



OPEN

## Structurally-constrained optical-flow-guided adversarial generation of synthetic CT for MR-only radiotherapy treatment planning

Rajat Vajpayee<sup>1,3</sup>, Vismay Agrawal<sup>1,3</sup> & Ganapathy Krishnamurthi<sup>1,2</sup>

The rapid progress in image-to-image translation methods using deep neural networks has led to advancements in the generation of synthetic CT (sCT) in MR-only radiotherapy workflow. Replacement of CT with MR reduces unnecessary radiation exposure, financial cost and enables more accurate delineation of organs at risk. Previous generative adversarial networks (GANs) have been oriented towards MR to sCT generation. In this work, we have implemented multiple augmented cycle consistent GANs. The augmentation involves structural information constraint (StructCGAN), optical flow consistency constraint (FlowCGAN) and the combination of both the conditions (SFCGAN). The networks were trained and tested on a publicly available Gold Atlas project dataset, consisting of T2-weighted MR and CT volumes of 19 subjects from 3 different sites. The network was tested on 8 volumes acquired from the third site with a different scanner to assess the generalizability of the network on multicenter data. The results indicate that all the networks are robust to scanner variations. The best model, SFCGAN achieved an average ME of 0.9–5.9 HU, an average MAE of 40.4–4.7 HU and 57.2–1.4 dB PSNR outperforming previous research works. Moreover, the optical flow constraint between consecutive frames preserves the consistency across all views compared to 2D image-to-image translation methods. SFCGAN exploits the features of both StructCGAN and FlowCGAN by delivering structurally robust and 3D consistent sCT images. The research work serves as a benchmark for further research in MR-only radiotherapy.

Magnetic Resonance (MR) imaging and Computed Tomography (CT) are two widely used methods for radiotherapy planning and have been proved to be effective diagnosis methods. Over the last decade, the use of MR imaging to support radiation therapy (RT) treatment planning has considerably increased, as MR provides high contrast of soft tissues compared to CT images, enabling better delineation of organs at risk (OAR) Schmidt and Payne<sup>23</sup>, Njeh<sup>20</sup>, Devic<sup>7</sup>. Moreover, MR imaging offers more information about lesion activities, such as cellular density and functional imaging of lesions. The limitation of MR in RT planning is its inability to provide the electron density information needed for dose calculation. In contrast, CT imaging provides electron density information, an essential component in dose calculation Van der Heide et al.<sup>26</sup>, Kazemifar et al.<sup>14</sup>, Ulin et al.<sup>25</sup>. Since there is no direct relationship between electron density information and MR intensity values, it is prevalent to find a combined pipeline that includes MR imaging for precise delineation of OARs and CT simulation for dose calculation. Involving CT acquisition in the workflow has drawbacks like increased radiation exposure and financial cost. In an attempt to reduce the radiation exposure to patients, Wang et al.<sup>27</sup> have introduced a method to utilize the merits of ATV (Anisotropic Total Variation) and reweighted technique to achieve better limited-angle CT reconstruction. Chen et al.<sup>4</sup> have also proposed an approach to enhance edge information in CBCT reconstruction, which uses information obtained by mapping edge contours from existing CT to CBCT. Moreover, the delineations on MR images had to be translated to CT domain using image registration, which introduces spatial uncertainties, and is considered the weakest link in the radiotherapy workflow Njeh<sup>20</sup>. To overcome these issues, several methods have been developed for the generation of synthetic CT (sCT) from MR

<sup>1</sup>Department of Engineering Design, Indian Institute of Technology Madras, Chennai, India. <sup>2</sup>Robert Bosch Center for Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, India. <sup>3</sup>These authors contributed equally: Rajat Vajpayee and Vismay Agrawal. ✉email: rajatvajpayee@alumni.iitm.ac.in

	Site 1	Site 2	Site 3
Number of patients	8	7	4
<b>CT</b>			
Manufacturer	Siemens	Toshiba	Siemens
Model	Somatom Definition AS+	Aquilion	Emotion 6
Slice thickness (mm)	3	2	2.5
Pixel size (mm <sup>2</sup> )	0.98 0.98	1 1	0.98 0.98
Kernel	B30f	FC17	B41s
<b>T2w images</b>			
Manufacturer	GE	Siemens	GE
Model	Discovery 750 w	–	Signa PET/MR
Field strength (T)	3	1.5	3
Sequence type	FRFSE	TSE	FRFSE
Slice thickness (mm)	2.5	2.5	2.5
Pixel size (mm <sup>2</sup> )	0.875 0.875	0.875 0.875–1.1 1.1	0.875 0.875
Bandwidth (Hz/pixel)	390	200	390
TR (ms)	6000–6600	12000–16000	9000
TE (ms)	97	91–102	65
Encoding direction	COL	ROW	COL

**Table 1.** Acquisition parameters for the different sites.

images, which can be coarsely divided into 4 categories, (1) Atlas-based methods, (2) Bulk density methods, (3) Probabilistic based methods and (4) machine learning methods Largent et al.<sup>16</sup>.

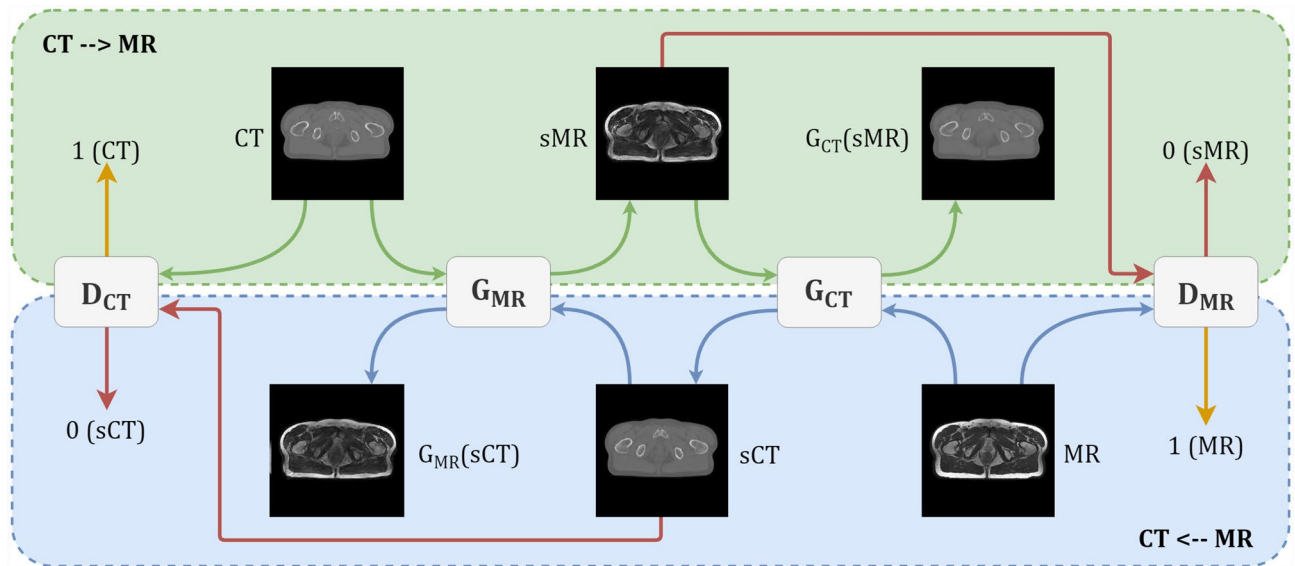
Hofmann et al.<sup>13</sup> was the first atlas-based approach for the computational conversion of MR to CT, achieving an MAE of 100.7 Hounsfield Units (HU). The bulk density method was initially introduced as a method for sCT generation. Anatomical contours from MR can be used to generate bulk anatomical density (BAD) maps Choi et al.<sup>6</sup>. Recent advancements in machine learning in the medical imaging field led to the development of sophisticated methods for synthetic image generation Ching et al.<sup>5</sup>. In 2013, Andreassen<sup>2</sup> proposed an approach to investigate the use of Random Forest regression (RaFR) and Gaussian Mixture Regression (GMR) for MR to CT translation. The introduction of deep learning techniques made it possible to perform image-to-image translation generalizable across datasets. In 2014, Goodfellow et al.<sup>10</sup> proposed a framework to estimate generative models using adversarial training of generative and discriminative networks. Since then, GANs have been consistently proved to be a promising path for achieving image-to-image translation tasks Alotaibi<sup>1</sup>. In Han<sup>11</sup>, a deep learning method was first introduced to generate sCT by training a generative model i.e. U-net to convert 2D MR to CT slice. They obtained an average MAE of 84.817.3 HU across test data. Maspero et al.<sup>18</sup> used phase, water and fat images as input to the GAN in pelvis data and obtained an average MAE of 61.09.0 HU on test data. In Wolterink et al.<sup>29</sup>, a generative model was trained on an unpaired brain dataset, and Nie et al.<sup>19</sup> employed a model on a paired dataset. Maspero et al.<sup>18</sup> has used cGAN to obtain prostate sCT for dosimetric calculations. So far, to our best knowledge, no deep learning based method has incorporated optical flow information for sCT generation.

This manuscript proposes deep learning models for MR to CT image generation in MR-only radiotherapy workflows for male pelvis data using a publicly available paired pelvis dataset. Test images have been taken from a scanner unseen to the trained models. This allows us to widen the scope of possibilities in training and improves the generalization of the conversion. Dosimetric (MAE and ME) and PSNR have been evaluated to quantify and compare the overall accuracy of sCT results with previously existing methods. The purpose of generating sCT from MR is to provide the electron density information for facilitating dose calculation in the absence of ground truth CT. The sCT encodes tissue information in HU which can be easily transformed to electron density in the treatment planning system (TPS) using a calibration curve. However, the dose calculation step is not incorporated in our research work, as the focus was to evaluate and compare the results among the proposed novel GANs.

## Materials and methods

**Dataset.** This study was conducted on a publicly available Gold Atlas project dataset Nyholm et al.<sup>21</sup>, which aimed to provide means for training and validation for segmentation and sCT algorithms. The data consists of male pelvic region T1- and T2-weighted MR and CT volumes of 19 patients acquired from 3 different sites. The dataset also includes consensus delineations of nine organs for each patient based on the T2-weighted MR images. The CT volumes were deformed to fit the anatomy of MR data, enabling the use of the organ delineations on the registered CT. Table 1 provides further details of the dataset acquisition parameters.

**Data preparation.** To remove the background noise from the scans, all the voxels outside the body in MR and CT images were set to 0 and -1024, respectively. The pelvis axial slices were converted to 16-bit grayscale images after truncating the intensity range of the original images. Similar to Fetty et al.<sup>9</sup>, the upper limit to truncate the image was kept as 1400 for CT, as any values greater than 1400 were identified in rare cases, mainly in the femoral bones, which is not considered in dose planning. For MR volumes, the upper limit was restricted to



**Figure 1.** Overall pipeline for all the models. The green section shows CT to MR translation and recovery back to input CT using MR and CT generators, respectively. Similarly, the blue part corresponds to MR to synthetic CT conversion steps. If the input of the discriminator is a synthesized image, then ideally discriminator should output 0, i.e. fake image. This is denoted by the red colour arrow.

99 percentile of all the intensity values to exclude outliers, followed by min-max normalization for each volume to normalize the intensity variations in MR scans. Inter-scan differences (air pockets and structures) have not been taken into account in this study.

The first and last two axial slices were removed from the corpus to avoid the aliasing effects in MR images. All the images were resized to (256, 256) dimension using bicubic interpolation. To facilitate the direct comparison of the results, we have used the split proposed by Boni et al.<sup>3</sup>, i.e. site 2 and site 3 acquisitions were used as the training set, whereas site 1 was used in the testing set. Such a split also tests the robustness of the network across scanners as the testing set contains data taken from a completely different scanner. In total, there are 1015 axial slices or 11 patient volumes in the training set and 659 axial slices or 8 volumes in the test set.

**Networks.** The study implements 4 different types of Generative Adversarial Networks (GANs) with a shared neural network architecture of generator and discriminator trained with different sets of loss functions, enabling the fair comparison of different loss functions to learn the given task. The first network used to obtain a baseline score in this study is the traditional CycleGAN Zhu et al.<sup>30</sup>. The other two networks add structural similarity (StructCGAN) and optical flow consistency (FlowCGAN) constraints during the optimization step of the generator in the GAN. The 4th network augments generator loss function with both structural similarity and optical flow constraints. All the networks were trained in a supervised fashion to establish a bidirectional mapping between CT and T2-weighted MR axial slices.

All the GANs implemented in this study have two generators and two discriminators. The generator,  $G_{CT}$  translated the MR image to produce synthetic CT, i.e.  $sCT = G_{CT}(MR)$ . Similarly, we have  $sMR = G_{MR}(CT)$ . The two discriminators,  $D_{CT}$  and  $D_{MR}$  aim to distinguish between real and synthetic images. The overall pipeline common to all 4 GANs is shown in the fig. 1.

The architecture of both the generators is ResNet He et al.<sup>12</sup> with 9 residual blocks. The discriminator has 5 convolutional blocks, each followed by instance normalization and ReLU non-linearity except the last layer. All ReLUs are leaky with a slope of 0.2. The first convolutional layer kernel has 64 channels, each with a size of (4, 4), and a stride of 2. Subsequently, in layer 2nd, 3rd and 4th, the number of channels in the kernel keep doubling without any changes in other parameters. The final convolutional block maps its input to a single number between 0 and 1, denoting the probability of prediction as a synthetic or real image.

**Loss functions.** *Cycle-consistent GAN (CycleGAN).* Traditionally, CycleGAN training has two types of terms in loss functions. The first corresponds to adversarial losses to match the data distribution of synthetic images with ground truth images. The second is cycle consistency loss which enables the model to learn a cycle-consistent mapping. Similar to Mao et al.<sup>17</sup>, least-square has been used instead of negative log-likelihood objective function to improve training stability and generate higher quality images. The generator is updated by fixing the discriminator parameters and vice-versa.

The least-squares GAN loss function is provided in Eq. (1). Upon backpropagation, it minimizes the possibility of the translated image being recognized as a synthesized image by the discriminator.

$$\mathcal{L}_{GAN} = \mathbb{E}_{CT \sim p_{data}(CT)} [(D_{MR}(sMR) - 1)^2] + \mathbb{E}_{MR \sim p_{data}(MR)} [(D_{CT}(sCT) - 1)^2] \quad (1)$$

The cycle loss function ensures the cyclic consistency between the input and generated images. The aim is to recover the input image when the synthesized image is reconstructed back to its original imaging modality. The loss function is as follows:

$$\mathcal{L}_{\text{cycle}} = \mathbb{E}_{\text{CT} \sim p_{\text{data}}(\text{CT})} [ \|G_{\text{CT}}(\text{sMR}) - \text{CT}\|_1 ] + \mathbb{E}_{\text{MR} \sim p_{\text{data}}(\text{MR})} [ \|G_{\text{MR}}(\text{sCT}) - \text{MR}\|_1 ] \quad (2)$$

The net generator loss function is provided in Eq. (3). Here,  $\lambda_{\text{cycle}}$  is taken as 10.

$$\mathcal{L}_{G_{\text{net}}^{\text{cycleGAN}}} = \mathcal{L}_{\text{GAN}} + \lambda_{\text{cycle}} * \mathcal{L}_{\text{cycle}} \quad (3)$$

The discriminator loss corresponding to CT images is provided in Eq. (4). The aim is to predict label 1 for a real imaging modality and label 0 for a synthetic image.

$$\mathcal{L}_{D_{\text{CT}}} = \mathbb{E}_{\text{MR} \sim p_{\text{data}}(\text{MR})} [(D_{\text{CT}}(\text{MR}) - 1)^2] + \mathbb{E}_{\text{MR} \sim p_{\text{data}}(\text{MR})} [(D_{\text{CT}}(\text{sCT}))^2] \quad (4)$$

Similarly, the loss corresponding to  $D_{\text{MR}}$  is calculated. The net discriminator loss is the average of the discriminator loss functions of individual imaging modalities, shown below.

$$\mathcal{L}_{\mathcal{D}_{\text{net}}} = \frac{(\mathcal{L}_{D_{\text{CT}}} + \mathcal{L}_{D_{\text{MR}}})}{2} \quad (5)$$

The above net discriminator loss function is the same for all of our models. In contrast, the net generator loss function changes case-by-case and is explicitly defined further in the manuscript.

*Structurally-Similar CycleGAN (StructCGAN).* MSE and L1 loss functions only measure the pixel-wise intensity difference. Structural Similarity (SSIM) index was introduced by Wang et al.<sup>28</sup> to assess the perceptual image quality of a distorted image given a reference image. The comparison is performed on 3 key features: 1) luminance, 2) contrast and 3) structure. The resultant is a decimal value in the range [0, 1], wherein higher values denote greater structural similarity. The equation is defined as:

$$\text{SSIM}_{\text{CT}} = \frac{(2\mu_{\text{CT}}\mu_{\text{sCT}} + C_1)(2\sigma_{\text{CT},\text{sCT}} + C_2)}{(\mu_{\text{CT}}^2 + \mu_{\text{sCT}}^2 + C_1)(\sigma_{\text{CT}}^2 + \sigma_{\text{sCT}}^2 + C_2)} \quad (6)$$

where,  $\mu_{\text{CT}}$  and  $\mu_{\text{sCT}}$  represent the average of the 2D axial image slice of the reference and synthesized CT images, respectively;  $\sigma_{\text{CT}}$  and  $\sigma_{\text{sCT}}$  are the variance of the reference and synthesized CT image respectively;  $\sigma_{\text{CT},\text{sCT}}$  is the covariance of the reference and synthesized CT image; and  $C_1$  and  $C_2$  are two hyper-parameters used to prevent instability, and were set as 0.0001 and 0.009 respectively, to stabilize the division with the weak denominator.

Similarly, the SSIM index for MR images is also calculated. The net SSIM index is the average SSIM index of individual imaging modalities. The structural similarity loss function is constructed by subtracting the net SSIM index from 1, shown below.

$$\mathcal{L}_{\text{SSIM}} = 1 - \frac{(\text{SSIM}_{\text{CT}} + \text{SSIM}_{\text{MR}})}{2} \quad (7)$$

The authors of the SSIM index asserts that it is more practical to apply SSIM locally rather than globally. Thus, the SSIM metric was applied regionally on windows of size (11, 11), and the overall mean was taken across all the regions to backpropagate the loss function. The generator net loss is shown in Eq. (8), wherein the  $\lambda_{\text{SSIM}}$  is taken as 2, and  $\lambda_{\text{cycle}}$  is taken as 8.

$$\mathcal{L}_{G_{\text{net}}^{\text{StructCGAN}}} = \mathcal{L}_{\text{GAN}} + \lambda_{\text{cycle}} * \mathcal{L}_{\text{cycle}} + \lambda_{\text{ssim}} * \mathcal{L}_{\text{ssim}} \quad (8)$$

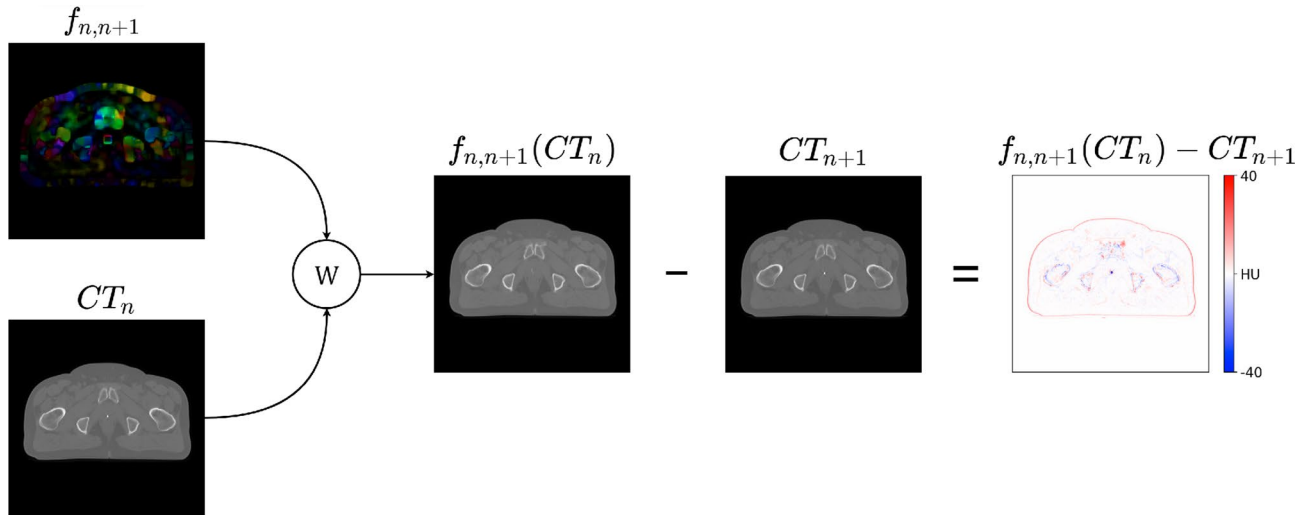
*Optical-Flow-Guided CycleGAN (FlowCGAN).* To the best of our knowledge, all the previous work in MR to sCT generation using adversarial methods used only 2D slices. Though it works quite well for individual 2D slices, there is an unavoidable discontinuity in the combined 3D volume. To facilitate interframe image consistency, we enforced L1 loss between a warped slice and a ground truth slice while training. The warping is done both forward and backward from the current slice, i.e. slices adjacent to a given slice in an epoch to enforce dual-sided optical flow consistency.

The optical flow is calculated in advance using the Farenback method Farneback<sup>8</sup> between the ground truth pairs adjacent slices. Figure 2 shows the dense optical flow field image and the corresponding warping from the current slice to the next for original CT pairs. The Farenback method for optical flow calculation is prone to negligible error, and the same is reflected in the rightmost image of Fig. 2. The optical flow loss equation for CT is as follows:

$$\mathcal{L}_{\text{flow}}^{\text{CT}} = \mathbb{E}_{\text{CT} \sim p_{\text{data}}(\text{CT})} [ \|f_{n,n-1}(\text{sCT}_n) - \text{CT}_{n-1}\|_1 ] + \mathbb{E}_{\text{CT} \sim p_{\text{data}}(\text{CT})} [ \|f_{n,n+1}(\text{sCT}_n) - \text{CT}_{n+1}\|_1 ] \quad (9)$$

Similarly, the optical flow loss for MR is also calculated. Furthermore, the optical flow loss of both the imaging modalities is averaged to obtain the net optical flow loss. The net generator loss for FlowCGAN is given in Eq. (10). Here, the  $\lambda_{\text{flow}}$  is set as 2 and  $\lambda_{\text{cycle}}$  as 8.

$$\mathcal{L}_{G_{\text{net}}^{\text{FlowCGAN}}} = \mathcal{L}_{\text{GAN}} + \lambda_{\text{cycle}} * \mathcal{L}_{\text{cycle}} + \lambda_{\text{flow}} * \mathcal{L}_{\text{flow}} \quad (10)$$



**Figure 2.** The warping of CT image from current to next slice using the flow information.

*Structurally-Similar Optical-Flow-Guided CycleGAN (SFCGAN).* The idea behind this network was to integrate the previous two networks so that images have high structural integrity and are coherent across slices. Therefore, both structural similarity and optical flow consistency loss were used together during the network training. The net generator loss equation is communicated in Eq. (11). The hyperparameters,  $\lambda_{SSIM}$  and  $\lambda_{flow}$  were taken as 2 each, and  $\lambda_{cycle}$  is taken as 6.

$$\mathcal{L}_{G_{net}^{SFCGAN}} = \mathcal{L}_{GAN} + \lambda_{cycle} * \mathcal{L}_{cycle} + \lambda_{SSIM} * \mathcal{L}_{SSIM} + \lambda_{flow} * \mathcal{L}_{flow} \tag{11}$$

**Objective function.** For every epoch, the generator was backpropagated followed by the discriminator. Moreover, following the strategy by Shrivastava et al.<sup>24</sup> to reduce model oscillation, the discriminator was updated using the buffer of 50 previously generated images, rather than the ones produced by the latest generators. The objective function is expressed as follows:

$$\begin{aligned} G_{MR}^*, G_{CT}^* &= \arg \min_{G_{MR}, G_{CT}} \mathcal{L}_{G_{net}}(G_{MR}, G_{CT}, D_{MR}, D_{CT}) \\ D_{MR}^*, D_{CT}^* &= \arg \min_{D_{MR}, D_{CT}} \mathcal{L}_{D_{net}}(G_{MR}, G_{CT}, D_{MR}, D_{CT}) \end{aligned} \tag{12}$$

**Training.** All the networks were trained for 200 epochs with a batch size of 4. Adam optimizer Kingma and Ba<sup>15</sup> was used with a learning rate of 0.0002. The learning rate was kept constant for the first 100 epochs and subsequently decayed linearly to zero over the next 100 epochs. The networks were implemented on PyTorch Paszke et al.<sup>22</sup> version 1.7.1 and trained on GeForce RTX 3090 GPU. The networks used approximately 11GB of GPU RAM for training. It took around 8 hrs to train each network.

**Comparison metrics.** Similar to other research work in this field Maspero et al.<sup>18</sup>, Fetty et al.<sup>9</sup>, we have estimated mean absolute error (MAE), mean error (ME) and peak signal-to-noise ratio (PSNR) between the prediction and ground truth volume inside the body contour. These parameters are calculated using these Equations:

MAE.

$$MAE(CT, sCT) = \frac{1}{n} \sum_{i=1}^n |CT(i) - sCT(i)| \tag{13}$$

ME.

$$ME(CT, sCT) = \frac{1}{n} \sum_{i=1}^n (CT(i) - sCT(i)) \tag{14}$$

Models	ME (HU)	MAE (HU)	PSNR (dB)
CycleGAN	− 9.3 4.9	49 4.4	56.1 1
StructCGAN	− 5.8 9.1	42.3 5.4	56.9 1.5
FlowCGAN	2 4.4	43.1 4.5	57 1.3
SFCGAN	− <b>0.9 5.9</b>	<b>40.4 4.7</b>	<b>57.2 1.4</b>
pix2pixHD*	− 18.3	48.5 6	−
pix2pix*	− 11.4 19	62.0 12	−

**Table 2.** The table shows the scores of different model averaged across test volumes inside the whole body contour. \*Taken from Boni et al.<sup>3</sup>. Results from SFCGAN are in bold font.

	Models	Anal canal	Penile Bulb	Prostate	Seminal vesicles	Bladder	Rectum	Femoral head L	Femoral head R
<b>ME (HU)</b>									
	CycleGAN	− 15.8 14.5	14.7 20.7	− <b>1.3 24.7</b>	− 11 54.9	21 32.3	− 66.4 80.4	30.1 70	− 3.5 41.3
	StructCGAN	− 15.3 16.7	49.7 17.4	24.4 20.6	− 3.3 52.7	63.7 24.1	− 104.7 86.2	28 64.1	24 39.5
	FlowCGAN	− 8.9 25.3	32.8 14.1	18.1 17.1	− 13.2 59.9	65.9 19.8	− 79.2 88.7	14.4 60.1	− <b>9.5 51.2</b>
	SFCGAN	− <b>3.8 16.8</b>	<b>4.8 17.4</b>	20.4 16.4	− <b>1.2 32.7</b>	41.9 23.3	− <b>63.1 90</b>	<b>12.3 58.1</b>	12.2 37.3
	pix2pixHD*	− 24.6 18	− 19.2 15	− 11.6 12	− 14.0 26	− 23.9 23	− 77.6 90	−	−
	pix2pix*	− 26.4 16	38.6 25	17.5 29	13.1 19	− <b>0.6 31</b>	− 85.2 80	−	−
<b>MAE (HU)</b>									
	CycleGAN	32.5 14	25.6 15.3	52.5 19.6	67.3 32.5	61.3 20.8	119.2 58.3	111.5 32.2	104.4 15.4
	StructCGAN	23.3 12.5	53.7 13.3	48.8 10.3	53.8 26.1	80.7 18.1	119.5 80.1	90 32.4	82.1 11.2
	FlowCGAN	23.4 19.7	43.3 10.3	47 10.4	63.8 30	83.4 18.9	116.7 73.4	<b>89.7 24.5</b>	<b>80.6 10.6</b>
	SFCGAN	<b>21.9 10.8</b>	<b>25.4 8.5</b>	<b>44.8 11.8</b>	<b>45.9 11.4</b>	61.1 18.5	107.9 68.2	98.1 25.7	93.8 11
	pix2pixHD*	30.3 14	28.1 9	47.1 6	44.7 15	<b>49.4 12</b>	<b>101.8 78</b>	−	−
	pix2pix*	36.0 13	56.5 16	62.3 9	54.8 11	61.6 10	109.8 78	−	−
<b>PSNR (dB)</b>									
	CycleGAN	62.4 5.7	67.1 4.2	55.7 4.1	58.2 3.8	58.9 2.6	52.4 3.7	53.5 2.3	53.8 1.3
	StructCGAN	66.6 5.2	60.8 1.8	56.1 3.6	60.1 4.3	57.4 2	52.7 4.6	<b>55.3 2.6</b>	55.7 1.1
	FlowCGAN	66.5 6.7	62.3 1.9	56.2 3.5	58.7 4.3	56.8 1.6	52.4 4.3	<b>55.3 2.3</b>	<b>56.1 1.3</b>
	SFCGAN	<b>67 4.5</b>	<b>66.8 2.8</b>	<b>56.3 3.5</b>	<b>60.9 2.6</b>	<b>59.5 2.4</b>	<b>53.6 4.8</b>	54.5 1.9	54.9 0.8

**Table 3.** ME, MAE and PSNR values of consensus delineations of organs. Best results are in bold font.

PSNR.

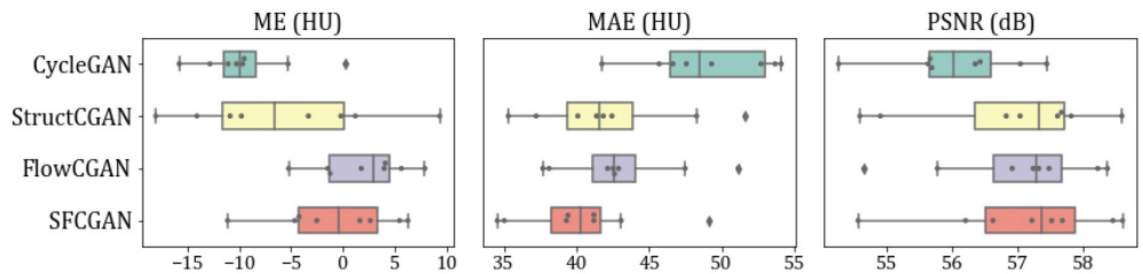
$$\text{PSNR}(CT, sCT) = 20 \log_{10} \frac{\text{MAX}_I}{\text{MSE}(CT, sCT)} \quad (15)$$

Where,  $n$  denotes the total number of pixels in a CT volume.  $CT(i)$  and  $sCT(i)$  are the intensity values (in HU) of the  $i$ th pixel in the ground truth and synthesized CT volumes, respectively. The  $\text{MAX}_I$  value in PSNR is taken as 65535, corresponding to the maximum intensity value of a 16-bit image. MAE and ME are calculated after denormalizing the 16-bit output image to HU.

## Results

Since only the generator without backpropagation is required during the test phase, hardly 1.5GB of GPU RAM was utilized. CT synthesis took an average of 8 secs on our GPU for all the 8 test patients.

Table 2 provides the comparison metrics scores of the whole volume for all the GANs implemented in our study. The results for pix2pixHD and pix2pix taken from Boni et al.<sup>3</sup> is also shown for comparison. The comparison is fair as they have also used the same dataset splits for training and testing purposes. SFCGAN outperforms all the networks with an average ME of 0.9 5.9 HU, an average MAE of 40.4 4.7 HU and 57.2 1.4 dB PSNR. The StructCGAN and FlowCGAN outperform both pix2pixHD and pix2pix network, but CycleGAN fails to exceed pix2pixHD.



**Figure 3.** Box plots illustrating the Mean Error (ME), Mean Absolute Error (MAE) and PSNR scores distribution across test dataset volumes for CycleGAN, StructCGAN, FlowCGAN and SFCGAN networks.

All the comparison metrics were also calculated for 8 segmented organs, and the results are provided in Table 3. Comparing MAE scores, SFCGAN performs the best in the case of anal canal, penile bulb, prostate and seminal vesicles. FlowCGAN outperforms all the networks in both left and right femoral heads. For urinary bladder and rectum, pix2pixHD performs the best.

In Fig. 3, boxplots of ME, MAE and PSNR have shown. It can be inferred that the MAE of FlowCGAN has the minimum IQR (Inter Quartile Range) followed by SFCGAN with a slight difference of 0.52 HU. Nevertheless, SFCGAN outperforms FlowCGAN as the difference between their MAE score is 2.26 HU. It can be seen that our baseline model CycleGAN has minimum variability, followed by SFCGAN and FlowCGAN. PSNR has been computed to evaluate the quality of sCT. SFCGAN achieved the highest median of 57.33 dB, followed by FlowCGAN of 57.28 dB.

Figure 4 shows example output images of CycleGAN and SFCGAN. The coronal cross-section in the figure is obtained by stacking elements from axial slices. The axial-coronal slice pairs are taken from different patients and correspond to different pelvis segments for diverse visual inspection.

## Discussion

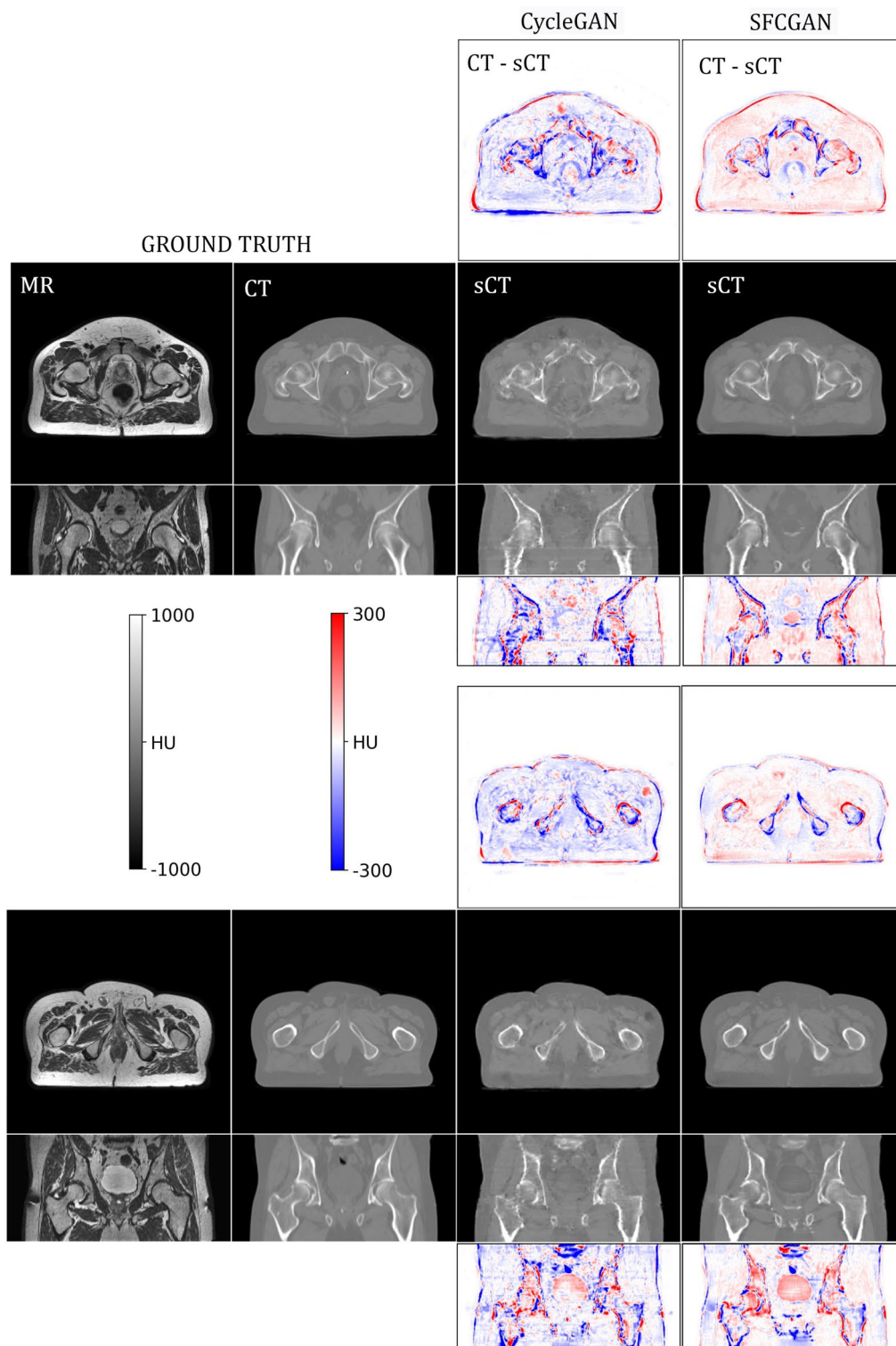
We inspected a new deep learning method for MR to sCT generation for dose calculation. Proposed methods consist of training 3 kinds of models (StructCGAN, FlowCGAN, SFCGAN) with CycleGAN as the base model trained with paired pelvis dataset. We evaluated (Table 2) the synthetic CT images generated by comparing their intensities with original CT using MAE and ME. The results showed that SFCGAN achieved high compliance between the HU maps for both CT images by showing the lowest ME and MAE of -0.9 5.9 and 40.4 4.7 respectively. Also, PSNR has been used to evaluate the quality of reconstructed images and SFCGAN achieved the highest 57.2 1.4, which is superior to our baseline model CycleGAN.

Figure 4 shows the output generated by SFCGAN. It can be seen that differences in the intensity values of soft tissues in original CT and sCT is minimal. Although sCT has brighter pixels at the edges compared to the CT, there is a considerable difference in intensity values at the bone and soft tissue boundaries. Additionally, incorporating optical flow information as an input enables our model to maintain the interframe consistency in sCT. The interframe consistency is reflected in the continuity of bones and tissues along the vertical axis in the coronal cross-section of SFCGAN results. Unsurprisingly, the same is absent in CycleGAN.

The study can be further improved by including a larger dataset. Unfortunately, to the best of our knowledge, there is no alternative publicly available paired pelvis dataset for research purposes. To minimize the limitation of the study posed by small dataset, the testing is performed on 8 pelvis volumes taken from a completely different site. The results are promising across unseen scanner data and indicate generalizability of the network across multiple scanners. These results show that our proposed model provides an accurate and reproducible sCT which can improve the robustness of MR-only radiotherapy workflow and reduce the clinical workload of performing a CT scan for the patients who are unable to undergo a CT scan due to radiation exposure.

## Conclusion

In this work, we developed different types of CycleGAN models for generating synthetic medical images and compared the results against state-of-the-art results Boni et al.<sup>3</sup>. FlowCGAN can serve as an approximation to provide the accurate motion of pixel values in sCT images to maintain interframe consistency. StructCGAN generates CT images that are structurally robust and can provide better accuracy for diagnosing fractures. SFCGAN exploits the features of both StructCGAN and FlowCGAN by proving structurally robust and frame consistency in the sCT images. The model performed well on the dataset acquired from entirely different sites, showing the model's generalizability. The model serves as a benchmark for further research in MR-only radiotherapy and dose estimation in pelvis scans.



**Figure 4.** MR to CT conversion. In the above figure, column (i) represents ground truth MR (ii) represents ground truth CT . In column (iii) and (iv), the black background images are synthesized CT by CycleGAN (baseline) and SFCGAN respectively. The images in the white background represents the corresponding error in synthesized CT.



## Data availability

The data that support the findings of this study are available from Tufve (record owner), but restrictions apply to the availability of these data, which were used under license for the current study. Data are however available from the owner upon reasonable request at <https://zenodo.org/record/583096> We hereby confirm that all methods were carried out in accordance with relevant guidelines and regulations as publicly available data is used in the study.

Received: 28 February 2022; Accepted: 8 August 2022

Published online: 01 September 2022

## References

- Alotaibi, A. Deep generative adversarial networks for image-to-image translation: A review. *Symmetry* **12**(10), 1705 (2020).
- Andreasen, D. Creating a Pseudo-CT from MRI for MRI-only based Radiation Therapy Planning. DTU supervisor: Koen Van Leemput, Ph. D., kvle@dtu.dk, DTU Compute. Matematiktorvet, Building (2013).
- Boni, B. K. B. *et al.* MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. *Phys. Med. Biol.* **65**(7), 075002 (2020).
- Chen, Y. *et al.* Low dose CBCT reconstruction via prior contour based total variation regularization (PCTV): A feasibility study. *Phys. Med. Biol.* <https://doi.org/10.1088/1361-6560/aab68d> (2018).
- Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J. Royal Soc. Interface* **15**(141), 20170387 (2018).
- Choi, JH. *et al.* Bulk anatomical density based dose calculation for patient-specific quality assurance of MRI-only prostate radiotherapy. *Frontiers in Oncology* (2019), p. 997.
- Devic, S. MRI simulation for radiotherapy treatment planning. *Med. Phys.* **39**(11), 6701–6711 (2012).
- Farneback, G. Two-frame motion estimation based on polynomial expansion. *Scandinavian conference on Image analysis*. Springer. 2003, pp. 363–370.
- Fetty, L. *et al.* Investigating conditional GAN performance with different generator architectures, an ensemble model, and different MR scanners for MR-sCT conversion. *Phys. Med. Biol.* **65**(10), 105004 (2020).
- Goodfellow, I., *et al.* Generative adversarial nets. *Adv. Neural Inform. Process. Syst.* **27** (2014).
- Han, X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med. Phys.* **44**(4), 1408–1419 (2017).
- He, K. *et al.* Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.
- Hofmann, M. *et al.* MRI-based attenuation correction for PET/MRI: A novel approach combining pattern recognition and atlas registration. *J. Nucl. Med.* **49**(11), 1875–1883 (2008).
- Kazemifar, S. *et al.* MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother. Oncol.* **136**, 56–63 (2019).
- Kingma, D. P., & Ba, J. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- Largent, A., & *et al.* “45 A comparison of pseudo-CT generation methods for prostate MRI-based dose planning: deep learning, patch-based, atlas-based and bulk-density methods”. *Physica Medica: European Journal of Medical Physics* **68** (2019), p. 28.
- Mao, X. *et al.* Least squares generative adversarial networks. *Proceedings of the IEEE International Conference On Computer Vision*. 2017, pp. 2794–2802.
- Maspero, M. *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), 185001 (2018).
- Nie, D. *et al.* Medical image synthesis with context-aware generative adversarial networks. *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer. 2017, pp. 417–425.
- Njeh, C. F. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J. Med. phys./Assoc. Med. Phys. India* **33**(4), 136 (2008).
- Nyholm, T. *et al.* MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project. *Med. Phys.* **45**(3), 1295–1300 (2018).
- Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inform. Process. Syst.* **32** (2019).
- Schmidt, M. & Payne, G. Radiotherapy planning using MRI. *Phys. Med. Biol.* **60**(22), R323 (2015).
- Shrivastava, A. *et al.* Learning from simulated and unsupervised images through adversarial training. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 2107–2116.
- Ulin, Kenneth, M Urie, Marcia, & M Cherlow, Joel. “Results of a multi-institutional benchmark test for cranial CT/MR image registration”. *International Journal of Radiation Oncology\* Biology\* Physics* **77**(5) (2010), pp. 1584–1589.
- Van der Heide, U. A. *et al.* Functional MRI for radiotherapy dose painting. *Mag. Reson. Imaging* **30**(9), 1216–1223 (2012).
- Wang, T. *et al.* Reweighted anisotropic total variation minimization for limited-angle CT reconstruction. *IEEE Trans. Nucl. Sci.* **64**(10), 2742–2760. <https://doi.org/10.1109/TNS.2017.2750199> (2017).
- Wang, Z. *et al.* Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- Wolterink, J. M. *et al.* Deep MR to CT synthesis using unpaired data. *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer. 2017, pp. 14–23.
- Zhu, J.-Y. *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2223–2232.

## Acknowledgements

We gratefully acknowledge the support of the engineering design department development fund.

## Author contribution

R.V., R.V. provided equal contribution in writing manuscript text. V.A., R.V. provided equal contribution in preparing figures and tables. G.K. reviewed the manuscript and provided guidance for research work.

## Competing interest

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to R.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022