Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data

Kévin N D Brou Boni[1, 2, a)], John Klein[2], Akos Gulyban[3], Nick Reynaert[1,3], David Pasquier[2,4]

[1] Centre Oscar Lambret, Department of Medical Physics, Lille, France
[2] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
[3] Institut Jules Bordet, Department of Medical Physics, Brussels, Belgium
[4] Centre Oscar Lambret, Department of Radiotherapy, Lille, France
[a)] Author to whom correspondence should be addressed. Electronic mail: k-brouboni@o-lambret.fr

---

Abstract

Purpose: MR-to-CT synthesis is one of the first steps in the establishment of an MRI-only workflow in radiotherapy. Current MR-to-CT synthesis methods in deep learning use unpaired MR and CT training images with a Cycle Generative Adversarial Network (CycleGAN) to minimize the effect of misalignment between paired images. However, this approach critically assumes that the underlying inter-domain mapping is approximately deterministic and one-to-one. In the current study, we use an Augmented CycleGAN (AugCGAN) model to create a robust model that can be applied to different scanners and sequences using unpaired data.

Materials and methods: This study included T2-weighted MR and CT pelvic images of 38 patients in treatment position from 5 different centers. The AugCGAN was trained on 2D transverse slices of 19 patients from 3 different sites. The network was then used to generate synthetic CT (sCT) images of 19 patients from the two other sites. Mean absolute errors (MAEs) for each patient were evaluated between real and synthetic CT images. Original treatment plans of 9 patients were re-calculated using sCT images to assess the dose distribution in terms of voxel-wise dose difference, gamma, and dose volume histogram analysis.

Results: The mean MAEs were 59.8 Hounsfield units (HU) and 65.8 HU for the first and second test sites, respectively. The maximum dose difference to the target was 1.2% with a gamma pass rate using the 3%, 3 mm criteria above 99%. The average time required to generate a complete sCT image for a patient on our GPU was 8.5 s.

Conclusion: This study suggests that our unpaired approach achieves good performance in generalization with respect to sCT image generation.

I. Introduction

Since its introduction in the late 80s, magnetic resonance imaging (MRI) is being increasingly used in radiotherapy. MRI has been particularly appreciated for its superior soft-tissue contrast over computed tomography (CT). As a result, MRI can better discern tumors than CT images, and has therefore been integrated into the radiotherapy treatment workflow. In practice, after MR acquisition, regions of interest (ROIs) are contoured directly on the image by a radiation oncologist. However, CT is still required since it serves as the basis for dose calculation because of the need for electron-density information. The contours drawn on an MR image are propagated to a CT image after co-registration. In the professional community, this step is considered as the weakest link in the radiotherapy workflow.[1] The image-registration process introduces spatial uncertainties whose significance depends on the localization.

To reduce these systematic spatial uncertainties introduced by inter-scan differences, interest has been rapidly growing in MRI-only radiotherapy.[2] However, the main challenge associated with this modality is that MRI intensities cannot be directly used to obtain information regarding electron density, which is required for dose calculation in radiation therapy treatment planning systems (TPS).

This problem can be partially solved by converting MRI into a so-called synthetic CT (sCT). Several methods for this purpose have recently emerged with the advent of deep learning, especially methods using Generative Adversarial Networks (GANs).[3,4] These networks rely on learning the relationship between MRI and CT images in a paired[5,6] or unpaired fashion.[7,8] (Figure 1). Learning in a paired fashion requires perfect registration of a set of MR and CT images, a time-consuming task that often involves deformable registration and consequently re-introduces geometrical uncertaintiy.[9,10] Due to their adaptability/flexibility, unpaired models using at most rigid registration and based on CycleGAN[11] have attracted more interest and offer the potential to improve sCT generation accuracy by using appropriate models such as a 3D model[12] or a deeper model.[13] Nevertheless, these models are not robust to the MR variability arising in multicentric workflows, highlighting the common generalization issues in machine learning when dealing with out-of-distribution samples. Since a cycleGAN only learns one-to-one mappings,[14] i.e., the model associates each CT with a single MR sequence, mapping of several MRI instances to the same CT image represents a bottleneck that will become a major limitation in the clinical implementation of this approach.[12]
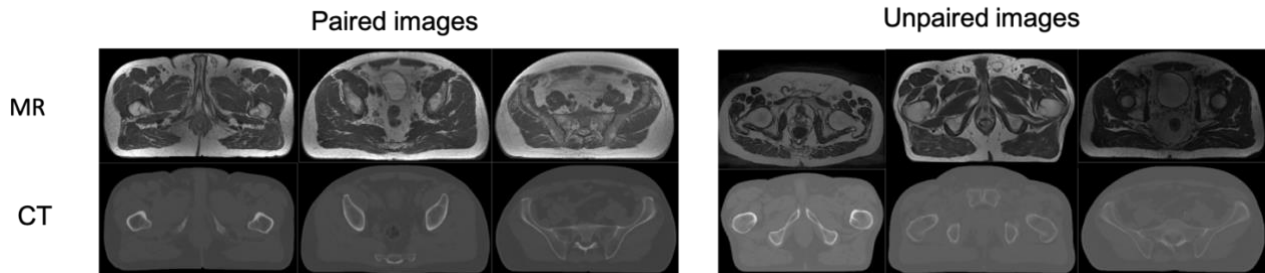

Figure 1: In the paired fashion, the input MR and ground truth CT slices correspond to the same patient at the same anatomical localization. In contrast, models based on the unpaired fashion use CT and MR slices of different patients at different anatomical localizations during training.

In this paper, we aimed to improve the generalizability of MR-to-CT synthesis with unpaired data. We devised an Augmented CycleGAN[14] (AugCGAN) that allowed the use of multicentric data in both training and testing phases. We evaluated the accuracy of sCT images obtained with this approach by performing both image and dosimetric comparisons in order to show that our model can learn from several MRI distributions.

II.   Materials and methods
A.   Unsupervised learning of one-to-many
1.   Unpaired learning

Given two distributions $p_{mr}(mr)$ from MRI and $p_{ct}(ct)$ from CT, we assume there exists a mapping between these two domains only if they are highly dependent, with the same localization as the instance. To recover this mapping, Zhu *et al* (2017) used two GANs with cycle consistency. In MR-to-CT, we can define the first GAN as $\{G_{mr2ct} : MRI \mapsto CT, D_{ct} : CT \mapsto \{0,1\}\}$ and the second as $\{G_{ct2mr} : CT \mapsto MRI, D_{mr} : MRI \mapsto \{0,1\}\}$. These networks follow the standard optimization described by Goodfellow *et al*[15] with an adversarial objective (marginal matching) formulated for the first CT-GAN as follows:

$$\mathcal{L}_{GAN}^{CT}(G_{mr2ct}, D_{ct}) = \mathbb{E}_{ct \sim p_{data}(ct)}[\log D_{ct}(ct)] + \mathbb{E}_{mr \sim p_{data}(mr)}[\log(1 - \tilde{ct})],$$

$$\widetilde{ct} = G_{mr2ct}(mr). \quad (1)$$

and similarly for the second MR-GAN $\{G_{ct2mr}, D_{mr}\}$. The goal here is that the output of each mapping should match the distribution of the target domain. Each GAN learns a generator (G) that is meant to fool a discriminator (D), i.e., a binary classifier that discriminates genuine samples from generated ones.

Unpaired learning is possible thanks to cycle consistency: transferring a modality to the other and then back into it produces a reconstructed image close to the original. The cycle-consistency loss starting from an MRI slice is given by:

$$\mathcal{L}_{cyc}^{MRI}(G_{mr2ct}, G_{ct2mr}) = \mathbb{E}_{mr \sim p_{data}(mr)} \|mr' - mr\|_1,$$
$$mr' = G_{ct2mr}(\widetilde{ct}). \quad (2)$$

Similarly, when starting from a CT slice.

2. Limitations of CycleGAN

The learning process cannot be performed correctly in the one-to-many case. Since the CycleGAN model generates deterministic mappings,[16] the generator G_{ct2mr} cannot be optimized properly when the MR outputs differ substantially (constructor, sequence); see Figure 2 for an illustration of these discrepancies in terms of voxel intensity distributions. The main difficulty is that the cycle consistencies when plugged together amount to require that $G_{mr2ct}(G_{ct2mr}(ct)) \approx ct'$; therefore, the generators must be inverse functions of one another and thus bijective, which contradicts the fact that many MR images must be mapped to the same CT image.

To overcome this learning problem, we propose to resort to an "augmentation" proposed by Almahari et al.[14] to capture any missing information when generating or reconstructing an MR image from G_{ct2mr}.
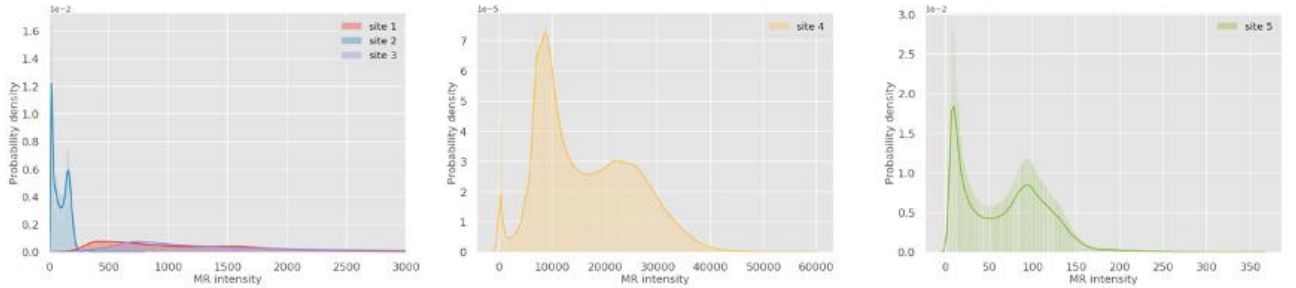

Figure 2: Distributions on MR intensity according different scanners. Zero values were not taken into account.

3. Augmented CycleGAN in MR-to-CT

Instead of learning a mapping $G_{ct2mr}: CT \mapsto MRI$ with a single input, this generator is augmented by including a latent space Z_{MRI} as an additional input space that captures missing information (i.e., "style" information characterizing the MRI distribution in terms of sequence type and manufacturer). The inverse mapping G_{mr2ct} causes the loss of this crucial information needed to synthetize or reconstruct the correct MRI style; thus, the latent space allows recovery of missing information when transforming an MRI into a CT.

The CT-to-MR mapping becomes a stochastic mapping $G_{ct2mr}: CT \times Z_{MRI} \mapsto MRI$ capable of correctly optimizing the cycle consistency loss. The proposed model captures Z_{MRI} by using an encoder $E_{mr}: CT \times MRI \mapsto Z_{MRI}$. In total, an AugCGAN is composed of five networks: two in the first GAN $\{G_{mr2ct}, D_{ct}\}$, two in the second GAN $\{G_{ct2mr}, D_{mr}\}$, and one encoder $\{E_{mr}\}$.

The authors adopt a marginal matching loss $\mathcal{L}_{GAN}^{Z_{MRI}}$ for the latent space in order to encourage the

encoder E_{mr} to produce realistic samples matching $p(z_{mr})$, a standard Gaussian prior over Z_{MRI}:

$$\mathcal{L}_{GAN}^{Z_{MRI}}\left(G_{mr2ct}, E_{mr}, D_{z_{mr}}\right) = \mathbb{E}_{z_{mr}\sim p(z_{mr})}\left[\log D_{z_{mr}}(z_{mr})\right] + \mathbb{E}_{mr\sim p_{data}(mr)}\left[\log\left(1 - D_{z_{mr}}(\widetilde{z_{mr}})\right)\right],$$
$$\widetilde{z_{mr}} = E_{mr}(\tilde{c}t, mr), (3)$$

this latent space helps the reconstruction of $mr' = G_{ct2mr}(\tilde{c}t, \widetilde{z_{mr}})$. Cycle consistency is also extended to the latent space:

$$\mathcal{L}_{cyc}^{Z_{MRI}}(G_{ct2mr}, E_{mr}) = \mathbb{E}_{\substack{mr\sim p_{data}(mr) \\ z_{mr}\sim p_{data}(z_{mr})}} \|z_{mr} - z'_{mr}\|_1,$$

$$z'_{mr} = E_{mr}(ct, \widetilde{mr}), \quad \widetilde{mr} = G_{ct2mr}(ct, z_{mr}). (4)$$

Training in the forward direction CT-to-MR is done by the following optimization:

$$\mathcal{L}_{GAN}^{MRI}(G_{ct2mr}, D_{mr}) + \mathcal{L}_{cyc}^{CT}(G_{ct2mr}, G_{mr2ct}) + \gamma.\mathcal{L}_{cyc}^{Z_{MRI}}(G_{ct2mr}, E_{mr}), (5)$$

with $\gamma = 0.025$, training in the backward direction (MR-to-CT) is based on the following optimization:

$$\mathcal{L}_{GAN}^{CT}(G_{mr2ct}, D_{ct}) + \mathcal{L}_{GAN}^{Z_{MRI}}\left(G_{mr2ct}, E_{mr}, D_{z_{mr}}\right) + \mathcal{L}_{cyc}^{MRI}(G_{mr2ct}, E_{mr}, G_{ct2mr}). (6)$$

The AugCGAN is trained on both objectives simultaneously; Figure 3 summarizes both mappings.
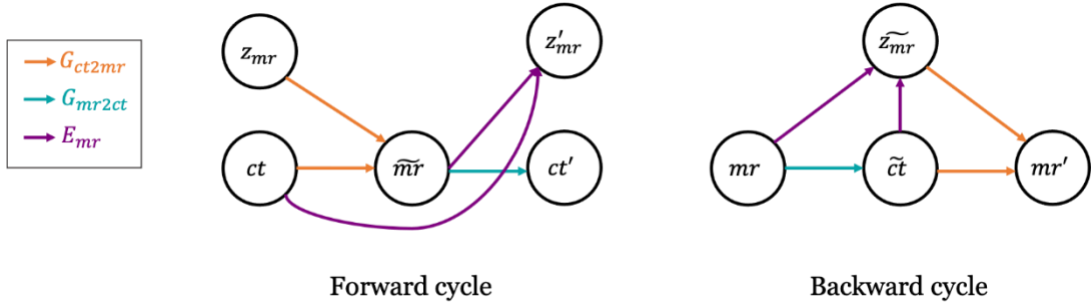

Figure 3: Augmented Cycle GAN model consists of a forward and backward cycle in which a generator is augmented with a latent space to allow a better reconstructed MR image.

B.   Patient data collection

This study included pelvic MR and CT images of 39 patients with prostate or rectal cancer obtained from 5 different sites. All scans were acquired in the radiotherapy treatment position. Half of the images were obtained from the public dataset named Gold Atlas project (GaP),[17] which aims to provide a source of training and validation images for segmentation as well as sCT generation. The remaining half were obtained from *Institut Jules Bordet (IJB)* and *Centre Oscar Lambret (COL)*.

Table 1 provides the acquisition settings for both CT and MRI and site-wise splitting of the data for network training and evaluation. Patients with hip implants were included in training and testing. Observe that different manufacturers and different parameters are used, thereby making this multicentric experimental setting inappropriate for cycle GAN without augmentation.

Table 1: Acquisition settings for the five sites. TSE stands for turbo spin echo, FRFSE for fast recovery

fast spin echo, and Col for columns.

	Training			Test	
	Site 1 (GaP)	Site 2 (GaP)	Site 3 (GaP)	Site 4 (IJB)	Site 5 (COL)
Patients	8	7	4	8	8
Slices	79–100	92–116	66–91	45–86	72–88
CT					
Manufacturer	Siemens	Toshiba	Siemens	Toshiba	Siemens
Model	Somatom Definition AS+	Aquilon	Emotion 6	Aquilon	Somaton Confidence
Slice thickness (mm)	3	2	2.5	3	3
Kernel	B30f	FC17	B41s	FC17	Br38f
T2-w					
Manufacturer	GE	Siemens	GE	Siemens	Siemens
Model	Discovery 750w 3T	- 1.5T	Signa PET/MR 3T	Magnetom Skyra 3T	Magnetom Sola 1.5T
Sequence type	FRFSE	TSE	FRFSE	TSE	AX T2 SPACE
Slice thickness (mm)	2.5	2.5	2.5	3	2
Bandwidth (Hz/pixel)	390	200	390	255	651
Encoding direction	Col	Row	Col	Row	Row
TR (ms)	6000–6600	12000–16000	6000–10000	3500	1770
TE (ms)	97	91–102	65	132	172

C.   Image pre-processing

To ensure a simple and smooth workflow, patient images underwent three pre-processing steps:

-Registration: CT images were rigidly registered and resampled to MR images. CT images were then cropped to match the field of view (FOV) of MRI. Registration for patients in site 5 was performed in Raystation (v 9A, RaySearch Laboratories, Sweden) by radiation oncologists, while the other patients were registered using Elastix.[18]

-Exclusion of surrounding air/noise: A mask excluding surrounding air or outside structures was obtained on the CT and MR images by using the external ROI option (threshold level based) on Raystation. Voxels outside the body were automatically assigned to -1024 HU for CT and 0 for MR.

-Normalization: Limit values were arbitrarily set at 1500 HU and depending on the MR constructor 10000 or 15000. HU values were normalized, MR intensities as well site-wise following the min-max feature scaling.

D.   Training of the Networks

We adapted the AugCGAN implementation to make it suitable for radiotherapy image specifications. Modifications over the vanilla AugCGAN have been made in order to use 16-bit greyscale images, modify input size, and use data augmentation.

The 39 patients were separated into a training set containing 19 patients from site 1 to 3. The 19 patients from sites 4 and 5 were used for testing. The network was trained using RMSProp optimizer with an initial learning rate of 0.0002 for 150 epochs, then for another 150 epochs with a linearly decay learning rate to zero.

Training took an average of 28 h on an Nvidia Tesla V100 SXM2 (GPU) with a batchsize of 4. Original images were resampled to 256×256. Then, 128×128 sub-images were randomly cropped during training. Data augmentation was performed by horizontal flip, increasing the size of the training set to 3328 image pairs.

E.   sCT evaluation

In the testing phase, the working image size was 256×256 pixels. sCTs were generated using only the generator on the GPU (Nvidia Quadro P6000). The created image files were then converted into DICOM format by using MICE Toolkit (v1.1.3, NONPI Medical AB, Sweden).

1.   Image comparison

sCT and CT were compared on a voxel-wise basis by using the mean absolute error (MAE) and the mean error (ME). Considering the voxels within the body contours, MAE and ME in HU were calculated for each patient on MICE Toolkit. Detailed statistics by organ are only available for site 5, where delineated structures were copied and resampled from CT to the sCT.

2.   Dose comparison

Halcyon (VMAT) treatment plans were optimized on each CT image in Raystation by using the Collapsed Cone (v5.1) algorithm on a $3 \times 3 \times 3 \ mm^3$ grid. Optimized clinical plans were then recalculated on the sCT image for dose comparison. Only the 8 patients of site 5 underwent this procedure.

A dose-volume histogram (DVH) analysis was performed after copying the structures (PTV, femoral heads, bladder, and rectum) to sCT. The chosen DVH points were D_{98}, D_{50} and D_2. Voxel-wise absolute dose differences in percentage were computed within a dose threshold of $90\%, 50\%$, and 10% of the prescribed dose D_p. Finally, two different gamma pass rates ($\gamma_{3\%,3mm}, \gamma_{2\%,2mm}$) of the dose distributions were calculated using four different lower dose thresholds (0%, $<10\%$, $<50\%$, and $<90\%$).

III.   Results

CT and MR scans were not necessarily acquired on the same day; the median interval between the scans was 2 days (range: 1 h to 19 days). Interscan differences were therefore not taken into account in this study. CT synthesis took on average 8.5 s on the GPU.

A. Image comparison

Figure 4 shows an example of two test patients from sites 4 and 5. Differences are most acute in the external geometry and bone structures.

The proposed method produced an average MAE of 65.8 ± 9 HU and an average ME of -5.8 ± 6 HU for site 4. Patients from site 5 showed an average MAE of 59.8 ± 11 HU and an average ME of -0.7 ± 3 HU, indicating low variability across sites. Table 2 provides the average MAE and ME by ROIs only for site 5.
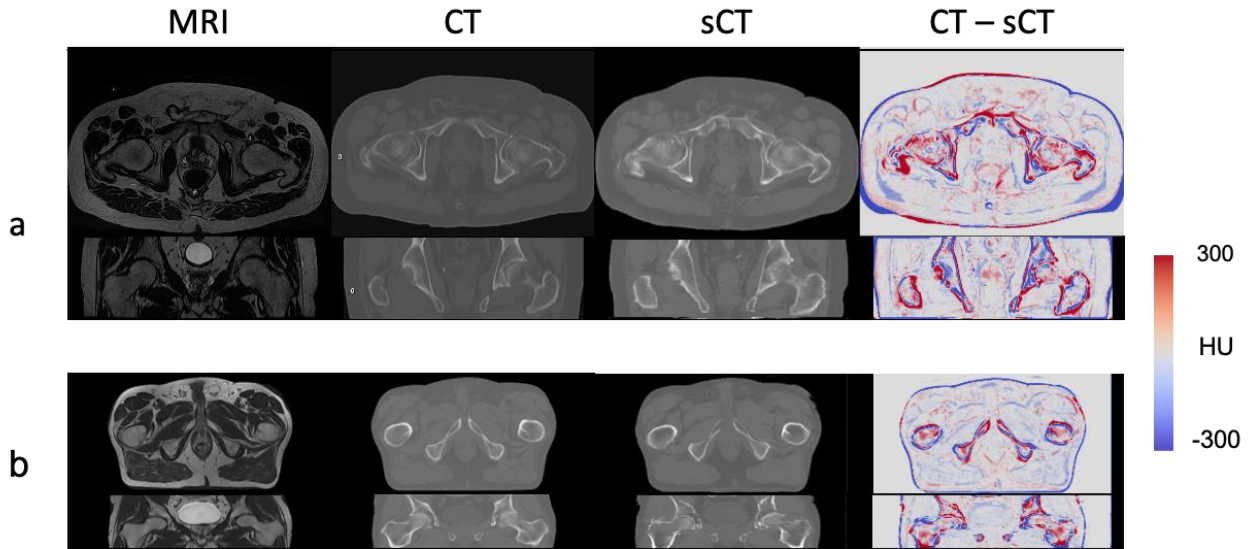

Figure 4: From left to right, MR, CT, and sCT images, and the difference (CT–sCT). The images on top represent the axial plane, while those on the bottom represent the frontal plane. (a) Site 5 and (b) site 4.

Table 2: Average MAE and ME in HU (±σ) between sCT and real CT for different locations in site 5.

	MAE	ME
Bladder	65.3 ± 29	-13.4 ± 20
Rectum	60.3 ± 25	-10.7 ± 29
Femoral Head L	178.9 ± 43	0.1 ± 49
Femoral Head R	197.1 ± 43	-1.0 ± 64

B. DVH analysis

The differences between several DVH points on sCT and CT were computed and presented as a boxplot in Figure 5. The values are rescaled to the prescribed dose.
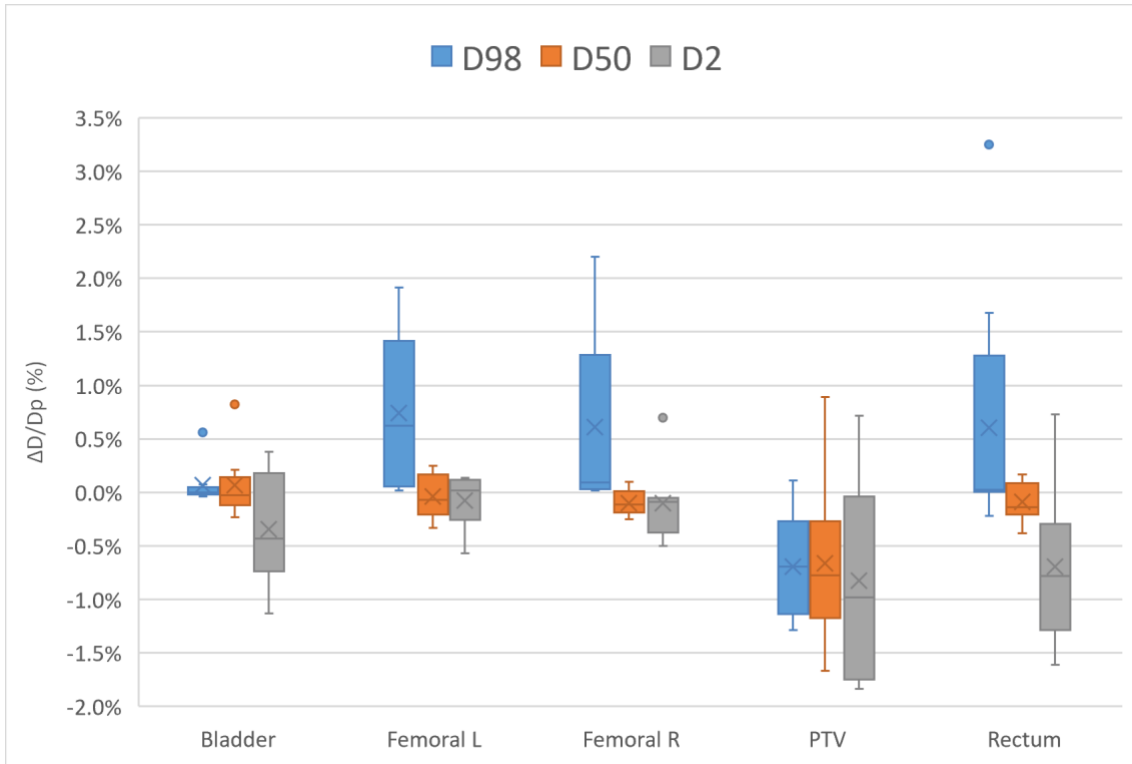
Figure 5: DVH parameter differences between dose on CT and sCT for the PTV and OARs for site 5.

C.   Dose difference

The doses calculated on CT (D_{CT}) and sCT (D_{sCT}) along with their relative differences are presented in Figure 6 for three patients. Mean absolute dose differences ($\frac{|D_{CT}-D_{sCT}|}{D_p}$) were computed with several dose thresholds. Table 3 reports the statistics in terms of the mean dose difference related to the prescribed dose calculated on thresholds of 10%, 50%, and 90% of the prescribed dose and the passing rates of local gamma.
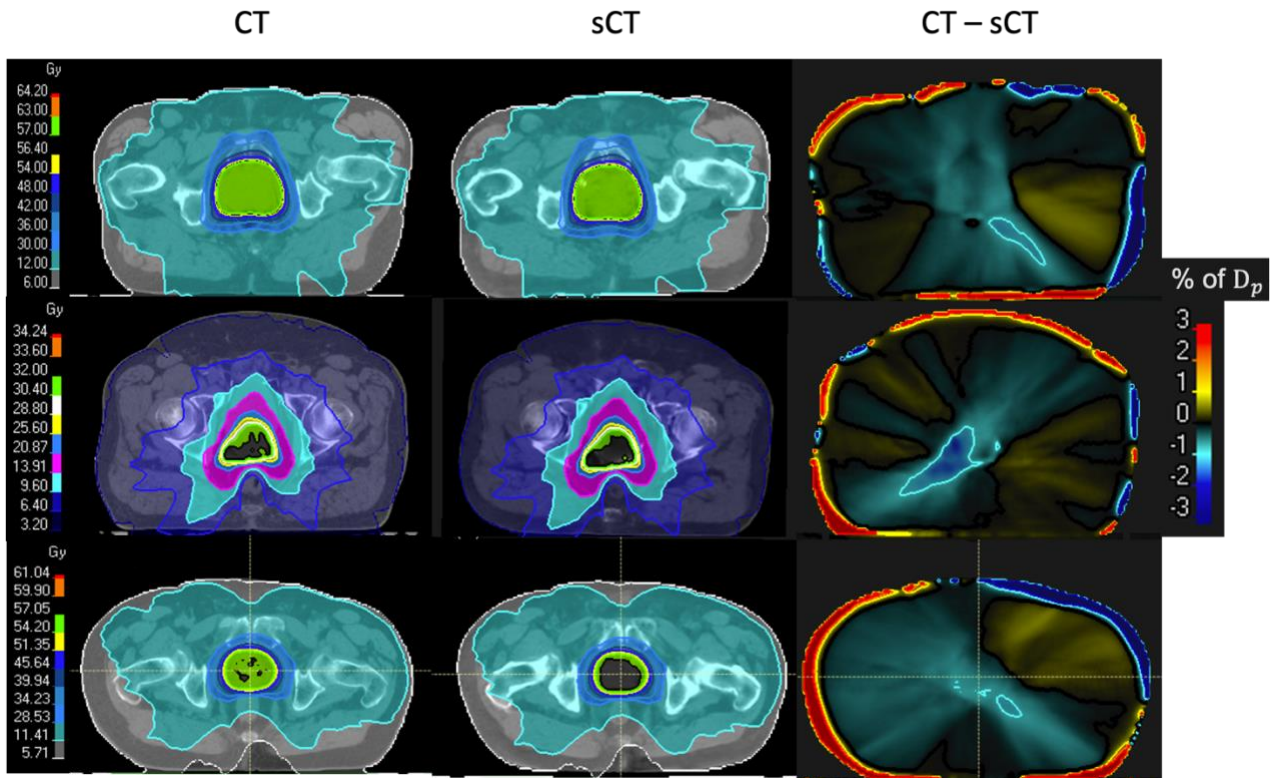
Figure 6: From left to right, doses calculated on CT and sCT and the relative dose difference (CT – sCT).

Table 3: Mean dose difference relative to the prescribed dose and gamma pass rate (±σ) between CT and sCT and the range of values

Volume	$Dose\ difference$ (%)	$\gamma_{3\%,3mm}$ (%)	$\gamma_{2\%,2mm}$ (%)
Body	0.13±0.07 [0.05; 0.28]	99.5±0.4 [98.8; 99.9]	95.5±2.2 [92.5; 98.3]
Dose>10%	0.20±0.07 [0.08; 0.30]	99.5±0.5 [98.7; 100]	94.8±2.3 [90.7; 97.9]
Dose>50%	0.48±0.24 [0.24; 0.94]	99.9±0.2 [99.3; 100]	97.6±1.3 [95.6; 99.5]
Dose>90%	0.70±0.38 [0.28; 1.22]	99.8±0.4 [99.8; 100]	99.8±0.4 [99.8; 100]

## IV.	Discussion

In this study, an MR-to-CT synthesis method was proposed to provide increased generalization performance in comparison with state-of-the-art approaches. We augmented the CycleGAN approach by adding a latent space to enable the reconstruction of several MR images, therefore making the training of this unsupervised model possible with multicentric data. The results suggest that our model can learn to synthetize a clinically acceptable sCT while using different scanners/parameters. With pre-training, this model can address the limitations of small datasets by finetuning it with a reduced sample.

The use of a vanilla CycleGAN in the same configuration produces unreliable results. A vanilla cycleGAN will not necessarily fail to convergence but will fail to learn an appropriate model because it simply does not have the capacity to learn one-to-many mappings while stochastic cycleGAN do

The average MAE of 65.8 HU and 59.8 HU for both sites were comparable to those obtained in a previous paired study (60 HU)[5] or unpaired study (50.8 HU).[12] However, several biases in training data may be observed on sCT. First, this method may show a tendency to generate higher HU values due to the different image value-to-density table. Our model occasionally tries to generate a fiducial marker in the prostate or shows a higher HU value in the bottom of the bladder due to the presence of contrast agent. In clinical practice, it will be essential to standardize the training data with a clinical protocol by performing additional modifications, such as directly using the electron density and not HU values as the output of the model.

This study is the first to show sCT images generated with a CycleGAN approach with unpaired multicentric data that underwent dosimetric evaluation in the pelvic area. Although interscan differences certainly influence the recalculation plan for the sCT images, all the DVH points recalculated on sCT were within $\pm 3\%$ with respect to the CT images. Maximum relative dose difference in the high-dose region is 1.2%, which is in consistent with previous studies.[5,19]

An advantage of increased generalization is that it will avoid the need to retrain a model for each new MR sequence or even a small change in sequence setting. Current CycleGAN-like models do not take into account MR variability. Collection of new data and retraining a model will substantially impede the use of deep learning in clinical practice.

## V.	Conclusion

We propose an augmentation of CycleGAN to generate synthetic CT images using multicentric data in an unpaired fashion without the need for a dedicated MR sequence. The possibility of using several sequences with this approach will allow the development of a single-body model instead of an anatomy-specific model. Our method was designed to improve the generalization of MR-to-CT synthesis and not directly improve the accuracy of HU. Future studies should aim to further develop AugCGAN (which shares much with CycleGAN) with deep learning methods such as 3D or dense blocks to improve accuracy as well as generalization.

References

1. Njeh C. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J Med Phys*. 2008;33(4):136. doi:10.4103/0971-6203.44472

2. Nyholm T, Jonsson J. Counterpoint: Opportunities and Challenges of a Magnetic Resonance Imaging–Only Radiotherapy Work Flow. *Semin Radiat Oncol*. 2014;24(3):175-180. doi:10.1016/j.semradonc.2014.02.005

3. Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. Least Squares Generative Adversarial Networks. November 2016. http://arxiv.org/abs/1611.04076.

4. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. June 2014. http://arxiv.org/abs/1406.2661.

5. Maspero M, Savenije MHF, Dinkla AM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. 2018;63(18):185001. doi:10.1088/1361-6560/aada6d

6. Brou Boni KND, Klein J, Vanquin L, et al. MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. *Phys Med Biol*. 2020;65(7):075002. doi:10.1088/1361-6560/ab7633

7. Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I. Deep MR to CT Synthesis Using Unpaired Data. In: ; 2017:14-23. doi:10.1007/978-3-319-68127-6_2

8. Hiasa Y, Otake Y, Takao M, et al. Cross-Modality Image Synthesis from Unpaired Data Using CycleGAN. In: ; 2018:31-41. doi:10.1007/978-3-030-00536-8_4

9. Thor M, Bentzen L, Elstrøm U V., Petersen JBB, Muren LP. Dose/volume-based evaluation of the accuracy of deformable image registration for the rectum and bladder. *Acta Oncol (Madr)*. 2013;52(7):1411-1416. doi:10.3109/0284186X.2013.814151

10. Nyholm T, Nyberg M, Karlsson MG, Karlsson M. Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. *Radiat Oncol*. 2009;4(1):54. doi:10.1186/1748-717X-4-54

11. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. March 2017. http://arxiv.org/abs/1703.10593.

12. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46(8):3565-3581. doi:10.1002/mp.13617

13. Lei Y, Wang T, Liu Y, et al. MRI-based synthetic CT generation using deep convolutional neural network. In: Angelini ED, Landman BA, eds. *Medical Imaging 2019: Image Processing*. SPIE; 2019:100. doi:10.1117/12.2512549

14. Almahairi A, Rajeswar S, Sordoni A, Bachman P, Courville A. Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data. February 2018. http://arxiv.org/abs/1802.10151.

15. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, eds. *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc.; 2014:2672-2680. http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

16. Zhu J-Y, Zhang R, Pathak D, et al. Toward Multimodal Image-to-Image Translation. November 2017. http://arxiv.org/abs/1711.11586.

17. Nyholm T, Svensson S, Andersson S, et al. MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project. *Med Phys*. 2018;45(3):1295-1300. doi:10.1002/mp.12748

18. Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging*. 2010;29(1):196-205. doi:10.1109/TMI.2009.2035616

19. Kemppainen R, Suilamo S, Tuokkola T, Lindholm P, Deppe MH, Keyriläinen J. Magnetic resonance-only simulation and dose calculation in external beam radiation therapy: a feasibility study for pelvic cancers. *Acta Oncol (Madr)*. 2017;56(6):792-798. doi:10.1080/0284186X.2017.1293290