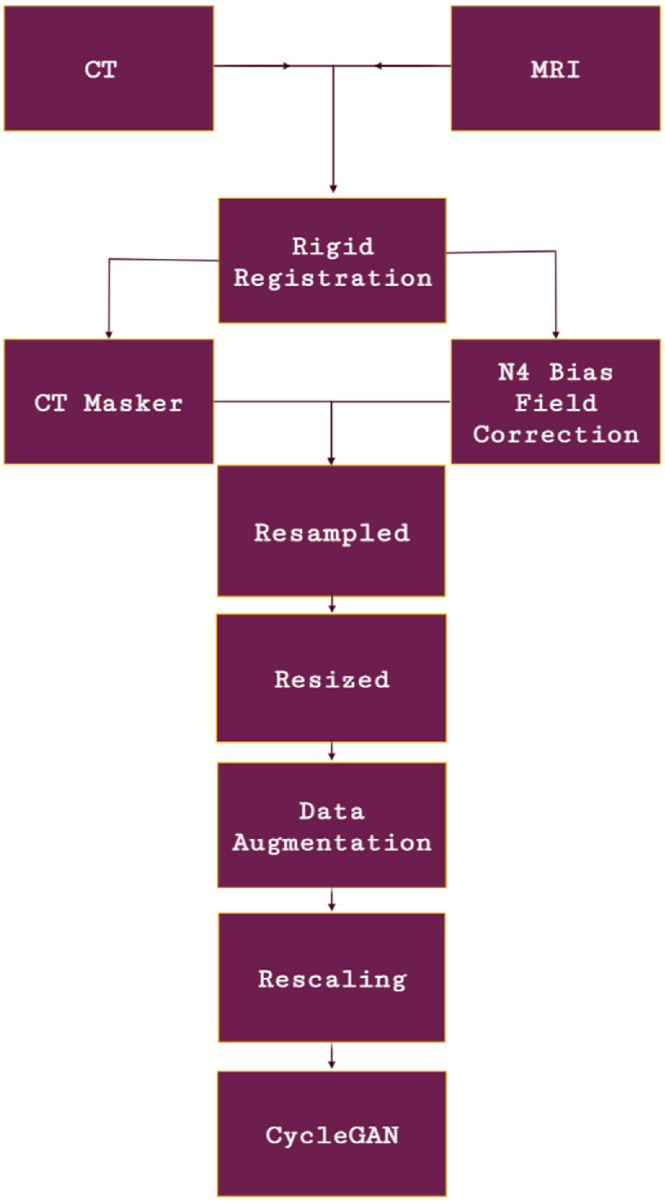

Fig. 2. Image pre-processing steps implemented in this study protocol.

function was developed to prepare the output image (the structure of the generator is illustrated in Fig. 3 (Gen G and Gen F)). All convolutional layers had 3 × 3 kernel size except the first and last layer which had a kernel size of 7 × 7. For each downsampling and up sampling layers zero padding was applied to preserve the size of output feature maps after convolution operations. Rectified linear activation unit (ReLU) was used in all layers of the generator, except tanh for the last layer in the generator.

In the discriminator, a convolutional layer, a leaky ReLU activation function and downsampling blocks followed by a final convolutional layer with one node were applied to the input image, respectively. A detailed illustration of the discriminator is presented in Fig. 3 (Disc X, Disc Y). The leaky ReLU was used as an activation function in all layers of the discriminator, except none in the discriminator. A number of important hyperparameters were optimized. In this context, different networks were trained to adopt an optimal number of residual blocks in the generator, down sampling blocks for the discriminator, filters, and loss function. A complete implementation of preprocessing steps and model architecture in TensorFlow library is available on the author's GitHub.

3. Evaluation

3.1. Image quality

The accuracy of pseudo CT images evaluation was performed using a number of 3D metrics, including the mean absolute error (MAE) defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{n} abs(CT_i - SynCT_i)$$

where i is the index of corresponding CT-MR slices and N is the number of slices for each patient CT volume. The peak signal-to-noise ratio (PSNR) was also calculated. It is defined as:

$$PSNR = 10 \times \log_{10}\frac{im_{max}^{2}}{MSE}$$

where, im_{max} is the maximum possible pixel value in the image and MSE is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{n} (CT_i - SynCT_i)^2$$

Another metric that was utilized to quantify the similarity between pCT and the ground truth image was the structure similarity index metric (SSIM) defined as:

$$SSIM\ (x,\ y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x, μ_y, σ_x^2, σ_y^2, and σ_{xy} are the average of x, average of y, variance of x, variance of y, and covariance of x and y, respectively. SSIM is an image metric relying on the luminance, contrast and degradation of structural information between two images [26]. Wilcoxon signed-rank two-sided test was performed on imaging metrics to assess whether the pCT images generated using T1 and T2 is statistically different with the standard of reference. For this purpose, the trained T1 model vs T2 model and T1 Masked model vs T2 Masked model were compared. In addition, the performance of the proposed CT Masker was validated by means of the Wilcoxon signed-rank two-sided test and the comparison of T1 vs T1 Masked model and T2 vs T2 Masked model.

3.2. Dosimetry analysis

Regarding the limited transverse FOV coverage in MR images, pCTs would comprise smaller volumes. Hence, CTs were cropped to the corresponding pCT volume to have the same scattering effect contributing to points of interest (POIs) on both images. Target volumes and OARs were contoured on CT images and reviewed by a radiation oncologist. The OAR and PTV contours delineated on CT images were exploited in this work. Beside to each patient's real structures, simple geometries, e.g. spheres and cubes were used as OARs to compare the dose-volume histograms (DVHs). Dose calculation was performed on Core Plan 3.5.05 as TPS with correction-based ETAR method, grid size of 3 mm, and Varian 2100CD as a Linac. All plans were prescribed with 40 Gy in 25 fractions. Different numbers of beams with 6 and 18 MV energies, wedges, and bolus were used to optimize each plan (optimized plans were used for both pCTs and CTs). Gamma analysis at 3%, 3 mm, 2%, 2 mm and 1%, 1 mm with 10% threshold of the prescribed dose was performed using the Slicer RT package [27]. Moreover, DVH key parameters dose discrepancies, including D_{90}, D_{10}, mean dose, min dose, and max dose were measured in terms of MAE and ME to verify the target dose coverage and critical structures sparing. Ultimately, dose differences statistical significance level was calculated by the Wilcoxon test. All statistical analysis was performed on Python.

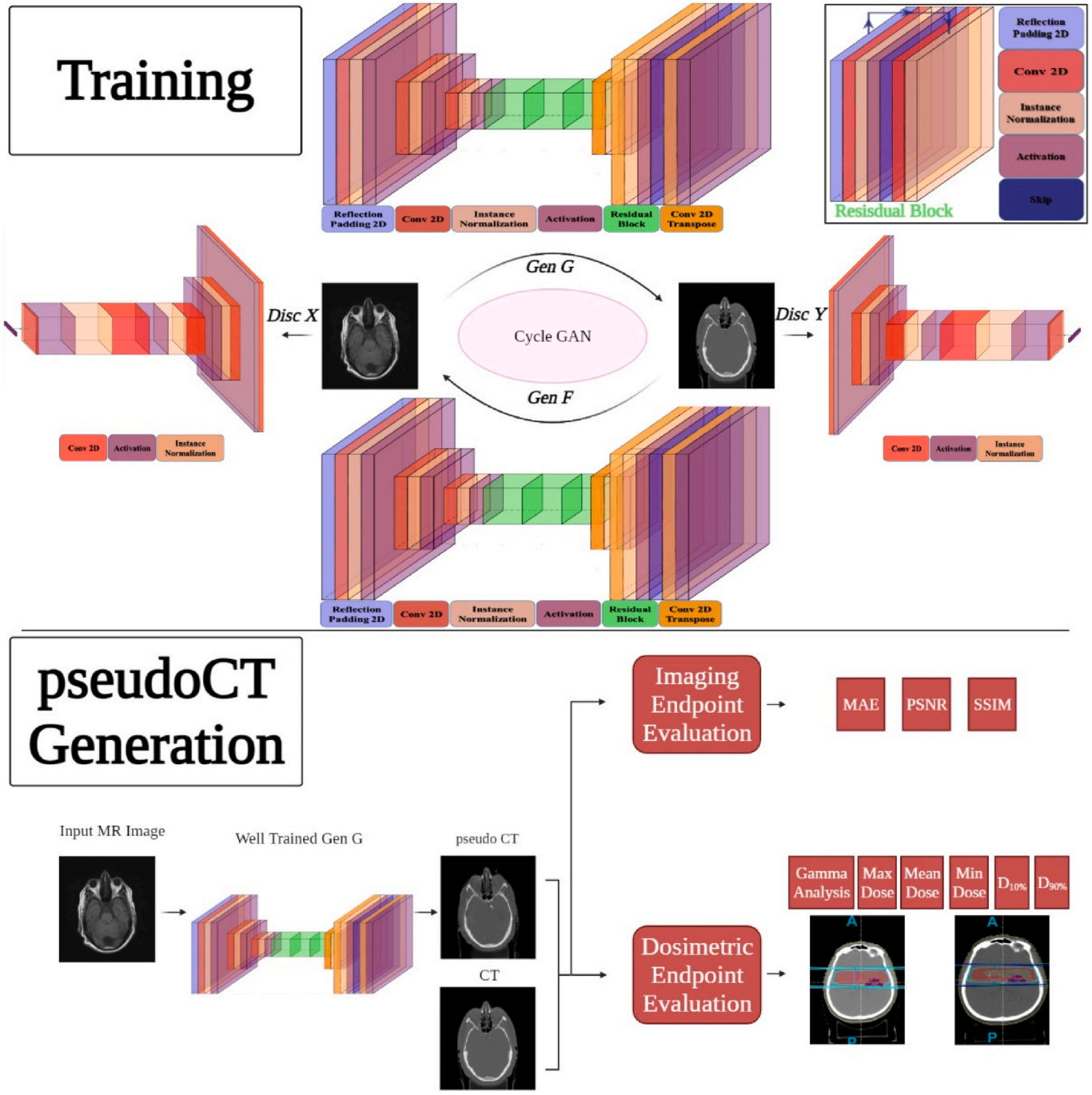

Fig. 3. Schematic workflow of the proposed CycleGAN. The upper panel of the figure depicts the training phase and architecture of the generators and discriminators. The lower panel depicts pCT generation using test patients and predicted images evaluation.

Table 4
Performance of models using MSE and BCE loss functions on imaging endpoint metrics, including MAE, PSNR, and SSIM. HU= Hounsfield unit; dB = decibels.

	T1 Model (MSE)	T1 Model (BCE)
MAE (HU)	79.51 ± 33.42	85.44 ± 31.02
PSNR (dB)	23.62 ± 2.35	22.01 ± 1.38
SSIM	0.84 ± 0.06	0.82 ± 0.07

4. Results

4.1. Image quality

The calculated metrics to select the best loss function are shown in Table 4. Considering the better performance of MSE loss function with MAE of 79 HU over binary cross entropy (BCE) with MAE of 85 HU, MSE was used in this study. Fig. 4 depicts violin plots of MAE, PSNR, and SSIM for each method. The mean (±standard deviation, SD) MAE, PSNR, SSIM for each of the methods are listed in Table 5. T1, T1 Masked, T2, and T2 Masked models scored 79.51 ± 33.42 HU, 62.65 ± 30.72 HU, 76.51 ± 18.41 HU, and 61.87 ± 22.58 HU on MAE, respectively. PSNR values were 23.62 ± 2.35 (dB), 26.95 ± 3.38 (dB), 23.43 ± 1.00 (dB),

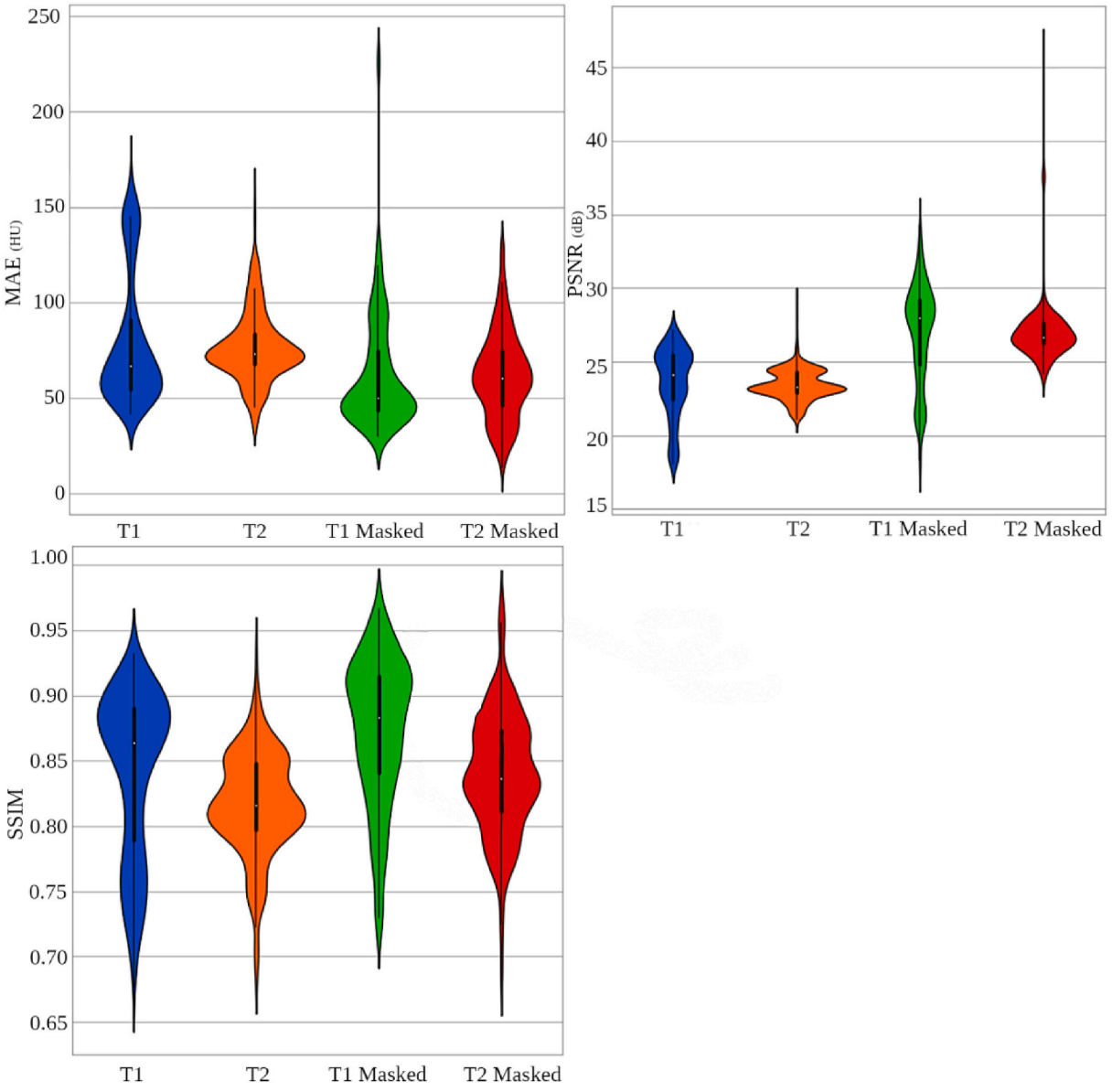

Fig. 4. Violin plots represent interquartile range and median of errors for MAE, PSNR, and SSIM calculated on the test set for the each model.

Table 5
Imaging endpoint metrics, including MAE, PSNR, and SSIM for the validation and test sets. T1/T2 = trained CycleGAN without CT Masker; T1/T2 Masked = trained CycleGAN with CT Masker; HU= Hounsfield unit; dB = decibels.

	T1	T1 Masked	T2	T2 Masked
MAE_{val} (HU)	80.11 ± 29.10	60.47 ± 34.65	75.00 ± 20.00	66.04 ± 19.00
MAE_{test} (HU)	79.51 ± 33.42	62.65 ± 30.72	76.51 ± 18.41	61.87 ± 22.58
$PSNR_{val}$ (dB)	23.06 ± 1.85	26.90 ± 3.26	23.63 ± 1.04	27.10 ± 2.48
$PSNR_{test}$ (dB)	23.62 ± 2.35	26.95 ± 3.38	23.43 ± 1.00	27.05 ± 2.25
$SSIM_{val}$	0.84 ± 0.07	0.88 ± 0.05	0.83 ± 0.07	0.82 ± 0.03
$SSIM_{test}$	0.84 ± 0.06	0.87 ± 0.05	0.81 ± 0.04	0.84 ± 0.05

and 27.05 ± 2.25 (dB), respectively. Ultimately, these models scored 0.84 ± 0.06, 0.87 ± 0.05, 0.81 ± 0.04, 0.84 ± 0.05 on SSIM.

Fig. 5 provides visual comparison of real and predicted pCTs by T1 and T1 Masked models in axial, sagittal, and coronal planes. The same information is provided in Fig. 6 for T2 and T2 Masked models. The patients presented in Fig. 6 for T2 and T2 Masked model correspond to an outlier and best achieved performance of the model, respectively. The results of the Wilcoxon test which assessed the influence of CT Masker by comparing T1 and T2 Masked and T2 and T2 Masked models for all metrics ($p < 0.0001$) are reported in Table 6. The findings of the Wilcoxon test comparing the quality of generated pCTs based on T1-and T2-weighted images are summarized in Table 7. Although the violin plots shown in Fig. 4 and imaging metrics summarized in Table 5 may seem to suggest that T2 sequence would slightly be a better option, statistical tests categorically refute this assumption since there is no meaningful differences between them.

The training of CycleGAN is somehow different from conventional deep learning models. Considering the adversarial training concept behind GANs, generator and discriminator are trained competitively, if either of them outcompete the other one, the training process would not be successful. Moreover, an ideal training should not have large variations in training metrics; otherwise, it is an indicator of failed training. The model was trained on an NVIDIA GeForce GTX 1080 GPU. On average, the model took 5.5 seconds to generate pCT images for each patient, which is very fast and suitable for deployment in clinical setting

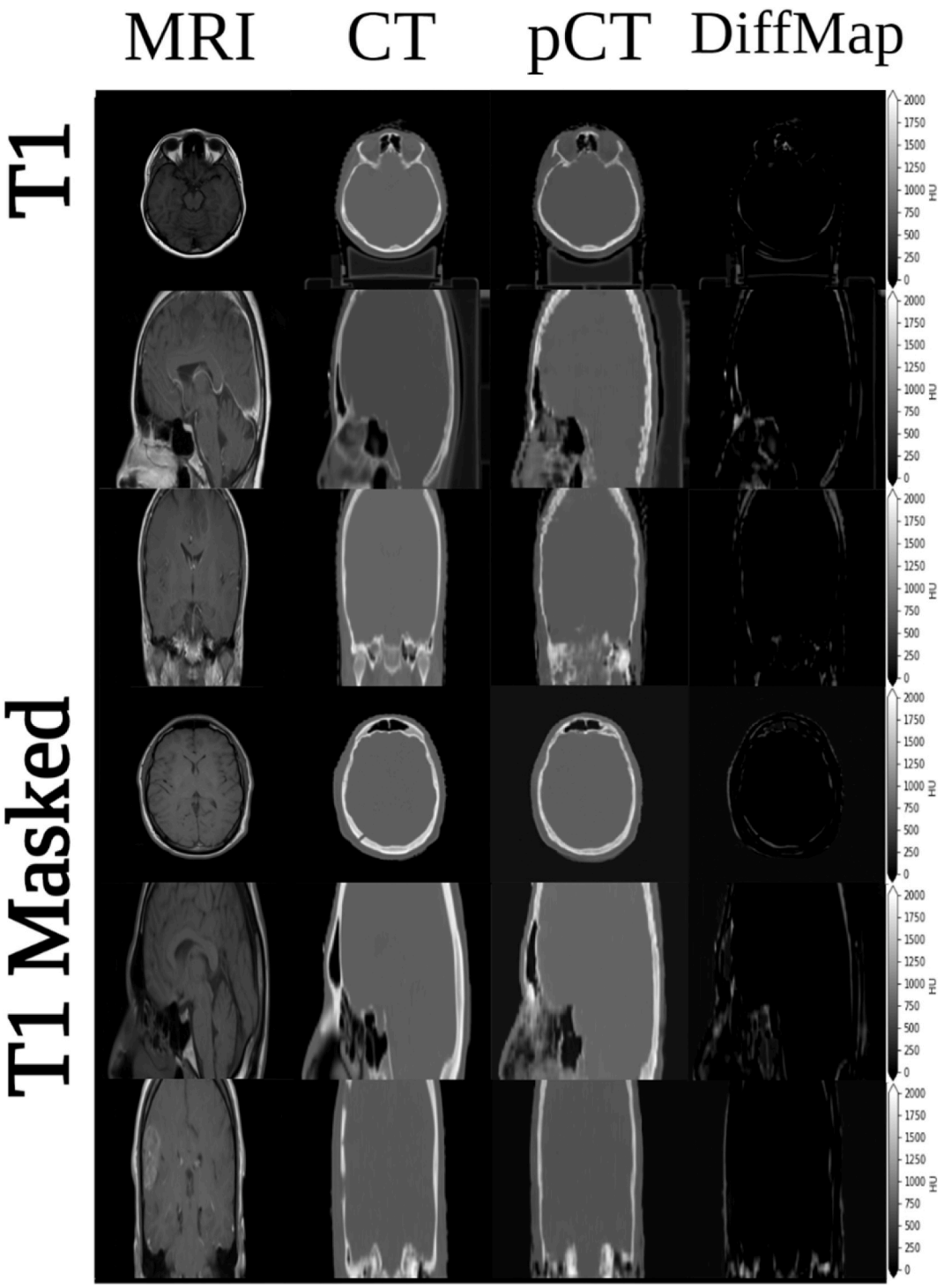

Fig. 5. Visual representation of model performance for T2-weighted MR images with and without CT Masker. From left to right: input MR images, pCT, actual CT image, and absolute difference map (CT-pCT).

[19]. Fig. 7 shows the training loss of CycleGAN.

4.2. Dosimetry analysis

There was no statistically significant differences between T1 and T2 images for the generation of pCT images. Therefore, the generated pCT images of the test set using T1 model was used for dosimetric assessment. Two patients with significant model failure depicted as outliers were excluded from the dosimetric comparison. Although few artifacts were noticed in some slices using T1 model, the dosimetric analysis was performed by avoiding the intersection between the beam and the artifacts. The mean dose differences (Mean ± SD) in PTV for the mean dose, minimum dose, maximum dose, D_{90}, and D_{10} were (0.61 ± 0.6)%, (−1.744 ± 1)%, (0.5 ± 0.45)%, (−0.943 ± 0.9)%, and (−0.3 ± 0.3)%, respectively. The mean absolute dose differences (Mean ± SD) in the PTV for mean dose, minimum dose, maximum dose, D_{90}, and D_{10} were (0.93 ± 0.40)%, (2.07 ± 0.33)%, (1.31 ± 0.39)%, and (0.71 ± 0.23)%, respectively. These metrics are shown in Table 8 for the other structures. The comparison of dose differences showed good agreement with actual CT images for various organs. Mean gamma pass rate of 1 mm/1%, 2

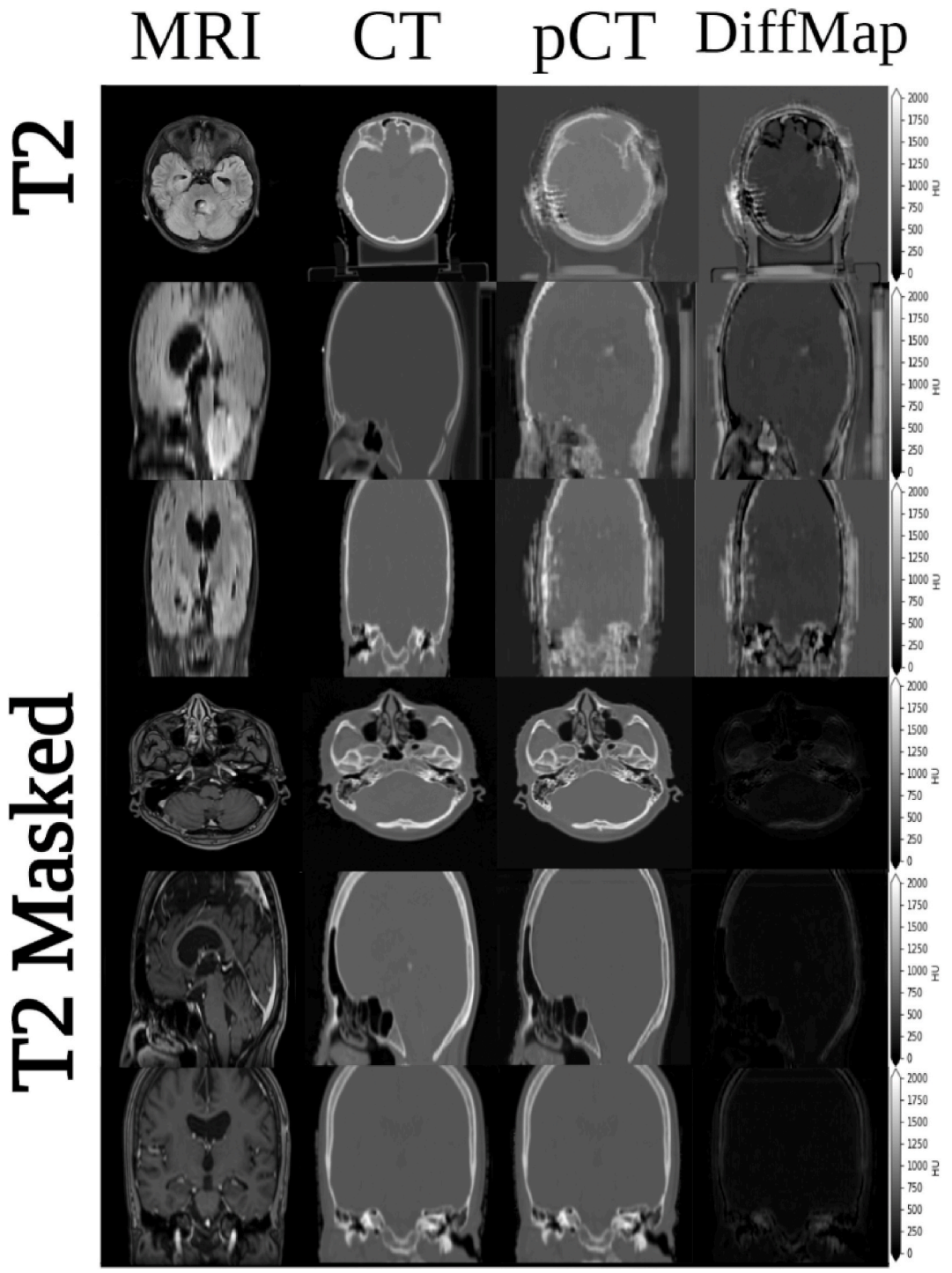

Fig. 6. Visual representation of model performance for T2-weighted MR images with and without CT Masker. From left to right: input MR images, pCT, actual CT image, and absolute difference map (CT-pCT). Note that the depicted patient for the T2 model is an outlier whereas the depicted patient for T2 Masked model reflects the best performance of the model.

Table 6
Statistical Wilcoxon signed-rank two-sided test performed for T1 vs. T1 Masked and T2 vs. T2 Masked models to evaluate the contribution of the proposed CT Masker.

Wilcoxon	T1	T1 Masked	T2	T2 Masked
MAE p-value	<0.0001		<0.0001	
PSNR p-value	<0.0001		<0.0001	
SSIM p-value	<0.0001		<0.0001	

Table 7
The quality of generated pCTs by T1 vs T2 and T1 Masked vs T2 Masked models was assessed using statistical Wilcoxon signed-rank two-sided test to analyze the suitability of T1 and T2 weighted images for pCT generation.

Wilcoxon	T1	T2	T1 Masked	T2 Masked
MAE p-Value	0.045		0.600	
PSNR p-Value	0.070		0.543	
SSIM p-Value	<0.0001		<0.0001	

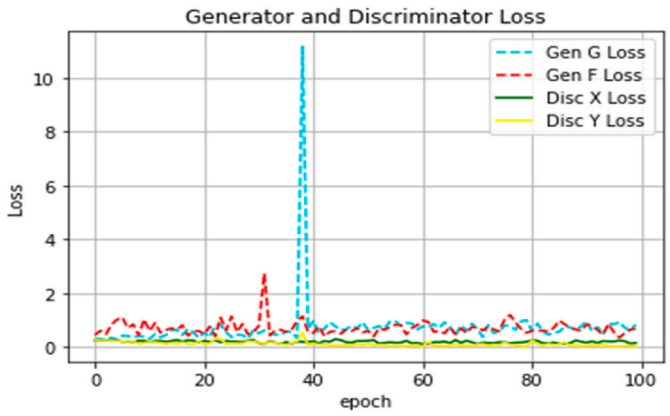

Fig. 7. Generators and discriminators adversarial training loss for 100 epochs.

Table 8
Mean absolute differences and mean differences for T1 model between some key DVH dosimetric points with respect to those calculated based on CT scans. * Wilcoxon test compared the dosimetric metrics between pCT and CT images. P-values for all metrics was insignificant. Note. D90% and D10% are defined as the minimum dose absorbed by 90% and 10% of volume, respectively.

Structure	Mean dose	Min dose	Max dose	$D_{90\%}$	$D_{10\%}$
Average dose discrepancies mean ± SD (range) (%) (p-Value)					
PTV_{ME}	0.61 ± 0.60	−1.74 ± 1.0	0.50 ± 0.45	−0.94 ± 0.90	−0.30 ± 0.30
PTV_{MAE}	0.93 ± 0.40	2.07 ± 0.33	0.72 ± 0.18	1.31 ± 0.39	0.71 ± 0.23
Right Eye_{ME}	0.20 ± 0.20	0.10 ± 0.10	0.20 ± 0.20	0.15 ± 0.10	0.17 ± 0.20
Right Eye_{MAE}	0.29 ± 0.12	0.18 ± 0.08	0.28 ± 0.13	0.21 ± 0.06	0.21 ± 0.11
Left Eye_{ME}	0.30 ± 0.20	0.1 ± 0.20	0.29 ± 0.30	0.30 ± 0.20	0.27 ± 0.30
Left Eye_{MAE}	0.38 ± 0.18	0.21 ± 0.11	0.37 ± 0.20	0.43 ± 0.16	0.33 ± 0.22

mm/2%, 3 mm/3% criteria with 10% dose threshold were 90.1% ± 6.05%, 95% ± 3.68%, 98.96% ± 1.1%, respectively. The upper panel of Fig. 8 depicts DVH differences for a representative patient. The treatment plan and Gamma map of the same patient are shown in the lower panel of Fig. 8.

5. Discussion

In the present study, we developed a novel and fast pCT generation method based on a CycleGAN architecture using heterogeneous multi-center datasets for brain tumor patients. Regarding the computation time needed to provide paired datasets without missing alignments between MR and CT images, the CycleGAN model would be a better and practical choice compared to other architectures. The model selection was based on previous works reporting the superiority of GANs, particularly CycleGAN. For instance, Kearney et al. [28] compared U-net and CycleGAN architectures and concluded that the later significantly outperformed the former. Emami et al. [29] reported the superiority of GAN over CNN. More importantly, CycleGAN is the model of choice because of its ability to work with heterogeneous (unpaired) data. In this study, we challenged the homogenous radiotherapy workflow by acquiring extremely diverse data sets. The inhomogeneity was introduced in every aspect to CycleGAN, i.e. scanners, T1-and T2-weighted images, MR sequences, voxel sizes, post/pre-operative radiotherapy patients, patient position on the couch, etc. Considering the above conditions, conventional deep learning models would not be an optimal choice for this complex scenario. This motivated us to opt for unsupervised learning. Supplementary Table 1 compares the image quality metrics, dose differences, and gamma pass rates with previous studies.

Models' performance was assessed by pixel-wise imaging and dosimetric endpoints. Although almost all previous research and clinical studies used T1-weighted MRI sequences for the brain [2], there is no preferred MR sequence for pCT generation purpose, which is consistent with our findings ($p > 0.05$) in Table 7. CycleGAN with nine residual blocks, two downsampling, and two upsampling blocks, achieved the best performance over other hyperparameter choices. The better performance of CycleGAN using MSE loss function is in agreement with Mao et al. [30]. MR images were rigidly registered to CT scans in order to provide aligned images to the CycleGAN. The motivation behind registration was to ease the training for the CycleGAN to concentrate on other important variations and advanced features present in the dataset for the sake of generalizability by reducing the simple variations in the preparation step as much as we could and benefit from the whole capacity of the models. Anyway, the CycleGAN benefited from unsupervised loss function term.

The qualitative comparison of pCT images generated with and without the applied CT Masker (Supplementary Figs. 1 and 2) shows that removing the headrest and fixator significantly contributes to higher pCT image quality, which is revealed by the statistical tests (Table 6). Since the images were collected from different hospitals with various scanners and acquisition settings, such as, couches, headrests, …etc, this tended to confuse the models by focusing on features that introduced artifacts randomly in some slices (shown in Supplementary Figs. 1 and 2). While removing outer objects (containing no useful information) is the potential approach to achieve good quality pCTs, in practice their presence is mandatory. Hence, the masked structures could be added to pCT at the end to address this issue. Although various masking algorithms were developed, known as skull stripping tools, they are not appropriate for pseudo CT generation. Previously developed algorithms were dedicated for brain tissue extraction without paying attention to outer structures. Yet, in pCT generation for radiotherapy, bony structures are crucial for accurate dose calculation. Therefore, we have developed a new method suitable for this application. The proposed CT masker not only improved pCT overall quality but also was promising in blurring reduction which was addressed in previous studies [2].

Although CT Masker contributed to improve the quality of the generated pCTs, we performed a dosimetric validation using T1 model without CT Masker to assess the clinical performance of the worst model and ensure whether it is within the acceptable range of clinical dosimetric uncertainties. Considering 3% error as an acceptable range in clinical setting, even the worst model would meet current standards. Given the evident contribution of CT Masker, better clinical dosimetric performance is expected for the models trained using CT Masker. However, future work will focus on the development of a deep learning-based CT masking model and perform a dosimetric validation for deep learning enabled masked models. The dosimetric evaluation performed for T1 was in agreement with previous studies [31]. Fig. 8 (upper panel) shows DVH differences for a patient where lower performance was achieved in terms of pCT generation in air bone interface (i.e. nasal cavity). While PTV, left eye, right eye, external structures, and pseudo-OAR2 show good agreement, there was a discrepancy in pseudo-OAR1 dose values. The pseudo-OAR1 dose values variation in this patient could be assigned to the lower performance of model in the corresponding region.

Compared with previous works, such as Kearney et al. [28] their attention aware CycleGAN (A-CycleGAN) trained with 90 resampled and normalized [0, 1] image volumes scored better in terms of MAE and PSNR. However, our CycleGAN model scored better on SSIM and generated comparable and even better pCTs visually. Wang et al. [32] achieved MAE of 131 ± 24 for the overall region using U-net trained with rigid and deformable registration followed by histogram matching. Their images suffered from blurring in air-bone interfering structures.

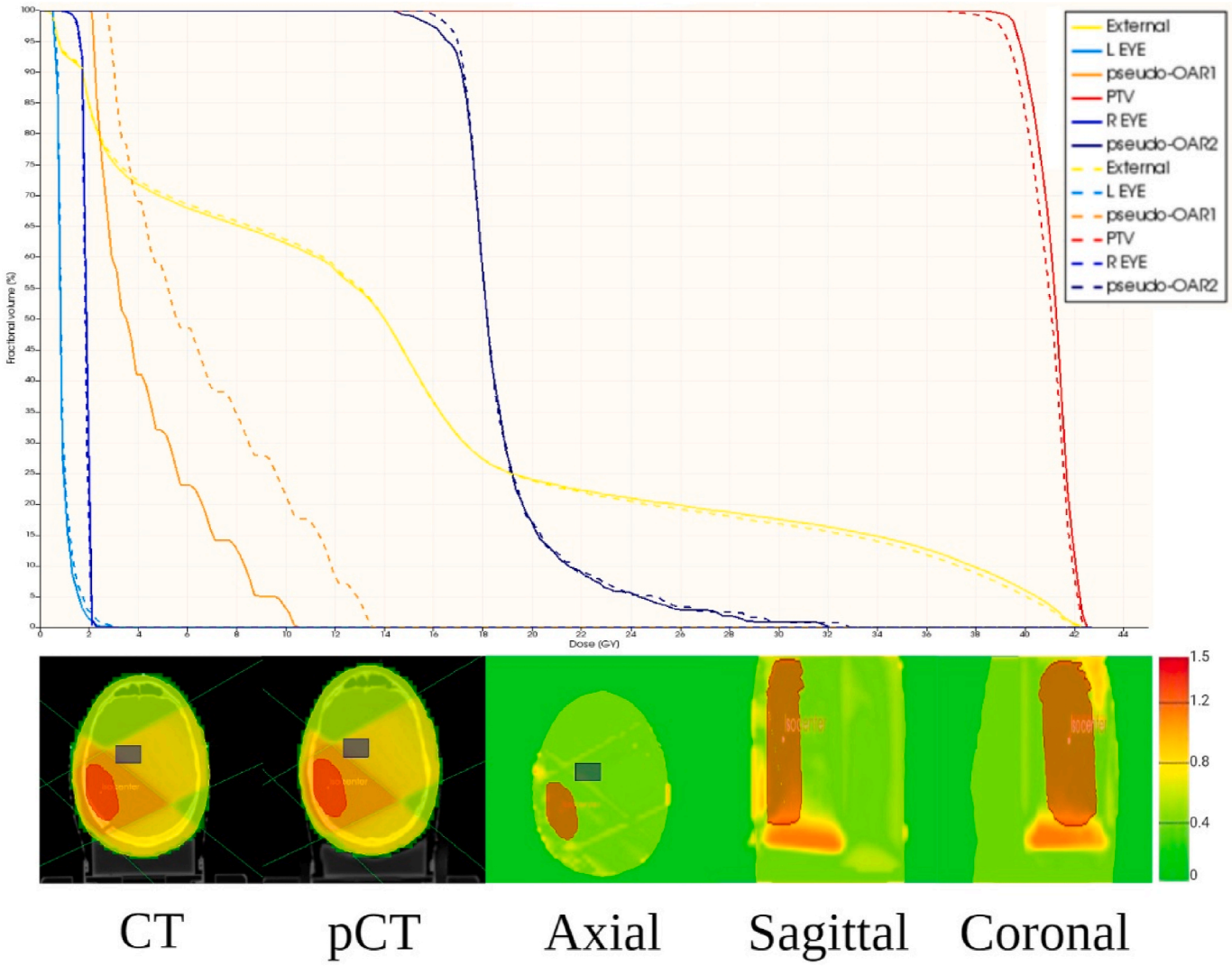

Fig. 8. Representative patient with a tumor in the right temporal brain lobe showing in the upper panel similar isodose curves between CT (solid line) and pCT (dashed line). In addition, simple geometrical structures were contoured at various sites, named pseudo-OAR1 and pseudo-OAR2. In the lower panel of the figure, the generated plan on CT and pCT is presented. The Gamma index map for this patient is also depicted.

Liu et al. [14] used 40 patients to train deep CNN model with rigid (Euler) and non-rigid (B-spline) transformations and assessed the quality of pCTs in VMAT planning. They tested their model on 10 patients and reported 75 ± 23 HU for MAE and 99.2% for 3%/3 mm gamma pass rate. Kazemifar et al. [22] used 70 patients resampled (equalized pixel and voxel spacing) data to train a GAN model and validate synCTs using VMAT. Their model averagely scored 48 HU for MAE, 98.7% ± 1.1% for 2%/2 mm and 93.6% ± 3.4% for 1%/1 mm gamma pass rates. Although their model scored lower in MAE, the visual quality of the generated images and the dosimetric metrics (0.61% compared to 0.7% mean dose and 0.5% compared to 0.6% maximum dose) were in agreement or even better.

There are two issues with some image quality metrics used for evaluation. The most important drawback of MAE is the lack of gold standard against which the results are to be compared, which might compromise its interpretability. Second, it is not clear if an increase in either of the metrics, such as MAE, would change the dose distribution. Addressing this issue would be quite challenging owing to differences in dose calculation algorithms and treatment modalities. However, as a simple comparison with previous studies, the worst and best reported MAE were 150 HU and 20 HU, respectively. Considering the heterogeneity of the dataset, the reported metric is quite reasonable. The DVH discrepancies are within the accepted range of 3% in radiation therapy, where more than 95% of the pixels passed a gamma index of 3% and 3 mm DTA (TG 119) [33]. Moreover, previous studies have shown significant variations in imaging metrics, which could be due to the different training data sets as well as models' non-repeatability. Due to ethical issues, patients' data sharing is not allowed to achieve consensus and objective comparison. Hence, two machine learning-based solutions, referred to as Federated learning [34] and reproducibility in deep learning [35] should be incorporated into pCT generation as well as other applications of AI in medicine. Federated learning introduced by google in 2016 [36] enables training models with decentralized data. As noted by Edmund et al. [2] in a previous review article, presently available performance metrics, such as MAE, do not embody the clinical endpoints. Hence, the development of objective metrics is needed. Considering the above-mentioned facts, dosimetric comparison is crucial. Currently, the best approach for comparing the results would be to focus on dosimetric metrics with respect to treatment modality and dose calculation algorithm.

As suggested by previous studies, we have resized initial 512 × 512 images to 256 × 256 matrix size. Converting back to the original size and comparing preprocessed CT volumes with reference CT volumes for each patient clarified that resizing as expected introduced dummy data

and consequently decreased pCT generation accuracy and treatment outcome. Nonetheless, training deep learning models with the original image size may be more precise. It is accompanied with higher computational costs, demanding GPUs with higher computational capability. To the best of our knowledge, we used the most heterogeneous data set consisting of 87 patients with T1-weighted and 102 patients with T2-weighted images. We employed T1 and T2 images comprising a wide range of MRI sequences and corresponding diverse CT scans from three radiation therapy centers. Altogether, images from more than twenty scanners were included in this study. Considering the wide range of CT and MRI acquisition parameters listed in Tables 1–3 and Fig. 1, the proposed method is promising in handling multicentric datasets. This ensures that the model is highly generic besides resulting in acceptable performance.

In addition, to the best of our knowledge, this is the first study implementing MRI-only radiotherapy in 3DCRT with correction-based ETAR dose calculation method since previous studies focused on intensity-modulated radiotherapy and volumetric modulated arc therapy with pencil beam, Monte Carlo, …etc. Future studies should investigate the validity of pCTs in tomotherapy with collapsed cone algorithm as a more accurate dosimetric approach. Although intensity correction is one of the rational and usual steps in medical image processing, Andres et al. [37] reported no improvement in dose metrics over models not incorporating any correction. In addition, comparing supervised models [37] with our unsupervised cycleGAN, it seems logical to expect the CycleGAN to handle intensity variations.

There are some outliers for the test dataset in each model, some of them are depicted in Supplementary Figs. 3 and 4 for each model. Despite the generalizability of the model, these model failures on rare data warns unconscious use in the clinic. The trained model had biases for some MRI systems and sequences. This can be attributed to two potential reasons. First and foremost, insufficient training of images from a specific scanner or sequence. Second, the presence of data originating from one or more inconsistent scanners/sequences with the rest of the datasets. The significant and wide range of discrepancies in voxel size of the initial MR images may hinder models' convergence and optimal update of weights and biases in neural networks to some extent. In this regard, the most important issue is combining several MR imaging protocols (for clinical and diagnostic purposes), which would alter the quantitative characteristics of images selectively in each anatomical region (tissue-based variation), e.g. white matter, gray matter, CSF, and air. This tissue-based variation seems to be challenging in pCT generation process. For instance, one patient scanned on the Genesis Signa using the FLAIR FIL protocol is presented as an outlier for the T2 model. This is a good example reflecting changing the protocol and lack of training data. Although there are some patients scanned with FLAIR and FIL protocols separately in the training data set, the lack of combination of these images seems challenging for the deep learning model.

Accordingly, analyzing the test results, the prospective challenges of feeding new data to the CycleGAN, from the easiest to the most challenging cases, could be sorted out as follows. First, working with new tumor sizes and types. Second, starting with same sequences from a new scanner model. In this case, there would be either slight overestimation or underestimation in electron density prediction. Third, inputting rarely seen MR images. Likewise, unseen MR protocols that resemble the training datasets, for instance, training with either T2 or T2-Gd and inputting the other one in the test phase. Forth, handling MR images with completely different train and test protocols. Last, MR protocols with relatively new acquisition parameters, and the combination of different protocols.

Adversarial training loss of the generator and discriminator are depicted in Fig. 7. The training of CycleGAN is different from conventional deep learning models. Considering the adversarial training concept behind GANs, the generator and discriminator are trained competitively, if either of them outcompete the other one, the training process would not be successful. Moreover, an ideal training should not have large variations in training metrics; otherwise, it is an indicator of failed training. Even though we have used axial slices as inputs and then sagittal and coronal planes were post generated using transverse planes, future studies should use either each or all other planes as inputs and even determine the best orthogonal plane for this task. Considering the errors present in soft-tissue, future studies could be devised to incorporate deep learning-based segmentation of CT images into bone, air, and soft-tissue to feed the network. Stacking each image with the corresponding Sobel filter to detect edges may be a practical solution to increase the accuracy in different tissue interferences. The small sample size and the lack of external validation dataset is one of the limitations of this study. We wished to have access to an independent dataset to objectively validate the results. Unfortunately, we had no access to extra datasets. However, we split the dataset and kept the test set unseen during modelling to reduce possible errors/biases.

6. Conclusion

A promising CycleGAN-based pCT generation algorithm was proposed which is capable of handling multicenter imaging datasets. All MR sequences performed competitively with no significant difference in pCT generation. The proposed CT Masker proved to be promising in improving the model's accuracy and robustness.

Declaration of competing interest

None Declared.

Acknowledgments

This work was supported by Iran University of Medical Sciences and the Swiss National Science Foundation under Grant No. SNSF 320030_176052.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105277.

References

[1] H. Emami, M. Dong, S.P. Nejad-Davarani, C.K. Glide-Hurst, Generating synthetic CTs from magnetic resonance images using generative adversarial networks, Med. Phys. 45 (8) (2018) 3627–3636, https://doi.org/10.1002/mp.13047.
[2] J.M. Edmund, T. Nyholm, A review of substitute CT generation for MRI-only radiation therapy, Radiat. Oncol. 12 (1) (2017) 1–15, https://doi.org/10.1186/s13014-016-0747-y.
[3] T. Huynh, et al., Estimating CT image from MRI data using structured random forest and auto-context model, IEEE Trans. Med. Imag. 35 (1) (2016) 174–183, https://doi.org/10.1109/TMI.2015.2461533.
[4] K. Ulin, M.M. Urie, J.M. Cherlow, Results of a multi-institutional benchmark test for cranial CT/MR image registration, Int. J. Radiat. Oncol. Biol. Phys. 77 (5) (2010) 1584–1589, https://doi.org/10.1016/j.ijrobp.2009.10.017.
[5] P.L. Roberson, P.W. McLaughlin, V. Narayana, S. Troyer, G.V. Hixson, M.L. Kessler, Use and uncertainties of mutual information for computed tomography/magnetic resonance (CT/MR) registration post permanent implant of the prostate, Med. Phys. 32 (2) (2005) 473–482, https://doi.org/10.1118/1.1851920.
[6] C.J. Dean, et al., An evaluation of four CT-MRI co-registration techniques for radiotherapy treatment planning of prone rectal cancer patients, Br. J. Radiol. 85 (1009) (2012) 61–68, https://doi.org/10.1259/bjr/11855927.
[7] A. Largent, et al., Pseudo-CT generation for MRI-only radiation therapy treatment planning: comparison among patch-based, atlas-based, and bulk density methods, Int. J. Radiat. Oncol. Biol. Phys. 103 (2) (Feb. 2019) 479–490, https://doi.org/10.1016/j.ijrobp.2018.10.002.
[8] C. Catana, et al., Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype, J. Nucl. Med. 51 (9) (2010) 1431–1438, https://doi.org/10.2967/jnumed.109.069112.
[9] J.A. Dowling, et al., Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences, Int. J. Radiat. Oncol. Biol. Phys. 93 (5) (2015) 1144–1153, https://doi.org/10.1016/j.ijrobp.2015.08.045.
[10] M. Kapanen, J. Collan, A. Beule, T. Seppälä, K. Saarilahti, M. Tenhunen, Commissioning of MRI-only based treatment planning procedure for external beam

radiotherapy of prostate, Magn. Reson. Med. 70 (1) (2013) 127–135, https://doi.org/10.1002/mrm.24459.
[11] X. Cao, J. Yang, Y. Gao, Y. Guo, G. Wu, D. Shen, Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis, Med. Image Anal. 41 (2017) 18–31, https://doi.org/10.1016/j.media.2017.05.004.
[12] X. Han, MR-based synthetic CT generation using a deep convolutional neural network method, Med. Phys. 44 (4) (2017) 1408–1419, https://doi.org/10.1002/mp.12155.
[13] A.M. Dinkla, et al., Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network, Med. Phys. 46 (9) (2019) 4095–4104, https://doi.org/10.1002/mp.13663.
[14] F. Liu, P. Yadav, A.M. Baschnagel, A.B. McMillan, MR-based treatment planning in radiation therapy using a deep learning approach, J. Appl. Clin. Med. Phys. 20 (3) (2019) 105–114, https://doi.org/10.1002/acm2.12554.
[15] D. Nie, et al., Medical image synthesis with context-aware generative adversarial networks, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 10435 (2017) 417–425, https://doi.org/10.1007/978-3-319-66179-7_48. LNCS.
[16] V. Bok, Deep Learning with Generative Adversarial Networks Jakub Langr. .
[17] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, Proc. IEEE Int. Conf. Comput. Vis. (2017) 2242–2251, https://doi.org/10.1109/ICCV.2017.244, 2017-Octob.
[18] Y. Liu, et al., MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method, Br. J. Radiol. 92 (1100) (2019), https://doi.org/10.1259/bjr.20190067.
[19] M. Maspero, et al., Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy, Phys. Med. Biol. 63 (18) (2018) aada6d, https://doi.org/10.1088/1361-6560/aada6d.
[20] A.M. Dinkla, et al., Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network, Med. Phys. 46 (9) (2019) 4095–4104, https://doi.org/10.1002/mp.13663.
[21] M.F. Spadea, et al., Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—application in brain proton therapy, Int. J. Radiat. Oncol. Biol. Phys. 105 (3) (2019) 495–503, https://doi.org/10.1016/j.ijrobp.2019.06.2535.
[22] S. Kazemifar, et al., MRI-only brain radiotherapy: assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach, Radiother. Oncol. 136 (2019) 56–63, https://doi.org/10.1016/j.radonc.2019.03.026.
[23] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J.P.W. Pluim, Elastix: A toolbox for intensity-based medical image registration, IEEE Trans. Med. Imag. 29 (1) (2010) 196–205, https://doi.org/10.1109/TMI.2009.2035616.
[24] D.P. Shamonin, E.E. Bron, B.P.F. Lelieveldt, M. Smits, S. Klein, M. Staring, Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease, Front. Neuroinf. 7 (JAN, 2014), https://doi.org/10.3389/fninf.2013.00050.
[25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. (2016) 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016-Decem.
[26] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, https://doi.org/10.1109/TIP.2003.819861.
[27] C. Pinter, A. Lasso, A. Wang, D. Jaffray, G. Fichtinger, SlicerRT: Radiation therapy research toolkit for 3D Slicer, Med. Phys. 39 (10) (2012) 6332–6338, https://doi.org/10.1118/1.4754659.
[28] V. Kearney, et al., Attention-aware discrimination for MR-to-CT image translation using cycle-consistent generative adversarial networks, Radiol. Artif. Intell. 2 (2) (2020), e190027, https://doi.org/10.1148/ryai.2020190027.
[29] H. Emami, M. Dong, S.P. Nejad-Davarani, C.K. Glide-Hurst, Generating synthetic CTs from magnetic resonance images using generative adversarial networks, Med. Phys. 45 (8) (2018) 3627–3636, https://doi.org/10.1002/mp.13047.
[30] X. Mao, Q. Li, H. Xie, R.Y.K. Lau, Z. Wang, S.P. Smolley, Least squares generative adversarial networks, Proc. IEEE Int. Conf. Comput. Vis. (2017) 2813–2821, https://doi.org/10.1109/ICCV.2017.304, 2017-Octob.
[31] Y. Peng, et al., Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning, Radiother. Oncol. 150 (2020) 217–224, https://doi.org/10.1016/j.radonc.2020.06.049.
[32] Y. Wang, C. Liu, X. Zhang, W. Deng, Synthetic CT generation based on T2 weighted MRI of nasopharyngeal carcinoma (NPC) using a deep convolutional neural network (DCNN), Front. Oncol. 9 (November, 2019), https://doi.org/10.3389/fonc.2019.01333.
[33] I. Human, H. Series, Iaea human health series publications, Accuracy Requir. Uncertainties Radiother 37 (2016) 1–2 [Online]. Available: http://www.iaea.org/Publications/index.html.
[34] M.J. Sheller, et al., Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data, Sci. Rep. 10 (1) (2020), https://doi.org/10.1038/s41598-020-69250-1.
[35] M. Hartley, T.S.G. Olsson, dtoolAI: reproducibility for deep learning, SSRN Electron. J. (2020), https://doi.org/10.2139/ssrn.3565984.
[36] H. Brendan McMahan, E. Moore, D. Ramage, S. Hampson, B. Agüera y Arcas, Communication-efficient learning of deep networks from decentralized data, Proc. 20th Int. Conf. Artif. Intell. Stat. AISTATS (2017), 2017.
[37] E. Alvarez Andres, et al., Dosimetry-driven quality measure of brain pseudo computed tomography generated from deep learning for MRI-only radiation therapy treatment planning, Int. J. Radiat. Oncol. Biol. Phys. 108 (3) (2020) 813–823, https://doi.org/10.1016/j.ijrobp.2020.05.006.