

A high-performance method of deep learning for prostate MR-only radiotherapy planning using an optimized Pix2Pix architecture

S. Tahri ^{a,*}, A. Barateau ^a, C. Cadin ^a, H. Chourak ^{a,b}, S. Ribault ^a, F. Nozahic ^a, O. Acosta ^a, J.A. Dowling ^b, P.B. Greer ^{c,d}, A. Largent ^e, C. Lafond ^a, R. De Crevoisier ^a, J.C. Nunes ^a

^a Univ. Rennes 1, CLCC Eugène Marquis, INSERM, LTSI – UMR 1099, F-35000 Rennes, France

^b CSIRO Australian e-Health Research Centre, Herston, Queensland, Australia

^c School of Mathematical and Physical Sciences, University of Newcastle, Newcastle, Australia

^d Department of Radiation Oncology, Calvary Mater, Newcastle, Australia

^e Developing Brain Institute, Department of Diagnostic Imaging and Radiology, Children's National Hospital, Washington, DC, USA



ARTICLE INFO

Keywords:

Synthetic CTs
Deep learning
Pix2Pix
MRI
Radiation therapy
Dose evaluation

ABSTRACT

Purpose: The first aim was to generate and compare synthetic-CT (sCT) images using a conditional generative adversarial network (cGAN) method (Pix2Pix) for MRI-only prostate radiotherapy planning by testing several generators, loss functions, and hyper-parameters. The second aim was to compare the optimized Pix2Pix model with five other architectures (bulk-density, atlas-based, patch-based, U-Net, and GAN).

Methods: For 39 patients treated by VMAT for prostate cancer, T2-weighted MRI images were acquired in addition to CT images for treatment planning. sCT images were generated using the Pix2Pix model. The generator, loss function, and hyper-parameters were tuned to improve sCT image generation (in terms of imaging endpoints). The final evaluation was performed by 3-fold cross-validation. This method was compared to five other methods using the following imaging endpoints: the mean absolute error (MAE) and mean error (ME) between sCT and reference CT images (rCT) of the whole pelvis, bones, prostate, bladder, and rectum. For dose planning analysis, the dose-volume histogram metric differences and 3D gamma analysis (local, 1 %/1 mm) were calculated using the sCT and reference CT images.

Results: Compared with the other architectures, Pix2Pix with Perceptual loss function and generator ResNet 9 blocks showed the lowest MAE (29.5, 107.7, 16.0, 13.4, and 49.1 HU for the whole pelvis, bones, prostate, bladder, and rectum, respectively) and the highest gamma passing rates (99.4 %, using the 1 %/1mm and 10 % dose threshold criterion). Concerning the DVH points, the mean errors were -0.2% for the planning target volume $V_{95\%}$, 0.1 % for the rectum $V_{70\text{Gy}}$, and -0.1 % for the bladder $V_{50\text{Gy}}$.

Conclusion: The sCT images generated from MRI data with the Pix2Pix architecture had the lowest image errors and similar dose uncertainties (in term of gamma pass-rate and dose-volume histogram metric differences) than other deep learning methods.

1. Introduction

Computed tomography (CT) is commonly used as the reference imaging method for radiation therapy treatment planning [1]. Indeed, CT values can be directly converted to electron densities for radiation dose calculation [1]. However, CT images have several drawbacks, particularly the limited soft-tissue contrast and ionizing radiation to the patient [2,3].

In the last few years, it has been shown that MRI could be valuable in

the radiation therapy treatment workflow because it is ionizing radiation-free and offers high soft-tissue contrast [4,5]. However, replacing CT by MRI for radiotherapy treatment planning is not easy. One of the main issues is that MRI intensity values are not directly related to electron densities. Several approaches have been proposed to estimate electron density for MRI-only radiotherapy treatment planning, such as bulk density-based [6–10], atlas-based [9–16], and machine learning-based (ML) methods (patch-based [9,17,18], and deep learning-based methods (DLMs) [19,20]).

* Corresponding author.

E-mail address: safaa.tahri@univ-rennes1.fr (S. Tahri).

DLMs are models that comprise multiple processing layers to learn multi-scale data representations with multiple levels of abstraction [19–21]. These methods are increasingly used in radiotherapy workflows, for instance for tissue segmentation, image processing and reconstruction, image registration, and treatment planning [19,20,22–26]. DLMs [27] for synthetic CT (sCT) image generation from MRI images are trained to model the relationships between the CT maps (in Hounsfield Units, HU) and MRI intensities. Once the optimal deep learning parameters are estimated, the model can be applied to a test MRI to generate the corresponding sCT. The main advantage of DLMs is the very short calculation time for sCT generation, an essential feature for implementation in clinical practice.

The DLMs for sCT generation from MRI can be divided into two classes: generator-only and generative adversarial network (GAN) [19].

In generator-only networks, the generator model can be considered as representing a complex end-to-end mapping function, that transforms an input MRI image into its corresponding CT image [19]. During the training phase, the generator tries to minimize an objective function called a loss function, which is an intensity-based similarity measurement between the generated image (sCT) and the reference CT image [19]. Several loss functions have been used in these generative models, for instance the mean square error (MSE) [18,19,28–30], mean absolute error (MAE) [19,20,31–33], and perceptual loss[18,19]. In sCT generation from MRI, the generator architectures are generally based on convolution encoder-decoder networks (CED) [19]. The most well-known and popular CED variant for biomedical image applications is the U-shaped CNN (U-Net) architecture proposed by Ronneberger et al. [34]. The U-Net [34] has a CED structure with direct skip connections between the encoder and decoder.

The GAN-based architecture was proposed by Goodfellow et al. [35] in 2014. The principle is to simultaneously train two separate neural networks that try to beat each other: the generator (one of the generator-only models described previously) and the discriminator [19,35,36]. The generator tries to produce a realistic image to fool the discriminator. Conversely, the discriminator tries to distinguish between the real data and those generated by the generator. The architecture of the generator is often the U-Net [34], or less frequently the ResNet [37]. The overall loss functions used in these GANs are L2-norm [18,38], perceptual loss [18], and the multiscale perceptual loss [18].

To allow explicit control of the actual data generation, Goodfellow et al. proposed the conditional GAN (cGAN) [35], an extension of the GAN architecture. It incorporates additional information, such as class labels and input image modality (MRI images in the radiotherapy workflow context), in the synthesis process to improve image generation. The cGAN architecture is interesting because it includes a direct correlation between the labels and the images, so it could permit to increase training performance as it was demonstrated by Isola et al. [39]. Indeed, thanks to this correlation, a major advantage should be noted compared to the GAN methods. The cGAN could require less training data than the GAN, to obtain similar performance. This improvement would therefore make it possible to use fewer data and focus more on the quality of the data which plays a major role in the sCT generation.

Usually, the generator architecture is a U-Net [34], but other generator architectures may be proposed, such as ResNet, SE-ResNet, Dense-Net, and Embedded Net [19]. Moreover, different loss functions can be used, such as binary cross-entropy, MAE, L2, L1, or mutual information [19]. In addition, Pix2Pix, a well-known cGAN implementation proposed by Isola et al. in 2016 [39], allows high-resolution image-to-image translation.

In recent years, the interest in generating sCT images from MRI images has been increasing in radiotherapy [20]. However, there is only limited data on the effects of different generator network architectures in this field [19,20]. Moreover, no study investigated at the same time, the effects of different generator network architectures, different loss functions, and different hyper-parameters on pelvis sCT image

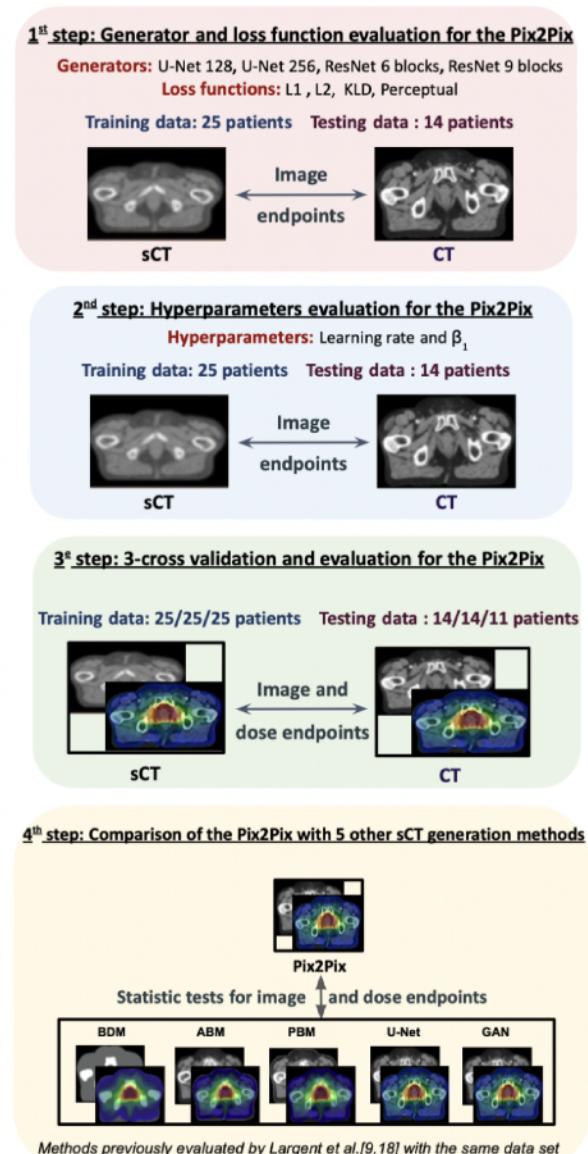


Fig. 1. Workflow of the study. The first step was the evaluation of the generator and the loss function with a training cohort of 25 patients and a test cohort of 14 patients. Four generators were evaluated: U-net 128, U-Net 256, ResNet 6 blocks and ResNet 9 blocks. Four loss functions were assessed: L1, L2, KLD and Perceptual. Imaging endpoints were used to compare the results. The second step was the evaluation of hyper parameters or the generator (learning rate and β_1) with a training cohort of 25 patients and a test cohort of 14 patients. Imaging endpoints were used to compare the results. The third step was an evaluation in terms of image and dose endpoints with the generator, loss function and hyper parameters obtained in the previous steps. This fourth step also was a comparison of results from the Pix2Pix method to results from five other methods (bulk-density, atlas-based, patch-based, U-Net, and GAN) previously published.

generation. Therefore, the first aim of this work was to identify the generator, the loss function, and the hyper-parameters that allow obtaining a high quality sCT images from MRI data with Pix2Pix, to perform accurate dose calculations. The second aim was to compare the sCT images generated with Pix2Pix model with the sCT images generated by five previously described methods (bulk-density, atlas-based, patch-based, U-Net, and GAN) [9,18] in terms of image and dose endpoints, using the same patient cohort.

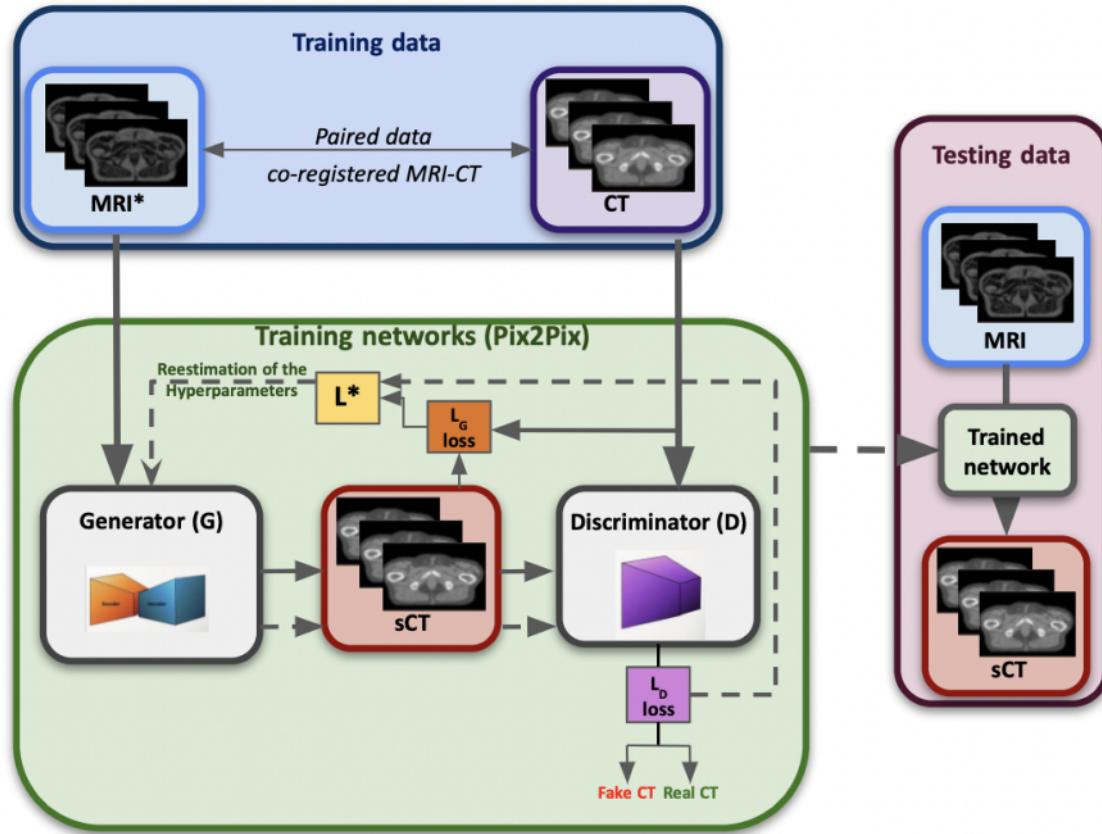


Fig. 2. Pix2Pix architecture to generate synthetic-CT (sCT) from MRI. For each patient from the training data base, the CT and MR training images were first nonrigidly coregistered. The training of the architecture consists of two competing multilayer networks: the generator and the discriminator. In the testing step, for a new given data, the MRI goes through the trained network to obtain the corresponding sCT. These sCT were compared to the CT with imaging and dosimetric endpoints. MRI* pre-processed with the following steps built using the Insight Segmentation and Registration Toolkit (ITK 4.4) ref i) N4 Bias field correction [41] (B-spline fitting: [160, 3, 0, 0.5]; convergence: [100x100x100, 0.001]; shrink factor: 3); ii) histogram equalization (levels: 1024; matchpoints: 7, threshold at mean intensity, iii) Non-linear filtering by anisotropic anisotropic diffusion (10 iterations; time step:0.03; conductance: 1.0) [19].

2. Materials and methods

Fig. 1 shows the workflow of the study. The first step was the evaluation of the generators and the loss functions. The second step was the evaluation of hyperparameters (learning rate and momentum β_1). Imaging endpoints were used to compare results from both steps. The third step was an evaluation in terms of image and dose endpoints with the generator, the loss function, and the hyperparameters obtained in the previous steps with the Pix2Pix. Finally, the fourth step was a comparison in terms of image and dose endpoints with the results from five other sCT generation methods (bulk-density, atlas-based, patch-based, U-Net, and GAN) previously published [9,18].

2.1. Patient dataset

Thirty-nine patients received volumetric modulated arc therapy for localized prostate cancer. The study protocol was approved by the local health ethics committee, and informed consent was obtained from all patients [12].

Patients had both an initial CT (reference CT) and 3 T MRI in the treatment position. CT images were acquired on a GE LightSpeedRT large-bore scanner (2.5 mm slices) or a Toshiba Aquilion (2.0 mm slices). Patients were scanned with a full bladder and empty rectum. All patients followed a bowel preparation protocol (high fiber diet and daily psyllium fiber metamucil intake). Patients were positioned supine on a rigid couch-top with knee and ankle immobilization stocks.

The MRI images were acquired with a Siemens Skyra 3 T scanner. All

patients were positioned by 2 radiation therapists. The MR acquisition was a 3-dimensional (3D), T2-weighted 1.6 mm isotropic SPACE sequence with field of view to cover the entire pelvis (including all the bladder), with the following parameters: field-of-view = 430x430x200 mm³, TE = 102 ms, TR = 1200 ms, flip angle = 135°. This MRI scanner was equipped with a dedicated radiation therapy flat couch and coil mounts supplied by CIVCO Medical Solutions and a laser bridge from LAP Laser. The manufacturer's 3D distortion was used.

The T2-weighted scans were preprocessed with the following steps by using the Insight Segmentation and Registration Toolkit (ITK) [40] to correct MRI non-uniformity and normalize the MRI contrast as in [12], namely:

(1) N4 bias field correction (B-spline fitting [41]: [160, 3, 0, 0.5]; convergence: [100 × 100 × 100, 0.001]; shrink factor: 3); (2) MRI contrast normalization using histogram -matching (levels: 1024, match points: 7, threshold at mean intensity); and (3) Nonlinear filtering via gradient anisotropic anisotropic diffusion [42] (10 iterations; time step: 0.03; conductance: 1.0).

Further, although the delay between CT and MRI acquisitions was kept as short as possible, the patient's anatomy was still slightly different between acquisitions, so, each CT was registered to their related MRI (of the same patient) (as described in [43] with a robust symmetric rigid registration [43], (metric, normalized cross-correlation; geometric transform, rigid), followed by structure-guided deformable registration to promote bone rigidity while allowing high-quality bladder and rectum deformable registration (metric, normalized mutual information with 64 bins; geometric transform, B-spline free-form deformation). This

registered CT was considered as the ground truth.

The study utilizes the same patient data as described in previous studies [9,18]. For all patients, organ delineation was performed on CT by a senior oncologist in agreement with the GETUG/RECORAD [44] group recommendation, including the whole pelvis, bones, prostate, bladder, and the rectum.”

2.2. Synthetic-CT generation with the Pi2Pix method

The Pix2Pix model [39] contains two neural networks: a generator and a discriminator. The generator creates sCT images from MRI images, while the discriminator evaluates the quality of the produced sCT images relative to the reference CT images (Fig. 2).

2.2.1. Pix2Pix network

2.2.1.1. Pix2Pix architecture. In all networks, the discriminator was PatchGAN [39,45]. It takes a sCT image as input and gives a probability value as output. This probability value is close to 1 if the sCT image looks like a true CT image, and close to 0 for a false CT image.

PatchGAN is a type of discriminator for generative antagonist networks that penalizes the structure only at the scale of local image patches, and not at the whole image level [45]. In an image, this discriminator tries to classify each patch $N \times N$ as real or false. This discriminator is run by image convolution, averaging all responses to provide a final output response. In our study, PatchGAN divided the generated images into patches of 70×70 voxels in dimension.

The adversarial loss function L_D (Fig. 2) of the discriminator D, which is the Binary cross-entropy (BCE), can be expressed as [39]:

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

where x is the real CT image, z is a random noise vector (Gaussian random variable), G is the generator network trained to synthesize realistic images (sCT) from random noise (z) to fool the discriminator D, where y is the extra information of the real data, while D is the discriminator which is a classifier trained by adversarial learning to distinguish fake images (generated by G: the sCT) from real CT images.

Traditionally, the L_G loss function (Fig. 2) used by the generator is a L1 loss, but for this work, also L2 [45], Kullback–Leibler Divergence (KLD), and Perceptual loss were tested.

For the Pix2Pix architecture, the aim of L^* (Equation (2)) is to minimize the L_G loss, whereas D aims to maximize it (Fig. 2).

$$L^* = \operatorname{argmin}_G \max_D L_D(G, D) + \lambda L_G(G) \quad (2)$$

where $\lambda > 0$.

2.2.1.2. Evaluation of generator and L_G loss function. For the first step of the Pix2Pix study (Fig. 1), four different generators of Pix2Pix architecture were evaluated: U-Net 128, U-Net 256, ResNet 6 blocks, and ResNet 9 blocks [38]. Additional Fig. 1 describes the generators. For each one, four loss functions were tested (L1, L2, KLD, Perceptual). The training was performed with the Adam optimization algorithm described by Isola et al. [39]: A learning rate (LR) of 2×10^{-4} , momentum parameters of 0.5 and 0.999 for β_1 and β_2 respectively, and 100 epochs.

Then, the architecture with the best results, based on imaging endpoints (Fig. 1), has been used for the next step of the study.

2.2.2. Evaluation of the hyperparameters

The second step of the study (Fig. 1) involves the evaluation of the hyperparameters for the Pix2Pix network. Several hyperparameters have been tested to select the ones which give the best results, based on the imaging endpoints used in the first step.

For this purpose, first, the learning rate has been varied (neighboring values of the Isola et al. [39] learning rate: 2×10^{-4}), with a fixed β_1 . Ten

learning rates have been tested, from 1×10^{-5} to 1×10^{-3} to keep the one with the best results. Then, with the fixed learning rate, five β_1 (neighboring values of the Isola et al. [39] β_1 : 0.5) have been tested, from 0.4 to 0.9.

2.2.3. Training and validation patient cohorts

The entire patient cohort (39 patients) was split into 3 training cohorts, with 25 patients in each one, and 3 validation cohorts of 14, 14, and 11 patients. To test the generator, loss function, and hyper-parameters, only one training was performed with 25 patient images, with a validation cohort of 14 patients (First and second steps, Fig. 1). After the generator, loss function, and hyper-parameters evaluation, a 3-fold cross-validation was performed with 300 epochs (Third step, Fig. 1).

2.3. Comparison of Pix2Pix method with five other synthetic-CT generation methods

The sCT images generated by the Pix2Pix model were compared with those generated by Dowling et al. [12] and Largent et al. [9,18] with the bulk density, atlas-based, patch-based, and two DL methods (U-Net and GAN) in previously published studies [9,12,18] in terms of image and dose endpoints (Fourth step, Fig. 1).

2.3.1. Bulk-density method (BDM)

The bulk density method consists of segmenting MRI images into several classes (corresponding to different interest regions: air pocket, soft tissue, and bones). From the MRI, the 3 types of structures were either manually delineated by an expert or segmented by automatic processing. As explained by Largent et al. [9] the air was defined by manually thresholding the inner part of the rectum, bones were manually delineated, and the soft tissue was obtained by subtracting the bone and the air, from the whole body contour. The assigned density was equal to 1 for soft tissues. The densities assigned to air and bone corresponded to the mean density values of the CTs of the training cohort (0.55 and 1.24, respectively).

2.3.2. Atlas-based method (ABM)

Dowling et al. [12] published the atlas-based method. It consists of non-rigid registrations of one or several co-registered MRI-CT atlases with a target MRI. Then a fusion step using local similarities between the target MRI and training MRI atlases has been done. A comparison between a small neighborhood around each voxel in the target MRI and the same region on the registered MRI atlases was performed. The outcomes of the patch differences were considered as weights and normalized.

Each voxel in the sCT was generated by applying the local weighting of the registered CT atlases in the same location.

2.3.3. Patch-based method (PBM)

The PBM used for comparison was published by Largent et al. in 2018 [9]. The PBM can be decomposed into four steps. The first step is interpatient rigid or affine registration with MR images, then a feature extraction step was performed to obtain spatial, textural, and gradient information from the registered MRI. The third step consists in selecting the training patches closest to the target MRI patches, and finally, a multipoint-wise aggregation scheme was conducted to generate the sCT patches. A 3-fold cross-validation was performed with 300 epochs for sCT generation.

2.3.4. Deep learning methods (DLMs)

The generated sCTs from Pix2Pix were compared with sCTs generated from two others: A U-Net (with L2 loss function) and a GAN (with U-Net generator and L2 loss function). The Adam algorithm was used (Learning rate = 1×10^{-4} , $\beta_1 = 0.9$, epochs = 300, batch size = 5, $\lambda_1 = 5$ and $\lambda_2 = 1$) [18].

Table 1

Reference CT values and imaging endpoints comparing the reference CT with the synthetic-CTs generated with Pix2Pix method for different generators and loss functions, with fixed hyperparameters (learning rate (LR) = 2x10-4, $\beta_1 = 0.5$, $\beta_2 = 0.999$, epochs = 100).

Reference CT values (HU)		Whole pelvis 3.9 ± 17.3		Bones 344.9 ± 67.8		Prostate 30.3 ± 11.1		Bladder 5.8 ± 11.7		Rectum -22.4 ± 79.9	
Loss function	Generator	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)
L1	ResNet 6	45.5 ±	-3.7 ±	199.1 ±	113.0 ±	20.6 ±	-6.4 ±	20.1 ±	-3.5 ±	93.3 ±	-32.5 ±
	blocks	9.4*	13.2*	40.0	55.4	5.7	12.3	9.0	13.4	85.2*	97.4*
	ResNet 9	46.2 ±	-1.0 ±	200.2 ±	124.2 ±	21.3 ±	-0.7 ±	22.5 ±	2.2 ±	84.0 ±	-43.6 ±
	blocks	9.6*	14.7	41.0*	55.4*	4.2*	13.8*	9.0*	15.8	86.6*	97.0*
	U-Net 128	48.1 ±	2.4 ±	219.6 ±	147.7 ±	21.7 ±	5.7 ±	19.1 ±	3.1 ±	76.5 ±	-27.8 ±
		14.8*	15.8*	69.7*	95.7	8.9*	19.0*	7.7	12.5	67.7*	76.4*
	U-Net 256	45.5 ±	-0.5 ±	208.1 ±	129.2 ±	21.6 ±	4.8 ±	17.7 ±	1.5 ±	79.8 ±	-39.1 ±
		8.4*	13.8	41.1*	59.8*	7.0*	17.2*	3.7	14.1	75.9*	88.4*
L2	ResNet 6	52.4 ±	-12.3 ±	211.6 ±	106.0 ±	23.4 ±	-4.5 ±	25.4 ±	-9.8 ±	85.4 ±	-47.0 ±
	blocks	8.3*	14.7*	39.1*	57.6	6.0*	14.3	12.0*	17.3	88.9*	94.2*
	ResNet 9	54.5 ±	-14.0 ±	209.5 ±	107.3 ±	22.6 ±	-0.8 ±	24.0 ±	-5.7 ±	98.5 ±	-34.0 ±
	blocks+	7.5*	14.4*	38.9*	59.2	6.9*	14.6*	6.3*	15.5	83.8*	96.0*
	U-Net 128	54.9 ±	5.3 ±	216.9 ±	139.0 ±	30.0 ±	4.0 ±	24.8 ±	-3.7 ±	160.2 ±	86.8 ±
		8.8*	18.4*	45.9*	66.2*	9.0*	22.2*	5.9*	15.4	101.5*	119.1*
	U-Net 256	54.4 ±	-6.8 ±	216.9 ±	136.0 ±	25.8 ±	-2.8 ±	27.9 ±	-0.9 ±	105.5 ±	-10.9 ±
		8.1*	13.7*	42.9*	62.3*	3.7*	14.0*	9.6*	13.2	80.5*	102.1*
KLD	ResNet 6	46.1 ±	-4.1 ±	197.6 ±	118.4 ±	20.8 ±	-4.6 ±	21.1 ±	-3.8 ±	81.1 ±	-48.1 ±
	blocks	9.1*	13.4*	41.0	57.0	5.1	14.0	7.5	13.2	87.4*	91.5*
	ResNet 9	45.0 ±	1.6 ±	197.4 ±	123.1 ±	22.1 ±	-4.2 ±	21.3 ±	-1.0 ±	76.9 ± 89.0	-54.0 ±
	blocks	8.7	13.9*	41.7	56.8	6.0*	14.9	7.6	13.5		92.9
	U-Net 128	46.1 ±	-1.5 ±	204.0 ±	130.2 ±	21.8 ±	2.2 ±	22.3 ±	5.9 ±	78.4 ± 82.8	-39.9 ±
		8.0*	13.2	41.6*	61.4*	5.6	15.8*	6.0*	13.6		89.0*
	U-Net 256	48.3 ±	-2.2 ±	203.9 ±	119.3 ±	21.8 ±	-1.1 ±	22.2 ±	0.1 ±	81.4 ±	-39.7 ±
		8.1*	14.5*	40.0*	60.8	5.5	16.1*	6.3*	12.9	83.6*	79.0*
Perceptual	ResNet 6	43.6 ±	-1.8 ±	198.5 ±	115.7 ±	18.8 ±	-5.0 ±	19.0 ±	-2.6 ±	80.0 ±	-52.5 ±
	blocks	9.7	12.8	40.2	56.6	5.2	12.9	7.7	15.2	87.4*	95.1
	ResNet 9	43.0 ±	-0.3 ±	197.7 ±	117.2 ±	20.1 ±	-9.3 ±	17.8 ±	-4.9 ±	75.7 ±	-56.5 ±
	blocks	9.4	12.9	40.5	55.7	6.4	12.8	6.6	14.0	90.1	93.2
	U-Net 128	43.8 ±	0.7 ± 13.2	207.6 ±	125.9 ±	23.3 ±	-5.7 ±	17.1 ±	-4.9 ±	74.7 ± 78.5	-50.6 ±
		8.5*		38.9*	58.6	5.4*	16.6	6.6	12.0		82.3
	U-Net 256	43.7 ±	1.2 ±	206.5 ±	125.1 ±	22.5 ±	-5.4 ±	17.2 ±	-6.6 ±	74.4 ± 80.9	-54.3 ±
		8.2*	12.7*	38.6*	59.0	4.9*	15.2	6.2	13.1		83.6

MAE = mean absolute error; ME = mean error; HU = Hounsfield Units.

The imaging endpoint values are expressed as mean ± standard deviation. Results were obtained with 14 patient data in evaluation.

The Wilcoxon test was used, first to compare the MAE of the Perceptual loss function with the generator ResNet 9 block to those of the other methods, and second to compare the ME of the methods to a null distribution.

*Significant differences were considered at p-value < 0.05.

2.4. Evaluation

2.4.1. Image endpoints

The organ delineations made by the radiation oncologist were rigidly propagated from the reference CT to the sCT images.

For each step of the study, the network accuracy was evaluated by comparing the generated sCT images to the reference CT images. Voxel-wise comparisons of the sCT and reference CT images, expressed in HU, were performed by calculating the Mean Absolute Error (MAE) and Mean Error (ME).

$$MAE = \frac{1}{n} \sum_{i=1}^n |HU_{CTref}(i) - HU_{sCT}(i)| \quad (4)$$

$$ME = \frac{1}{n} \sum_{i=1}^n HU_{CTref}(i) - HU_{sCT}(i) \quad (5)$$

MAE and ME values close to 0 indicate good quality of the generated sCT. MAE and ME were calculated for each organ (whole pelvis, bones, prostate, bladder, and rectum).

2.4.2. Dose endpoints

Volumetric modulated arc therapy (VMAT) was planned on the CT images with the Pinnacle v.9.10 (Philips) treatment planning system for prostate and seminal vesicles. The collapsed cone convolution algorithm was used for dose calculation. Sequential treatment was delivered with a total dose of 50 Gy to the prostate and seminal vesicles, followed by a

boost of 28 Gy in the prostate (2 Gy per fraction). The GETUG dose-volume constraints were applied to the organs at risk [44].

The beam parameters used to compute the dose from the CT images were also used to calculate the dose from the sCT images. The sCT dosimetric accuracy was evaluated by comparing the dose distribution (3D gamma indices and Dose-Volume Histogram, DVH, metrics), calculated using the sCT and the reference CT images. Briefly, a spatial dose evaluation was performed using 3D gamma analysis (criteria: local, 1%/1 mm, dose thresholds = 10 % and 30 % of the prescription dose) and the dose distributions from the reference CT and sCT images. Then, the method accuracy was evaluated by computing the dose uncertainty (MAE) defined by the differences in mean absolute values across DVHs calculated from the dose on the reference CT and sCT images.

2.4.3. Statistical analyses

Wilcoxon signed-rank tests were performed to compare the endpoints. For MAE (image and dose), this test was used to compare the lowest MAE among all the methods to the MAE of each other method. For the ME (image), this test was used to compare the ME of each method to 0 (null distribution). For the DVH comparisons across the sCT generation methods, a nonparametric permutation test was performed to control the presence of false positives in case of multiple statistical tests (5 cGy DVH bin-wise). The details of permutation tests were previously explained in the Largent et al. [18] and Mylona et al. [46] studies.

Table 2

Reference CT values and imaging endpoints comparing the reference CT with the synthetic-CTs generated with Pix2Pix method, for different learning rate and momentum β_1 .

Reference CT values (HU)	Whole pelvis		Bones		Prostate		Bladder		Rectum		
	3.9 ± 17.3		344.9 ± 67.8		30.3 ± 11.1		5.8 ± 11.7		-22.4 ± 79.9		
LR	β_1	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)
1×10^{-05}	0.5	50.4 ± 9.5*	-28.6 ± 11.8	182.0 ± 18.4*	7.5 ± 68.0	20.1 ± 5.8*	0.3 ± 14.1	21.1 ± 8.6*	-9.4 ± 16.3	77.1 ± 94.2*	-68.6 ± 96.5
1×10^{-04}	0.5	41.7 ± 9.4*	-11.9 ± 14.2	156.1 ± 26.5*	-5.9 ± 54.7	23.4 ± 15.8*	3.7 ± 20.5	49.8 ± 103.7*	31.9 ± 108.9*	95.3 ± 94.5*	-34.8 ± 118.8*
2×10^{-04}	0.5	42.4 ± 9.4*	-9.5 ± 15.2*	150.5 ± 25.6*	-5.6 ± 52.5	22.9 ± 15.8*	3.0 ± 21.2	45.3 ± 91.3*	26.7 ± 94.5*	97.9 ± 104.6*	-23.5 ± 133.1*
3×10^{-04}	0.5	37.1 ± 8.6*	-15.3 ± 12.2*	138.3 ± 18.5*	-15.8 ± 48.8*	18.5 ± 4.6*	-3.8 ± 13.0	19.6 ± 6.1*	1.4 ± 16.2	75.7 ± 89.9	-57.1 ± 95.1
4×10^{-04}	0.5	38.6 ± 7.5	-7.0 ± 13.4*	144.1 ± 18.7*	-1.9 ± 47.5	23.1 ± 15.6*	7.0 ± 20.0	51.1 ± 90.1*	33.3 ± 86.8*	86.7 ± 89.0*	-37.7 ± 105.5*
5×10^{-04}	0.5	36.6 ± 7.2	-7.8 ± 13.5*	139.8 ± 19.2*	-5.9 ± 53.6	20.0 ± 4.8*	4.1 ± 15.4	20.6 ± 5.0*	3.1 ± 16.7	74.1 ± 89.1	-44.7 ± 96.6
6×10^{-04}	0.4	37.0 ± 8.9	-13.8 ± 11.9*	133.9 ± 21.1	0.1 ± 48.6*	18.2 ± 4.4	-3.0 ± 12.9	18.2 ± 6.8	-1.3 ± 14.9	75.2 ± 88.7*	-53.3 ± 94.3
	0.5	36.1 ± 8.3	-16.0 ± 12.0	134.7 ± 19.9	-7.9 ± 49.8	18.3 ± 4.2	-2.8 ± 12.8	18.5 ± 5.9	-3.2 ± 14.6	74.1 ± 89.3	-50.5 ± 95.1
	0.6	36.3 ± 8.9	-14.8 ± 11.7*	131.9 ± 23.6	8.1 ± 49.0	18.0 ± 4.7	-4.9 ± 12.2	18.7 ± 7.2	-2.9 ± 14.7	74.2 ± 89.5	-55.4 ± 94.8
	0.7	36.5 ± 7.6	-11.1 ± 12.7*	139.3 ± 18.0	-16.1 ± 52.2	19.0 ± 5.2	-2.3 ± 14.5	19.9 ± 6.2*	-1.4 ± 15.5	75.7 ± 89.6*	-44.9 ± 97.2
	0.9	92.6 ± 4.1*	-15.5 ± 15.5	338.8 ± 58.7*	332.0 ± 61.8*	31.6 ± 10.5*	26.9 ± 14.4*	16.3 ± 5.3*	1.2 ± 14.2	76.3 ± 86.3*	-35.7 ± 94.6
7×10^{-04}	0.5	37.5 ± 7.4	-12.6 ± 11.5*	140.7 ± 16.5*	-22.0 ± 48.4*	20.2 ± 9.4*	-0.4 ± 13.8	24.6 ± 21.6*	3.9 ± 26.2	76.0 ± 88.3*	-50.4 ± 94.3
9×10^{-04}	0.5	37.1 ± 9.5*	-15.1 ± 12.0*	134.4 ± 22.2	7.5 ± 51.3*	17.9 ± 4.6	-3.6 ± 13.0	22.6 ± 16.5*	-2.2 ± 18.8	75.0 ± 89.5	-49.6 ± 96.6
1×10^{-03}	0.5	38.1 ± 9.3*	-17.0 ± 12.2*	136.9 ± 20.7*	-2.0 ± 51.3	19.5 ± 6.1*	-8.5 ± 13.5*	19.3 ± 8.8*	-4.1 ± 15.5	76.6 ± 87.8*	-54.9 ± 93.2

LR = learning rate; MAE = mean absolute error; ME = mean error; HU = Hounsfield Units.

The imaging endpoint values are expressed as mean ± standard deviation. The tested learning rate and β_1 are shown in the row. Results were obtained with 14 patient data in evaluation.

The Wilcoxon test was used, first to compare the MAE of the Perceptual loss function with the generator ResNet 9 block with LR = 6×10^{-04} and $\beta_1 = 0.5$ to those of the other hyperparameters, and second to compare the ME of the methods to a null distribution.

*Significant differences were considered at p-value < 0.05.

Table 3

Reference CT values and imaging endpoints comparing the reference CT with the synthetic-CTs obtained by each method.

Reference CT values (HU)	Whole pelvis		Bones		Prostate		Bladder		Rectum	
	3.9 ± 17.3		344.9 ± 67.8		30.3 ± 11.1		5.8 ± 11.7		-22.4 ± 79.9	
Network	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)	MAE (HU)	ME (HU)
Atlas-based method	43.9 ± 13.9*	-11.5 ± 17.8	130.8 ± 26.7*	20.9 ± 50.6*	25.9 ± 5.0*	-5.6 ± 13.7	17.9 ± 6.4*	6.1 ± 12.4*	75.1 ± 59.0*	-1.6 ± 76.6*
Bulk -density method	87.4 ± 13.3*	-16.4 ± 21.8*	239.5 ± 29.5*	18.06 ± 70.1	33.5 ± 9.0*	30.3 ± 11.3*	15.7 ± 5.9*	5.1 ± 11.7*	115.5 ± 75.5*	62.9 ± 92.5
Patch Based method	44.7 ± 11.4*	6.0 ± 19.0	143.6 ± 27.8*	58.3 ± 45.5*	20.6 ± 6.0*	8.2 ± 15.0	21.1 ± 9.0*	10.7 ± 14.0*	78.0 ± 60.5*	7.0 ± 73.2*
U-Net	34.4 ± 7.7*	-1.0 ± 14.2	125.3 ± 22.0*	20.2 ± 42.3	18.1 ± 5.2*	0.8 ± 12.9	18.6 ± 7.4*	3.4 ± 13.6	65.0 ± 65.7*	-24.0 ± 72.5*
GAN	34.1 ± 7.5*	-1.1 ± 13.7	123.9 ± 20.6*	19.4 ± 41.4	17.7 ± 4.5*	0.3 ± 12.0	18.8 ± 8.9*	3.7 ± 14.6	68.3 ± 64.4*	-20.5 ± 73.6*
Pix2Pix	29.5 ± 7.9	-9.0 ± 9.5	107.7 ± 24.9	-12.9 ± 34.6	16.0 ± 3.8	0.6 ± 9.9	13.4 ± 5.7	0.2 ± 10.1	49.1 ± 61.0	63.7 ± 22.5

MAE = mean absolute error; ME = mean error; HU = Hounsfield Units.

The imaging endpoint values are expressed as mean ± standard deviation. The tested methods are shown in the row. Results were obtained with 39 patients data in evaluation.

The Wilcoxon test was used, first to compare the MAE of the Pix2Pix to those of the other methods, and second to compare the ME of the methods to a null distribution.

*Significant differences were considered at p-value < 0.05.

3. Results

3.1. First step: Pix2Pix generator and loss function evaluation

The Additional Fig. 2 shows a summary of the results for each step of the study.

Table 1 shows the reference CT values and comparison of the two investigated metrics (MAE and ME) in HU for all the ROIs. The generator Resnet 9 blocks with the perceptual loss function showed the lowest

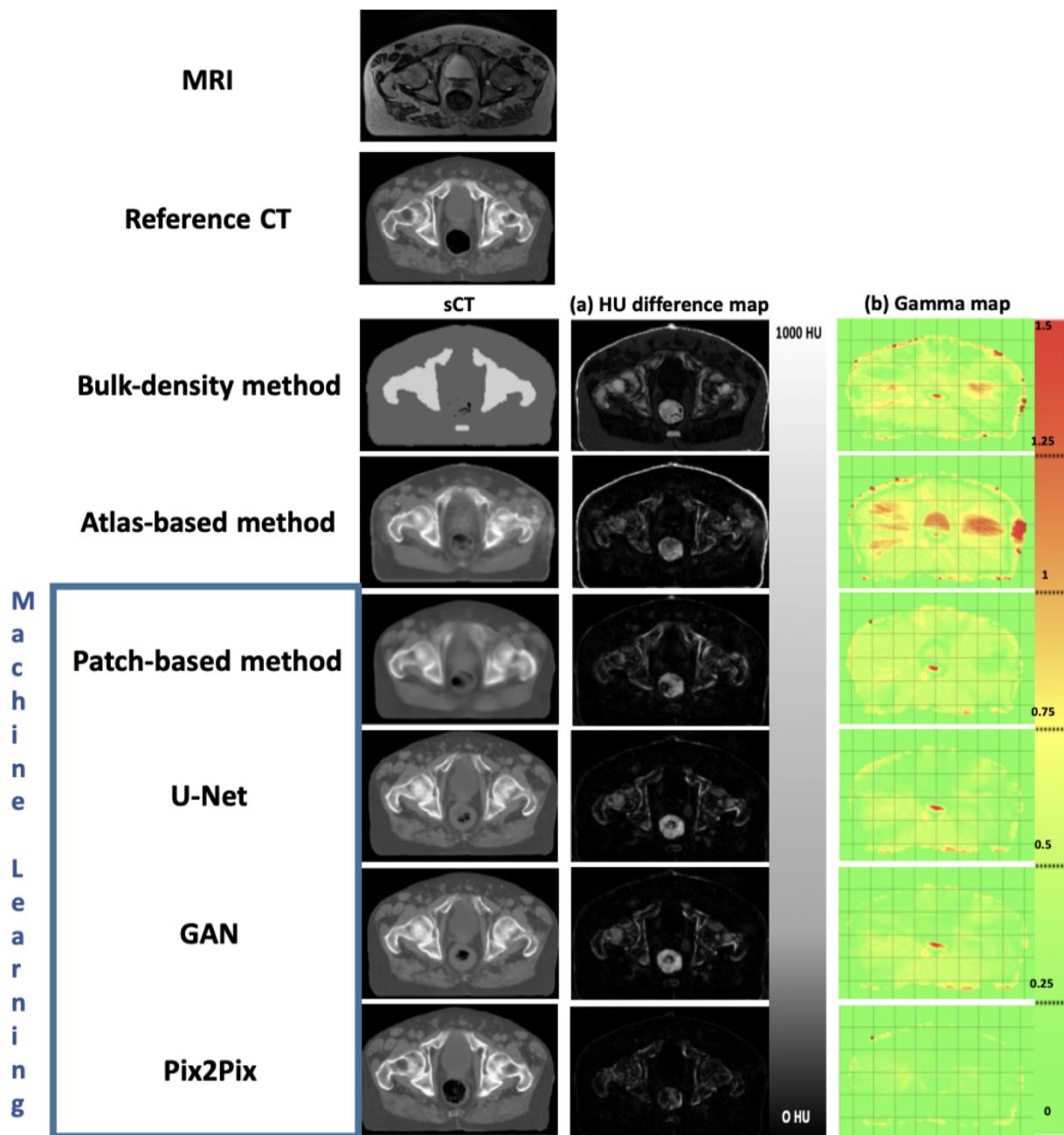


Fig. 3. MRI, Reference CT, synthetic-CT (sCT), (a) Hounsfield units difference maps and (b) gamma maps for bulk-density, atlas-based and machine learning methods (patch-based, U-Net, GAN and Pix2Pix) for one patient. Gamma analyses were performed with criteria: local, 1 %/1mm and 10 % dose threshold. MAE values for this patient for the whole pelvis, were 85 HU, 52 HU, 50 HU, 38 HU, 37 HU, and 26 HU, for, respectively, the BDM, ABM, PBM, U-Net, GAN and Pix2Pix.

MAE in the whole pelvis contour.

3.2. Second step: Pix2Pix hyperparameters evaluation

Table 2 presents the results of the hyperparameters evaluation in terms of MAE and ME. The lowest MAE values were obtained with a learning rate of 6×10^{-4} and a momentum β_1 of 0.5.

3.3. Third step: Pix2Pix image and dose evaluation

For the final generation of sCT with the Pix2Pix method using 3-fold cross-validation, an MAE of 29.5 ± 7.9 HU was obtained. The mean gamma pass rate was 99.4 ± 0.9 %, using the 1 %/1mm and 10 % dose threshold criterion.

3.4. Fourth step: Comparison of synthetic-CT generation methods

Table 3 summarizes the results of the image evaluation (MAE and ME) in HU for the different methods. The Pix2Pix architecture showed the lowest MAE (29.5 ± 7.9 HU). In Fig. 3a the reference CT, sCT, and difference map for bulk-density, atlas, patch, U-Net, GAN, and Pix2Pix method are shown for a patient (MAE = 27 HU).

Figure 3b illustrates the gamma map for a patient for all the methods. Fig. 5 presents the mean gamma and the gamma pass-rate values for the six methods. The highest gamma pass rates (99.4 %) values were obtained for the Pix2Pix architecture.

Fig. 4 presents the dose uncertainties (MAE) for all sCTs generation methods along with the entire dose-volume histogram (DVH) for the prostate, bladder, rectum, and femoral heads. It shows, especially for the

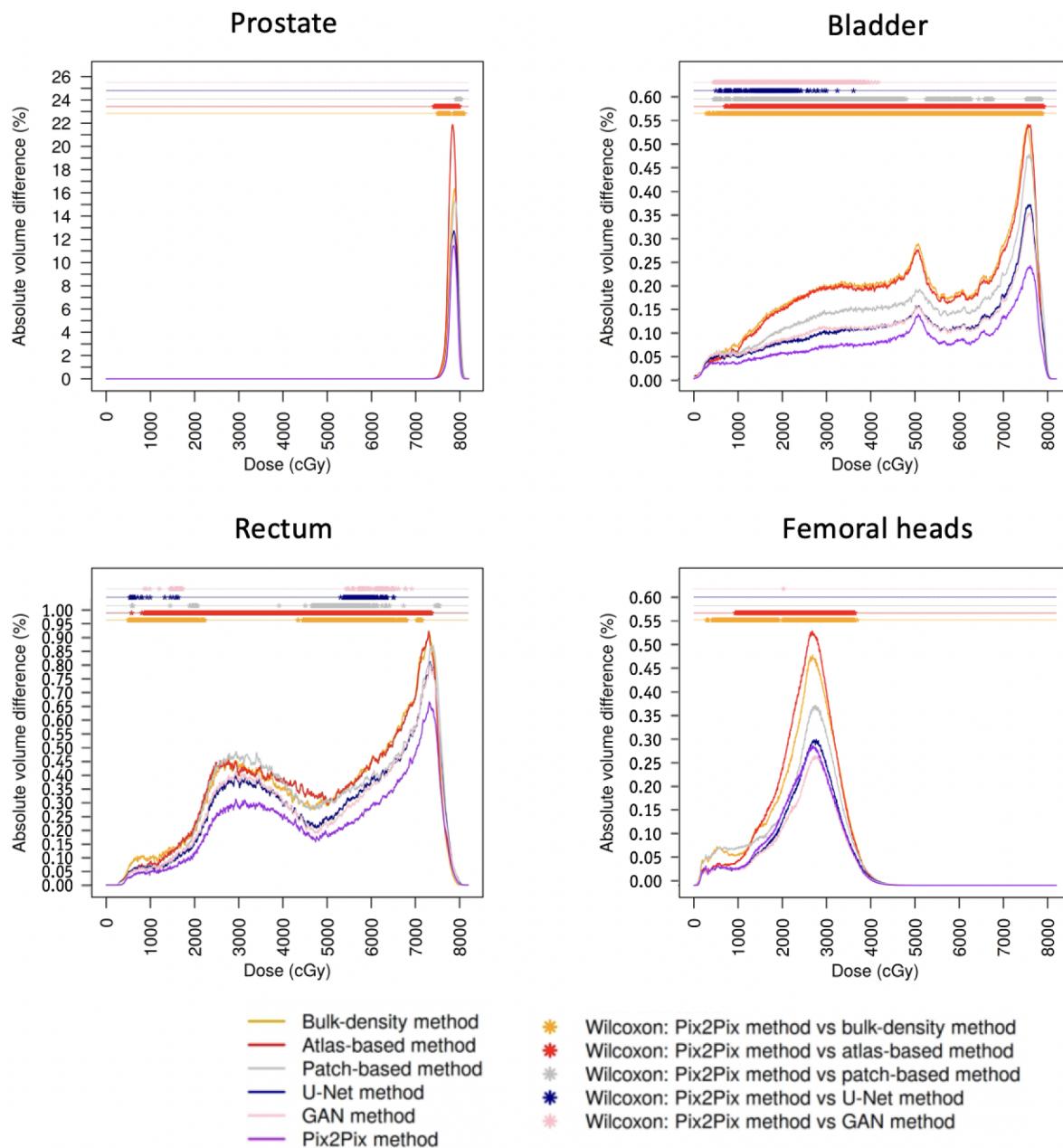


Fig. 4. Dose uncertainty of each synthetic-CT generation method (mean absolute DVH differences between the reference CT and the synthetic-CT for all volumes-of-interest). The dose uncertainty is defined as the mean absolute DVH differences between the reference CT and the sCT corresponding to each method. Permutation tests were used to compare the absolute DVH differences of Pix2Pix network to those of the other methods. Significant differences (p -value < 0.05) are displayed on the top of each figure using the symbol *.

rectum and bladder, a significant improvement of the dose uncertainties with the Pix2Pix method.

3.5. Calculation time

Two parameters can be considered for calculation time. The training computation time and the testing computation time (i.e., the computation time to generate a sCT).

The atlas-based method (ABM) was implemented in C++ using the Insight ToolKit (ITK 4.4) library. The mean computation time to generate one sCT was 160 min (without the use of cluster architecture or graphical processing unit parallelization).

The non-local mean patch-based method (PBM) was implemented in C++ using the Insight ToolKit library. The training computation time was approximately 24 h (without the use of cluster architecture or GPU

parallelization) for the whole 3-fold cross-validation. Mean computation time to generate one sCT was 62 min.

The DLMs were implemented in Python using Keras (V.59), TensorFlow (U-Net and GAN), PyTorch (Pix2Pix), and a Nvidia Quadro RTX6000 (24 GB GDDR6) GPU card. The training computation times (whole 3-fold cross-validation) of the U-Net, GAN, and Pix2Pix were respectively 17, 57 and 39 h. The sCT generation computation time (per MRI scan) of the U-Net, GAN, and Pix2Pix were respectively 9, 9, and 5 s.

The computation time for the DL-based approaches (UNet, GAN, Pix2Pix) was lower than those of the ABM and PBM. However, DL-based methods are built upon already optimized Hardware and Software Parallel Architectures (GPU, Convolutional Networks), whereas for the others the codes on CPU were not optimized.

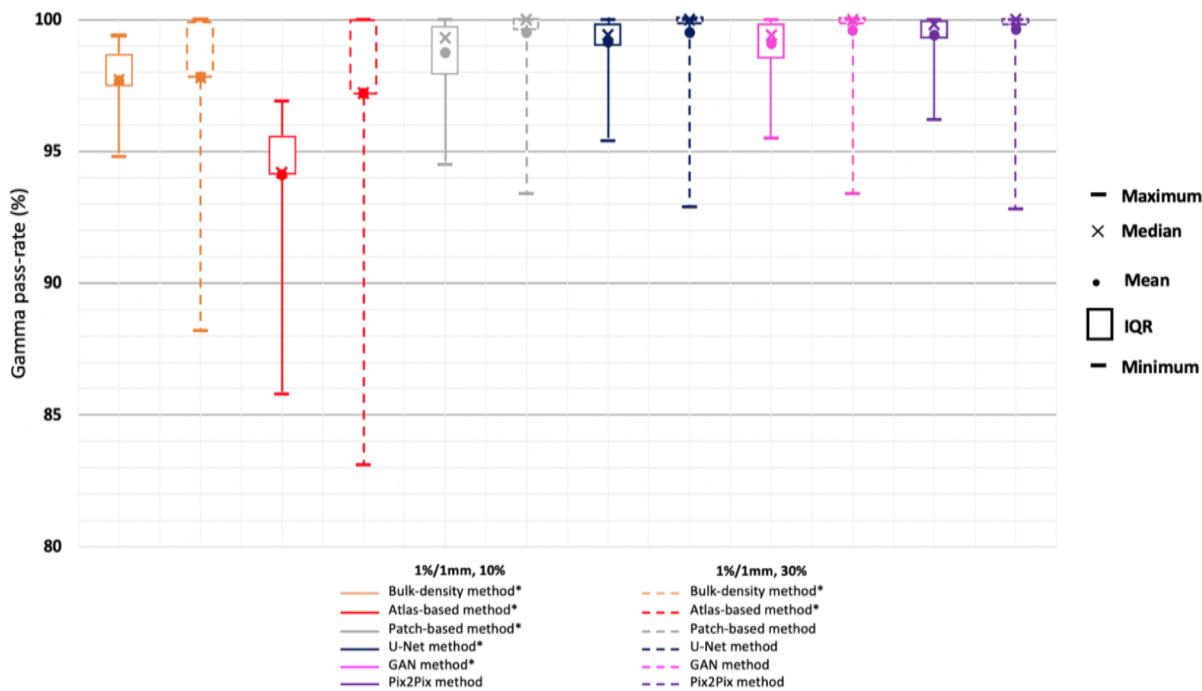


Fig. 5. Gamma pass-rate boxplot for each synthetic-CT generation method calculated from the reference CT and synthetic-CT dose distributions IQR: Inter quartile range Solid line represents the gamma pass-rate for the 1 %/1mm, 10 % dose threshold criteria, and dotted line, the gamma pass-rate for the 1 %/1mm, 30 % dose threshold criteria. The Wilcoxon test was used to compare the mean gamma of the Pix2Pix to those of the other methods, and to compare the gamma pass-rate of the Pix2Pix to those of the other methods. *Significant differences were considered at p-value < 0.05.

4. Discussion

This study investigated a Pix2Pix DLM for pelvis sCT image generation from MRI data. Generators, loss functions, and hyperparameters were selected according to image endpoints (Fig. 1). First, the generators and loss functions were evaluated with fixed hyperparameters: A learning rate (LR) of 2×10^{-4} , momentum parameters of 0.5 and 0.999 for β_1 and β_2 respectively, and 100 epochs. Then, the variation of the learning rate and the momentum β_1 has been studied with the generator and loss function with the best results based on imaging endpoints, respectively ResNet 9 blocks and Perceptual loss (Fig. 1). This variation was therefore not carried out for each generator and loss function. Then, it is not excluded that with other hyperparameters there would have been better results with other generators and loss functions. However, testing exhaustively all combination of hyperparameters for each generator and loss functions, would have been extremely time-consuming, for a gain of sCT accuracy in terms of image, probably minimal, and a gain in terms of dose even lower.

Overall, compared with the reference CT images, the sCT images generated by the Pix2Pix method provided low dose uncertainties, thereby making them clinically acceptable for MRI-based prostate cancer treatment planning. The Pix2Pix method provides the lowest image errors (image endpoints) compared to the five previously published methods with the same patient data (Table 3). The Pix2Pix methods provided similar dose errors than the U-Net and GAN methods previously published (Fig. 6).

DLMs have been used for sCT generation from pelvis MRI data in at least 18 studies [19,20,47]. The comparison between the literature and our study can only be indirect because the data sets are different. Differences in dataset size, MRI sequence(s), artifacts, populations, and misalignment between sCT and CT images could influence the evaluation [28].

In the state of the art, several architectures have been used, such as fully convolutional network, deep embedding convolutional neural

network, convolutional neural network U-Net, GAN, CycleGAN (cGAN) [19,20]. Fig. 6 summarizes the MAE values (in HU) for the whole pelvis obtained in the literature (with different DLMs) and in the present study.

Only five studies have used a cGAN architecture for prostate sCT generation [45,48–51], with two applying the Pix2Pix method [50,51]. The first Pix2Pix study used a cohort of 70 patients, including prostate, cervix, and male pelvis. Four generators were tested, and the lowest MAE was 41.2 HU in the whole pelvis. The second study on the Pix2Pix architecture [51] for pelvis imaging included a cohort of 60 patients, and the whole pelvis MAE was 54.3 HU. Our Pix2Pix architecture (ResNet 9 Blocks, perceptual loss) compares favorably, as indicated by its MAE value of 29.5 HU for the whole pelvis.

Our present results can be compared directly only with those of two previous studies by our research team [9,18] on the same patient cohort. In these two studies, the lowest MAE results were obtained with a GAN architecture that led to values of 34.1 HU in the whole pelvis versus 29.5 HU, with the Pix2Pix architecture tested in the present study (Table 3).

Half of the DLM studies in the pelvic region evaluated dose uncertainties [19]. Considering the gamma pass-rate, the values were between 89.3 % and 99.2 (mean value: 94.7 %) using the 1 %/1mm criterion. Cusumano et al. [48] is the only study that has evaluated the gamma pass-rate for a Pix2Pix architecture (89.3 %; for the 1 %/1mm, 10 % dose threshold criteria). But the type of gamma index is not specified. In the present study, the mean gamma passing rate was 99.4 % (for the local, 1 %/1mm, 10 % dose threshold criteria) with the Pix2Pix method (Fig. 5).

The MAE values for the rectum were three times higher than those for other soft tissues, such as the prostate and bladder (Table 3). In addition, dose uncertainties for rectum were 10 times higher than for other soft tissues. Moreover, the patients with the lowest gamma values are systematically those with the biggest differences in gas volume in the rectum between the CT and the MRI. For all the sCT generation methods, it is always the same patient with the lowest gamma values. For this patient almost the entire volume of the rectum was plenty gas on the CT,

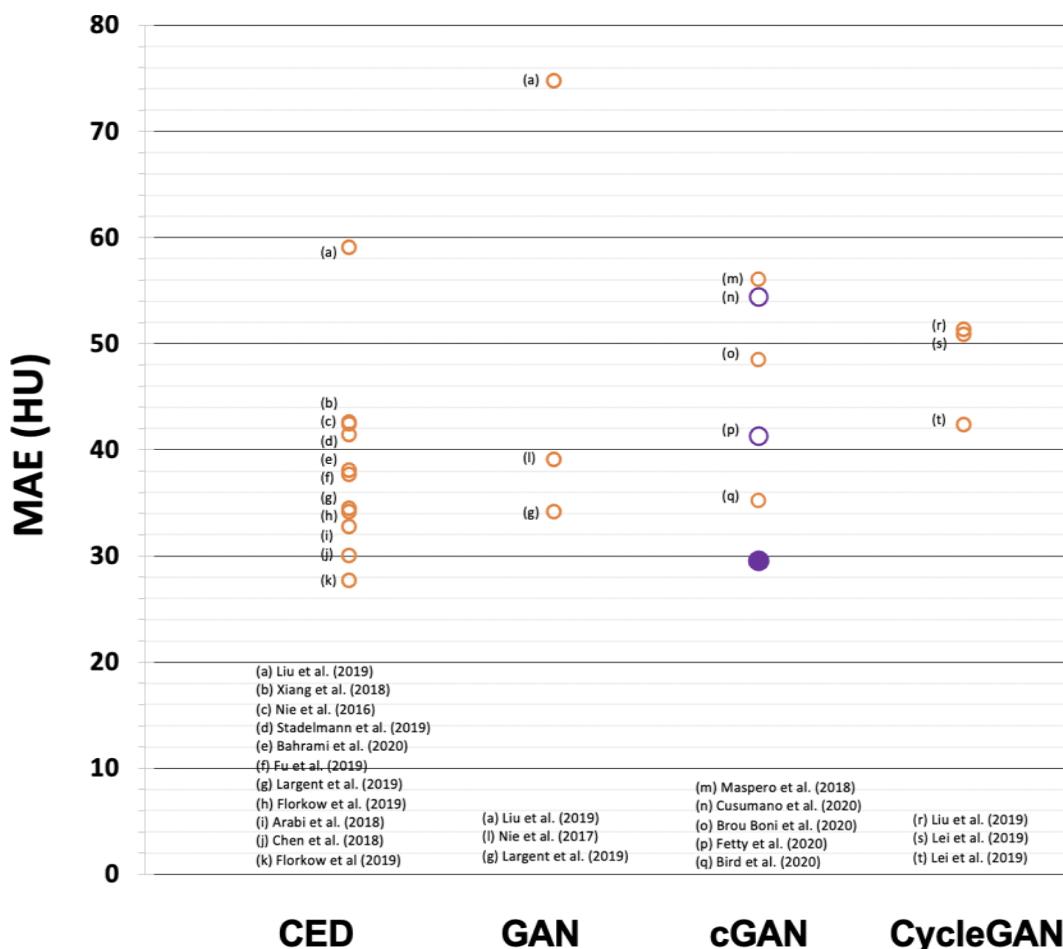


Fig. 6. State of the art of mean absolute error of synthetic-CT from MRI in the whole pelvis obtained by deep learning methods for prostate radiotherapy CED: convolutional encoder-decoder; GAN: generative adversarial network; cGAN: conditional GAN; MAE = mean absolute error; ME = mean error; HU = Hounsfield Units Each data point represents a MAE value reported in the literature for the whole pelvis contour. These values are classified in four methods: CED (generator-only methods in deep learning such as U-Net, CNN), GAN, cGAN, and CycleGAN. Full purple circles represent the MAE for this study. Empty purple circles represent the MAE values for studies using the Pix2Pix method. Empty orange circles represent MAE values using CED, GAN, cGAN (excepted the Pix2Pix one), or CycleGAN methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

but not on the MR image.

Indeed, a known issue with sCT generation in the pelvic area is the location of gas in rectum between CT and MR images. Therefore, these high MAE values could be related to the difference in gas pockets between the MRI and reference CT images.

Lastly, the proposed Pix2Pix method can be clinically used only if it allows the generation of sCT images from different MRI vendors and models. To this aim, the next step is the evaluation of multicenter impacts on sCT generation.

5. Conclusion

To generate sCT images for MRI-based prostate cancer dose planning, the cGAN (Pix2Pix) architecture (with a generator ResNet 9 blocks and the perceptual loss function) appears to be particularly promising in clinical practice, due to the low dose uncertainty and fast calculation time. The proposed method (Pix2Pix) showed the lowest imaging error (MAE = 29.5 HU) compared to previously published sCT generation methods (bulk, atlas, or deep learning methods) with the same dataset. However, similar dose errors were obtained to previously published deep learning methods (U-Net or GAN).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by, the Departmental Committees CD22, CD29, CD35, CD44, and CD49 of the “Ligue Inter-Régionale Contre Le Cancer Du GrandOuest” (CSIRGO 2020), and Défis scientifiques émergents de l’université de Rennes 1 AAP 2020 et 2021.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmp.2022.10.003>.

References

- [1] Seco J, Evans PM. Assessing the effect of electron density in photon dose calculations: Effect of electron density in photon dose calculations. *Med Phys* 2006; 33:540–52. <https://doi.org/10.1118/1.2161407>.
- [2] Pathmanathan AU, McNair HA, Schmidt MA, Brand DH, Delacroix L, Eccles CL, et al. Comparison of prostate delineation on multimodality imaging for MR-guided radiotherapy. *Br J Radiol* 2019;92(1096):20180948.

- [3] Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol* 2015;60: R323–61. <https://doi.org/10.1088/0031-9155/60/22/R323>.
- [4] Cusumano D, Boldrini L, Dhont J, Fiorino C, Green O, Güngör G, et al. Artificial Intelligence in magnetic Resonance guided Radiotherapy: Medical and physical considerations on state of art and future perspectives. *Phys Med* 2021;85:175–91.
- [5] Kajikawa T, Kadoya N, Tanaka S, Nemoto H, Takahashi N, Chiba T, et al. Dose distribution correction for the influence of magnetic field using a deep convolutional neural network for online MR-guided adaptive radiotherapy. *Phys Med* 2020;80:186–92.
- [6] Lambert J, Greer PB, Menk F, Patterson J, Parker J, Dahl K, et al. MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning. *Radioter Oncol* 2011;98(3):330–4.
- [7] Chen L, Price RA, Wang L, Li J, Qin L, McNeely S, et al. MRI-based treatment planning for radiotherapy: Dosimetric verification for prostate IMRT. *Int J Radiat Oncol* 2004;60(2):636–47.
- [8] Lee Y. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radioter Oncol* 2003;66:203–16. [https://doi.org/10.1016/S0167-8140\(02\)00440-1](https://doi.org/10.1016/S0167-8140(02)00440-1).
- [9] Largent A, Barateau A, Nunes J-C, Lafond C, Greer PB, Dowling JA, 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* 2019;103(2):479–90.
- [10] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol* 2017;12:28. <https://doi.org/10.1186/s13014-016-0747-y>.
- [11] Sjölund J, Forsberg D, Andersson M, Knutsson H. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Phys Med Biol* 2015; 60:825–39. <https://doi.org/10.1088/0031-9155/60/2/825>.
- [12] Dowling JA, Sun J, Pichler P, Rivest-Hénault D, Ghose S, Richardson H, 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* 2015;93(5):1144–53.
- [13] Greer PB, Dowling JA, Lambert JA, Fripp J, Parker J, Denham JW, et al. A magnetic resonance imaging-based workflow for planning radiation therapy for prostate cancer. *Med J Aust* 2011;194(S4). <https://doi.org/10.5694/j.1326-5377.2011.tb02939.x>.
- [14] Uh J, Merchant TE, Li Y, Li X, Hua C. MRI-based treatment planning with pseudo CT generated through atlas registration: MRI-based treatment planning with atlas approach. *Med Phys* 2014;41:051711. <https://doi.org/10.1118/1.4873315>.
- [15] Sivesson C, Nordström F, Nilsson T, Nyholm T, Jonsson J, Gunnlaugsson A, et al. Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm: MRI-only radiotherapy planning using the statistical decomposition algorithm. *Med Phys* 2015;42(10):6090–7.
- [16] Demol B, Boydev C, Korhonen J, Reynaert N. Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images: MRI-only treatment planning in atlas-based pseudo-CT images. *Med Phys* 2016;43:6557–68. <https://doi.org/10.1118/1.4967480>.
- [17] Andreasen D, Van Leemput K, Edmund JM. A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis: A patch-based pseudo-CT approach. *Med Phys* 2016;43:4742–52. <https://doi.org/10.1118/1.4958676>.
- [18] Largent A, Barateau A, Nunes J-C, Mylona E, Castelli J, Lafond C, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int J Radiat Oncol* 2019;105(5):1137–50.
- [19] Boulanger M, Nunes J-C, Chourak H, Largent A, Tahri S, Acosta O, et al. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. *Phys Med* 2021;89:265–81.
- [20] Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: A review. *Med Phys* 2021;48(11):6537–66.
- [21] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44. <https://doi.org/10.1038/nature14539>.
- [22] Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. *Med Phys* 2019;46(1):e1–36.
- [23] Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. *Br J Radiol* 2019;92:20190001. <https://doi.org/10.1259/bjr.20190001>.
- [24] Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. *Comput Biol Med* 2018;98:126–46. <https://doi.org/10.1016/j.combiomed.2018.05.018>.
- [25] Feng M, Valdes G, Dixit N, Solberg TD. Machine learning in radiation oncology: opportunities, requirements, and needs. *Front Oncol* 2018;8:110. <https://doi.org/10.3389/fonc.2018.00110>.
- [26] Papadimitroulas P, Brocki L, Christopher Chung N, Marchadour W, Vermet F, Gaubert L, et al. Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization. *Phys Med* 2021;83:108–21.
- [27] Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, et al. Artificial intelligence and machine learning for medical imaging: A technology review. *Phys Med* 2021;83:242–56.
- [28] Arabi H, Dowling JA, Burgos N, Han X, Greer PB, Koutsouvelis N, et al. Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region. *Med Phys* 2018;45:5218–33. <https://doi.org/10.1002/mp.13187>.
- [29] Neppi S, Landry G, Kurz C, Hansen DC, Hoyle B, Stöcklein S, et al. Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans. *Acta Oncol* 2019;58(10):1429–34.
- [30] Fu J, Yang Y, Singhrao K, Ruan D, Chu F-I, Low DA, et al. Deep learning approaches using 2D and 3D convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging. *Med Phys* 2019;46(9):3788–98.
- [31] Dinkla AM, Florkow MC, Maspero M, Savenije MHF, Zijlstra F, Doornaert PAH, et al. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network. *Med Phys* 2019;46(9):4095–104.
- [32] Florkow MC, Zijlstra F, Willemsen K, Maspero M, Berg CAT, Kerkmeijer LGW, et al. Deep learning-based MR-to-CT synthesis: The influence of varying gradient echo-based MR images as input channels. *Magn Reson Med* 2020;83(4):1429–41.
- [33] Thummerer A, de Jong BA, Zaffino P, Meijers A, Marmitt GG, Seco J, et al. Comparison of the suitability of CBCT- and MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients. *Phys Med Biol* 2020;65: 235036. <https://doi.org/10.1088/1361-6560/abb1d6>.
- [34] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv150504597 Cs* 2015.
- [35] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. *ArXiv14062661 Cs Stat* 2014.
- [36] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. *Med Image Anal* 2019;58:101552. <https://doi.org/10.1016/j.media.2019.101552>.
- [37] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR. Las Vegas, NV, USA: IEEE; 2016. p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.
- [38] Liu X, Emami H, Nejad-Davarani SP, Morris E, Schultz L, Dong M, et al. Performance of deep learning synthetic CTs for MR-only brain radiation therapy. *J Appl Clin Med Phys* 2021;22(1):308–17.
- [39] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. *ArXiv161107004 Cs* 2018.
- [40] Ibanez L, Schroeder W. ITK Software Guide n.d.:836.
- [41] Tustison NJ, Avants BB, Cook PA, Yuanjie Zheng, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 2010;29(6): 1310–20.
- [42] Weickert J. Anisotropic Diffusion in Image Processing n.d.:184.
- [43] Rivest-Hénault D, Dowson N, Greer PB, Fripp J, Dowling JA. Robust inverse-consistent affine CT–MR registration in MRI-assisted and MRI-alone prostate radiation therapy. *Med Image Anal* 2015;23:56–69. <https://doi.org/10.1016/j.media.2015.04.014>.
- [44] Beckendorf V, Guerif S, Le Prisé E, Cosset J-M, Bougnoux A, Chauvet B, et al. 70 Gy Versus 80 Gy in localized prostate cancer: 5-year results of GETUG 06 randomized trial. *Int J Radiat Oncol* 2011;80:1056–63. <https://doi.org/10.1016/j.ijrobp.2010.03.049>.
- [45] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jungenleim-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol* 2018; 63:185001. <https://doi.org/10.1088/1361-6560/aada6d>.
- [46] Mylona E, Acosta O, Lizée T, Lafond C, Crehange G, Magné N, et al. Voxel-based analysis for identification of urethrovesical subregions predicting urinary toxicity after prostate cancer radiation therapy. *Int J Radiat Oncol* 2019;104(2):343–54.
- [47] Bahrami A, Karimian A, Arabi H. Comparison of different deep learning architectures for synthetic CT generation from MR images. *Phys Med* 2021;90: 99–107. <https://doi.org/10.1016/j.ejmp.2021.09.006>.
- [48] Bird D, Nix MG, McCallum H, Teo M, Gilbert A, Casanova N, et al. Multicentre, deep learning, synthetic-CT generation for ano-rectal MR-only radiotherapy treatment planning. *Radiother Oncol* 2021;156:23–8.
- [49] Brou Boni KND, Klein J, Vanquin L, Wagner A, Lacornerie T, Pasquier D, et al. MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. *Phys Med Biol* 2020;65:075002. <https://doi.org/10.1088/1361-6560/ab7633>.
- [50] Fettig L, Löfstedt T, Heilemann G, Furtado H, Nesvacil N, Nyholm T, et al. Investigating conditional GAN performance with different generator architectures, an ensemble model, and different MR scanners for MR-sCT conversion. *Phys Med Biol* 2020;65:105004. <https://doi.org/10.1088/1361-6560/ab857b>.
- [51] Cusumano D, Lenkowicz J, Votta C, Boldrini L, Placidi L, Catucci F, et al. A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases. *Radiother Oncol* 2020;153:205–12.