IOP Publishing Phys. Med. Biol. 65 (2020) 075002 https://doi.org/10.1088/1361-6560/ab7633

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
30 September 2019

REVISED
10 February 2020

ACCEPTED FOR PUBLICATION
13 February 2020

PUBLISHED
2 April 2020

MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network

Kévin N D Brou Boni[1,2,5], John Klein[2], Ludovic Vanquin[1], Antoine Wagner[1], Thomas Lacornerie[1], David Pasquier[2,3] and Nick Reynaert[1,4]

[1] Department of Medical Physics, Centre Oscar Lambret, Lille, France
[2] University Lille, CNRS, Centrale Lille, UMR 9189 – CRIStAL, Lille, France
[3] Department of Radiotherapy, Centre Oscar Lambret , Lille, France
[4] Department of Medical Physics, Institut Jules Bordet, Brussels, Belgium
[5] Author to whom any correspondence should be addressed.

E-mail: k-brouboni@o-lambret.fr

Keywords: CT synthesis, Generative Adversarial Networks, radiotherapy, MRI, dose evaluation

Abstract

The establishment of an MRI-only workflow in radiotherapy depends on the ability to generate an accurate synthetic CT (sCT) for dose calculation. Previously proposed methods have used a Generative Adversarial Network (GAN) for fast sCT generation in order to simplify the clinical workflow and reduces uncertainties. In the current paper we use a conditional Generative Adversarial Network (cGAN) framework called pix2pixHD to create a robust model prone to multicenter data.

This study included T2-weighted MR and CT images of 19 patients in treatment position from 3 different sites. The cGAN was trained on 2D transverse slices of 11 patients from 2 different sites. Once trained, the network was used to generate sCT images of 8 patients coming from a third site. The Mean Absolute Errors (MAE) for each patient were evaluated between real and synthetic CTs. A radiotherapy plan was optimized on the sCT series and re-calculated on CTs to assess the dose distribution in terms of voxel-wise dose difference and Dose Volume Histograms (DVH) analysis.

It takes on average of 7.5 s to generate a complete sCT (88 slices) for a patient on our GPU. The average MAE in HU between the sCT and actual patient CT (within the body contour) is 48.5 ± 6 HU with our method. The maximum dose difference to the target is 1.3%.

This study demonstrates that an sCT can be generated in a multicentric context, with fewer pre-processing steps while being fast and accurate.

1. Introduction

Interest has been rapidly growing in complementing and even replacing Computed Tomography (CT) with Magnetic Resonance Imaging (MRI) in the field of radiation therapy thanks to a superior soft-tissue contrast. In addition, an MRI-only workflow avoids extra radiation to the patient and reduces errors related to inter-modality registration. Currently, the main challenge is that MRI pixel values are not directly related to electron density, which is needed in radiation therapy treatment planning systems (TPS) for dose calculation.

This problem is solved by converting an MRI to a so-called synthetic CT (sCT) or pseudo CT. Many different sCT generation methods have been proposed in the literature. These techniques recently underwent significant changes with the emergence of deep learning. Accuracy and velocity have dramatically increased (Han 2017, Dinkla *et al* 2018). Generative Adversarial Networks (GAN) have boosted this trend with their ability to learn generating any data distribution in a paired (Nie *et al* 2017, Maspero *et al* 2018) or unpaired fashion (Wolterink *et al* 2017). So far, to the best of our knowledge no deep learning-based method described in the scientific literature, has included data from different medical imaging centers using different CT and MRI.

Table 1. Acquisition settings for the three sites. TSE stands for turbo spin echo and FRFSE for fast recovery fast spin-echo, COL for columns.

	Site 1	Site 2	Site 3
Number of patients	8	7	4
CT			
Manufacturer	Siemens	Toshiba	Siemens
Model	Somatom Definition AS+	Aquilion	Emotion 6
Slice thickness (mm)	3	2	2.5
Kernel	B30f	FC17	B41s
T2-w			
Manufacturer	GE	Siemens	GE
Model	Discovery 750 w 3 T	− 1.5 T	Signa PET/MR 3T
Sequence type	FRFSE	TSE	FRFSE
Slice thickness (mm)	2.5	2.5	2.5
Bandwidth (Hz/pixel)	390	200	390
Encoding direction	COL	ROW	COL
TR (ms)	6000–6600	12 000–16 000	6000–10 000
TE (ms)	97	91–102	65

In this paper, we discuss a new multi-scale approach by using an existing conditional GAN (cGAN) (Wang *et al* 2018) with paired data coming from different sites. A proof of concept study is conducted by creating a test set with images coming from a site not used in the train set. This will allow to cover a wide range of possibilities (artifact, anatomical malformation, MRI intensity variability) in the training and thus improve the generalizability of MRI to CT conversion. Finally, a dosimetric evaluation is performed to assess the dose accuracy on the sCT.

2. Materials and methods

2.1. Patients data collection

This study included pelvic MR and CT images of 19 male patients with prostate or rectal cancer. Images were taken from the public dataset named the Gold Atlas project (Nyholm *et al* 2018) aimed to provide a source of training and validation for segmentation as well as sCT generation methods. Patients with locally advanced tumors were not included in this database. Radiotherapy planning for prostate cancer was carried out for all patients. Indeed, these were early stage rectal cancers that did not deform the pelvic anatomy and allowed realistic planning of prostate cancer radiotherapy.

Nineteen patients coming from three sites were selected and scanned in radiotherapy treatment position, T2-weigthed MR and CT images were acquired following clinical protocol. Table 1 provides the acquisition settings.

9 organs were segmented by five experts based on MRI, and consensus contours among the experts are also available. The open source library ITK was used to perform a deformable registration on the CT to fit the anatomy of the MRI, enabling the use of the delineations on the registered CT.

2.2. Image pre-processing

A mask excluding surrounding air was obtained on the CT and MRI using the external ROI option (threshold level based) on Raystation (v7.0). Voxels outside the body were automatically assigned to -1024 *HU* for CT and 0 for MR. Inter-scan differences (air pockets and structures) have not been taken into account in this study. HU were normalized, MR intensities as well patient-wise. Finally, all dicom files were converted to 16-bit grayscale images compatible with current deep learning frameworks. The first and last slices were not taken into account for the training due to aliasing in MRI. This allowed the use of this dataset consisting of aligned MR-CT as part of an image-to-image translation problem.

2.3. Network

2.3.1. cGAN baseline

GANs are characterized by two networks: the generator $G(z)$ with z a noise vector and a discriminator $D(y)$. For the current application, y represents a CT image. All CT images are distributed according to an unknown probability distribution p_y. G attempts to transform the vector z into images so that a sample of size n, $\left\{G\left(z^{(1)}\right),\ldots,G\left(z^{(n)}\right)\right\}$ follows the probability distribution p_y. D attempts to separate the images actually distributed according to p_y from those produced by his opponent G. Actually, $D(x)$ is understood as the probability that image x is a true CT.

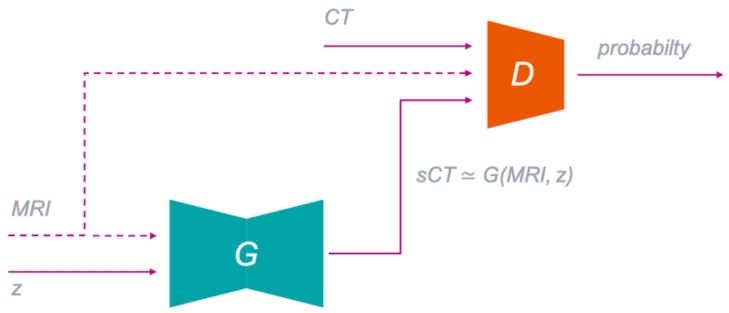

Figure 1. The discriminator D learns to classify a real CT from a synthetic CT while the generator G learns to fool D following a min-max game.

To convert an MRI into a CT, the networks have to be conditioned with an MR image x. A simple way to achieve this objective is to feed these two networks with x (as additional input). The generator and the discriminator therefore become $G(x, z)$ and $D(x, y)$ respectively (figure 1). As the training progresses, G must be able to generate samples that are more and more faithful to the distribution p_y, making it more and more difficult for D to detect fakes CT images. G and D are trained alternately and share the same objective function. The discriminator tries to maximize it while the generator tries to minimize it. The objective function L_{cGAN} is the following expected cross-entropy:

$$L_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z)))]. \quad (1)$$

This network is optimized following the standard approach of Goodfellow *et al* (2014) by alternating the gradient ascent/descent steps between the generator and the discriminator. z is induced by dropout (Hinton *et al* 2012) in both the training and test phases.

2.3.2. The pix2pixHD network

The cGAN introduced by Wang *et al* (2018) used in this work improves photorealism and resolution on four important aspects.

- Coarse-to-fine generator: the generator which has an encoder-decoder architecture is separated in two sub-networks $G = \{G_{global}, G_{local}\}$. The first one is the center of an encoder-decoder architecture and is thus itself a (smaller) encoder-decoder. It is pre-trained on low resolution images. The local generator (the entire encoder-decoder structure) is then fine-tuned on high resolution images.
- Multi-scale discriminators: G has to fight against several discriminators $D = \{D_1, D_2, D_3\}$. Each of these discriminators works at a different image scale.
- A feature matching loss L_{FM} (Wang *et al* 2018) is added in order to stabilize the training of the generator by matching intermediate representations (feature maps) in the different layers of the discriminators from real and synthesized images. The idea behind this additional loss term is that the generator will be forced to produce images with more natural statistics at different scales. If we denote $D_k^{(i)}$ the i-th layers of D_k, L_{FM}[6] is then calculated as:

$$L_{FM}(G, D_k) = \sum_{i=1}^{N} MAE(D_k^{(i)}(x,y), D_k^{(i)}(x, G(x,z))), \quad (2)$$

- where N is the total number of layers.
- Instead of the usual cross-entropy cGAN loss, the authors recommend the Least Square GAN (LSGAN) loss (Mao *et al* 2017), a quadratic version. This loss address the problem of vanishing gradient when updating the generator (Arjovsky *et al* 2017) for sample lying on the 'True' decision boundary but still far from the real data distribution. LSGAN loss penalises these samples enabling faster convergence and more realistic image generation.

In our sCT generation implementation, pre-training the smaller resolution generator (G_{local}) proved to be counterproductive and led to poorer results. The generator G used here follows the architecture proposed

[6]Note that for $y, \hat{y} \in \mathbb{R}^n$, $MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} |y - \hat{y}|$.

by Johnson *et al* (2016) and learns to synthesize a CT. We chose to work with $K = 2$ discriminators working at different scales, both of them being trained to differentiate real and synthesized CT images. The first discriminator D_1 operates at standard scale while the second D_2 operates with downsampled images by a factor 2. These discriminators have identical architectures with different receptive fields. They follow the PatchGAN architecture (Isola *et al* 2016) forcing the generator to produce consistent images while encouraging finer details. Training this model tends to produce realistic CT images but regarding HU, performances do not seem as good as they visually do. To overcome this difficulty without adding a post-processing step, we propose to add an additional L_1 reconstruction loss (MAE) term between the generated sCT and the true CT. The full objective function is then calculated as:

$$\begin{aligned} &\min_D \sum_{k=1,2} \mathbb{E}_{x,y}[D_k(x,y)-1]^2 + \mathbb{E}_{x,z}[D_k(x,G(x,z))]^2, \\ &\min_G \mathbb{E}_{x,y,z}\left[\lambda.MAE(y,G(x,z))\right] + \sum_{k=1,2}[D_k(x,G(x,z))-1]^2 + \frac{\mu}{K}.L_{FM}(G,D_k), \end{aligned} \tag{3}$$

with $\lambda = 10$ and $\mu = 5$ are two hand-tuned hyperparameters.

2.3.3. Training of the network

The 19 patients were separated into a training set containing 7 patients from site 2 and 4 patients from the third one. The 8 patients coming from the site 1 were used as testing set. The network was trained using Adam optimizer with an initial learning rate of 0.0002 for 100 epochs, then for another 100 epochs with a linearly decay learning rate to zero.

Training took on average 17 h on an Nvidia Quadro P6000 with a batchsize of 1. Data augmentation was performed by horizontal flip increasing the size of the training set to 2008 image pairs.

2.4. sCT evaluation

Once the network was trained, each sCT was generated using only the generator on the GPU. The images files created are then converted to a DICOM format, allowing their use on a treatment planning system.

2.4.1. Image comparison

Synthetic CT and registered CT were compared on a voxel-wise basis using the *MAE* and the Mean Error[7] (*ME*). Considering the voxels within the body contours, *MAE* and in *HU* were calculated for each patient.

A 16-bit implementation of a vanilla pix2pix (Isola *et al* 2016, Maspero *et al* 2018) was trained in the same multicentric configuration. *MAE* and *ME* of the sCT generated by pix2pix is also calculated for each patient.

2.4.2. Dose comparison

Tomotherapy treatment plans were optimized on each sCT in Raystation (v7.0) using the Collapsed Cone (v3.5) algorithm on a grid of $1 \times 1 \times 1\,\text{mm}^3$. The prescription was $39 \times 2\,Gy$ to the planning target volume (PTV) (prostate with 5 mm uniform margin). The resulting plans were then recalculated on the CT for dose comparison.

A dose volume histogram (DVH) analysis was performed after copying the structures (PTV, femoral heads, bladder wall and rectum wall) to CT. 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.

3. Results

3.1. Image comparison

CT synthesis took on average 7.5 *s* on GPU. Figure 2 shows an example of one of our test patients. As expected, differences are most pronounced in the bone structures. Staircase patterns are visible on the bone in the frontal view. This may be due to the 2D generation technique used that does not take into account adjacent slices.

The proposed method produced an average *MAE* of $48.5 \pm 6\,HU$ and an average ME of $-18.3 \pm 9\,HU$ for our 8 patients. Vanilla pix2pix produced an average *MAE* of $62.0 \pm 12\,HU$ and an average ME of $-11.4 \pm 19\,HU$. Table 2 provides the average *MAE* and *ME* for target volumes and organs at risk (OAR) for pix2pixHD and pix2pix.

[7]Note that for $y, \hat{y} \in \mathbb{R}^n$, $ME(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n} y - \hat{y}$.

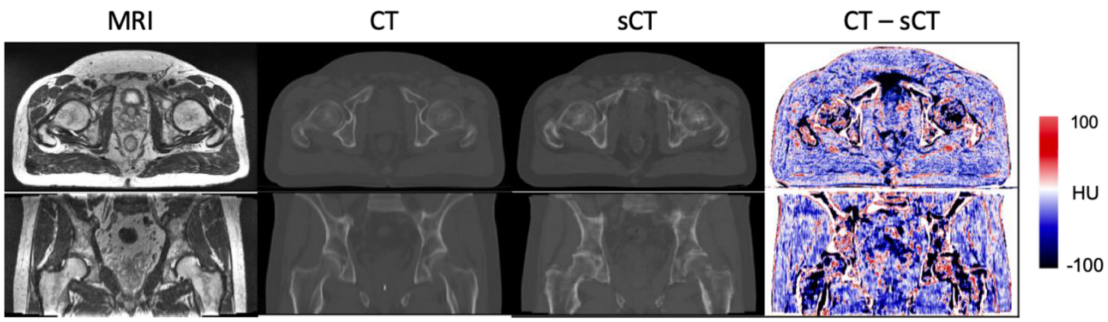

Figure 2. From left to right, MR image, CT, sCT and difference (CT—sCT). The images on top represent the axial plane, on the bottom, the frontal plane.

Table 2. Average MAE and ME in HU (± σ) between sCT and real CT for different locations when training with pix2pixHD based model and pix2pix.

	MAE		ME	
	pix2pixHD	pix2pix	pix2pixHD	pix2pix
Bladder wall	49.4 ± 12	61.6 ± 10	− 23.9 ± 23	− 0.6 ± 31
Rectum wall	101.8 ± 78	109.8 ± 78	− 77.6 ± 90	− 85.2 ± 80
Anal canal	30.3 ± 14	36.0 ± 13	− 24.6 ± 18	− 26.4 ± 16
Penile bulb	28.1 ± 9	56.5 ± 16	− 19.2 ± 15	38.6 ± 25
Femoral Heads	90.5 ± 9	112.7 ± 23	− 25.9 ± 47	45.7 ± 44
Seminal Vesicles	44.7 ± 15	54.8 ± 11	− 14.0 ± 26	13.1 ± 19
Prostate	47.1 ± 6	62.3 ± 9	− 11.6 ± 12	17.5 ± 29

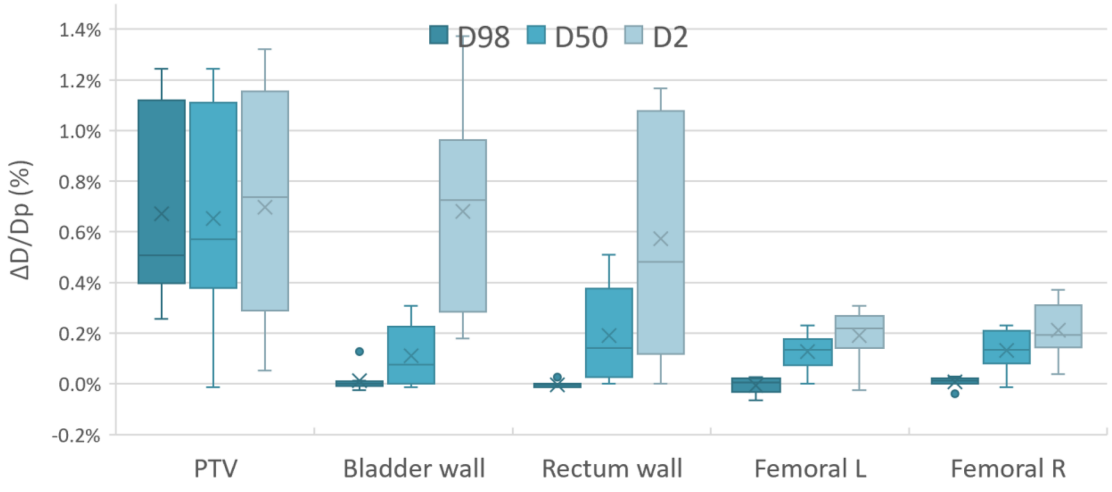

Figure 3. DVH parameter differences between dose on CT and sCT for the PTV and OARs.

3.2. DVH analysis

The absolute difference between the DVH points on sCT and CT were always below 1.4%. Figure 3 shows a boxplot of the DVH point difference for the PTV and the OARs.

3.3. Dose difference

Mean absolute dose differences were computed with several dose thresholds. Differences only appear in high dose regions and the body contour as shown on figure 4. The sCTs tend to have higher Hounsfield units (HU) resulting a global decreased dose inside the body. Inner negative dose differences are often due to lower HU on the sCT in bone area or air pocket not generated in sCT.

Table 3 reports the statistics in terms of mean dose difference related to the prescribed dose calculated on a threshold of 10%, 50% and 90% of the prescribed dose.

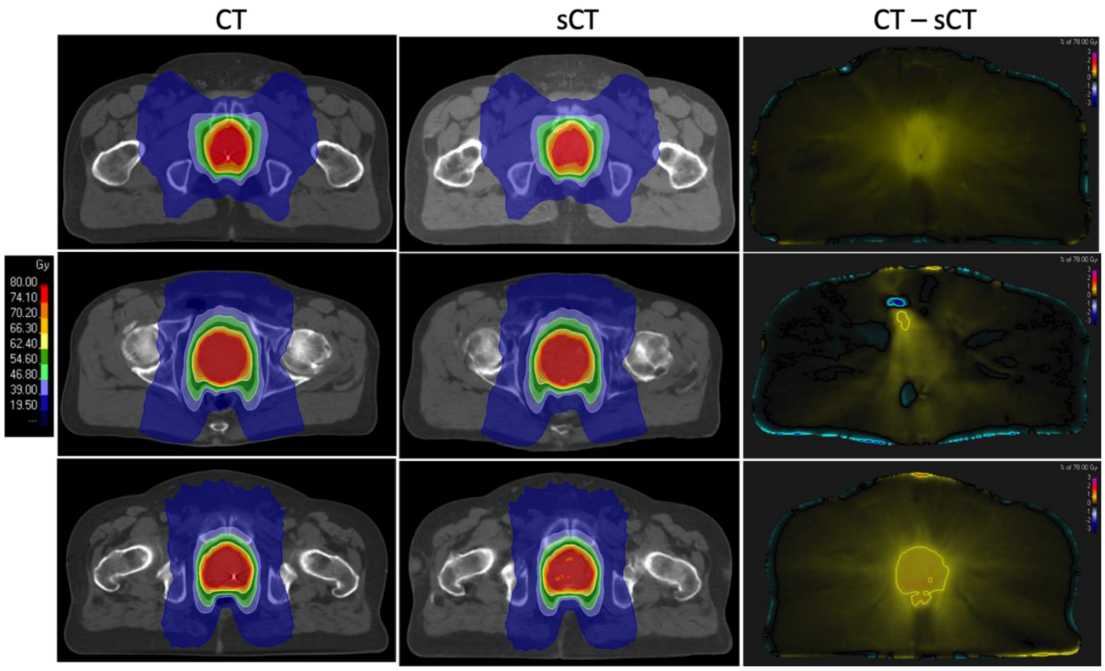

Figure 4. From left to right, dose calculated on CT, sCT and dose difference (CT − sCT).

Table 3. Mean dose difference (± σ) between CT and sCT and range of values.

Volume	$\frac{\lvert D_{CT} - D_{sCT} \rvert}{D_{Presc}}$ (%)
Body	0.01 ± 0.01 [0.01; 0.03]
Dose > 10%	0.12 ± 0.07 [0.00; 0.22]
Dose > 50%	0.49 ± 0.29 [0.03; 0.92]
Dose > 90%	0.68 ± 0.35 [0.19; 1.23]

4. Discussion and conclusion

Maspero *et al* (2018) showed that conditional GANs can synthetize CT from MRI. In the current work, a good performance is achieved with a limited dataset with a coarse-to-fine approach, by incorporating a feature matching loss and the use of the Least Square GAN loss.

This paper shows for the first time a robust neural network trained and tested with data coming from different medical imaging centers. Without ever having seen an image from the test site, our model learns to synthetize a clinically acceptable sCT, which may be generalized to different MRI manufacturers. This process has the capability to tackle the images variability problem in clinical practices, since changes can happen in image acquisition parameters or with machine replacement for instance. This study was done using standard morphological sequence (T2-w Spin Echo) without the need of any dedicated sequences.

Results look promising although a presence of artifact patterns can be noted. This may be partially due to the low amount of data and to the transposed convolutions used in the decoder part in the generator. The use of a third discriminator seems to get rid of this problem without improving quantitative results. The average MAE (48.5 *HU*) and the dosimetric evaluation (dose differences within 1.4%) obtained in this study compare similarly with other state-of-the-art single center results (Nie *et al* 2017, Maspero *et al* 2018) in the literature for the pelvic area. These small differences would be suitable for clinical implementation. It is a well-known fact that deep learning models can benefit from more training data, which leads to the expectation that better results will be obtained when feeding our algorithm with more datasets. A direct comparison with other studies is sensitive and complex since distinct datasets are used. The size of the dataset, the sequence(s) used, the diversity (artifact, specific case, etc) and the misalignment between the sCT and the CT are some of the numerous factors that make a direct comparison difficult.

Improvements need to be introduced in order to mitigate the discontinuity across the slices and therefore improve image quality. The use of 3D convolution leads to questionable results in the community, since they are greedy and not so effective. As a future perspective, we plan to improve sCT generation via Recurrent Neural Contextual Learning. Such models are expensive, and their benefits will have to be balanced with their increased complexity.

A multi-center study based on the conversion of MR intensities to HU includes uncertainties related to the different image value to density table (IVDT). Direct conversion to electron density would avoid these errors but the benefit remains to be studied.

Acknowledgment

This project was supported by the fund grant contract no. 2S04-022 under the Interreg 2 Seas 2014–2020 programme co-financed by the European Development Fund.

References

Arjovsky M, Chintala S and Bottou L 2017 Wasserstein generative adversarial networks *Int. conf. on Machine Learning* pp 214–23

Dinkla A M, Wolterink J M, Maspero M, Savenije M H F, Verhoeff J J C, Seravalli E, Išgum I, Seevinck P R and van den Berg C A T 2018 MR-only brain radiation therapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network *Int. J. Radiat. Oncol.* **102** 801–12

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 *Generative Adversarial Networks* (arXiv:1406.2661)

Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method *Med. Phys.* **44** 1408–19

Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R 2012 Improving neural networks by preventing co-adaptation of feature detectors (arXiv:1207.0580)

Isola P, Zhu J-Y, Zhou T and Efros A A 2016 *Image-to-Image Translation with Conditional Adversarial Networks* (arXiv:1611.07004)

Johnson J, Alahi A and Fei-Fei L 2016 *Perceptual Losses for Real-Time Style Transfer and Super-Resolution (Lecture Notes in Computer Science*, vol 9906) (Berlin: Springer) pp 694–711

Mao X, Li Q, Xie H, Lau R Y K, Wang Z and Smolley S P 2017. Least squares generative adversarial networks *2017 IEEE Int. Conf. on Computer Vision (ICCV)* (Piscataway, NJ: IEEE) pp 2813–21

Maspero M, Savenije M H F, Dinkla A M, Seevinck P R, Intven M P W, Jurgenliemk-Schulz I M, Kerkmeijer L G W and van den Berg C A T 2018 Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy *Phys. Med. Biol.* **63** 185001

Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q and Shen D 2017 *Medical Image Synthesis with Context-Aware Generative Adversarial Networks (Lecture Notes in Computer Science*, vol 10435) (Berlin: Springer) pp 417–25

Nyholm T *et al* 2018 MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project *Med. Phys.* **45** 1295–300

Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J and Catanzaro B 2018 High-resolution image synthesis and semantic manipulation with conditional GANs *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 8798–807

Wolterink J M, Dinkla A M, Savenije M H F, Seevinck P R, van den Berg C A T and Išgum I 2017 *Deep MR to CT Synthesis Using Unpaired Data* (*Lecture Notes in Computer Science*, vol 10435) (Berlin: Springer) pp 14–23