

Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review

M Boulanger, Jean-Claude Nunes, H Chourak, A Largent, S Tahri, O Acosta, R de Crevoisier, C Lafond, A Barateau

▶ **To cite this version:**

M Boulanger, Jean-Claude Nunes, H Chourak, A Largent, S Tahri, et al.. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. Physica Medica, 2021, 89, pp.265-281. 10.1016/j.ejmp.2021.07.027 . hal-03336875

HAL Id: hal-03336875
https://hal.science/hal-03336875

Submitted on 30 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Deep learning methods to generate synthetic CT from MRI in radiotherapy: a literature review

M. Boulanger[1], J.-C. Nunes[1], H. Chourak[1,3], A. Largent[2], S. Tahri[1], O. Acosta[1], R. De Crevoisier[1], C. Lafond[1], A. Barateau[1]

1. Univ. Rennes 1, CLCC Eugène Marquis, INSERM, LTSI - UMR 1099, F-35000 Rennes, France
2. Developing Brain Institute, Department of Diagnostic Imaging and Radiology, Children's National Hospital, Washington, DC, United States
3. CSIRO Australian e-Health Research Centre, Herston, Queensland, Australia

Running title: Deep learning sCT generation from MRI

ACCEPTED MANUSCRIPT - CLEAN COPY

Highlights:

- Review of deep learning approaches to generate synthetic-CTs for MRI-based dose calculation.
- Overview and discussion of image and dose metrics for synthetic-CT evaluation.
- Review of synthetic-CT image and dose accuracy per anatomical localization.

ACCEPTED MANUSCRIPT - CLEAN COPY

Abstract

Purpose: In radiotherapy, MRI is used for target volume and organs-at-risk delineation for its superior soft-tissue contrast as compared to CT imaging. However, MRI does not provide the electron density of tissue necessary for dose calculation. Several methods of synthetic-CT (sCT) generation from MRI data have been developed for radiotherapy dose calculation. This work reviewed deep learning (DL) sCT generation methods and their associated image and dose evaluation, in the context of MRI-based dose calculation.

Methods: We searched the PubMed and ScienceDirect electronic databases from January 2010 to March 2021. For each paper, several items were screened and compile in figures and tables.

Results: This review included 57 studies. The DL methods were either generator-only based (45% of the reviewed studies), or generative adversarial network (GAN) architecture and its variants (55% of the reviewed studies). The brain and pelvis were the most commonly investigated anatomical localizations (39% and 29% of the reviewed studies, respectively), and more rarely, the head-and-neck (H&N) (15%), abdomen (10%), liver (5%) or breast (3%). All the studies performed an image evaluation of sCTs with a diversity of metrics, with only 36 studies performing dosimetric evaluations of sCT.

Conclusions: The median mean absolute errors were around 76 HU for the brain and H&N sCTs and 40 HU for the pelvis sCTs. For the brain, the mean dose difference between the sCT and the reference CT was <2%. For the H&N and pelvis, the mean dose difference was below 1% in most of the studies. Recent GAN architectures have advantages compared to generator-only, but no superiority was found in term of image or dose sCT uncertainties. Key challenges of DL sCT generation methods from MRI in radiotherapy is the management of movement for abdominal and thoracic localizations, the standardization of sCT evaluation, and the investigation of multicenter impacts.

Keywords: Deep learning, MRI, synthetic-CT, radiation therapy, dose calculation

ACCEPTED MANUSCRIPT - CLEAN COPY

Introduction

In radiation therapy, computed tomography (CT) is the standard imaging modality for treatment planning. Magnetic resonance imaging (MRI) is a complementary modality to CT providing better soft-tissue contrast without irradiation. MRI improves the delineation accuracy of the target volume and/or organs at risk (OARs) in the brain, head-and-neck (H&N), and lung or prostate radiotherapy [1–3]. However, MRI does not provide information on the electron density of the tissue, require for accurate dose calculation. Most of the literature has proposed the generation of synthetic-CT (sCT) images for MRI-based dose planning. sCT (or pseudo-CT) is a synthetic image in Hounsfield Units (HU) generated from MRI data.

The methods for generating sCTs can be divided into three categories: bulk density, atlas-based and machine learning (ML) methods (including classical ML methods and deep learning methods [DLMs]). The bulk density methods consist of segmenting MRI images into several classes (usually air, soft-tissue, and bone). Each of these delineated volumes is assigned a homogeneous electron density, and the dose can then be calculated. This method has several drawbacks: it is tedious, time-consuming, operator-dependent, and does not consider tissue heterogeneity [4–8]. The atlas-based methods involve complex, non-rigid registrations of one or several co-registered MRI-CT atlases with a target MRI. This registration step is followed by a fusion step to generate the sCT. The drawbacks of this method are the lack of robustness in the case of large anatomical variations and the need for computationally intensive pairwise registrations [4,5,9,10]. Among the classical ML methods, the patch-based methods (such as [4]) can be decomposed into four steps. The first step is interpatient rigid and affine registration with MR images. These methods involve inter-patient registration, feature extraction, and patch partitioning during the training step. The training patches closest to the patches of the target MRI are then selected for aggregation to generate the sCT [4]. The main drawbacks of this method are the imprecise interpatient registration and calculation time.

DLMs are models comprising multiple processing layers that learn multiscale representations of data through multiple levels of abstraction [11]. These methods have recently been introduced in radiotherapy for applications, including image segmentation, image processing and reconstruction, image registration, treatment planning, and radiomics [12–19]. DLMs have been proposed for sCT generation from MRI. They were trained to model the relationships between HU CT values and MRI intensities. Once the optimal DL parameters are estimated, the model can be applied to a test MRI to generate its corresponding sCT. DLMs have the advantage of being fast for sCT generation, and some do not require deformable inter-patient registration (only intra-patient registration) such as in [20].

Two reviews, both published in 2018, have already summarized sCT generation methods from MRI [21,22], they focused only on the bulk density, atlas-based, and voxel methods and did not include

recent DLMs. Other studies have listed sCT generation methods from MRI in the context of MR-only radiotherapy [2,23–25]. More recently, Wang et al. [26] proposed a review on medical imaging synthesis using DL and Spadea and Maspero et al. [27] a review on sCT generation with DLM from MR, CBCT and PET images.

This study aimed to review literature studies using DLMs for MRI-based dose calculation in radiation therapy. This paper reviews the DL networks (with the loss functions), the image and dose endpoints for evaluation and the results per anatomical localization.

Materials and methods

We searched the PubMed and ScienceDirect electronic databases from January 2010 to March 2021 (date of first online release) using the following keywords: “deep learning”, “substitute CT” or “pseudo CT” or “computed tomography substitute” or “synthetic CT”, “MRI” or “MR” or “magnetic resonance imaging”, “radiation therapy” or “radiotherapy”. Mesh terms used in PubMed were: “radiotherapy”, “Magnetic Resonance Imaging”, and “deep learning”. The search string on PubMed was: "MRI" AND "radiotherapy" AND ("GAN" OR "CNN" OR "deep learning" OR "machine learning" OR "U-Net" OR "neural network") NOT "radiomics" NOT "chemotherapy" NOT "brachytherapy" NOT "Positron Emission Tomography Computed Tomography" NOT "chemoradiotherapy" NOT "segmentation" NOT "reconstruction". We only retained original research papers (no abstract, no review paper) that reported data obtained from humans, were written in English, and addressed DL sCT generation from MRI in radiotherapy.

For each paper, we screened: anatomical localization, MR device, MR sequence, pre or post-treatment, use of registration, number of patients included in the study, type of DL network, loss functions, number of patients for training step, number of patients for evaluation step, main image and dose evaluation results. Tables per anatomical localization (brain, H&N, breast-liver-abdomen, and pelvis) were created to compile these information.

Results

Figure 1 summarizes the number of DLM studies for sCT generation from MRI in radiation therapy per year and anatomical localization. The first study was published in 2016 [28] and, at the time of manuscript submission, a total of 57 articles meeting the selection criteria had been published. Some studies investigated sCT generation for several anatomical localizations [29–33].

ACCEPTED MANUSCRIPT - CLEAN COPY

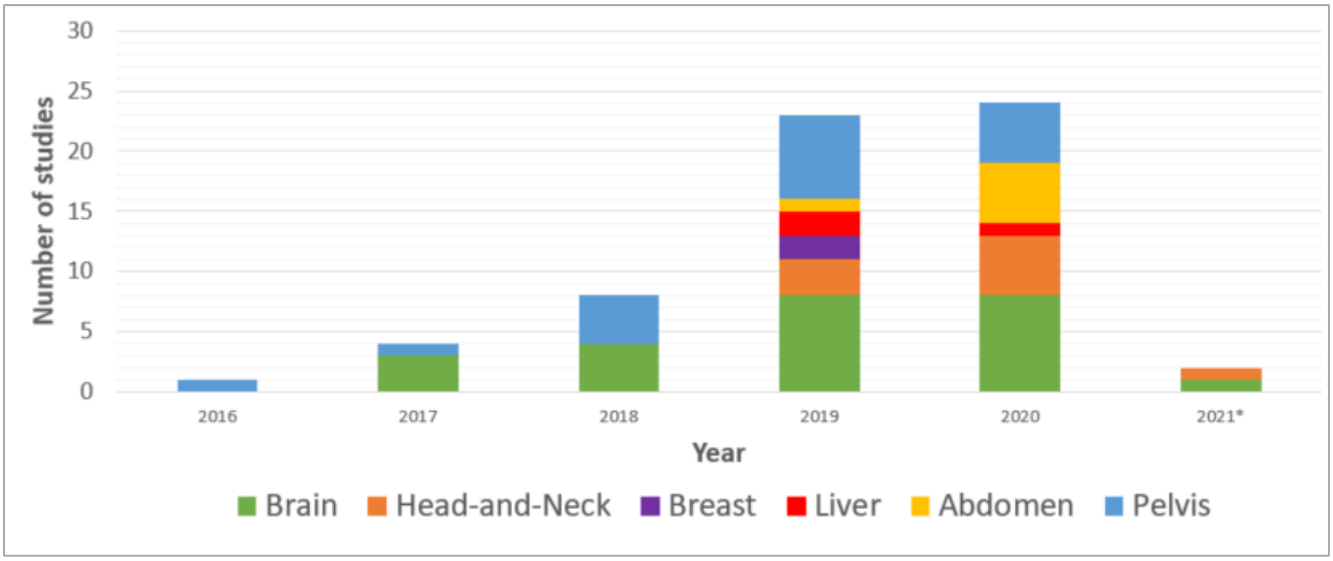


Figure 1: Numbers of publications on deep learning methods for synthetic-CT generation from MRI in radiation therapy per year and anatomical localization.
**: ongoing year, number of studies at the time of publication*

In total, 24 studies were based on brain data, 9 on H&N data, 2 on breast data, 3 on liver data, 6 on abdomen data, and 18 on pelvic data.

A. Common deep learning networks for sCT generation from MRI

Deep learning, as a mainstream of machine learning method, uses trainable computational models containing multiple processing components with adjustable parameters to learn a representation of data. Many DL network architectures have been developed, depending on specific applications or learning data. Several reviews have detailed the DL network architectures for radiotherapy or medical imaging [12,26,27,34–37]. The DL architecture for sCT generation from MRI can be roughly divided into two classes: generator-only and generative adversarial network (GAN) and its variants (such conditional-GAN, Least square GAN and cycle-GAN). Figure 2 shows the hierarchy of the DL architectures.

ACCEPTED MANUSCRIPT - CLEAN COPY

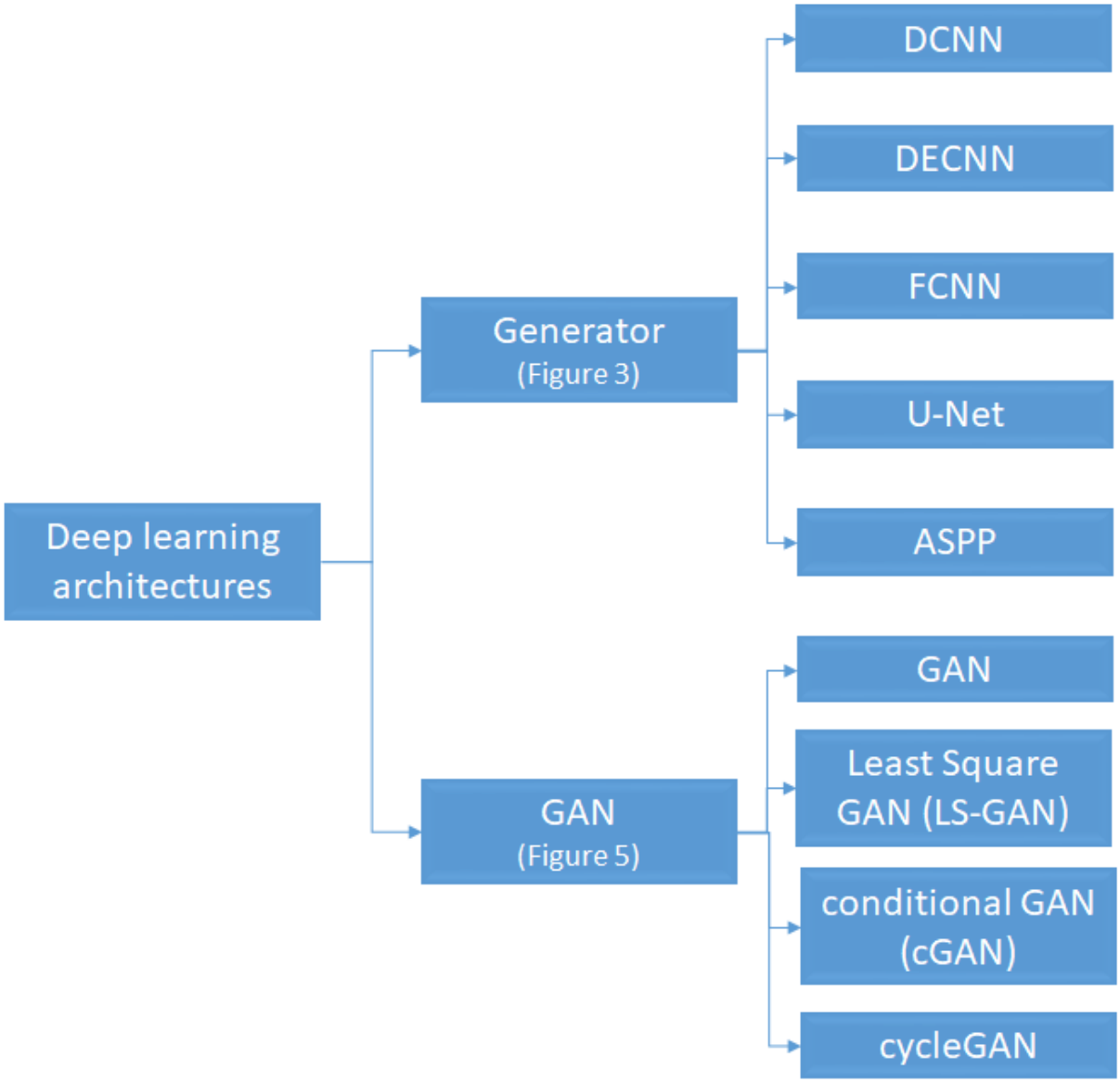


Figure 2: Hierarchy of deep learning architectures

Deep learning (DL) architectures can roughly be divided into generator-only and generative adversarial network (GAN). In generator-only different DL architectures are included such as deep convolutional neural network (DCNN), deep embedding CNN (DECNN), fully CNN (FCNN), U-Net, or atrous spatial pyramid pooling (ASPP). GAN family includes GAN, and its most popular variants: Least Square GAN (LS-GAN) conditional GAN (cGAN), and cycle-GAN.

1. Generator-only models

i. Basic concepts of convolutional neural networks (CNN)

For image applications, a convolutional neural network (CNN, or ConvNet) is a popular class of deep neural networks using a set of convolution kernels/filters for detecting image features. A CNN consists

of an input layer, multiple hidden layers and an output layer. The hidden layers include layers that perform convolutions with trainable kernels. Nonlinear activation functions (Rectified Linear Units (ReLU)[38], Leaky-RELU [39], Parametric-ReLU (PreLU) or exponential linear unit (ELU) [40]) play a crucial role in discriminative capabilities of the deep neural networks. The ReLU layer preserves the input otherwise is the most commonly used activation layer due to its computational simplicity, representational sparsity, and linearity. It is commonly to periodically insert a pooling layer between successive convolutional layers in a CNN architecture. Pooling layers allow to reduce the dimension (subsampling) of the feature maps. These maps are generated by following the convolutional operations. The pooling methods performs down-sampling by dividing the input into rectangular pooling regions and computing the average, the maximum, or the minimum of each region represented by the filter (mean pooling, max-pooling, min-pooling). Batch normalization [41] layers are inserted after a convolutional or fully connected layer to improve the convergence of the loss function during gradient descent (optimizer). It prevents the problem of vanishing gradient from arising and significantly reduces the time required for network convergence. After several convolution and pooling layers, the CNN generally ends with several fully connected layers. Dropout is one of the most promising techniques for regularization of CNN. Softmax layer is typically the final output layer in a neural network that performs multi-class classification (for example: object recognition).

ii. Generator-only models

The generator model can be considered as representing a complex end-to-end mapping function that transforms an input MR image to its corresponding CT image. During the training phase, the generator tries to minimize an objective function called a loss function (voxel-wise loss function L_G), which is an intensity-based similarity measurement between the generated image (sCT) and the corresponding ground truth image (real CT). Figure 3 presents the global architecture of generator-only model.

ACCEPTED MANUSCRIPT - CLEAN COPY

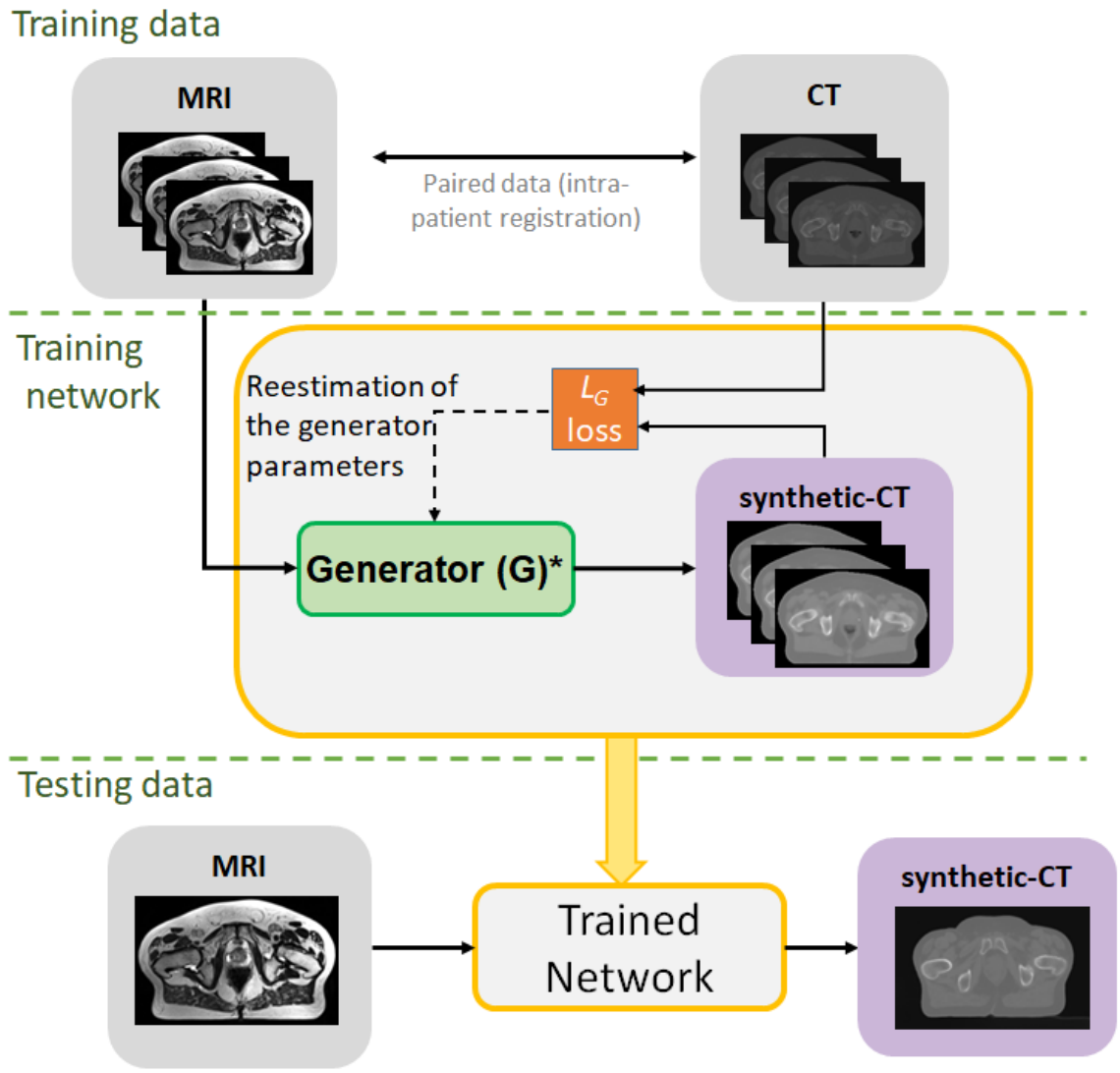


Figure 3: Illustration of generator-only model

**: Generator model varies according to networks.*

The generator models often based on convolution encoder-decoder networks (CED) are trained to produce synthetic CTs (sCTs) from MRI. For this purpose, a single loss function L_G between MRI and registered CT images is computed. In the testing step, for a new given test patient, the MRI goes through the trained network to obtain the corresponding sCT.

In sCT generation from MRI, the generator architectures are generally based on convolution encoder-decoder networks (CED). In the literature, the variants of generator model include deep CED network [42], deep embedding CNN (DECNN) or Embedded Net [30], fully convolutional network (FCN) [28], U-Net [20,42–57] [56,58,59], efficient CNN (eCNN) model [60], ResNet [61], SE-ResNet [61,62], and DenseNet [63]. Figure 4 presents some architectures of CED-based generators.

ACCEPTED MANUSCRIPT - CLEAN COPY

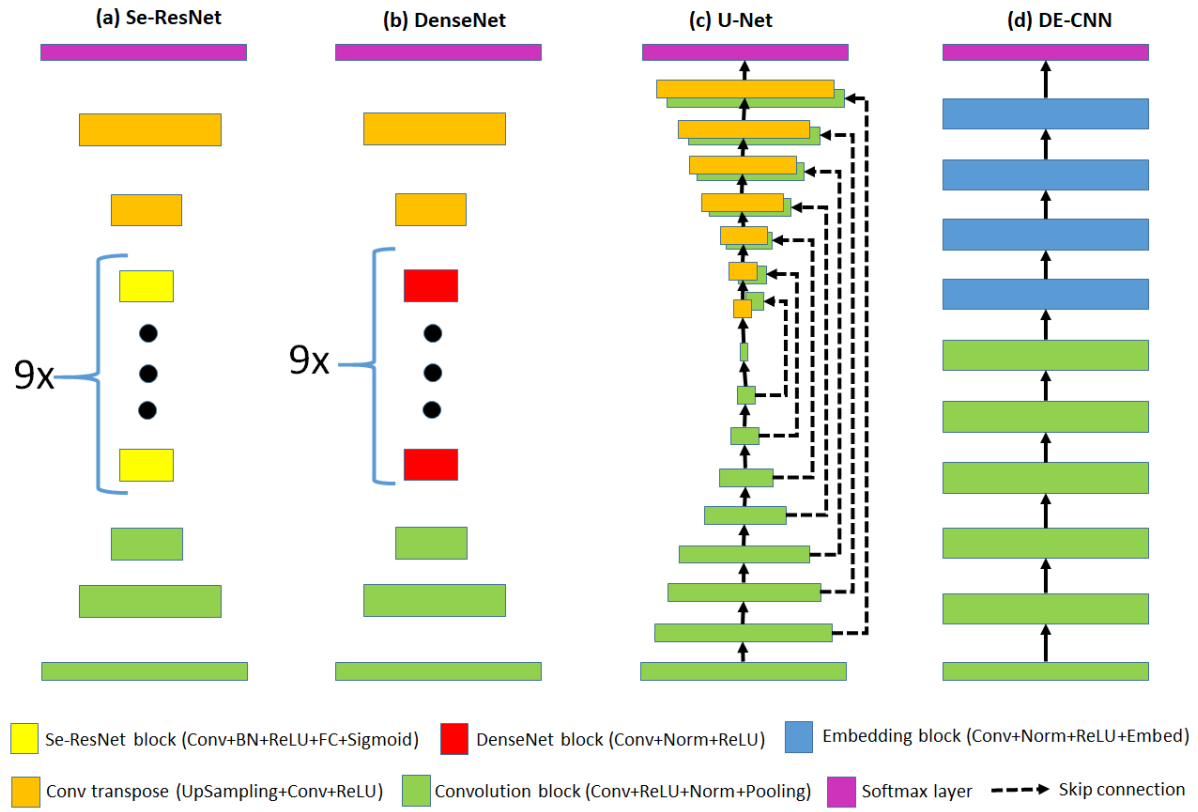


Figure 4: Representation of generator architecture for U-Net [64] **and adapted implementations of DenseNet** [63]**, SE-ResNet** [65]**, and Embedded Net** [30]**.**
The size of the boxes indicates the relative resolutions of the feature maps. The green boxes represent convolutional layers and orange boxes represent transposed convolutional layers. Yellow, red and blue boxes represent the SE-ResNet, DenseNet, and Embedded blocks.

The CED network consists of a paired encoder and decoder networks. CED have been extensively used in DL literatures thanks its excellent performance. In the encoding part, low-level feature maps are down-sampled to high-level feature maps. In the decoding part, the high-level feature maps are upsampled to low-level feature maps using the transposed convolutional layer to construct the prediction image (sCT).

The encoder network uses a set of combined 2D convolution filtering (no dilated convolutions) for detecting image features, followed by normalization (instance [66] or batch normalization [41]), a nonlinear activation function (ReLU [38], LeakyRELU [39], or PreLU), and max-pooling.

The decoder path combines the feature and spatial information through a sequence of symmetrical transpose convolutional layers (up-convolutions), up-sampling operators, concatenate layer (concatenations with high-resolution features), and convolutional layers with a ReLU activation function.

The most well-known and popular CED variants for biomedical image applications is the U-shaped CNN (U-Net) architecture proposed by Ronneberger et al. [67]. The U-Net [67] has a CED structure with

ACCEPTED MANUSCRIPT - CLEAN COPY

direct skip connections between the encoder and decoder. Han et al. were the first to publish a sCT study with a U-Net architecture [44] that is similar to Ronneberger's model. This 2D U-net model directly learns a mapping function to convert a 2D MR grayscale image to its corresponding 2D sCT image. Han et al. study [44] differs from the original U-net since the three fully connected layers were removed. Thus, the number of parameters is reduced by 90%, and the final model is easier to train. In Wang et al. [46], the U-net model used batch normalization [41] and leaky ReLU, which was different from the classical U-net [67].

The DECNN model proposed by Xiang et al. [30] is derived by inserting multiple embedding blocks into the U-net architecture. This embedding strategy helps to backpropagate the gradients in the CNN and also provides easier and more effective training of the end-to-end mapping from MR to CT with faster convergence.

The efficient CNN (eCNN) model [60] was built based on the encoder-decoder networks in the U-Net model [67] where the convolutional layers were replaced with the building structures (aiming at extracting image features from the input MRI).

Some generative models use dilated convolutions called "atrous convolution" (rather than conventional convolutions) that expands the receptive field without loss of resolution or coverage [68]. Wolterink et al. [68] used a dilated CNN capturing larger anatomical context to differentiate between tissues with similar intensities on MR.

The ResNet architecture [61] has three convolutional layers (containing convolution operations, a batch normalization layer, a ReLU) activation function, followed by nine residual blocks (containing convolutional layers, batch normalization layers, and ReLU activation function) with fully connected layers. HighRes-net [69] consists of a CED architecture with residual connections, normalization layers, and rectified linear unit (ReLU) activations [38] using high-resolution ground truth (no pooling layers) as supervision with few trainable parameters [43]. The atrous spatial pyramid pooling (ASPP) generator [56] employs atrous or dilated convolution and is implemented in a similar U-Net architecture. The ASPP module permits a reduction in the total number of trainable parameters (almost divided by 4).

FCN better preserves the neighborhood information in the generated sCT images [28]. Compared to the conventional CNN, the pooling layers are not used in this task of image-to-image translation [28]. FCNs can simplify and speed network learning and inference and make the learning problem much easier. However, Fully connected layers are incredibly computationally expensive.

The deep CED network [42] consists of a combined encoder network (the popular Visual Geometry Group [VGG] 16-layer net model) and a decoder network (reversed VGG16) with multiple symmetrical shortcut connections between layers.

ACCEPTED MANUSCRIPT - CLEAN COPY

Twenty-nine state-of-the-art sCT image generation methods have adopted a generator-only network [20,28,30,42–57,70–79]. The loss functions L_G evaluating sCT and real CTs used in these generative models are:

- the mean square error (MSE), the L2-norm, or the Euclidean norm: only for sCT [20,42,46,47,55,57,78], for sCT and embedding blocks [30],
- the MAE, mean absolute deviation (MAD), or L1-norm [43–45,49,52,53,70,71],
- a combined MAE and MSE loss [48],
- perceptual loss [20] based on VGG (the output of the 7th VGG16 convolutional layer).

The use of L2 distance as a loss function tends to produce blurry results. Perceptual loss is used to capture the discrepancy between the high frequency components within an image.
One limitation of generative models based on CNN is that they may lead to blurry results due to generally misalignment between MR and CT [80].

2. Generative adversarial network (GAN)

The following section summarizes GAN-based architectures to generate sCT from MRI. We introduce the GAN architecture and three most popularGAN-based extensions: least squares-GAN, conditional-GAN, and cycle-GAN.

i) GAN

The adversarial learning strategy was proposed by Goodfellow et al. [81] to generate better sCT images than previous generator-only models. The original way is to simultaneously train two separate neural networks, the generator G (one of the generator-only models described in i) and figure 4) and the discriminator D. These two neural networks form a two-player min-max game where G tries to produce realistic images to fool D while D tries to distinguish between real and synthetic data [82,83]. Compared to generator-only models, GAN introduces a data-driven regularizer, the adversarial loss, to ensure that the learned distribution approaches the ground truth.

ACCEPTED MANUSCRIPT - CLEAN COPY

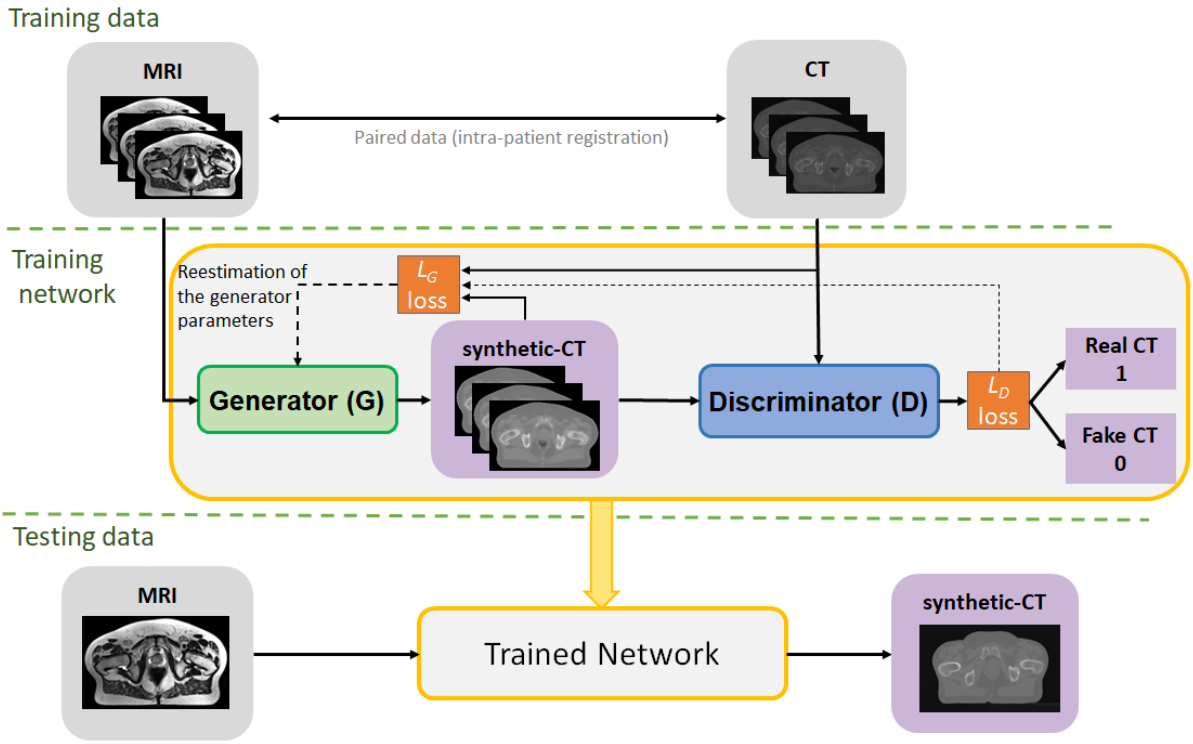


Figure 5: Generative adversarial network (GAN) architecture

GAN consists of two adversarial CNNs. The first CNN, called the generator (such illustrated in figure 3), trained to synthetize images that resemble real images (such as real CT). The second CNN, called discriminator trained to differentiate fake image (synthetic image) from real images (which is considered a binary classification problem). The loss function L_D of the discriminator (called adversarial loss) is generally the binary cross-entropy. G and D are trained alternatively and share the same objective function of adversarial loss. The overall loss function L_G combined the adversarial loss and a voxel-wise loss function (measuring the similarity between the real CT and synthetic-CT voxels).

In the original version [83], the discriminator and generator are implemented as multilayer perceptrons (MLPs) and more recently implemented as CNNs. The architecture of the generator is often the conventional U-Net. Another proposed generator architecture in a GAN is ResNet [62] which is easy to optimize and can gain accuracy from considerably increased depth. The discriminator of the GAN [83] consists of six convolutional layers with different filter sizes but the same kernel sizes and strides, followed by five fully connected layers. ReLU was used as the activation function and a batch normalization layer for the convolutional layers. The dropout layer was added to the fully connected layers, and a sigmoid activation function was used in the last fully connected layer.

The discriminator used in [64] a convolutional "PatchGAN" classifier (markovian discriminator) models high frequency image structure in local patches and only penalizes structure at the scale of image patches.

ACCEPTED MANUSCRIPT - CLEAN COPY

Using adversarial loss, the classical GAN model can generate high-quality sCT images with less blurry results [29,80] than generator-only models. The discriminator tries to maximize it while the generator tries to minimize it.

In this review, six studies used classical GAN-based architectures to generate sCT from MRI [20,29,62,77,84,85]. The adversarial loss functions of the generator evaluating sCT and the original CTs used in these GANs are :

- L2-norm alone [20,85],
- perceptual loss [20,84] and the multiscale perceptual loss [20].

The adversarial loss function L_D of the discriminator used in these GANs was generally the binary cross-entropy [29].

Perceptual regularization, used by Largent et al. [20], helps to prevent images over-smoothing and loss of structure details. The perceptual loss functions are based on high-level features extracted from pre-trained VGG network (7th VGG16 in [20]).

As shown by several studies [29,62,86], (1) the adversarial network prevents the generated images from blurring and better preserve details, especially edge features; (2) the accuracy of sCT within the bone region is increased; and (3) the discriminator detects patch features in both real and fake images, mitigating misregistration problem caused by an imperfect alignment between multi-parametric MRI and CT. General convergence in GANs is heavily dependent on hyperparameter tuning to avoid vanishing [87] or exploding gradients, and they are prone to mode collapse. To tackle the training instability of GANs, a plethora of extensions and subclasses have been proposed.

ii) Least Squares-GAN (LS-GAN)

Most GANs use the binary cross-entropy as the discriminator loss function. However, this cross-entropy loss function leads to the saturation problem in GANs learning (the well-known problem of vanishing gradients [87]). Least square loss function strongly penalized the fake samples away from decision boundary and improve the stability of learning process. Mao et al. [88] adopted the least-squares loss function for the discriminator and showed that minimizing the objective function of LS-GAN minimizes the Pearson χ^2 divergence [89]. Emami et al. [62] replaced the negative log-likelihood objective with a least square loss function (L2 loss), which was more stable during training and generated better sCT quality.

iii) Conditional-GAN (cGAN)

Since the original GAN allows no explicit control on the actual data generation, Goodfellow et al. [83] proposed the conditional GAN (cGAN) to incorporate additional information such as class labels in the synthesis process. cGAN is an extension of the GAN model in which both the generator and the

ACCEPTED MANUSCRIPT - CLEAN COPY

discriminator are conditioned on some additional information. The sCT image output is conditioned on the MR image input.

Different generator architectures in a cGAN have been proposed, including SE-ResNet [61,62], DenseNet [63], U-Net [56,58,59], Embedded Net [30], and the atrous spatial pyramid pooling (ASPP) method [56]. Fetty et al. [90] evaluated four different generator architectures: SE-ResNet, DenseNet, U-Net, and Embedded Net in a cGAN to generate sCT from T2 MRI. Olberg et al. [56] explored two generators: the conventional U-Net architecture implemented in the Pix2Pix framework [64] and the ASPP method [91,92]. The discriminator of the GAN framework was similar in both implementations.

Twenty studies used a cGAN architecture to generate sCT from MRI [31,33,50,56–59,89,90,93–102]. The overall loss functions L_G evaluating sCT and real CTs used in these cGANs were as follows:

- adversarial loss function (binary cross entropy) [59,102],
- L1-norm (MAE) [93],
- least squares loss function (L2 loss) [89,102],
- mutual information (MI) [58,59],
- focal regression loss [103] used in [100],
- the combination of adversarial (binary cross-entropy) and L2-norm [56],
- the combination of L1-norm and PatchGAN loss (as proposed by Isola et al. [64]) used in [50,90,94],
- the combination of adversarial (binary cross-entropy) and term derived from the log-likelihood of the Laplace distribution [96],
- the combination of L*p*-norm, adversarial and gradient [33],
- the combination of multiscale L1-norm, L1 norm and PatchGAN loss [64] used in [89].

The loss functions L_D of the discriminator evaluating sCT and real CTs used in these cGANs areas follows:

- the mostly used adversarial loss (binary cross entropy) [56,58,59,89,94],
- least squares loss function (L2 loss) [89,95,102],
- L1-norm [102].

The L2-based loss function of the generator can cause image blurring. To alleviate blurriness and improve the prediction accuracy, the L1 norm [46] makes the learning more robust to outliers in the training data, such as noise or other artifacts in the images or due to imperfect matching between MR and CT images. The Markovian Discriminator loss or Patch-GAN loss [64], which can be understood as

ACCEPTED MANUSCRIPT - CLEAN COPY

a form of texture/style loss, effectively models the image as a Markov random field, assuming independence between pixels separated by more than a patch diameter.

Pix2Pix proposed by Isola et al. [64] is a successful cGAN variant for high-resolution image-to-image translation. Pix2Pix model generally uses Unet generator and PatchGAN discriminator. As investigated by Isola et al. [64], the use of a loss function based on L1 alone leads to reasonable but blurred results; while cGAN alone leads to sharp results but introduces image artifacts. The authors showed that training in an adversarial setting together with an L1 norm generated sharp images with few artefacts (tissue-classification errors, especially for bone and air differentiation).

In Hemsley et al. [96], the L1 term in cGAN loss function [64] is replaced by a term derived from the log-likelihood of the Laplace distribution to capture data dependent uncertainty.

To overcome MR/CT registration issues, Kazemifar et al. [58,59] used a generator loss function based on the mutual information (MI) in cGAN. The MI loss allows the cGAN to use unregistered data to generate sCT and seems to accurately distinguish between air and bone regions.

Instead of the usual cross-entropy cGAN loss, Mao et al. [88] recommend the quadratic version of the least square GAN. Olberg et al. [56] evaluated a Pix2Pix framework with two different generators: the conventional U-net and a proposed generator composed of stacked encoders and decoders separated by dilated convolutions applied to increase rates in parallel to encode large-scale features. The overall loss function was composed of adversarial (sigmoid cross-entropy) and MAE losses.

Twelve studies used a Pix2Pix architecture [31,50,56,89,90,93–95,98,99,102,104]. Most of these Pix2Pix frameworks used only one MRI sequence as input and generated one sCT as output (called single-input single-output, SISO). A variant of Pix2Pix architecture proposed by Sharma et al. [104] is multi-input and multiple-output (MIMO) combining information from all available MRI sequences and synthesizes the missing ones.

One of the main advantages of cGANs is that the networks learn reasonable image-to-image translations even if the training dataset size is small. However, cGANs require coregistered MR-CT image pairs for training except with mutual information as loss function [58,59].

iv) Cycle-GAN

For image-to-image translations between two modalities, the principles of the cycle-GAN are to extract characteristic features of both modalities and discover the underlying relationship between them [105]. The cycle-GAN involved two GANs: one to generate sCT from MRI and a second to generate p-MRI from sCT (the output of the first GAN). These dual GANs learn simultaneously and a cyclic loss function minimizes the discrepancy between the original CT and the sCT obtained from the chained generators.

ACCEPTED MANUSCRIPT - CLEAN COPY

Cycle GAN-based framework does not require paired MRI/CT images [80,106]. Wolterink et al. [80] found that training using unpaired images could, in some cases, outperform a GAN-model on paired images.

Eleven studies used a cycle-GAN architecture to generate sCT from MRI [32,33,57,77,80,101,102,106–109]. The overall loss functions L_G comparing the generated sCT and real CTs used in these cycle-GANs were:

- the combination of adversarial loss (cross-entropy) and L1-norm [33,102],
- the combination of the adversarial loss based on cross-entropy, the cycle consistency loss based on L1-norm, and the structural consistency loss based on L1-MIND [106] (the modality-independent neighborhood descriptor, MIND, introduced in [110]),
- the combination of L2-norm, adversarial loss (binary cross-entropy), the gradient difference loss and cycle consistency loss (based on L1 norm) [80],
- the combination of L*p*-norm (mean P distance, MPD), adversarial loss and gradient loss [32,33,77,108].

Loss functions L_D of the discriminator used in cycle-GAN are:

- L2-norm (least squares loss) [62] as proposed in [88,111],
- MAD (L1-norm) [32,77,102,108],
- L*p*-norm (MPD) [109].

Since L2-based loss functions tend to generate blurry images and L1-based loss functions may introduce tissue-classification errors, some authors [32,33,77,108,109] used an l*p*-norm (p = 1.5) distance, the MPD (Mean P distance). Using the MPD-based loss term, the authors also integrated an image gradient difference (GD) loss term (proposed in [29]) into the loss function [32,33,77,108,109], to retain sharpness in synthetic images, which maintain zones with strong gradients, such as edges. Cycle-GAN-based methods use MSE loss as distance loss function, which often leads to blurring and over-smoothing.

ACCEPTED MANUSCRIPT - CLEAN COPY

B. Data for sCT generation from MRI

1. MRI/CT image preprocessing and post-processing

In eighteen studies an MRI bias correction [20,30,32,33,43,44,47,49,78,84,90,93,95,102,106–109] was reported. In [30,32,44,47], intensity inhomogeneity (or non-uniformity) correction was performed in all MR images using the N3 bias field correction algorithm [112,113] to correct the bias field before training or synthesis. In [33,43,78,84,90,93,95,106–109], the authors reported that the intensity inhomogeneity of the MRI was corrected using the N4 bias field correction algorithm.

A 2D or 3D MRI geometry correction provided by the vendor was sometimes reported [49,57,70,106]. We can think that most of MR images had a geometry correction, but that it was not mentioned.

In [30,33,78,84,95], all MR images were normalized using a histogram-based intensity normalization [114] to minimize the inter-patient MR intensity variation. Intensity normalization was also used in [30,32]. In [44], all MR images were then histogram-matched to a randomly chosen template to help standardize image intensities across different patients using the method described by Cox et al. [115]. All MR volumes were normalized by aligning the white matter peak identified by fuzzy C-means in [106]. In [49,102], histogram standardizations performed using vendor-provided software (CLEAR) were applied as provided by the vendor.

In the study by Maspero et al. [94], the voxel intensity of CT was clipped within the interval HU to avoid an excessively large discretization step and the MR images were normalized to their 95% intensity interval over the whole patient. All the images were converted to 8-bits to conform to the Pix2Pix implementation [64]. Before training, the air cavities were filled in CT images and bulk-assigned (−1000 HU) as located in MR images using an automated method.

2. Training data characteristics

Compared to 2D CNN, 3D CNN can better model 3D spatial information (neighborhood information) owing to the use of 3D convolution operations [28] solving the discontinuity problem across slices, which are suffered by 2D CNN. However, the input type to DL models is mainly in 2D because fully 3D networks are much more difficult to train due to a large numbers of trainable parameters and requires exponentially more (GPU) memory and more data [28,44]. With the 2.5 D approach, Dinkla et al. [70] added 3D contextual information while maintaining a manageable number of trainable parameters. Furthermore, discontinuities across slices present in 2D methods, were decreased. Besides, the 2.5D approaches [45,70,71] include average axial, sagittal, and coronal images as input to train the CNN. In 3D (patch-based) CNN [28,32], an input MR image is first partitioned into overlapping patches. For

ACCEPTED MANUSCRIPT - CLEAN COPY

each patch, the CNN is used to predict the corresponding CT patch and all predicted CT patches are merged into a single CT image by averaging the intensities of overlapping CT regions.

Most of the reviewed studies used one MRI sequence as input and generated one sCT as output; an architecture generally called single-input single-output (SISO). Four studies used several MRI sequences as input to generate one sCT in output [50,72,93,104], these architectures referred to as multi-input single-output (MISO) [50,72,93,104] or multi-input multiple-output (MIMO) [104]. Moreover, most studies used training and evaluation data from one MRI device while eight studies used multi-device MRI. One study reported use of MRI data from different centers [97] and two studies [89,90] used data from the Gold Atlas Data set [116]. Five studies used low MR field (0.35 T) as input images [31,33,56,73,90].

3. Training and evaluation of data size

The studies included in this review used several training strategies including k-fold cross-validation, single-fold validation, or leave-one-out. In k-fold cross-validation, the dataset is divided into *k* subsets, and the holdout method is repeated *k* times. Each time, one of the *k* subsets is used as the test set and the other *k-1* subsets are combined to form a training set. The average error across all *k* trials is then computed. In single-fold validation, the dataset is separated into two sets, the training and testing sets. The leave-one out strategy consists on k-fold cross-validation taken to its logical extreme, with k equal to N, the number of data patients in the set.

Data size is a fundamental challenge for DL approaches. There is no reported minimal or optimal data size for DL training. In the head area, four studies assessed sCT image quality (with MAE) as a function of the number of available images for training, from 15 to 242 patients for Alvares Andres et al. [43], from 5 to 47 patients for Gupta [48], from 34 to 135 patients for Peng et al. [101], and from 1 to 40 patients for Maspero et al. [97]. Better image results were found for higher numbers of available images. A minimum of 10 patients seems to be needed since it has shown similar performance than a training of 20, 30 or 40 patients. One effective way to improve model robustness is to enhance the diversity of the training dataset. Data augmentation is essential to teach the network the desired invariance and robustness properties when only a few training samples are available. One common augmentation technique [32,44,93] is to apply random translations, rotations, zooms, and elastic deformations and adding low-level random noise to training images.

ACCEPTED MANUSCRIPT - CLEAN COPY

C. Evaluation metrics

sCT evaluation can be performed in terms of intensity, geometric fidelity, or dose metrics. A sCT evaluation was performed using intensity-based metrics for all reviewed studies and through dose criteria in 63% of the reviewed studies. The metrics used in the reviewed studies are listed in Table 1.

ACCEPTED MANUSCRIPT - CLEAN COPY

	Type of metrics	Metric	Definition	Ideal value
Image evaluation	Intensity-based metrics	ME: mean error	$ME = \frac{1}{N}\sum_{i=1}^{N} pCT_i - CT_i$	0 HU
		MAE: mean absolute error	$MAE = \frac{1}{N}\sum_{i=1}^{N} \lvert pCT_i - CT_i \rvert$	0 HU
		PSNR: peak signal to noise ratio	$PSNR = 10log_{10}\left(\frac{Q^2}{MSE}\right)$	Maximum of dB
		SSIM: structural similarity metric	$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\delta_{xy} + C_2)}{(\mu\lvert x^2 + \mu\rvert y^2 + C_1)(\delta\lvert x^2 + \delta\rvert y^2 + C_2)}$	1
		MSE: mean square error	$MSE = \frac{1}{N}\sum_{i=1}^{N} (pCT_i - CT_i)^2$	0
		RMSE: rot mean square error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (pCT_i - CT_i)^2}$	0 HU
		NCC: normalized cross-correlation	$NCC = \frac{1}{N}\sum_{x,y,z} \frac{(I_{CT}(x,y,z) - \mu_{CT})(I_{pCT}(x,y,z) - \mu_{pCT})}{\sigma_{CT}\sigma_{pCT}}$	
	Geometric fidelity metrics	DSC: dice score coefficient	$DSC = \frac{2(V_{CT} \cap V_{pCT})}{V_{CT} + V_{pCT}}$	1
		HD: Hausdorff distance	$H(pCT, CT_{ref}) = max\left(h(pCT, CT_{ref}), h(CT_{ref}, pCT)\right)$	0 mm
		MASD: mean absolute surface distance	$MASD(A,R) = \frac{d_{ave}(S_A, S_R) + d_{ave}(S_R, S_A)}{2}$	0 mm
Dose evaluation	Dose difference metrics	Voxel-to-voxel dose differences	Difference between the dose distribution computed on the reference CT and on the sCT	0 Gy or 0%
		DVH difference	Dose differences on DVH specific points (D_{max}, D_{70Gy}, etc.), for a given structure	0 Gy or 0%
	Gamma analysis metrics	Mean gamma	Value of the mean gamma	0
		Gamma pass-rate	Percentage of pixels/voxels with a gamma value lower than 1	100%

Table 1: Imaging and dose metrics used for the evaluation of synthetic-CT generation from MRI

ACCEPTED MANUSCRIPT - CLEAN COPY

Abbreviations: N: number of voxels; MSE: Mean square error; Q: range of voxel value of sCT and reference CT; x: reference CT; y: sCT; μ_x: mean value of x; μ_y: mean value of y; δ_x^2: variance of x; δ_y^2; variance of y; C_1 and C_2 are expressed as $(k_1Q)^2$ and $(k_2Q)^2$; I_{CT}: HU value of the reference CT, I_{sCT}: HU value of the sCT, μ_{CT}: mean intensity value of the reference CT, μ_{sCT}: mean intensity value of the sCT, σ_{CT} and σ_{sCT}: standard deviation of the reference CT and sCT; V: volume on CT and sCT; d_{ave}: absolute Euclidean distance; S_A: surface of the automated segmentation volume; S_R: surface of the reference organ delineation.

ACCEPTED MANUSCRIPT - CLEAN COPY

1. Intensity-based evaluation

Only three studies of sCT generation from MRI did not reported MAE [56,73,99]. Some articles reported MAE in bone or soft tissue while others reported MAE in anatomical structures such as the kidneys, bladder, or rectum [20,45,47,76,89]. Figure 6 summarizes the MAEs of the studies on brain, H&N, liver, abdomen, and pelvis sCT generation from MRI.

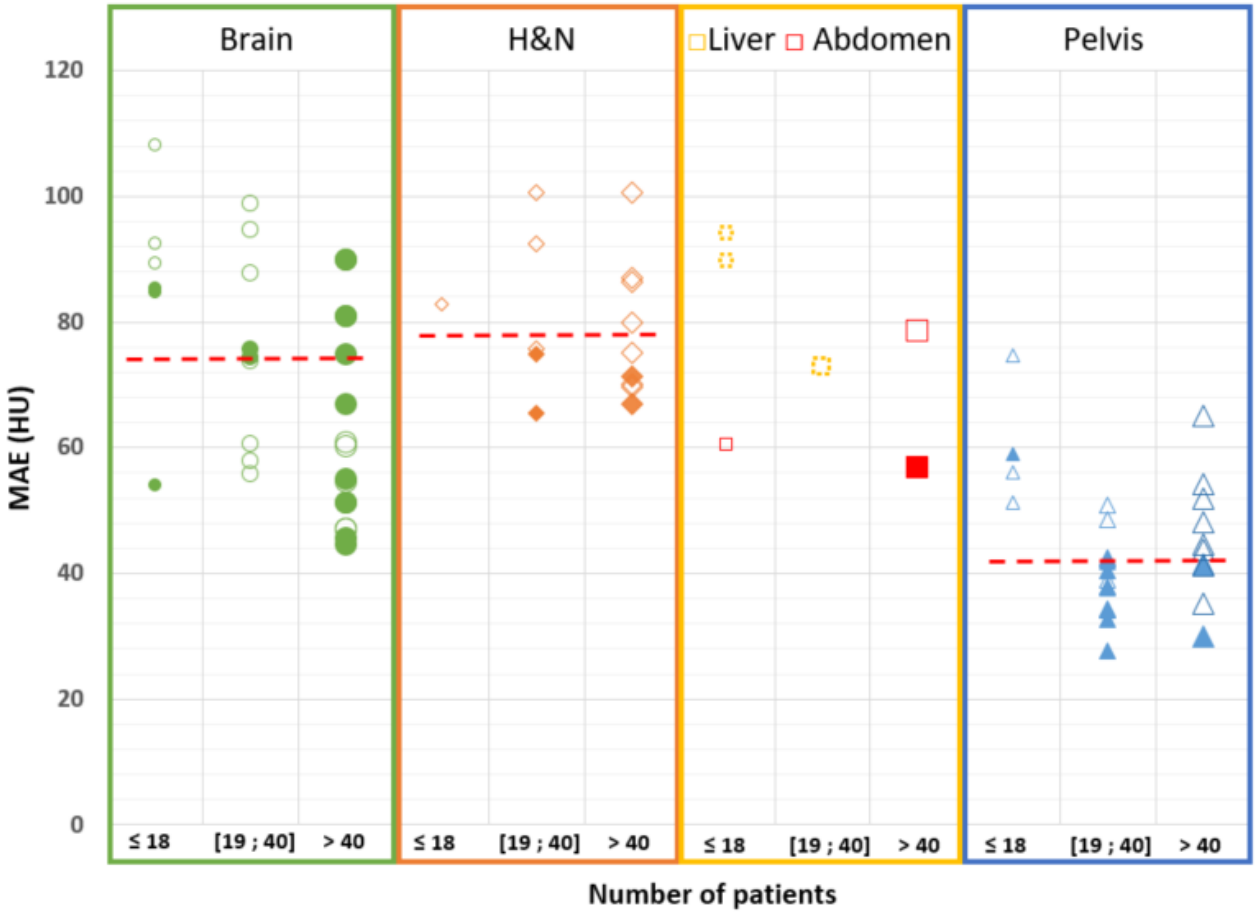


Figure 6: Mean absolute error (MAE) results for body structure between reference CT and sCT generated with a deep learning method for studies including the brain, H&N, liver, abdomen, and pelvis

Each marker represent a study result. Full markers represent generator-only models and empty markers generative models with adversarial. Results are divided into three categories: studies including less than 18 patients, studies including 19 to 40 patients and studies including more than 40 patients.
Red dotted lines represent the median values. The median values are: 74.2 HU for the brain, 77.9 HU for H&N, and 42.4 HU for the pelvis.
The selected values are listed in the Additional tables 1 to 4.

Eighteen studies reported ME results. Figure 7 details the MEs of the studies on brain, H&N, abdomen, and pelvis sCT generation from MRI. For the pelvis, three studies provided ME values for the bladder, rectum and soft tissue [20,47,89]. Some studies have illustrated MAE or ME for one or several slices.

ACCEPTED MANUSCRIPT - CLEAN COPY

Such difference maps allow for qualitative comparisons and spatial analyses.

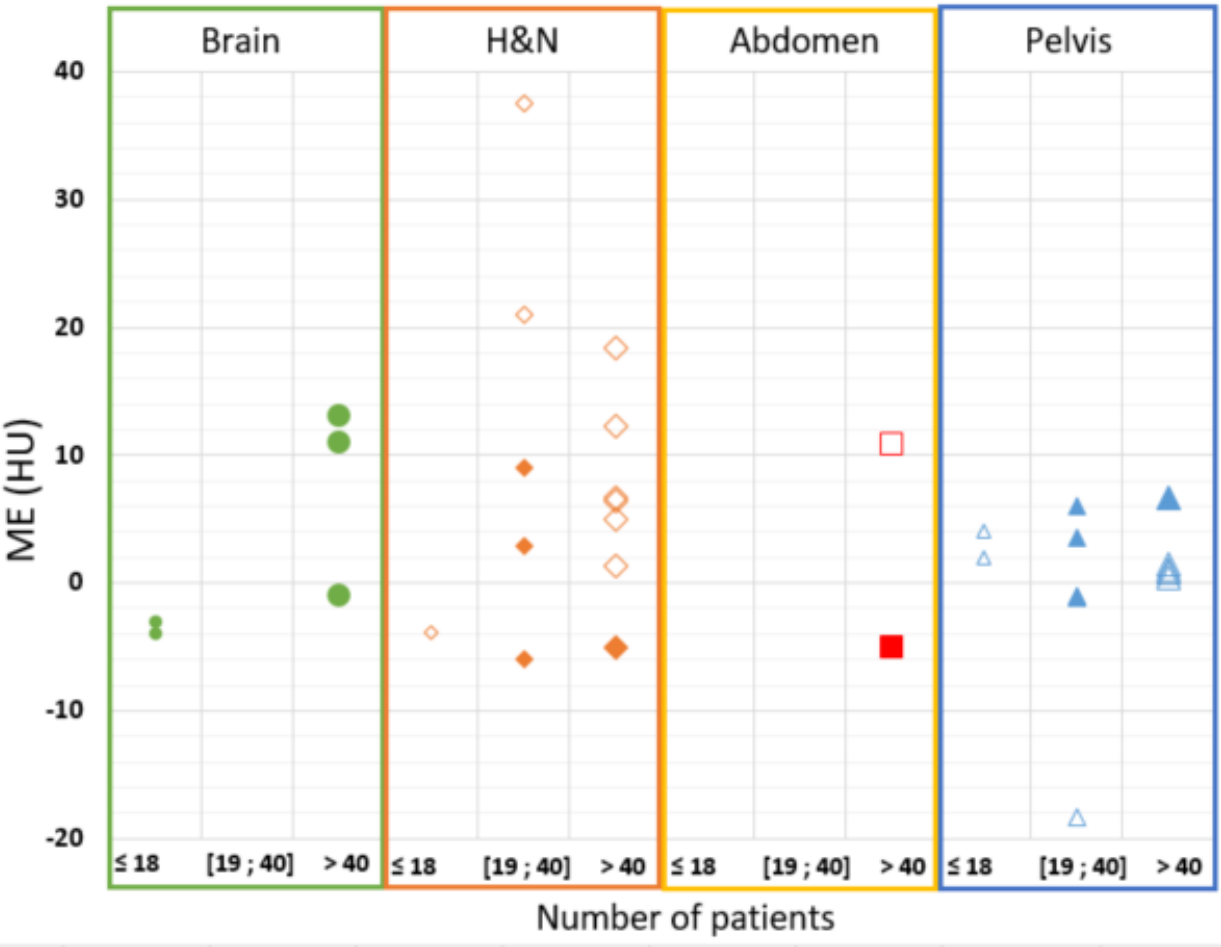


Figure 7: Mean error (ME) results between reference CT and sCT generated with a deep learning method for studies including the brain, head and neck, abdomen, and pelvis

Each marker represent a study result. Full markers represent generator-only models and empty markers generative models with adversarial. Results are divided into three categories: studies including less than 18 patients, studies including 19 to 40 patients and studies including more than 40 patients. The selected values are listed in the Additional tables 1 to 4.

The peak signal to noise ratio (PSNR) is the simplest and most widely used fidelity measure (full-reference quality metric), which is related to the distortion metric, the MSE. Twenty-two studies on sCT generation from MRI reported PSNR results. Figure 8 details PSNR results for the brain, H&N, and pelvis sCT generation from MRI studies.

ACCEPTED MANUSCRIPT - CLEAN COPY

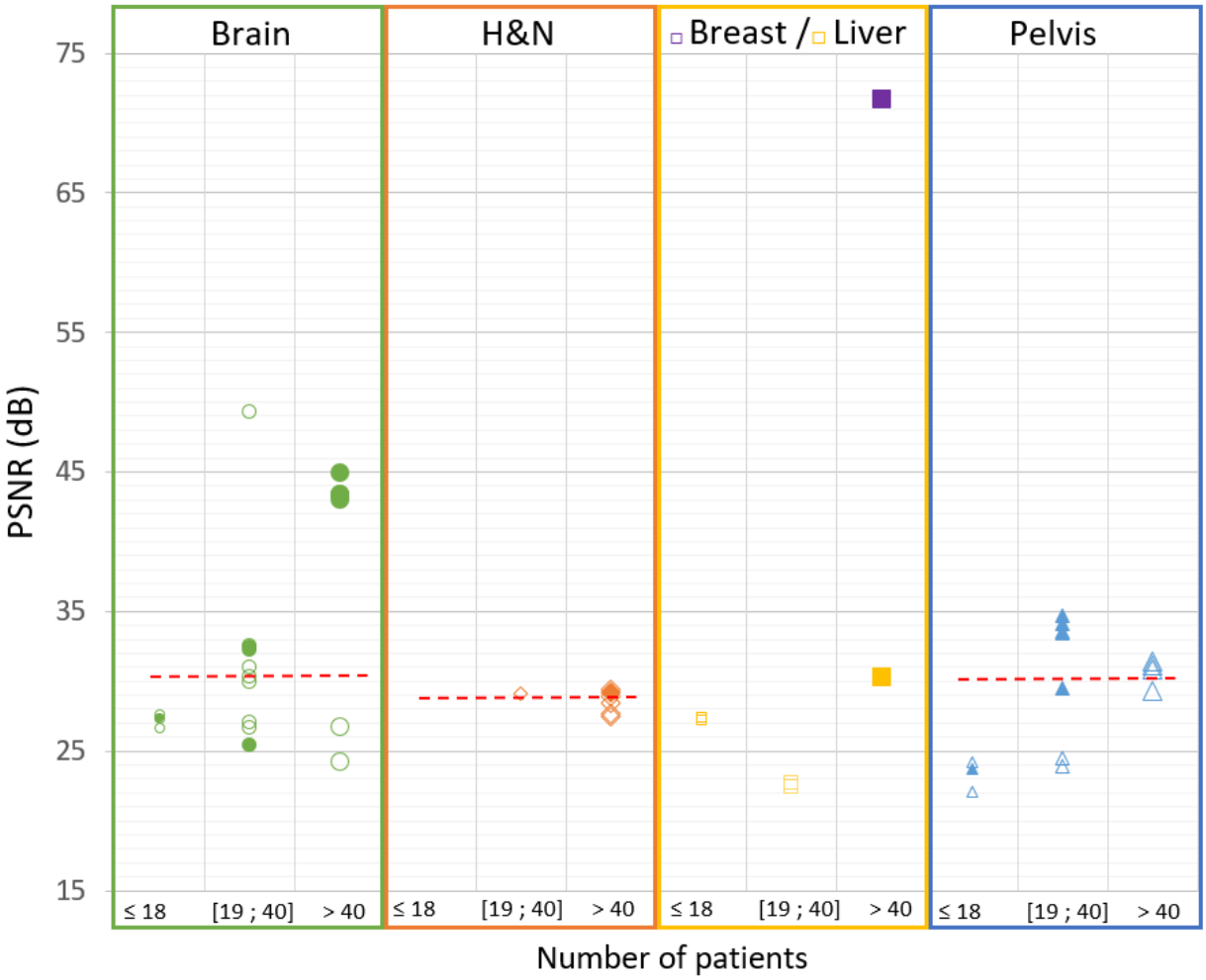


Figure 8: Peak signal-to-noise ratio (PSNR) results between reference CT and sCT generated with deep learning methods in studies on the brain, head and neck, breast, liver, and pelvis

Full markers represent generator-only models and empty markers generative models with adversarial. Results are divided into three categories: studies including less than 18 patients, studies including 19 to 40 patients and studies including more than 40 patients.

Red dotted lines represent the median values. The median values are: 30.3 dB for the brain, 28.9 dB for H&N, and 30.2 dB for the pelvis.

The selected values are listed in the Additional tables 1 to 4.

Four studies reported MSE values in the brain and pelvis. Only three studies reported the root mean square error (RMSE) value in the brain [99], the breast [56], and the abdomen [50]. Although the MSE is an attractive measure due to its simplicity of calculation, MSE/PSNR can be a poor predictor of visual fidelity in images [117].

More sophisticated measures have been developed to take advantage of the known characteristics of the human visual system (HVS). Wang et al. [118] proposed a structural similarity metric (SSIM) to capture the loss of image structure due to variations in lighting (contrast or brightness changes). The

ACCEPTED MANUSCRIPT - CLEAN COPY

SSIM captures image distortion as a combination of three types of distortion: correlation, contrast, and luminance.

2. Geometric fidelity evaluation

Geometric fidelity is based on delineated structures. Nineteen articles reported dice score coefficients (DSCs) between sCT and reference CT for bone, air, or body structures. One study reported DSCs for the bladder and rectum [47]. DSCs were between 0.85 and 0.99 for body and were higher than 0.68 and up to 0.93 for bone structure.

Only two studies reported Hausdorff distance (HD) values for the H&N area [79] and the pelvis [77]. Only one study reported mean absolute surface distance (MASD) values for body, bone, bladder, and rectum volumes [47]. Five studies reported normalized cross-correlation (NCC) values in the brain, liver, and pelvis [32,77,107–109].

The penultimate columns of *Additional tables 1, 2, 3, and 4* list the image results of sCT generated from MRI.

3. Dose evaluation

In MRI-only workflows for radiotherapy, a sCT is generated to perform dose calculation. In this context, studies have proposed dosimetric evaluation of dose calculation from sCT with DVH, voxel-to-voxel dose differences or gamma index analysis. Most studies evaluated dose calculation with photon particles, while nine studies investigates sCT dose uncertainties with protons [45,55,59,71,74,77,97,108,109].

v) Dose-volume histogram (DVH)

DVH is a widely used tool in routine clinical radiotherapy. All treatment planning systems (TPS) allow for the analysis of dose distributions through DVHs. Twenty-two sCT studies reported dose differences at DVH specific points. Eighteen studies reported mean dose differences in selected volume (PTV, CTV, OAR).

vi) Voxel-to-voxel dose difference

The dose difference is defined as the difference between the dose distribution computed on the reference CT and the sCT. The dose difference can be expressed as absolute value (Gy) or relative to the reference dose (%).

Several studies reported mean absolute dose error to express dose uncertainties and mean dose error to express systematic dose uncertainties [47,49,50,70,77,97,102]. Some studies have provided dose

ACCEPTED MANUSCRIPT - CLEAN COPY

differences using dose thresholds such as doses higher than 90% of the prescribed dose, while others have illustrated dose difference maps that allow qualitative and spatial analyses.

vii) Gamma index analysis

Gamma analyses allow spatial analysis (through gamma maps) of dose distributions calculated from sCT compared to those calculated from a reference CT [119]. Gamma analysis can be performed in two or three dimensions. This analysis combines dose and spatial criteria. Several parameters need to be set to perform a gamma analysis, including dose criteria, distance-to-agreement criteria, local or global analysis, and dose threshold. Interpretation and comparison between studies of gamma index results are challenging because they depend on the chosen parameters, dose grid size, and voxel resolution [120]. The gamma results can be expressed as gamma pass-rate (percentage of pixels/voxels with a gamma value lower than 1) or mean gamma. Twenty-eight articles reported gamma pass-rate results. Only one study reported mean gamma values in the pelvis [20].

Figure 9 summarizes the gamma pass-rate results between reference dose distribution and sCT dose distribution for several anatomical localizations. The mean gamma pass rates were above 89% for all localizations and up to 100%, depending on gamma criteria.

ACCEPTED MANUSCRIPT - CLEAN COPY

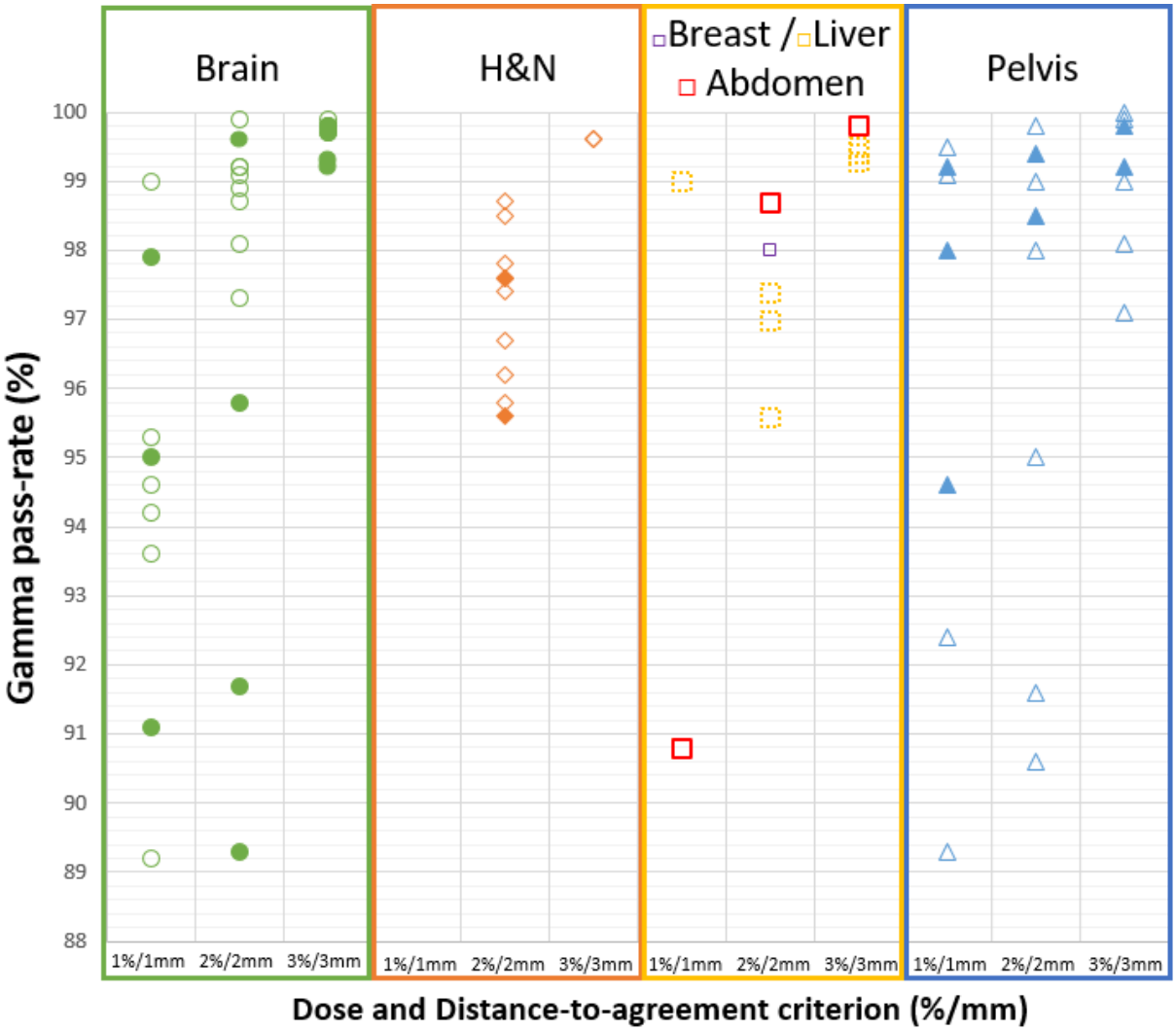


Figure 9: Gamma pass-rate results between reference CT and sCT dose distributions with deep learning methods for studies including the brain, breast, liver, abdomen, and pelvis

Full markers represent generator-only models and empty markers generative models with adversarial. The selected values are listed in the Additional tables 1 to 4.

viii) Specific metrics for proton dose calculation

Proton ranges along the beam paths were compared for dose distributions on the reference CT and sCT. In protontherapy, the range of the proton beam strongly depends on the stopping power ratio (SPR) of a given tissue relative to water, which can be determined using the electron density and effective atomic number through the Bethe-Bloch equation. The range is defined as at the 80% distal dose falloff along each beam direction. Several studies have reported the results of range shift or range difference (in mm) per beam [45,55,71,109].

The last columns of *Additional tables 1, 2, 3, and 4* summarize the dose results of the DL sCT generation studies for MRI dose calculation.

ACCEPTED MANUSCRIPT - CLEAN COPY

D. Image and dose results per anatomical localization

1. Brain

Twenty-four studies of the brain were performed between 2017 and 2021. Additional table 1 summarizes the DL networks and the image and dose metrics results of these studies on brain sCT generation from MRI in radiotherapy. T1-weighted (T1w) sequences were mostly used for generating sCTs 88% of the reviewed studies).

For sCT evaluation, all brain studies reported MAEs, which varied from 44 to 129 HU for the whole brain (Figure 6). For the brain, the MAEs for bone structure were above 174 HU and up to 399 HU in one study. Koike et al. [93] trained a cGAN network with only a T1 sequence or a combination of several sequences (T1w, T2w and FLAIR). The multi-sequence training showed a decrease in MAEs results for the body, soft tissue and bone (between -8 and -33 HU) [93]. Alvarez Andres et al. reported MAE values for CNN and U-Net networks, with higher values for the U-Net network than for the CNN network for the head (an increase of 9 HU) [43]. They also investigated the influence of several sequences (T1, T1-Gd, and T2 FLAIR images) as input in the CNN network. The MAE values were higher with a FLAIR sequence as input in the CNN network than those with T1 sequence (increase of 34 HU). The MAEs were also higher with contrast-enhanced T1-weighted MRI (T1-Gd) than those for T1w MRI (from +3 to +32 HU). Only four studies reported ME values [44,45,55,70], which ranged between -4 HU and 13 HU for the whole brain (Figure 7). The PSNR values were above 24 dB for all brain studies [29,30,32,57,62,72,80,97,98,106] (Figure 8). The DSCs were above 0.96 for the body, 0.69 for bone, and 0.70 for air structures [42,45,58,70,72]. For SSIM, values varied from 0.63 to 0.94 [57,62,72,97,106]. For NCC, two studies reported values of 0.96 [32,109]. Massa et al. [72] trained models on four different MR sequence: CUBE-FLAIR, T1, T1 post contrast and T2 fatsat. No sequence was statistically better on all the metrics (MAE, PSNR, SSIM, DSC).

Among the 24 brain studies, only 14 reported a dose evaluation [42,43,45,48,55,58,59,70,85,93,97–99,109]. Five studies reported results for proton dose planning [45,55,59,97,109]. All reported DVH mean dose differences were below 2% [42,45,48,59,70,85,93,97–99,109].

For most of the studies, gamma pass-rates were above 89% with the most restrictive criterion (1%/1 mm), and except for one study above 95% for other criteria (Figure 9). One study [55] reported a mean gamma pass rate of 89% with 2%/2 mm criteria. With the multi-sequence training, Koike et al. showed an increase in gamma pass-rates (between 0.1% and 1.1%), compared to single sequence training [93].

ACCEPTED MANUSCRIPT - CLEAN COPY

2. Head and Neck (H&N)

Nine DL sCT generation studies were performed in H&N radiotherapy. Additional table 2 summarizes the DL networks and the image and dose metrics results of these studies. The MRI sequences used in the H&N sCT studies were T1 and T2. Four studies used the Dixon reconstruction [49,50,79,102].

MAE and ME metrics have been widely reported in the literature. MAEs varied from 65 to 131 HU for the body or head structures (Figure 6). For the H&N, the MAEs for bone were above 166 HU and up to 357 HU in one study [46]. Qi et al. [50] used multi-sequence input (T1w, T2w, contrast-enhanced T1, and contrast-enhanced T1 Dixon water) images to train a cGAN. The multi-sequence training showed a decrease in MAE for the body, soft tissue, and bone and an increase in PSNR, SSIM, and DSC. They also compared cGAN and U-Net networks. With cGAN and single-sequence input, the MAE, PSNR, and DSC were higher than those obtained using U-Net. For the body, the MEs were mostly around 0 HU, above -6 HU and up to 37 HU (Figure 7). Five studies reported ME for air, bone, or soft tissue [9,46,49,79,101]. For bone structure, MEs were higher up to 247 HU. PSNR and SSIM were only reported in two studies [50,95]. The PSNR results were approximately 28 dB (Figure 8). The SSIMs were between 0.78 and 0.92. For bone structure, DSC values were between 0.70 and 0.89 [50,70,71,79,95]. DVH dose difference was performed in nine studies, with a mean difference < 1.6%. Klages et al. reported a mean dose (D_{mean}) to the parotid glands below 1% and the maximum dose (D_{max}) to the spinal cord below 1.5% [102]. In the only one protontherapy study, the dose differences reached 8% for some OARs [71].

The gamma pass rates were above 95% for the most restrictive criterion (2%/2mm) and above 98% for the other criterion (3%/3 mm). With the multi-sequence training, Qi et al. showed non-significant gamma pass-rate results [50]. In the same study, they also found higher gamma pass-rates for cGAN than for U-Net architectures.

3. Breast

Two DL sCT generation studies were carried out for breast radiotherapy. Additional table 3 summarizes the DL networks and the image and dose metrics results of the studies for breast sCT generation from MRI in radiotherapy. These two studies were based on MR images from a low field (0.35 T) MRI device. Jeon et al. [73] only reported DSC values for two patients. Olberg et al. reported a PSNR of 72 dB, an SSIM of 0.999, and an RMSE of 17 HU [56]. For dose results, they reported gamma passrate higher than 98% with 2%/2 mm criteria.

ACCEPTED MANUSCRIPT - CLEAN COPY

4. Liver

Three DL sCT generation studies were carried out for liver radiotherapy. Additional table 3 summarizes the DL networks and the image and dose metrics results of the studies for liver sCT generation from MRI in radiotherapy.

Two of the three liver studies used T1 sequence. The MAEs varied from 72 to 94 HU for body structure between studies. Fu et al. [33] compared cGAN and cycle-GAN DLMs for data from three patients. The MAEs were higher for cycle-GAN than for cGAN. The PNSR values were above 22 dB. The NCC values were 0.92 in two liver studies [107,108].

DVH dose difference were calculated in the three liver studies, with the mean differences below 1%. In one study, the dose difference in OARs was less than 0.6% [33]. The dose difference in PTV ($D_{95\%}$) was less than 1.1% [107,108]. The gamma pass rates were above 90% for the most restrictive criterion (1%/1mm) and above 95% for the other criteria (Figure 9). In the study by Fu et al. [33], the gamma pass-rates were higher for a cGAN DLM than those for a cycle-GAN DLM.

5. Abdomen

Six DL sCT generation studies were carried out for abdomen radiotherapy. Additional table 3 summarizes the DL networks and the image and dose metrics results of the studies for abdomen sCT generation from MRI in radiotherapy.

Acquisitions were performed in breath hold inspiration for two studies on 0.35 T MRI device [31,121]. The MAEs varied from 55 to 94 HU for body structure between abdomen studies. MAEs in lungs were 105 HU in two studies [74,76]. In Florkow et al. [74] PSNR value was 30 dB and DSC values were 0,76 for bone and 0.92 for lungs. Mean dose differences were lower than 1% and gamma passrate above 98% in 2%/2mm.

6. Pelvis

Eighteen DL sCT generation studies for the pelvis were performed between 2016 and 2020. Additional table 4 summarizes the DL networks and the image and dose metrics results of these studies.

Most MRI sequences in these pelvis studies were T2 sequences. T1 sequences [30,53,78] or Dixon reconstruction [52,54,94] were also used to generate sCT from MRI in the pelvis.

The reported MAEs were 27-65 HU for body structure (Figure 6) and around 120 HU and up to 250 HU for bone [20,31,47,51,53,78,100]. Fu et al. compared training in 2D and 3D in four patients [78], reporting higher MAEs for 2D than for 3D training (+2-5 HU). Largent et al. compared U-Net and GAN

ACCEPTED MANUSCRIPT - CLEAN COPY

networks with different loss functions [20]. With the L2 loss function, U-Net showed lower MAEs than those for GAN. For all studies, the MEs were generally near to 0 HU for the whole body structure (Figure 7). One multicenter study reported an ME of -18 HU [89]. For the pelvis, the MEs in the bone were up to 141 HU in one study [31]. The reported PSNRs were between 24 and 34 dB (Figure 8). Only one study reported SSIM in the pelvic area [75]. Only two studies reported DSC for the body (0.85 and 0.99) [47,77]. The DSCs for bone ranged between 0.70 and 0.93 [47,52,53,75,78]. Only one study reported a DSC for the bladder and rectum of 0.9 [47].

Among the 18 pelvic studies, only nine reported dose evaluations. Liu et al. [108] performed proton dose planning. Most studies reported dose differences below 1.5% for target volumes and OARs. Arabi et al. reported maximum dose differences below 0.5% for the bladder and rectum between 1.1% and 2.9% for the CTV and PTV. Some studies reported a very low dose difference (<0.6%) for PTV, bladder, rectum, and femoral heads [20,51]. However, Liu et al. reported dose differences up to 5% in the rectum and up to 11% in the bladder [77].

The gamma pass-rates were above 89% for the most restrictive criterion and generally above 95% for the other criteria (Figure 9). In a study using prostate data in training to generate sCT of the rectum and cervix [94], the gamma pass-rates were around 91% for gamma criteria of 2%/2 mm.

Discussion

This article reviewed deep learning methods used to generate sCT from MRI in radiation therapy, and their associated image and dose uncertainties. Two types of DL architectures are widely used; generator-only, and GAN. The most recent DLMs were cGAN and cycle-GAN. A variety of metrics for image evaluation (image intensity and geometric fidelity) has been proposed. All sCT studies include in this review reported MAE values. The median MAE results were 76 HU for head localization (brain and H&N) and liver, and 42 HU for the pelvic area. Dose evaluations consisted in DVH comparisons, voxel-to-voxel dose differences, or gamma index analyses. The mean dose differences were below 1% in the H&N, liver, breast, and pelvis sCT studies. In brain sCT studies, the mean dose difference was below 2%. For most of the studies, the gamma pass-rates were above 95% (with 2%/2 mm and 3%/3 mm criteria) (Figure 9).

In radiotherapy, the first sCT generation methods from MRI were bulk density and atlas-based. Other machine learning methods (non-DLM) have also been investigated, including patch-based or random forest [10,122]. This review focused on DLMs which are the most recent methods with the first study in the pelvic area reported in 2016. Different neural network architectures have been used in the literature with multiple parameters to be set. Compared to other sCT generation methods, DLMs have

fast computation times, and do not necessarily require deformable inter-patient registration. sCT generation DLMs have just been commercially available for a clinical use [79,123]. To our knowledge, no open source software is available for sCT generation from MRI with a DLM. Each research team has developed his own DLM with hyperparameter tuning. This review was not able to identify the most "accurate" DL architecture. Although GAN DLMs are the most recent, for now they do not outperform generator-only DLMs (Figures 6, 7, 8, and 9). Moreover, we acknowledge that studies are not directly comparable due to the great disparities in input data (imaging protocol, scanner parameters, etc.), training cohort sizes, evaluation cohort sizes, and methods of evaluation. Same data should be used to directly compare the results, such as data in open access from the Gold Atlas Project [116]. Two studies used these data [89,90]. Some studies directly compared several DLMs with the same data (Additional table 1, 2, 3, and 4). Size of patient cohort (training + evaluation) varied according to study and anatomical localization (Additional table 1, 2, 3, and 4). The median number of patients were 45 for the brain, 33 for H&N and 23 for pelvic localization. Studies including few patients (less than 19) did not show the better results (Figures 6, 7, and 8). But studies with more than 40 patients did not outperform compared with studies with 19 to 40 patients (Figure 6). Training strategy depends on the number of available data. If you have few data (less than 20 patient data), a leave-one-out strategy is recommended. Image quality of data training is important. Image with artefacts must be removed of the training. A first step of quality image optimization (MR sequence or CT acquisition parameters) can be useful.

Although cycle-GAN or other networks do not require paired data for training step, paired data are required for the evaluation step. For the brain a rigid registration can be sufficient [32] but not for H&N or pelvis area. Few studies used unpaired data in training, even with cycle-GAN [80,106] and with Mutual Information as loss function in GAN [58,59]. To have paired data, a deformable registration is needed, with additional uncertainties. Florkow et al. [52] quantified the uncertainties due to MRI-CT registration.

To perform the evaluation, sCTs generated from MRI are compared to reference CT. Even if the time between acquisitions is kept as short as possible, MRI acquisition and reference CT can differ even after non-rigid registration, due to gas volatility and bowel loops displacement in the abdomen, artifacts (teeth, hip prosthesis, fiducials, contrast agent, etc.) or internal movements (bladder and rectum filling) between CT and MR images. Maspero et al. [94] proposed to override gas in the rectum as in the reference CT and performed the imaging evaluation in the intersection volume of the body contours (reference CT $\cap$ sCT). Cusumaro et al. excluded some patients from their studies because of artifacts (artificial implants) or difference of air pocket locations between CT and MR images [31].

Most of studies reported global imaging evaluation metrics, without local or spatial analysis. Hemsley et al. [96] proposed a detailed sCT evaluation with uncertainty heatmaps. Models were proposed to

ACCEPTED MANUSCRIPT - CLEAN COPY

spatially quantify intrinsic and parameter sCT uncertainties [124]. With this method, uncertainty maps are a second output of the DL network [125]. Moreover, an analysis based on image gradient could be performed.

Reviewed articles aimed to show accuracy of sCT generation from MRI compared to a reference CT. In the future, we can imagine an MRI-only workflow without any CT acquisition. In this case, image evaluation metrics without reference need to be developed. Before any use in clinical practice, commissioning and quality assurance process must be implemented [126]. Practical guidelines on the use of MRI for external radiotherapy treatment planning were recently proposed by a multidisciplinary working group of the Institute of Physics and Engineering in Medicine (IPEM) [127]. This document overviews all the aspects of MRI implementation for radiotherapy are described (MR safety, training and education, patient set-up, MRI sequence, MR quality assurance, etc.).

To date, few DL studies have been carried out on abdomen, liver, breast, or H&N radiotherapy. This limited number may be due to the small number of patients undergoing MRI for liver or abdomen radiotherapy compared to brain or prostate radiotherapy. Moreover, standard acquisition of breast MRI is not in radiotherapy treatment position. Number of breast studies are increasing with the availably of MR acquisitions from MRI-linac device. The lack of sCT generation from H&N and abdomen MRI may be due to the complexity of these anatomical localizations with the large part of heterogeneities. MRI in the treatment position can be challenging for H&N acquisitions because specific coils are used [84]. An image quality optimization must be carried out to obtain better acquisition parameters. No study has yet investigated lung sCT generation from MRI with DLM. Movement is a huge challenge for MR imaging.

Several MRI sequences have been used to generate sCT from MRI in radiotherapy, with T2w sequences the most common. Some studies used specific reconstruction techniques such as mDixon or FLAIR. The FLAIR sequence is an inversion-recovery sequence. This sequence improves the detection of lesions of the cerebral parenchyma and enables visualization of edemas. It also facilitates the detection of white matter pathologies (softening, demyelination process), which appear as hypersignals.

Three studies investigated the impact of MISO, compared to a SISO [50,93,104]. MISO has the advantage of better tissue description. Koike et al. reported that MISO decreased the MAE and improved gamma pass-rate results compared to SISO [93]. Qi et al. used four sequences individually and combined them as input. The combination of sequences improved the sCT accuracy and robustness [50]. Sharma et al. [104] proposed a MIMO method generalizing to any combination of available and missing MRI sequence.

Moreover, three studies evaluated the impact of generating a sCT from a device other than the one used during training [89,90,97]. Such “multidevice” or “multicenter” impact is a key challenge to a commercial development.

ACCEPTED MANUSCRIPT - CLEAN COPY

The emergence of linacs combining MRI in the treatment room (MRI-linacs) increase the willingness of MRI-only workflow radiotherapy [128]. Some reviewed studies already used DLM for sCT generation from this device [31,33,56,73]. In this context, dose planning need to consider the presence of magnetic field, with the electron return effect [129,130]. Moreover, on MRI-linac, an MR image is acquired for each fraction. This image could be used to perform dose monitoring or replanning with the use of a DLM, in the context of MR-guided adaptive radiotherapy [131].

Conclusions

The emergence of DL allows the fast and accurate generation of sCT from MRI in radiotherapy. In the literature, a variety of DLMs have been applied, mainly for brain and pelvis cancer and also for H&N and liver. In each study, DLM has showed particularities in terms of hyperparameters or loss functions. Different MRI sequences are used depending on the anatomical location. Many metrics are used for image (voxel intensity and geometric fidelity) evaluation of the generated sCT. The MAE results were around 76 HU for head localization (brain and H&N) and liver, and 40 HU for pelvis. Dosimetric evaluation showed uncertainties below 2% for brain radiotherapy and lower than 1% for H&N, liver, abdomen, and pelvic areas. A better sCT quality was obtained with multiple inputs compared to single input of a DLM. Key challenges of the sCT generation for MRI in radiotherapy with DLMs is the standardization of sCT evaluation, and multicentrer impact.

Conflict of Interest Statement:

The authors have no relevant conflicts of interest to disclose.

ACCEPTED MANUSCRIPT - CLEAN COPY

References

[1] 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. BJR 2019;92:20180948. https://doi.org/10.1259/bjr.20180948.

[2] Kerkmeijer LGW, Maspero M, Meijer GJ, van der Voort van Zyp JRN, de Boer HCJ, van den Berg CAT. Magnetic Resonance Imaging only Workflow for Radiotherapy Simulation and Planning in Prostate Cancer. Clinical Oncology 2018;30:692–701. https://doi.org/10.1016/j.clon.2018.08.009.

[3] Jonsson J, Nyholm T, Söderkvist K. The rationale for MR-only treatment planning for external radiotherapy. Clin Transl Radiat Oncol 2019;18:60–5. https://doi.org/10.1016/j.ctro.2019.03.005.

[4] 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. International Journal of Radiation Oncology*Biology*Physics 2019;103:479–90. https://doi.org/10.1016/j.ijrobp.2018.10.002.

[5] 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. International Journal of Radiation Oncology • Biology • Physics 2015;93:1144–53. https://doi.org/10.1016/j.ijrobp.2015.08.045.

[6] Cusumano D, Placidi L, Teodoli S, Boldrini L, Greco F, Longo S, et al. On the accuracy of bulk synthetic CT for MR-guided online adaptive radiotherapy. Radiol Med 2020;125:157–64. https://doi.org/10.1007/s11547-019-01090-0.

[7] Choi JH, Lee D, O'Connor L, Chalup S, Welsh JS, Dowling J, et al. Bulk Anatomical Density Based Dose Calculation for Patient-Specific Quality Assurance of MRI-Only Prostate Radiotherapy. Front Oncol 2019;9. https://doi.org/10.3389/fonc.2019.00997.

[8] Kemppainen R, Suilamo S, Ranta I, Pesola M, Halkola A, Eufemio A, et al. Assessment of dosimetric and positioning accuracy of a magnetic resonance imaging-only solution for external beam radiotherapy of pelvic anatomy. Physics and Imaging in Radiation Oncology 2019;11:1–8. https://doi.org/10.1016/j.phro.2019.06.001.

[9] Chen S, Quan H, Qin A, Yee S, Yan D. MR image-based synthetic CT for IMRT prostate treatment planning and CBCT image-guided localization. Journal of Applied Clinical Medical Physics 2016;17:236–45. https://doi.org/10.1120/jacmp.v17i3.6065.

[10] Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, et al. Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model. IEEE Transactions on Medical Imaging 2016;35:174–83. https://doi.org/10.1109/TMI.2015.2461533.

[11] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.

[12] Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. Computers in Biology and Medicine 2018;98:126–46. https://doi.org/10.1016/j.compbiomed.2018.05.018.

[13] Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. Medical Physics 2019;46:e1–36. https://doi.org/10.1002/mp.13264.

[14] Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. BJR 2019;92:20190001. https://doi.org/10.1259/bjr.20190001.

[15] Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: advantages, potential, and challenges. Phys Med Biol 2020;65:05TR01. https://doi.org/10.1088/1361-6560/ab6f51.

[16] Boldrini L, Bibault J-E, Masciocchi C, Shen Y, Bittner M-I. Deep Learning: A Review for the Radiation Oncologist. Frontiers in Oncology 2019;9. https://doi.org/10.3389/fonc.2019.00977.

ACCEPTED MANUSCRIPT - CLEAN COPY

[17] Feng M, Valdes G, Dixit N, Solberg TD. Machine Learning in Radiation Oncology: Opportunities, Requirements, and Needs. Frontiers in Oncology 2018;8. https://doi.org/10.3389/fonc.2018.00110.
[18] Bibault J-E, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Letters 2016;382:110–7. https://doi.org/10.1016/j.canlet.2016.05.033.
[19] Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? Radiotherapy and Oncology 2018;129:421–6. https://doi.org/10.1016/j.radonc.2018.05.030.
[20] 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. International Journal of Radiation Oncology*Biology*Physics 2019;105:1137–50. https://doi.org/10.1016/j.ijrobp.2019.08.049.
[21] 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.
[22] Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging–Only Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics 2018;100:199–217. https://doi.org/10.1016/j.ijrobp.2017.08.043.
[23] Bird D, Henry AM, Sebag-Montefiore D, Buckley DL, Al-Qaisieh B, Speight R. A Systematic Review of the Clinical Implementation of Pelvic Magnetic Resonance Imaging (MR)-Only Planning for External Beam Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics 2019. https://doi.org/10.1016/j.ijrobp.2019.06.2530.
[24] Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. Physics in Medicine and Biology 2018. https://doi.org/10.1088/1361-6560/aaaca4.
[25] Wafa B, Moussaoui A. A review on methods to estimate a CT from MRI data in the context of MRI-alone RT. Mèd Technologies J 2018;2:150–78. https://doi.org/10.26415/2572-004X-vol2iss1p150-178.
[26] Wang T, Lei Y, Fu Y, Wynne JF, Curran WJ, Liu T, et al. A review on medical imaging synthesis using deep learning and its clinical applications. Journal of Applied Clinical Medical Physics 2020. https://doi.org/10.1002/acm2.13121.
[27] Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning-based synthetic-CT generation in radiotherapy and PET: a review. ArXiv:210202734 [Physics] 2021.
[28] Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. In: Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al., editors. Deep Learning and Data Labeling for Medical Applications, vol. 10008, Cham: Springer International Publishing; 2016, p. 170–8. https://doi.org/10.1007/978-3-319-46976-8_18.
[29] Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017, vol. 10435, Cham: Springer International Publishing; 2017, p. 417–25. https://doi.org/10.1007/978-3-319-66179-7_48.
[30] Xiang L, Wang Q, Nie D, Zhang L, Jin X, Qiao Y, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. Medical Image Analysis 2018;47:31–44. https://doi.org/10.1016/j.media.2018.03.011.
[31] 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. Radiotherapy and Oncology 2020. https://doi.org/10.1016/j.radonc.2020.10.018.
[32] Lei Y, Harms J, Wang T, Liu Y, Shu H-K, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Medical Physics 2019;46:3565–81. https://doi.org/10.1002/mp.13617.

ACCEPTED MANUSCRIPT - CLEAN COPY

[33] Fu J, Singhrao K, Cao M, Yu V, Santhanam AP, Yang Y, et al. Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. Biomed Phys Eng Express 2020;6:015033. https://doi.org/10.1088/2057-1976/ab6e1f.
[34] Kazeminia S, Baur C, Kuijper A, van Ginneken B, Navab N, Albarqouni S, et al. GANs for Medical Image Analysis. Artificial Intelligence in Medicine 2020.
[35] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical Image Analysis 2017;42:60–88. https://doi.org/10.1016/j.media.2017.07.005.
[36] Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, et al. A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises. ArXiv:200809104 [Cs, Eess] 2020.
[37] Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: advantages, potential, and challenges. Physics in Medicine & Biology 2020. https://doi.org/10.1088/1361-6560/ab6f51.
[38] Nair V, Hinton G. Rectified linear units improve restricted boltzmann machines. ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning 2010:807–14.
[39] Maas AL, Hannun AY, Ng AY. Rectifier Nonlinearities Improve Neural Network Acoustic Models 2013.
[40] Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). ICLR 2016.
[41] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning - 2015:448–56.
[42] Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. Journal of Applied Clinical Medical Physics 2019;20:105–14. https://doi.org/10.1002/acm2.12554.
[43] Andres EA, Fidon L, Vakalopoulou M, Lerousseau M, Carré A, Sun R, et al. Dosimetry-driven quality measure of brain pseudo Computed Tomography generated from deep learning for MRI-only radiotherapy treatment planning. International Journal of Radiation Oncology*Biology*Physics 2020:S0360301620311305. https://doi.org/10.1016/j.ijrobp.2020.05.006.
[44] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Medical Physics 2017;44:1408–19. https://doi.org/10.1002/mp.12155.
[45] Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep Convolution Neural Network (DCNN) Multiplane Approach to Synthetic CT Generation From MR images—Application in Brain Proton Therapy. International Journal of Radiation Oncology • Biology • Physics 2019;105:495–503. https://doi.org/10.1016/j.ijrobp.2019.06.2535.
[46] Wang Y, Liu C, Zhang X, Deng W. Synthetic CT Generation Based on T2 Weighted MRI of Nasopharyngeal Carcinoma (NPC) Using a Deep Convolutional Neural Network (DCNN). Front Oncol 2019;9. https://doi.org/10.3389/fonc.2019.01333.
[47] 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.
[48] Gupta D, Kim M, Vineberg KA, Balter JM. Generation of Synthetic CT Images From MRI for Treatment Planning and Patient Positioning Using a 3-Channel U-Net Trained on Sagittal Images. Front Oncol 2019;9. https://doi.org/10.3389/fonc.2019.00964.
[49] 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. Medical Physics 2019;46:4095–104. https://doi.org/10.1002/mp.13663.

ACCEPTED MANUSCRIPT - CLEAN COPY

[50] Qi M, Li Y, Wu A, Jia Q, Li B, Sun W, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. Medical Physics 2020;47:1880–94. https://doi.org/10.1002/mp.14075.
[51] Chen S, Qin A, Zhou D, Yan D. Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. Medical Physics 2018;45:5659–65. https://doi.org/10.1002/mp.13247.
[52] Florkow MC, Zijlstra F, M.d LGWK, Maspero M, Berg CAT van den, Stralen M van, et al. The impact of MRI-CT registration errors on deep learning-based synthetic CT generation. Medical Imaging 2019: Image Processing, vol. 10949, International Society for Optics and Photonics; 2019, p. 1094938. https://doi.org/10.1117/12.2512747.
[53] Florkow MC, Zijlstra F, Willemsen K, Maspero M, Berg CAT van den, Kerkmeijer LGW, et al. Deep learning–based MR-to-CT synthesis: The influence of varying gradient echo–based MR images as input channels. Magnetic Resonance in Medicine 2020;83:1429–41. https://doi.org/10.1002/mrm.28008.
[54] Stadelmann JV, Schulz H, Heide UA van der, Renisch S. Pseudo-CT image generation from mDixon MRI images using fully convolutional neural networks. Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 10953, International Society for Optics and Photonics; 2019, p. 109530Z. https://doi.org/10.1117/12.2512741.
[55] Neppl 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 Oncologica 2019;58:1429–34. https://doi.org/10.1080/0284186X.2019.1630754.
[56] Olberg S, Zhang H, Kennedy WR, Chun J, Rodriguez V, Zoberi I, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. Med Phys 2019;46:4135–47. https://doi.org/10.1002/mp.13716.
[57] Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. Quant Imaging Med Surg 2020;10:1223–36. https://doi.org/10.21037/qims-19-885.
[58] Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiotherapy and Oncology 2019;136:56–63. https://doi.org/10.1016/j.radonc.2019.03.026.
[59] Kazemifar S, Barragán Montero AM, Souris K, Rivas ST, Timmerman R, Park YK, et al. Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors: Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors. Journal of Applied Clinical Medical Physics 2020;21:1–11. https://doi.org/10.1002/acm2.12856.
[60] V. Badrinarayanan, A. Kendall, R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2017;39:2481–95. https://doi.org/10.1109/TPAMI.2016.2644615.
[61] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Computer Vision and Pattern Recognition 2015.
[62] Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Medical Physics 2018;45:3627–36. https://doi.org/10.1002/mp.13047.
[63] Huang G, Liu Z, Maaten VD, Weinberger KQ. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, p. 2261–9. https://doi.org/10.1109/CVPR.2017.243.

ACCEPTED MANUSCRIPT - CLEAN COPY

[64] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE; 2017, p. 5967–76. https://doi.org/10.1109/CVPR.2017.632.
[65] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018;42:7132–41. https://doi.org/10.1109/TPAMI.2019.2913372.
[66] Ulyanov D, Vedaldi A, Lempitsky V. Instance Normalization: The Missing Ingredient for Fast Stylization. Computer Vision and Pattern Recognition 2017.
[67] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv:150504597 [Cs] 2015.
[68] Wolterink JM, Leiner T, Viergever MA, Išgum I. Dilated Convolutional Neural Networks for Cardiovascular MR Segmentation in Congenital Heart Disease. Reconstruction, Segmentation, and Analysis of Medical Images, RAMBO 2016, HVSMR 2016 Lecture Notes in Computer Science 2017;10129:95–102. https://doi.org/10.1007/978-3-319-52280-7_9.
[69] Li W, Wang G, Fidon L, Ourselin S, Cardoso MJ, Vercauteren T. On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap P-T, et al., editors. Information Processing in Medical Imaging, Cham: Springer International Publishing; 2017, p. 348–60.
[70] Dinkla AM, Wolterink JM, Maspero M, Savenije MHF, Verhoeff JJC, Seravalli E, et al. MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network. International Journal of Radiation Oncology*Biology*Physics 2018;102:801–12. https://doi.org/10.1016/j.ijrobp.2018.05.058.
[71] 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. Physics in Medicine & Biology 2020. https://doi.org/10.1088/1361-6560/abb1d6.
[72] Massa HA, Johnson JM, McMillan AB. Comparison of deep learning synthesis of synthetic CTs using clinical MRI inputs. Phys Med Biol 2020;65:23NT03. https://doi.org/10.1088/1361-6560/abc5cb.
[73] Jeon W, An HJ, Kim J, Park JM, Kim H, Shin KH, et al. Preliminary Application of Synthetic Computed Tomography Image Generation from Magnetic Resonance Image Using Deep-Learning in Breast Cancer Patients. J Radiat Prot Res 2019;44:149–55. https://doi.org/10.14407/jrpr.2019.44.4.149.
[74] Florkow MC, Guerreiro F, Zijlstra F, Seravalli E, Janssens GO, Maduro JH, et al. Deep learning-enabled MRI-only photon and proton therapy treatment planning for paediatric abdominal tumours. Radiotherapy and Oncology 2020;153:220–7. https://doi.org/10.1016/j.radonc.2020.09.056.
[75] Bahrami A, Karimian A, Fatemizadeh E, Arabi H, Zaidi H. A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI. Med Phys 2020;47:5158–71. https://doi.org/10.1002/mp.14418.
[76] Liu L, Johansson A, Cao Y, Dow J, Lawrence TS, Balter JM. Abdominal synthetic CT generation from MR Dixon images using a U-net trained with 'semi-synthetic' CT data. Physics in Medicine & Biology 2020;65:125001. https://doi.org/10.1088/1361-6560/ab8cd2.
[77] Liu Y, Lei Y, Wang Y, Shafai-Erfani G, Wang T, Tian S, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. Phys Med Biol 2019;64:205022. https://doi.org/10.1088/1361-6560/ab41af.
[78] 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. Medical Physics 2019;46:3788–98. https://doi.org/10.1002/mp.13672.
[79] Palmér E, Karlsson A, Nordström F, Petruson K, Siversson C, Ljungberg M, et al. Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic

resonance imaging only workflow for head and neck radiotherapy. Physics and Imaging in Radiation Oncology 2021;17:36–42. https://doi.org/10.1016/j.phro.2020.12.007.
[80] Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, Berg CAT van den, Isgum I. Deep MR to CT Synthesis using Unpaired Data. ArXiv:170801155 [Cs] 2017.
[81] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. ArXiv:14062661 [Cs, Stat] 2014.
[82] Yi X, Walia E, Babyn P. Generative Adversarial Network in Medical Imaging: A Review. Medical Image Analysis 2019;58:101552. https://doi.org/10.1016/j.media.2019.101552.
[83] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. ArXiv:14062661 [Cs, Stat] 2014.
[84] Largent A, Marage L, Gicquiau I, Nunes J-C, Reynaert N, Castelli J, et al. Head-and-Neck MRI-only radiotherapy treatment planning: From acquisition in treatment position to pseudo-CT generation. Cancer/Radiothérapie 2020:S1278321820300615. https://doi.org/10.1016/j.canrad.2020.01.008.
[85] 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;22:308–17. https://doi.org/10.1002/acm2.13139.
[86] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Computer Vision and Pattern Recognition 2017.
[87] Bengio Y, Simard P, Frasconi P. Learning Long-Term Dependencies with Gradient Descent is Difficult. IEEE Transactions on Neural Networks 1994.
[88] Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. Least Squares Generative Adversarial Networks. IEEE International Conference on Computer Vision 2017:9.
[89] 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 era using a conditional generative adversarial network. Physics in Medicine & Biology 2020. https://doi.org/10.1088/1361-6560/ab7633.
[90] Fetty 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. https://doi.org/10.1088/1361-6560/ab857b.
[91] Chen L-C, Zhu Y, Papandreou G, Schrof F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Computer Vision and Pattern Recognition 2018.
[92] Chen LC, Papandreou G, Schroff F, Adam F. Rethinking atrous convolution for semantic image segmentation. Computer Vision and Pattern Recognition 2017.
[93] Koike Y, Akino Y, Sumida I, Shiomi H, Mizuno H, Yagi M, et al. Feasibility of synthetic computed tomography generated with an adversarial network for multi-sequence magnetic resonance-based brain radiotherapy. J Radiat Res 2020;61:92–103. https://doi.org/10.1093/jrr/rrz063.
[94] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-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.
[95] Tie X, Lam S, Zhang Y, Lee K, Au K, Cai J. Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients. Med Phys 2020;47:1750–62. https://doi.org/10.1002/mp.14062.
[96] Hemsley M, Chugh B, Ruschin M, Lee Y, Tseng C-L, Stanisz G, et al. Deep Generative Model for Synthetic-CT Generation with Uncertainty Predictions. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, vol. 12261, Cham: Springer International Publishing; 2020, p. 834–44. https://doi.org/10.1007/978-3-030-59710-8_81.

ACCEPTED MANUSCRIPT - CLEAN COPY

[97] Maspero M, Bentvelzen LG, Savenije MHF, Guerreiro F, Seravalli E, Janssens GO, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. Radiotherapy and Oncology 2020;153:197–204. https://doi.org/10.1016/j.radonc.2020.09.029.
[98] Tang B, Wu F, Fu Y, Wang X, Wang P, Orlandini LC, et al. Dosimetric evaluation of synthetic CT image generated using a neural network for MR-only brain radiotherapy. J Appl Clin Med Phys 2021:acm2.13176. https://doi.org/10.1002/acm2.13176.
[99] Bourbonne V, Jaouen V, Hognon C, Boussion N, Lucia F, Pradier O, et al. Dosimetric Validation of a GAN-Based Pseudo-CT Generation for MRI-Only Stereotactic Brain Radiotherapy. Cancers 2021;13:1082. https://doi.org/10.3390/cancers13051082.
[100] 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. Radiotherapy and Oncology 2021;156:23–8. https://doi.org/10.1016/j.radonc.2020.11.027.
[101] Peng Y, Chen S, Qin A, Chen M, Gao X, Liu Y, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiotherapy and Oncology 2020;150:217–24. https://doi.org/10.1016/j.radonc.2020.06.049.
[102] Klages P, Bensilmane I, Riyahi S, Jiang J, Hunt M, Deasy JO, et al. Comparison of Patch-Based Conditional Generative Adversarial Neural Net Models with Emphasis on Model Robustness for Use in Head and Neck Cases for MR-Only Planning 2020:27. https://doi.org/arXiv:1902.00536.
[103] M. Weber, M. Fürst, J. M. Zöllner. Automated Focal Loss for Image based Object Detection. 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, p. 1423–9. https://doi.org/10.1109/IV47402.2020.9304830.
[104] Sharma A, Hamarneh G. Missing MRI Pulse Sequence Synthesis Using Multi-Modal Generative Adversarial Network. IEEE Transactions on Medical Imaging 2020;39:1170–83. https://doi.org/10.1109/TMI.2019.2945521.
[105] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. IEEE International Conference on Computer Vision 2017.
[106] Yang H, Sun J, Carass A, Zhao C, Lee J, Xu Z, et al. Unpaired Brain MR-to-CT Synthesis Using a Structure-Constrained CycleGAN. In: Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, et al., editors. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Cham: Springer International Publishing; 2018, p. 174–82. https://doi.org/10.1007/978-3-030-00889-5_20.
[107] Liu Y, Lei Y, Wang T, Kayode O, Tian S, Liu T, et al. MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method. BJR 2019;92:20190067. https://doi.org/10.1259/bjr.20190067.
[108] Liu Y, Lei Y, Wang Y, Wang T, Ren L, Lin L, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. Phys Med Biol 2019;64:145015. https://doi.org/10.1088/1361-6560/ab25bc.
[109] Shafai-Erfani G, Lei Y, Liu Y, Wang Y, Wang T, Zhong J, et al. MRI-Based Proton Treatment Planning for Base of Skull Tumors. International Journal of Particle Therapy 2019;6:12–25. https://doi.org/10.14338/IJPT-19-00062.1.
[110] Heinrich MP, Jenkinson M, Bhushan M, Matin T, Fergus V. Gleeson, Brady SM, et al. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Medical Image Analysis 2012.
[111] Mao X, Li Q, Xie H. Multi-class Generative Adversarial Networks with the L2 Loss Function. Computer Vision and Pattern Recognition 2016.
[112] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29:1310–20. https://doi.org/10.1109/TMI.2010.2046908.
[113] Sled, JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data 1998:87–97.

ACCEPTED MANUSCRIPT - CLEAN COPY

[114] Nyul LG, Udupa JK, Zhang X. New Variants of a Method of MRI Scale Standardization. IEEE Trans Med Imag 2000:143–50.
[115] Cox I, Roy S, Hingorani SL. Dynamic histogram warping of image pairs for constant image brightness. Proc Int Conf Image Proc 1995:366–9.
[116] Nyholm T, Svensson S, Andersson S, Jonsson J, Sohlin M, Gustafsson C, et al. MR and CT data with multiobserver delineations of organs in the pelvic area—Part of the Gold Atlas project. Medical Physics 2018;45:1295–300. https://doi.org/10.1002/mp.12748.
[117] Girod B. What's wrong with mean-squared error. Digital Images and Human Vision (A B Watson, Ed) 1993:207–20.
[118] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing 2004;13:600–12. https://doi.org/10.1109/TIP.2003.819861.
[119] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Medical Physics 1998;25:656–61.
[120] Hussein M, Clark CH, Nisbet A. Challenges in calculation of the gamma index in radiotherapy–towards good practice. Physica Medica 2017;36:1–11.
[121] Fu J, Singhrao K, Cao M, Yu V, Santhanam AP, Yang Y, et al. Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. Biomedical Physics & Engineering Express 2020;6:015033. https://doi.org/10.1088/2057-1976/ab6e1f.
[122] Yang X, Lei Y, Shu H-K, Rossi P, Mao H, Shim H, et al. Pseudo CT estimation from MRI using patch-based random forest. Medical Imaging 2017: Image Processing, vol. 10133, International Society for Optics and Photonics; 2017, p. 101332Q. https://doi.org/10.1117/12.2253936.
[123] Lerner M, Medin J, Jamtheim Gustafsson C, Alkner S, Siversson C, Olsson LE. Clinical validation of a commercially available deep learning software for synthetic CT generation for brain. Radiat Oncol 2021;16:66. https://doi.org/10.1186/s13014-021-01794-6.
[124] Bragman FJS, Tanno R, Eaton-Rosen Z, Li W, Hawkes DJ, Ourselin S, et al. Uncertainty in multitask learning: joint representations for probabilistic MR-only radiotherapy planning. ArXiv:180606595 [Cs] 2018;11073:3–11. https://doi.org/10.1007/978-3-030-00937-3_1.
[125] Tanno R, Worrall D, Kaden E, Ghosh A, Grussu F, Bizzi A, et al. Uncertainty Quantification in Deep Learning for Safer Neuroimage Enhancement. ArXiv:190713418 [Cs, Eess, Stat] 2019.
[126] Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. Radiotherapy and Oncology 2020. https://doi.org/10.1016/j.radonc.2020.09.008.
[127] Speight R, Dubec M, Eccles CL, George B, Henry A, Herbert T, et al. IPEM topical report: guidance on the use of MRI for external beam radiotherapy treatment planning *. Phys Med Biol 2021;66:055025. https://doi.org/10.1088/1361-6560/abdc30.
[128] Kurz C, Buizza G, Landry G, Kamp F, Rabe M, Paganelli C, et al. Medical physics challenges in clinical MR-guided radiotherapy. Radiat Oncol 2020;15:93. https://doi.org/10.1186/s13014-020-01524-4.
[129] Cusumano D, Teodoli S, Greco F, Fidanzio A, Boldrini L, Massaccesi M, et al. Experimental evaluation of the impact of low tesla transverse magnetic field on dose distribution in presence of tissue interfaces. Physica Medica 2018;53:80–5. https://doi.org/10.1016/j.ejmp.2018.08.007.
[130] Raaijmakers AJE, Raaymakers BW, Lagendijk JJW. Experimental verification of magnetic field dose effects. Physics in Medicine and Biology 2007;52:4283–91. https://doi.org/10.1088/0031-9155/52/14/017.
[131] Otazo R, Lambin P, Pignol J-P, Ladd ME, Schlemmer H-P, Baumann M, et al. MRI-guided Radiation Therapy: An Emerging Paradigm in Adaptive Radiation Oncology. Radiology 2021;298:248–60. https://doi.org/10.1148/radiol.2020202747.

ACCEPTED MANUSCRIPT - CLEAN COPY

First author, year [reference]	*No. of patients*	*Magnetic field, system / MRI sequence*	*Preprocessing of MRI images*	*Deep learning method / RR or DR* / 2D or 2.5D or 3D training	*Overall loss function (for GANs) or generator loss function*	*Adversarial loss used by discriminator (for GANs)*	*No. of patients in training / evaluation*	*Image and geometry fidelity results*			*Dose results*
Han et al., 2017 [1]	18	1.5 T Siemens / 3D T1	N3 bias field correction algorithm, histogram matching	DCNN: U-Net / RR / 2D	L1 loss	-	15 (6-fold cross validation)/ 18	**MAE [HU]** 84.8 ± 17.3	**ME [HU]** -3.1 ± 21.6	**MSE [HU²]** 188.6 ± 33.7	/
Nie et al., 2017 [2]	16	3 T Siemens / ❖	/	GAN (with FCN for generator and CNN for discriminator) / ❖/ 3D	L2 loss + gradient loss + adversarial loss	Binary cross entropy	15 (leave-one-out) / 16	**MAE [HU]** 92.5 ± 13.9		**PSNR [dB]** 27.6 ± 1.3	/
Wolterink et al., 2017 [3]	24	1.5 T Philips / 3D T1	/	Cycle-GAN / RR / 2D	Least squares loss + adversarial loss + gradient difference loss + cycle consistency loss (based on L1 norm)	Binary cross entropy	18 / 6	**MAE [HU]** 73.7 ± 2.3		**PSNR [dB]** 32.5 ± 0.7	/
Dinkla et al., 2018 [4]	52	1.5 T Philips / pre contrast 3D T1	3D geometry correction provided by the vendor + intensity normalization	Dilated CNN / RR 2.5D	L1 loss		26 (two-fold cross validation)/ 52	**MAE [HU]** Body = 67 ± 11 Soft tissue = 22 ± 3 Bone = 174 ± 29 Air = 159 ± 22	**ME [HU]** Body = 13 ± 9 Soft tissue = -2 ± 3 Bone = 75 ± 41 Air = -71 ± 27	**DSC** Body contour = 0.98 ± 0.01 Bone = 0.85 ± 0.04	Mean dose difference < 1% γ_{*} = 91.1% γ_{*}^{*} = 95.8% γ_{*}^{**} = 99.3%

ACCEPTED MANUSCRIPT - CLEAN COPY

										Air = 0.71 ± 0.07	
Emami et al., 2018 [5]	15	1 T Philips / Post-Gadolinium T1	/	GAN (ResNet for generator and CNN for discriminator) / RR / 2D	L1 loss + adversarial loss	Least square loss	12 (5-fold cross validation) / 15	**MAE [HU]** Head = 89.3 ± 10.3 Soft tissue = 41.9 ± 8.6 Bone = 255.2 ± 47.7 Air = 240.9 ± 60.2	**PSNR [dB]** Head = 26.6 ± 1.2	**SSIM** Head = 0.83 ± 0.03	/
Xiang et al., 2018 [6]	16	3T Siemens / T1	N3 bias field correction algorithm, intensity histogram matching, intensity normalization	DECNN / RR / 3D	L2 norm + L2-norm of embedding blocks	-	15 (leave-one-out) / 16	**MAE [HU]** 85.4 ± 9.2	**PSNR [dB]** 27.3 ± 1.1		/
Yang et al., 2018 [7]	45	1.5 T Siemens / ❖	3D distortion correction algorithm (vendor) + N4 correction + intensity normalization	Cycle-GAN / ❖ / 2D	Adversarial loss, cycle-consistency loss (L1 norm), structure consistency loss (L1 based on MIND)	Binary cross entropy	27 / 15	**MAE [HU]** 129	**PSNR [dB]** 24.2	**SSIM** 0.78	/
Gupta et al., 2019 [8]	60	3T Siemens / T1 with Dixon	/	U-Net / RR / 2D	Combination of L1 loss: MAE + MSE	-	47 / 13	**MAE [HU]** All voxels = 81.0 ± 14.6 Soft tissue = 17.6 ± 3.4 Bone = 193.1 ± 38.3 Air = 233.8 ± 28.0			For 7 patients: D_{mean} PTV difference = 2.3%

ACCEPTED MANUSCRIPT - CLEAN COPY

Kazemifar et al., 2019 [9]	77	1.5 T ❖ / Post-Gadolinium 2D T1	/	cGAN / RR / 2D	Mutual information	Binary Cross entropy	70% of 77 / 1% of 77	**Average MAE over cross validation sets [HU]** 47.2 ± 11.0		**DSC** Head = 0.96 ± 0.02 Bone = 0.80 ± 0.06 Air = 0.70 ± 0.07	Dose differences < 2.4% for PTV and OARs $\gamma_{*} = 94.6\%$ $\gamma_{*}^{*} = 99.2\%$
					MAE			60.2 ± 22.0		***I***	/
Koike et al., 2019 [10]	15	❖ Philips, GE and Siemens / T1, T2 and FLAIR	N4 bias field correction algorithm	cGAN / DR / 2D	Adversarial loss + L1 loss	Binary cross entropy	12 (5-fold cross validation) / 15	**MAE [HU]** *T1w sequence* Body = 120.1 ± 20.4 Soft tissue = 46.3 ± 9.3 Bone = 399.4 ± 51.8 *All sequences* Body = 108.1 ± 24.0 Soft tissue = 38.9 ± 10.7 Bone = 366.2 ± 62.0			Dose difference < 1% *T1w sequence* $\gamma_{*} = 94.2\%$ $\gamma_{*}^{*} = 98.9\%$ $\gamma_{**}^{*} = 99.7\%$ *All sequences* $\gamma_{*} = 95.3\%$ $\gamma_{*}^{*} = 99.2\%$ $\gamma_{**}^{*} = 99.8\%$
Lei et al., 2019 [11]	24	❖ GE / T1-BRAVO	N3 bias field correction algorithm + intensity normalization	Cycle-GAN / RR / 3D	Adversarial loss + distance loss (combination of Mean P Distance and Gradient Difference)	MAD	23 (leave-one-out) / 24	**MAE [HU]** 55.7 ± 9.4	**PSNR [dB]** 26.6 ± 2.3	**NCC** 0.96 ± 0.01	**DSC** Air = 0.90 ± 0.12 Bone = 0.83 ± 0.06	/
							24/10	57.7 ± 8.4	27.0 ± 2.8	0.96 ± 0.01	/	
Liu et al., 2019 [12]	50	1.5 T GE / 3D T1 BRAVO + 1,5 T GE / post-contrast T1	/	DCNN / DR / 2D	MSE	-	40 / 10	**MAE [HU]** 75 ± 23		**DSC** Soft tissue = 0.94 ± 0.02 Bone = 0.85 ± 0.02	Absolute mean dose difference < 1.5% $\gamma_{**}^{*} = 99.2\%$

ACCEPTED MANUSCRIPT - CLEAN COPY

									Air = 0.95 ± 0.01	
Neppl et al., 2019 [13]	89	1.5 T Siemens / T1 MPRAGE	/	U-Net / RR / 2D	❖	-	57 / 4	**MAE [HU]** 55 ± 10	**ME [HU]** -1 ± 4	$\gamma* = 95\%$ $\gamma^{*}_{*} = 98\%$ *Protontherapy: Range difference of 3 mm < 5% of the profiles* $\gamma^{*}_{*} = 91.7\%$
				U-Net / RR/ / 3D				90 ± 20	11 ± 9	$\gamma* = 95\%$ $\gamma^{*}_{*} = 98\%$ *Protontherapy: Range difference of 3 mm < 5% of the profiles* $\gamma^{*}_{*} = 89.3\%$
Shafai-Erfani et al., 2019 [14]	50	1.5 T Siemens / 3D T1	N4 bias correction filter	Cycle-GAN / RR / 3D	Adversarial loss, MAD (L1 norm), Mean P distance and gradient-difference loss	Mean P distance : lp norm	25 / 25	**MAE [HU]** 54.6 ± 6.8	**NCC** 0.96 ± 0.01	*Protontherapy:* Dose difference < 0.5% Mean distal range: 1.1 ± 0.9 mm $\gamma* = 89.2\%$ $\gamma^{*}_{*} = 98.1\%$ $\gamma^{*}_{**} = 99.9\%$
Spadea and Maspero et al., 2019 [15]	15	3 T Siemens / 3D T1	/	DCNN: U-Net / RR / 2.5D	L1 loss	-	14 (leave-one-out) / 15	**MAE [HU]** Body = 54 ± 7 Cerebrospinal fluid = 10 ± 3 Gray matter = 8 ± 2 White matter = 6 ± 2	**ME [HU]** Body = -4 ± 17 Cerebrospinal fluid = 0 ± 9 Gray matter = 0 ± 6 White matter = 0 ± 4	**DSC** Bone = 0.93 ± 0.02 Air = 0.92 ± 0.03	*Protontherapy:* PTV dose difference < 0.1 Gy Relative range shift: 0.14 ± 1.11%

ACCEPTED MANUSCRIPT - CLEAN COPY

								Air = 53 ± 32	Air = -37 ± 39		
Alvarez Andres et al., 2020 [16]	402	1.5 T + 3 T GE / T1 and FLAIR	N4 filter	CNN (HighResNet)/ RR / 2D	-		242 / 79	**MAE [HU]** Head = 81 ± 22 Water = 38 ± 11 Bone = 228 ± 63 Air = 274 ± 63			Dose differences: PTV ($D_{2\%}$, $D_{50\%}$, $D_{95\%}$, $D_{98\%}$) < 0.3% $\gamma* = 97.9\%$ $\gamma^{*}_{*} = 99.6\%$ $\gamma^{*}_{**} = 99.8\%$
				U-Net / RR / 3D	MAE	-		**MAE [HU]** Head = 90 ± 21			$\gamma^{*}_{**} = 99.7\%$
Hemsley et al., 2020 [17]	105	1.5 T Philips / T1 and FLAIR	MR normalization	cGAN (pix2pix)/ RR / 2D	adversarial + log-likelihood of the Laplace distribution	Binary cross entropy	85 / 20	**MAE [HU]** Normal soft tissue = 6 ± 3 Air-bone interface = 237 ± 31			/
Kazemifar et al., 2020 [18]	77	1.5 T ❖ / 2D T1	/	cGAN / no registration / 2D	Mutual information	Binary cross entropy	54 / 11	**Average MAE over all-cross validation sets [HU]** 47.2 ± 11.0			*Protontherapy:* Mean dose difference < 1.8% Up to 5.1% for brainstem $D_{2\%}$
Li et al., 2020 [19]	34	1.5 T Siemens / T2	Distorsion correction	U-Net / RR / 2D	L1	-	28 / 6	**MAE [HU]** 75.5 ± 7.9	**PSNR [dB]** 25.4 ± 0.6	**SSIM** 0.94 ± 0.01	**MSE [HU]** $4.9{\cdot}10^4 \pm 7.7{\cdot}10^3$	/
					L2			75.5 ± 11.7	32.2 ± 1.1	0.94 ± 0.01	$1.1{\cdot}10^4 \pm 3.1{\cdot}10^3$	
					L1 + L2			74.2 ± 12.8	32.4 ± 1.2	0.94 ± 0.01	$1.0{\cdot}10^4 \pm 3.1{\cdot}10^3$	
				cGAN (pix2pix) / RR / 2D	❖	❖		94.6 ± 17.2	30.3 ± 1.2	0.91 ± 0.01	$1.6{\cdot}10^4 \pm 4.9{\cdot}10^3$	
				Paired Cycle-Gan / RR / 2D	❖	❖		87.7 ± 7.9	30.9 ± 0.6	0.92 ± 0.00	$1.4{\cdot}10^4 \pm 2.1{\cdot}10^3$	

ACCEPTED MANUSCRIPT - CLEAN COPY

				Unpaired Cycle-Gan	❖	❖		98.7 ± 12.7	29.9 ± 0.9	0.91 ± 0.01	$1.7 \cdot 10^4 \pm 3.3 \cdot 10^3$	
Liu et al., 2020 [20]	15	1.0 T Philips / Postgadolinium T1	/	GAN (ResNet for generator and CNN for discriminator) / RR / 2D	L1 loss + adversarial loss	L2 loss	12 / 3 (5-fold cross validation)	**Emami et al., 2018** [5]				Dose differences: PTV and OARs < 0.13 Gy γ_{*}= 99.0% γ_{*}^{*}= 99.9%
Maspero et al., 2020 [21]	60 (paediatric patients)	1.5 T and 3 T Philips / 3D T1	MRI was normalised and clipped to their 99th percentile intensity over the whole volume	cGAN / RR / 2D	L1 loss	❖	30 / 20 (4-fold cross-validation)	**MAE [HU]** 61.0 ± 14.1	**PSNR [dB]** 26.7± 1.9 dB	**SSIM** 0.86±0.03	**DSC** Body = 0.984 ± 0.004 Bone = 0.92 ± 0.05	*Photontherapy:* Dose differences < 0.9% γ_{*}^{*} > 93.9% γ_{**}^{*} >98.4% *Protontherapy:* Dose differences < 1% γ_{*}^{*} > 92.6% γ_{**}^{*} >97.2%
Massa et al., 2020 [22]	92	1.5 T GE / T1w, T2-FatSat, T1 Post contrast, T2 CUBE FLAIR	/	U-net with Inception V3 inspired blocks / DR / 2D	❖	-	81 / 11	**MAE [HU]** T1w: Brain = 51.2 ± 4.5 Bone = 31.1 ± 7.0 T2-FatSat: Brain = 45.7 ± 8.8 Bone = 30.3 ± 7.1 T1 Post contrast: Brain = 44.6 ± 7.5 Bone = 30.2 ± 6.0	**PSNR [dB]** T1w: Brain = 43.0 ± 2.0 Bone = 43.2 ± 1.9 T2-FatSat: Brain = 43.4 ± 1.2 Bone = 43.7 ± 1.2	**SSIM** T1w: Brain = 0.65 ± 0.05 Bone = 0.87 ± 0.03 T2-FatSat: Brain = 0.63 ± 0.03 Bone = 0.86 ± 0.02	**DSC** T1w: Soft tissue = 0.91 ± 0.03 Bone = 0.76 ± 0.12 T2-FatSat: Soft tissue = 0.91 ± 0.02 Bone = 0.77 ± 0.07	/

ACCEPTED MANUSCRIPT - CLEAN COPY

								T2 CUBE FLAIR: Brain = 51.2 ± 4.5 Bone = 36.1 ± 3.3	T1 Post contrast: Brain = 43.4 ± 1.2 Bone = 43.7 ± 1.2 T2 CUBE FLAIR: Brain = 44.9 ± 1.2 Bone = 45.2 ± 1.1	T1 Post contrast: Brain = 0.64 ± 0.03 Bone = 0.86 ± 0.03 T2 CUBE FLAIR: Brain = 0.61 ± 0.04 Bone = 0.84 ± 0.02	T1 Post contrast: Soft tissue = 0.90 ± 0.02 Bone = 0.75 ± 0.07 T2 CUBE FLAIR: Soft tissue = 0.88 ± 0.03 Bone = 0.69 ± 0.07	
Tang et al., 2020 [23]	37	3 T Siemens / T1 TIRM	/	cGAN (pix2pix)/ RR / ❖	MAE + adversarial loss	Least square loss	27 (5 fold cross-validation) / 10	**MAE [HU]** 60.5 ± 13.3	**PSNR [dB]** 49.2 ± 1.9			Dose differences: PTV < 0.13% Brainstem, optic chiasma, optic nerve < 0.77% γ^{*}_{*} = 97.3% γ^{*}_{**} = 99.8%
Bourbonne et al., 2021 [24]	184	1.5 T Siemens / T2 and post gadolinium 3D T1	MRI voxel intensities were clipped to be inside the [1-99%] quantile range.	cGAN (pix2pix) / ❖ / 2D	Adversarial loss and MAD (L1 norm)	PatchGAN	20 / 164	**RMSE [HU]** Soft-tissue = 13.54 ± 1.96 Bone = 175.50 ± 63.15				Dose differences: PTV < 0.4 Gy Local γ^{*}_{*} = 99.1%

Additional table 1: Synthetic-CT generation from brain MRI in the literature: summary of data, deep learning architecture, and image and dose evaluations

$\gamma*$ = 1%/1 mm gamma pass-rate 1%/1mm; γ^{*}_{*} = 2%/2 mm gamma pass-rate; γ^{*}_{**} = 3%/3 mm gamma pass-rate; ❖ = not specified in the study

ACCEPTED MANUSCRIPT - CLEAN COPY

Abbreviations: GAN = generative adversarial network, MAD = mean absolute distance, MAE = Mean Absolute Error; ME = Mean Error; DSC = Dice Similarity Coefficient; PSNR = Peak Signal-to-Noise Ratio; SSIM = Structural Similarity; NCC = Normalized Cross-Correlation; MSE = Mean Square Error; RR = Rigid Registration; DR = Deformable Registration; DECNN = Deep Embedding Convolutional Neural Network; DCNN = Deep CNN.

ACCEPTED MANUSCRIPT - CLEAN COPY

First author, year [reference]	*No. of patients*	*Magnetic field, system / MRI sequence*	*Preprocessing of MRI images*	*Deep learning method / RR or DR / 2D or 2.5D or 3D training*	*Overall loss function (for GANs) or generator loss function*	*Adversarial loss used by discriminator (for GANs)*	*No. of patients in training / evaluation*	*Image and geometric fidelity results*			*Dose results*
Dinkla et al., 2019 [25]	34	3 T Philips / T2 - Dixon	2D geometry correction + uniformity correction (CLEAR)	U-Net / DR / 3D	L1	-	22 (3-fold cross validation) / 34	**MAE [HU]** Body = 75 ± 9	**ME [HU]** Body = 9 ± 11	**DSC** Bone = 0.70 ± 0.07 Air = 0.79 ± 0.08	Dose difference < 1% γ^{*}_{*} = 95.6% γ^{*}_{**} = 98.7%
Klages et al., 2019 [26]	20	3 T Philips / mDixon T1 Fast Field Echo (FFE)	Intensity inhomogeneity correction (CLEAR)	cGAN (Pix2Pix) / DR / 2D	Adversarial loss + Absolute differences (L1)	Binary cross entropy	10 / 10	**MAE [HU]** Body = 92.4 ± 13.5	**ME [HU]** Body = 21 ± 11.8		Dose differences: PTV_{70Gy} < 0.8% OARs < 2%
				cycle-Gan / DR / 2D				Body = 100.7 ± 14.6	Body = 37.5 ± 14.9		Dose differences: PTV_{70Gy}< 1.6% OARs < 1.7%
Wang et al., 2019 [27]	33	1.5 T Siemens / T2 TSE	Histogram matching	DCNN: U-Net / RR and DR / 2D	L2: MSE	-	23 / 10	**MAE [HU]** Body = 131 ± 24	**ME [HU]** Body = −6 ± 13		For one patient: PTV $D_{98\%}$ difference < 1%

ACCEPTED MANUSCRIPT - CLEAN COPY

Largent et al., 2020 [28]	8	1.5 T GE / 3D T2	N4 bias-field correction and histogram matching	GAN / RR and DR / 2D	Perceptual loss	Cross entropy	7 (leave-one-out) / 8	**MAE [HU]** 82.8 ± 48.6			**ME [HU]** -3.9 ± 12.8		/
Peng et al., 2020 [29]	173	3 T Philips / T1	Intensity of each MR volume image was normalized as zero mean and unit variance and then scaled to a similar numeric range	cGAN (registered pairs) / DR / 2D	Adversarial loss + L1 loss	Least square loss	135 pairs / 28 pairs	**MAE [HU]** Body = 69.7 ± 9.3			**ME [HU]** Body = 18.4 ± 16.4		Dose differences: PTV ($D_{95\%}$) < 0.7% Parotids (D_{mean}) < 0.9% Mandible (D_{max}) < 1.5% γ^{*}_{*} = 98.7% γ^{*}_{**} = 99.6%
				cycle-Gan (unregistered pairs) / - / 2D	Cycle-consistency loss (MAE loss) + adversarial loss + L1 loss	PatchGAN loss		Body = 100.6 ± 7.7			Body = 6.7 ± 19.4		Dose differences: PTV ($D_{95\%}$) < 0.9% OARs (D_{mean}) < 1.2% Mandible (D_{max}) < 1.6% γ^{*}_{*} = 98.5% γ^{*}_{**} = 99.6%
Qi et al., 2020 [30]	45	3 T Philips / T1, T2, contrast-	Normalization to [-1, 1] based on the minimum	cGAN / RR / 2D	Adversarial loss + L1 loss	PatchGAN loss	30 / 15	**MAE [HU]** Body :	**ME [HU]** Body :	**PSNR [dB]** T1 = 28.9 ± 1.1	**SSIM** T1 = 0.84 ± 0.02	**DSC** Bone:	Dose difference < 1%

ACCEPTED MANUSCRIPT - CLEAN COPY

		enhanced T1 (T1C) and contrast-enhanced T1 with Dixon (T1DixonC)	and maximum intensity values					T1 = 75.2 ± 11.5 T2 = 87.0 ± 10.8 T1C = 80.0 ± 10.9 T1DixonC = 86.3 ± 10.8 Multiseq. = 70.0 ± 12.0	T1 = 1.3 ± 14.9 T2 = 12.3 ± 16.2 T1C = 5.0 ± 16.0 T1DixonC = 6.4 ± 16.7	T2 = 27.5 ± 0.9 T1C = 28.4 ± 1.1 T1DixonC = 27.7 ± 0.9 Multiseq. = 29.4 ± 1.3	T2 = 0.78 ± 0.03 T1C = 0.82 ± 0.03 T1DixonC = 0.79 ± 0.03 Multiseq. = 0.85 ± 0.03	T1 = 0.74 ± 0.05 T2 = 0.68 ± 0.05 T1C = 0.72 ± 0.05 T1DixonC = 0.69 ± 0.06 Multiseq. = 0.77 ± 0.05	T1: $\gamma^{\ddagger}_{*}$ = 97.4% T2: $\gamma^{\ddagger}_{*}$ = 95.8% T1C: $\gamma^{\ddagger}_{*}$ = 96.7% T1DixonC: $\gamma^{\ddagger}_{*}$ = 96.2% Multiseq.: $\gamma^{\ddagger}_{*}$ = 97.8%
				U-Net / RR / 2D	L1 loss	-		71.3 ± 12.4	/	29.2 ± 1.3	/	0.76 ± 0.05	Dose difference < 0.9% $\gamma^{\ddagger}_{*}$ = 97.6%
Thummerer et al., 2020 [31]	27	3 T Siemens / 3D spoiled gradient recalled echo	/	DCNN / DR / 2.5D	L1 loss	-	18 (3-fold cross validation) / 27	**MAE [HU]** Body = 65.4 ± 3.6		**ME [HU]** Body = 2.9 ± 9.4			*Protontherapy: Range errors > 3%* $\gamma^{\ddagger}_{*}$ = 97.6%
Tie et al., 2020 [32]	32	1.5 T Siemens / T1 precontrast + T1 postcontr	N4 bias correction algorithm + histogram normalization	cGAN (Pix2Pix) / RR / 2D	Combination of Pearson f-divergence + adversarial loss + L1 loss	Least square loss	28 (8-fold cross validation) / 32	**MAE [HU]** Head = 75.7 ± 14.6	**PSNR [dB]** 29.1 ± 1.6	**DSC** Bone = 0.86 ± 0.03	**SSIM** 0.92 ± 0.02		/

ACCEPTED MANUSCRIPT - CLEAN COPY

		ast and T2										
Palmér et al., 2021 [33]	44	1.5 T Siemens / T1w Dixon Vibe	❖	Deep CNN / RR + DR / 2D	❖	-	80 (multicenter database) / 44	**MAE [HU]** Body = 67 ± 14	**ME [HU]** Body = -5 ± 10	**DSC** Bones = 0.80 ± 0.07 Air =0.81 ± 0.1	**HD [mm]** Bones = 4.6 ± 1.2 Air = 2.8 ± 0.8	Mean dose differences PTV and OARs < 0.3% γ (2%/1 mm)=99.4%

Additional table 2: Synthetic-CT generation from head and neck MRI in the literature: summary of data, deep learning architecture, and image and dose evaluations

γ ⁑ = 2%/2 mm gamma pass-rate; γ ⁂ = 3%/3 mm gamma pass-rate; ❖ = not specified in the study.

Abbreviations: GAN = generative adversarial network, MAE = Mean Absolute Error; ME = Mean Error; HD = Hausdorff distance, DSC = Dice Similarity Coefficient; PSNR = Peak Signal-to-Noise Ratio; SSIM = Structural Similarity; RR = Rigid Registration; DR = Deformable Registration; MSE = mean square error, DCNN = Deep Convolutional Neural Network.

ACCEPTED MANUSCRIPT - CLEAN COPY

First author, year [reference]	*Anatomical site*	*No. of patients*	*Magnetic field, system / MRI sequence*	*Preprocessing of MRI images*	*Deep learning method / RR or DR / 2D or 2.5D or 3D training*	*Overall loss function (for GANs) or generator loss function*	*Adversarial loss used by discriminator (for GANs)*	*No. of patients in training / evaluation*	*Image and geometric fidelity results*		*Dose results*
Jeon et al., 2019 [34]	Breast	16	0.35 T, MRIdian (Viewray) / ❖	/	U-Net / DR / 2D	Binary cross-entropy	-	14/2	**DSC (%)** Patient 1 = 76.4 Patient 2 = 73.5		/
Liu et al., 2019 [35]	Liver	21	3 T Skyra / 3D T1 + 3 T TrioTim Siemens / T1 VIBE and 1.5 T GE / 2D T1	N4 bias field correction algorithm	cycle-GAN / DR / 3D	Weighted summation of adversarial loss and mean p distance (lp norm) + gradient difference	MAD	20 (leave-one-out) / 21	**MAE [HU]** Body = 72.9 ± 18.2	**PSNR [dB]** 22.4 ± 3.6	*Protontherapy:* PTV_{45Gy} and bowel dose difference < 1 Gy PTV $D_{95\%}$ < 1.1% Mean of absolute maximum range shift: 1.86 ± 1.55 mm $\gamma_{*} = 90.8\%$ $\gamma_{*}^{*} = 97.0\%$ $\gamma_{**}^{*} = 99.4\%$
Liu et al., 2019 [36]	Liver	21	3 T Skyra/ 3D T1 + 3 T TrioTim Siemens / T1 VIBE and 1.5 T GE / 2D T1	N4 bias field correction filter	cycle-GAN / RR and DR / 3D	Weighted summation of MAD (adversarial loss) and mean p distance (lp norm) +	MAD	20 (leave-one-out) / 21	**MAE [HU]** Body = 72.9 ± 18.2	**PSNR [dB]** 22.7 ± 3.6	Mean dose difference: PTV $D_{95\%}$ = 0.2% Dose difference to OARs < 0.06 Gy

ACCEPTED MANUSCRIPT - CLEAN COPY

						gradient difference						γ_{*} = 99.0%
Olberg et al., 2019 [37]	Breast	48	0.35 T Viewray / T1	Histogram matching	U-Net and ASPP / DR / 2D	Adversarial (Cross entropy loss) + MAE	Cross entropy loss	48 / 12	**RMSE [HU]** 17.7 ± 4.3	**PSNR [dB]** 71.7 ± 2.3	**SSIM** 0.9995 ± 0.0003	For 4 patients PTV $D_{95\%}$ < 0.9% $\gamma^{*}_{\ddagger} \geq 98.0\%$
Xu et al., 2019[38]	Abdomen	10	❖ / mDixon	❖	MCRcGAN (ResNet + multichannel cGAN) / DR / 2D	L1 loss + cGAN loss (cross entropy loss)	Cross entropy loss	9 (leave-one-out) / 10	**MAE [HU]** 60.4			/
Cusumano et al., 2020 [39]	Pelvis and abdomen	60 pelvis / 60 abdominal	0.35 T MRIdian (Viewray) / T2/T1	❖	cGAN (pix2pix) / DR / 2D	$L_{GAN}+\lambda \cdot L1$ with $\lambda = 100$ L_{GAN} : adversarial loss function	❖	40 pelvis, 40 abdomen/20 abdomen	**MAE [HU]** Body = 78.7 ± 18.5		**ME (HU)** Body = 10.8 ± 12.9	γ_{*} = 90.8% $\gamma^{*}_{\ddagger}$ = 98.7% γ^{*}_{**} = 99.8%
Florkow et al., 2020 [40]	Wilms' tumour (24) and Neuroblastoma (42)	66	1.5 T Philips / 3D T1 and 3D T2	Intensities clipped beyond the 95th percentile + resulting intensities linearly mapped to [-1; 1].	U-Net / DR / 3D	L1	-	54/12 (3- fold cross validation)	**MAE [HU]** Body = 57 ± 12 Lungs = 105 ± 34	**ME [HU]** Body = -5 ± 12 Lungs = -9 ± 67	**PSNR (dB)** Body = 30.3 ± 1.6	**DSC (%)** Bones = 76 ± 8 Lungs = 92 ± 9	Photontherapy: Mean dose difference <0.5% $\gamma^{*}_{\ddagger}$ > 99% Protontherapy: $\gamma^{*}_{\ddagger}$ > 96%
Fu et al., 2020 [41]	Liver + Abdomen	12 (8 liver + 1 pancreas + 1 adrenal gland + 2 middle abdomen tumors)	0.35 T, MRIdian (Viewray) / True-FISP	N4 bias field correction algorithm + histogram normalization	cGAN / DR / 2D	cGAN + L1	Binary Cross entropy loss	9 (4-fold cross validation)/ 12	**MAE [HU]** Body = 89.8 ± 18.7		**PSNR [dB]** 27.4 ± 1.6	For 8 liver patients Dose difference:< 0.6% $\gamma^{*}_{\ddagger}$ = 97.4% γ^{*}_{**} = 99.5%

ACCEPTED MANUSCRIPT - CLEAN COPY

					cycle-GAN / DR / 2D	Adversarial loss + cycle consistency	Binary Cross entropy loss		Body = 94.1 ± 30	27.2 ± 2.2	For 8 liver patients Dose difference < 1% γ^{*}_{*} = 95.6% γ^{*}_{**} = 99.3%
Liu et al., 2020 [42]	Abdomen	46	3 T Skyra Siemens/ 3D T1 ECHO	N4 Intensity inhomogeneity correction	U-Net / RR / 2.5D	❖	-	31/46 (3 fold cross validation)	**MAE [HU]** Liver = 24.1 ± 11.5 Kidneys = 47.1 ± 15 Spinal Cord = 29.8 ± 9.4 Lungs = 105.7 ± 35.0		Mean dose difference < 0.15 Gy
Qian et al., 2020 [43]	Abdomen	10	❖ / mDixon	❖	RU-ACGAN : hybrid network associated with AC-GAN and cGAN and combined with ResNet and U-Net simultaneously / DR /	L1 loss + adversarial loss	Binary cross entropy loss + Softmax loss	9/10 (leave-one-out)	**MAE [HU]** 55.6 ± 5.6	**RMSE [HU]** 106.4 ± 12.2	/

Additional table 3: Synthetic-CT generation from breast, liver, and abdomen MRI in the literature: summary of data, deep learning architecture, and image and dose evaluations

$\gamma *$ = 1%/1 mm gamma pass-rate 1%/1mm, $\gamma \overset{*}{*}$ = 2%/2 mm gamma pass-rate, $\gamma \overset{*}{**}$ = 3%/3 mm gamma pass-rate, ❖ = not specified in the study

Abbreviations: MAE = Mean Absolute Error, PSNR = Peak to Signal-to-Noise Ratio, SSIM = Structural Similarity, NCC = Normalized Cross-Correlation, RMSE = Root Mean Square Error, MAD = mean absolute distance, GAN = generative adversarial network, ASPP = atrous spatial pyramid pooling, RU-ACGAN = Auxiliary classifier-augmented GAN.

ACCEPTED MANUSCRIPT - CLEAN COPY

First author, year [reference]	*Anatomical site*	*No. of patients*	*Magnetic field, system / MRI sequence*	*Preprocessing of MRI images*	*Deep learning method / RR or DR / 2D or 2.5D or 3D training*	*Overall loss function (for GANs) or generator loss function*	*Adversarial loss used by discriminator (for GANs)*	*No. of patients in training / evaluation*	*Image and geometric fidelity results*				*Dose results*
Nie et al., 2016 [44]	Pelvis	22	❖	/	FCN / Manual alignment / 3D	❖	-	21 (leave-one-out) / 22	**MAE [HU]** Body = 42.4 ± 5.1		**PSNR [dB]** Body = 33.4 ± 1.1		/
Nie et al., 2017 [2]	Pelvis	22	❖	❖	GAN (with FCN for generator and CNN for discriminator)/ Manual alignment / 3D	L2 loss + gradient loss + adversarial loss	Binary cross entropy	21 (leave-one-out) / 22	**MAE [HU]** Body = 39.0 ± 4.6		**PSNR [dB]** Body = 34.1 ± 1.0		/
Arabi et al., 2018 [45]	Prostate	39	3 T Siemens / 3D T2 SPACE	N3 bias field correction algorithm	DCNN: U-Net / RR and DR / 2D	MAE	-	(4-fold cross validation) ¾ of 39 / 39	**MAE [HU]** Body = 32.7 ± 7.9	**ME [HU]** Body = 3.5 ± 11.7	**DSC** Bone = 0.93 ± 0.02	**MASD [mm]** Body = 1.8 ± 0.6	Mean dose difference < 0.5% Dose difference: CTV (D_{max}) = 2.9% OARs (D_{max})= 0.5% γ_{*} = 94.6% γ_{*}^{*} = 98.5% γ_{**}^{*} = 99.2%

ACCEPTED MANUSCRIPT - CLEAN COPY

Chen et al., 2018 [46]	Prostate	51	3 T Philips / T2 TSE	No bias field correction and MR intensity histogram normalization	U-Net / DR / 2D	MAE	-	36 / 15	**MAE [HU]** Body = 30.0 ± 4.9	**ME [HU]** Body = 6.7 ± 5.4	Dose differences: 0.2% (PTV/OARs) γ_{*} = 98.0% γ_{*}^{*} = 99.4% γ_{**}^{*} = 99.8%
Maspero et al., 2018 [47]	Prostate	59	3 T Philips / 3D Echo with Dixon	Normalized to their 95% intensity interval over the whole patient + converted to 8-bits	cGAN / RR / 2D	Adversarial loss +λ·L1 with λ = 100	PatchGAN loss	32 / 27	**MAE [HU]** Body = 65 ± 10	**ME [HU]** Body = 1 ± 6	For 10 patients /27 Dose difference: PTV ($D_{98\%}$)< 1% γ_{*}^{*} = 95.0% γ_{**}^{*} = 98.1%
	Rectum	18	3 T Philips / 3D Echo with Dixon					32 (prostate patients) / 18	**MAE [HU]** Body = 56 ± 5	**ME [HU]** Body = 2 ± 9	For 10 patients /18 Dose difference: PTV ($D_{98\%}$)< 1.2% γ_{*}^{*} = 91.6% γ_{**}^{*} = 97.1%
	Cervix	14	❖ / 3D Echo with Dixon					32 (prostate patients) / 14	**MAE [HU]** Body = 59 ± 6	**ME [HU]** Body = 4 ± 10	For 10 patients /14 Dose difference: PTV ($D_{98\%}$) < 1% γ_{*}^{*} = 90.6% γ_{**}^{*} = 97.1%

ACCEPTED MANUSCRIPT - CLEAN COPY

Xiang et al., 2018 [6]	Prostate	22	1.5 T Siemens / T1	N3 correction algorithm + histogram matching + intensity normalization	DECNN/ DR and RR / 2D	L2	-	21 (Leave-one-out) / 22	**MAE [HU]** Body = 42.5 ± 3.1		**PSNR [dB]** Body = 33.5 ± 0.8	/
Florkow et al., 2019 [48]	Prostate	23	3 T Philips / T1	/	U-Net / DR / 2D	L1	-	16 (3-fold cross-validation) / 23	**MAE [HU]** Body = 34.1 ± 7.9	**ME [HU]** Body = -1.2 ± 4.6	**PSNR [dB]** Body = 34.7 ± 1.7	**DSC** Bone = 0.78 ± 0.10 /
Florkow et al., 2019 [49]	Pelvis	24	3 T Philips / echo sequence with Dixon	/	U-Net / RR and DR / 2D	L1	-	11 (3-fold cross-validation) / 24	**MAE [HU]** Body = 27.6 ± 2.6		**DSC** Bone = 0.89 ± 0.02	/
Fu et al., 2019 [50]	Prostate	20	1.5 T Siemens / 2D T1 TSE	N4 bias field correction algorithm + histogram normalization	CNN / DR / 2D	MAE	-	16 / 4	**MAE [HU]** Body = 40.5 ± 5.4		**DSC** Bone = 0.81 ± 0.04	/
					CNN / DR / 3D				Body = 37.6 ± 5.1		Bone = 0.82 ± 0.04	
					Han's model [1]				Body = 41.9 ± 6;5		Bone = 0.80 ± 0.05	
Largent et al., 2019 [51]	Prostate	39	3 T Siemens / 3D T2 SPACE	Normalization and correction of image nonuniformity	U-Net / RR and DR / 2D	L2	-	25 (3-fold cross-validation) / 39	**MAE [HU]** Body = 34.4 ± 7.7		**ME [HU]** Body = -1.0 ± 14.2	Mean dose difference < 0.6% (PTV and OARs) $\gamma^* = 99.2\%$

ACCEPTED MANUSCRIPT - CLEAN COPY

					GAN / RR and DR / 2D	L2 + adversarial loss	Binary cross entropy		Body = 34.1 ± 7.5		Body = -1.1 ± 13.7		Mean dose difference < 0.6% (PTV and OARs) $\gamma^* = 99.1\%$
Lei et al., 2019 [11]	Prostate	20	❖ Siemens / 3D T2 SPACE	N3 correction algorithm + intensity normalization	Cycle-GAN / RR / 3D	Combination mean P distance of adversarial loss and distance loss gradient difference	MAD	19 (Leave-one-out) / 20	**MAE [HU]** 50.8 ± 15.5	**PSNR [dB]** 24.5 ± 2.6	**DSC** Air = 0.75 ± 0.06 Bone = 0.81 ± 0.05		/
								20/10	42.3 ± 15.5	23.9 ± 2.0	/		
Liu et al., 2019 [52]	Prostate	17	1.5 T Siemens / T2	/	DCNN / RR / 3D	Combination mean P distance of adversarial loss and distance loss gradient difference	MAD	16 (Leave-one-out) / 17	**MAE [HU]** Body = 59.0 ± 18.6	**PSNR [dB]** Body = 23.7 ± 3.0	**DSC** Bone = 0.81 ± 0.06	**HD 95% [mm]** Bone = 4.9 ±1.5	*Proton therapy* Mean absolute dose difference = 0.23% PTV DVH points difference < 1.5% Rectum DVH difference < 5% Bladder DVH
					GAN / RR / 3D	Combination mean P distance of adversarial loss and distance loss gradient difference			Body = 74.7 ± 20.0	Body = 22.1 ± 2.7	Bone = 0.81 ± 0.06	Bone = 4.7 ± 1.4 mm	
					cycle-GAN / RR / 3D	Combination of mean P distance		16 (leave	Body = 51.3				

ACCEPTED MANUSCRIPT - CLEAN COPY

						(MPS: lp norm) and gradient difference		one out) / 17	± 16.9 HU	24.2 ± 2.46 dB	Bone = 0.85 ± 0.05	4.2 ± 1.0 mm	difference < 11% Mean of absolute maximum range shift: 2.3 ± 2.5 mm γ_{*} = 92.4% γ_{**} = 98.0% γ_{***} = 99.0%
Stadelmann et al., 2019 [53]	Pelvis	42	3 T Philips / 3D FFE mDixon	/	U-Net and LinkNet / RR / 2D	L1 loss	-	27 / 15	**MAE [HU]** 41.4				/
Bahrami et al., 2020 [54]	Male Pelvis	15 (+4 additional patient for evaluation)	3 T Siemens Skyra / 3D T2	N4 correction algorithm + intensity normalization	eCNN / DR / 2D	MAE		15 (5-fold cross validation) / 4	**MAE [HU]** Body = 38 ± 5.6	**ME [HU]** Body = 6 ± 13.4	**PSNR [dB]** Body = 29.5 ± 1.3	**SSIM** Body = 0.96	**DSC** Bone = 0.77 ± 0.03	/
					U-Net	MAE			**MAE [HU]** Body = 46 ± 5.7	**ME [HU]** Body = 6.2 ± 17.9	**PSNR [dB]** Body = 27.	**SSIM** Body = 0.95	**DSC** Bone = 0.7 ± 0.09	/

ACCEPTED MANUSCRIPT - CLEAN COPY

										4 ± 0.6			
Bird et al., 2020 [55]	Ano-rectal	90 (73 rectum and 17 anus)	1.5 T Siemens / T2 SPACE	CT & MR voxels outside the patient external contour set to an intensity of 1024 and 0 respectively	cGAN / DR and RR / 2D	Focal regression loss	Focal regression loss	46 / 44	**MAE [HU]** Deformable registration: Body = 35.1 Rigid registration: Body = 44.5		**ME [HU]** Deformable registration: Body = 0.4 Bone = -95.5 Rigid registration: Body = 0.8		Dose differences: PTV Rectum ($D_{95\%}$, $D_{50\%}$ and $D_{2\%}$) < 0.7% PTV Anus ($D_{95\%}$, $D_{50\%}$ and $D_{2\%}$) < 0.5% $\gamma* = 99.5\%$ $\gamma^{*}_{*} = 99.8\%$ $\gamma^{*}_{**} = 100\%$
Brou Boni et al., 2020 [56]	Male pelvis	19 male pelvis from Gold Atlas data set [57]	1.5 T Siemens / T2 TSE, 3 T GE Discovery / T2 FRFSE + 3 T Signa GE / T2 FRFSE	/	cGAN / DR / 2D	L1+ Pearson divergence + adversarial loss	Least square loss	From Gold Atlas data set 11 / 8	**MAE [HU]** Body = 48.5 ± 6		**ME [HU]** Body = -18.3		DVH points difference for PTV, rectum wall, bladder wall, femoral heads < 1.4%
Cusumano et al., 2020 [39]	Pelvis and abdomen	60 pelvis / 60 abdominal	0.35 T MRIdian / T2/T1 image: TrueFISP	/	cGAN (pix2pix) / DR / 2D	adversarial loss +λ·L1 with λ = 100	PatchGAN loss	80/20	**MAE [HU]** Body = 54.3 ± 11.9		**ME [HU]** Body = 1.4 ± 8.6		Dose difference: PTV ($D_{98\%}$,$D_{50\%}$, $D_{2\%}$) < 0.07 Gy Rectum ($D_{98\%}$,$D_{50\%}$, $D_{2\%}$) < 0.05 Gy $\gamma* = 89.3\%$

ACCEPTED MANUSCRIPT - CLEAN COPY

											γ^{*}_{*} = 99.0% γ^{*}_{**} = 99.9%
Fetty et al., 2020 [58]	Pelvis (male pelvis + cervix)	40 prostate + 11 cervix + 19 male pelvis from Gold Atlas data set [57]	0.35 T Siemens Magnetom / 2D T2 TSE + 1.5 T Siemens / T2 TSE, 3 T GE Discovery / T2 FRFSE + 3 T Signa GE / T2 FRFSE	N4 bias field correction algorithm	cGan (Pix2Pix): 4 tested networks: SE-ResNet, DenseNet, U-Net, Embedded Net / RR and DR / 2D	L1 + adversarial loss	PatchGAN loss	25 / 10 +18 (19 with 1 patient excluded) from Gold Atlas data set [57]	**MAE [HU]** 4 tested networks *0.35 T* Body = 41.2 ± 3.7 *1.5 T* Body = 52.0 ± 5.5 *3 T Discovery* Body = 43.7 ± 6.2 *3 T Signa* Body = 48.2 ± 4.9	**PSNR [dB]** 4 tested networks *0.35 T* Body = 31.4 ± 1 *1.5 T* Body = 29.3 ± 1.0 *3 T Discovery Body* = 31.1 ± 1.1 *3 T Signa* Body = 30.8 ± 1.2	**MSE [100 HU²]** 4 tested networks *0.35 T* Body = 124.9 ± 29.6 *1.5 T* Body = 201.4 ± 53.3 *3 T Discovery* Body = 133.1 ± 35.0 *3 T Signa* Body = 146.2 ± 45.9	DVH points difference: $D_{98\%}$, $D_{50\%}$, $D_{2\%}$ for PTV, rectum, bladder and femoral heads < 1.5%

Additional table 4: Synthetic-CT generation from pelvis MRI in the literature: summary of data, deep learning architecture, and image and dose evaluations

$\gamma*$ = 1%/1 mm gamma pass-rate 1%/1mm; γ^{*}_{*} = 2%/2 mm gamma pass-rate; γ^{*}_{**} = 3%/3 mm gamma pass-rate; ❖ = not specified in the study

Abbreviations: LF = Loss Function; MAE = Mean Absolute Error; ME = Mean Error; DSC = Dice Similarity Coefficient; PSNR = Peak to Signal-to-Noise Ratio; NCC = Normalized Cross-Correlation; MSE = Mean Square Error; MASD = Mean Absolute Surface Distance; HD = Hausdorff Distance; MAD = mean absolute distance, RR = Rigid Registration; DR = Deformable Registration; DCNN = Deep CNN ; FCN = Fully convolution network.

ACCEPTED MANUSCRIPT - CLEAN COPY

References

[1] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Medical Physics 2017;44:1408–19. https://doi.org/10.1002/mp.12155.

[2] Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017, vol. 10435, Cham: Springer International Publishing; 2017, p. 417–25. https://doi.org/10.1007/978-3-319-66179-7_48.

[3] Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, Berg CAT van den, Isgum I. Deep MR to CT Synthesis using Unpaired Data. ArXiv:170801155 [Cs] 2017.

[4] Dinkla AM, Wolterink JM, Maspero M, Savenije MHF, Verhoeff JJC, Seravalli E, et al. MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network. International Journal of Radiation Oncology*Biology*Physics 2018;102:801–12. https://doi.org/10.1016/j.ijrobp.2018.05.058.

[5] Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Medical Physics 2018;45:3627–36. https://doi.org/10.1002/mp.13047.

[6] Xiang L, Wang Q, Nie D, Zhang L, Jin X, Qiao Y, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. Medical Image Analysis 2018;47:31–44. https://doi.org/10.1016/j.media.2018.03.011.

[7] Yang H, Sun J, Carass A, Zhao C, Lee J, Xu Z, et al. Unpaired Brain MR-to-CT Synthesis Using a Structure-Constrained CycleGAN. In: Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, et al., editors. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Cham: Springer International Publishing; 2018, p. 174–82. https://doi.org/10.1007/978-3-030-00889-5_20.

[8] Gupta D, Kim M, Vineberg KA, Balter JM. Generation of Synthetic CT Images From MRI for Treatment Planning and Patient Positioning Using a 3-Channel U-Net Trained on Sagittal Images. Front Oncol 2019;9. https://doi.org/10.3389/fonc.2019.00964.

[9] Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiotherapy and Oncology 2019;136:56–63. https://doi.org/10.1016/j.radonc.2019.03.026.

[10] Koike Y, Akino Y, Sumida I, Shiomi H, Mizuno H, Yagi M, et al. Feasibility of synthetic computed tomography generated with an adversarial network for multi-sequence magnetic resonance-based brain radiotherapy. J Radiat Res 2020;61:92–103. https://doi.org/10.1093/jrr/rrz063.

[11] Lei Y, Harms J, Wang T, Liu Y, Shu H-K, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Medical Physics 2019;46:3565–81. https://doi.org/10.1002/mp.13617.

[12] Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. Journal of Applied Clinical Medical Physics 2019;20:105–14. https://doi.org/10.1002/acm2.12554.

[13] Neppl 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 Oncologica 2019;58:1429–34. https://doi.org/10.1080/0284186X.2019.1630754.

[14] Shafai-Erfani G, Lei Y, Liu Y, Wang Y, Wang T, Zhong J, et al. MRI-Based Proton Treatment Planning for Base of Skull Tumors. International Journal of Particle Therapy 2019;6:12–25. https://doi.org/10.14338/IJPT-19-00062.1.

[15] Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep Convolution Neural Network (DCNN) Multiplane Approach to Synthetic CT Generation From MR images—

ACCEPTED MANUSCRIPT - CLEAN COPY

Application in Brain Proton Therapy. International Journal of Radiation Oncology • Biology • Physics 2019;105:495–503. https://doi.org/10.1016/j.ijrobp.2019.06.2535.

[16] Andres EA, Fidon L, Vakalopoulou M, Lerousseau M, Carré A, Sun R, et al. Dosimetry-driven quality measure of brain pseudo Computed Tomography generated from deep learning for MRI-only radiotherapy treatment planning. International Journal of Radiation Oncology*Biology*Physics 2020:S0360301620311305. https://doi.org/10.1016/j.ijrobp.2020.05.006.

[17] Hemsley M, Chugh B, Ruschin M, Lee Y, Tseng C-L, Stanisz G, et al. Deep Generative Model for Synthetic-CT Generation with Uncertainty Predictions. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, et al., editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, vol. 12261, Cham: Springer International Publishing; 2020, p. 834–44. https://doi.org/10.1007/978-3-030-59710-8_81.

[18] Kazemifar S, Barragán Montero AM, Souris K, Rivas ST, Timmerman R, Park YK, et al. Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors: Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors. Journal of Applied Clinical Medical Physics 2020;21:1–11. https://doi.org/10.1002/acm2.12856.

[19] Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. Quant Imaging Med Surg 2020;10:1223–36. https://doi.org/10.21037/qims-19-885.

[20] 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;22:308–17. https://doi.org/10.1002/acm2.13139.

[21] Maspero M, Bentvelzen LG, Savenije MHF, Guerreiro F, Seravalli E, Janssens GO, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. Radiotherapy and Oncology 2020;153:197–204. https://doi.org/10.1016/j.radonc.2020.09.029.

[22] Massa HA, Johnson JM, McMillan AB. Comparison of deep learning synthesis of synthetic CTs using clinical MRI inputs. Phys Med Biol 2020;65:23NT03. https://doi.org/10.1088/1361-6560/abc5cb.

[23] Tang B, Wu F, Fu Y, Wang X, Wang P, Orlandini LC, et al. Dosimetric evaluation of synthetic CT image generated using a neural network for MR-only brain radiotherapy. J Appl Clin Med Phys 2021:acm2.13176. https://doi.org/10.1002/acm2.13176.

[24] Bourbonne V, Jaouen V, Hognon C, Boussion N, Lucia F, Pradier O, et al. Dosimetric Validation of a GAN-Based Pseudo-CT Generation for MRI-Only Stereotactic Brain Radiotherapy. Cancers 2021;13:1082. https://doi.org/10.3390/cancers13051082.

[25] 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. Medical Physics 2019;46:4095–104. https://doi.org/10.1002/mp.13663.

[26] Klages P, Bensilmane I, Riyahi S, Jiang J, Hunt M, Deasy JO, et al. Comparison of Patch-Based Conditional Generative Adversarial Neural Net Models with Emphasis on Model Robustness for Use in Head and Neck Cases for MR-Only Planning 2020:27. https://doi.org/arXiv:1902.00536.

[27] Wang Y, Liu C, Zhang X, Deng W. Synthetic CT Generation Based on T2 Weighted MRI of Nasopharyngeal Carcinoma (NPC) Using a Deep Convolutional Neural Network (DCNN). Front Oncol 2019;9. https://doi.org/10.3389/fonc.2019.01333.

[28] Largent A, Marage L, Gicquiau I, Nunes J-C, Reynaert N, Castelli J, et al. Head-and-Neck MRI-only radiotherapy treatment planning: From acquisition in treatment position to pseudo-CT generation. Cancer/Radiothérapie 2020:S1278321820300615. https://doi.org/10.1016/j.canrad.2020.01.008.

ACCEPTED MANUSCRIPT - CLEAN COPY

[29] Peng Y, Chen S, Qin A, Chen M, Gao X, Liu Y, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiotherapy and Oncology 2020;150:217–24. https://doi.org/10.1016/j.radonc.2020.06.049.
[30] Qi M, Li Y, Wu A, Jia Q, Li B, Sun W, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. Medical Physics 2020;47:1880–94. https://doi.org/10.1002/mp.14075.
[31] 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. Physics in Medicine & Biology 2020. https://doi.org/10.1088/1361-6560/abb1d6.
[32] Tie X, Lam S, Zhang Y, Lee K, Au K, Cai J. Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients. Med Phys 2020;47:1750–62. https://doi.org/10.1002/mp.14062.
[33] Palmér E, Karlsson A, Nordström F, Petruson K, Siversson C, Ljungberg M, et al. Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic resonance imaging only workflow for head and neck radiotherapy. Physics and Imaging in Radiation Oncology 2021;17:36–42. https://doi.org/10.1016/j.phro.2020.12.007.
[34] Jeon W, An HJ, Kim J, Park JM, Kim H, Shin KH, et al. Preliminary Application of Synthetic Computed Tomography Image Generation from Magnetic Resonance Image Using Deep-Learning in Breast Cancer Patients. J Radiat Prot Res 2019;44:149–55. https://doi.org/10.14407/jrpr.2019.44.4.149.
[35] Liu Y, Lei Y, Wang Y, Wang T, Ren L, Lin L, et al. MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method. Phys Med Biol 2019;64:145015. https://doi.org/10.1088/1361-6560/ab25bc.
[36] Liu Y, Lei Y, Wang T, Kayode O, Tian S, Liu T, et al. MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method. BJR 2019;92:20190067. https://doi.org/10.1259/bjr.20190067.
[37] Olberg S, Zhang H, Kennedy WR, Chun J, Rodriguez V, Zoberi I, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. Med Phys 2019;46:4135–47. https://doi.org/10.1002/mp.13716.
[38] Xu K, Cao J, Xia K, Yang H, Zhu J, Wu C, et al. Multichannel Residual Conditional GAN-Leveraged Abdominal Pseudo-CT Generation via Dixon MR Images. IEEE Access 2019;7:163823–30. https://doi.org/10.1109/ACCESS.2019.2951924.
[39] 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. Radiotherapy and Oncology 2020. https://doi.org/10.1016/j.radonc.2020.10.018.
[40] Florkow MC, Guerreiro F, Zijlstra F, Seravalli E, Janssens GO, Maduro JH, et al. Deep learning-enabled MRI-only photon and proton therapy treatment planning for paediatric abdominal tumours. Radiotherapy and Oncology 2020;153:220–7. https://doi.org/10.1016/j.radonc.2020.09.056.
[41] Fu J, Singhrao K, Cao M, Yu V, Santhanam AP, Yang Y, et al. Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. Biomed Phys Eng Express 2020;6:015033. https://doi.org/10.1088/2057-1976/ab6e1f.
[42] Liu L, Johansson A, Cao Y, Dow J, Lawrence TS, Balter JM. Abdominal synthetic CT generation from MR Dixon images using a U-net trained with 'semi-synthetic' CT data. Physics in Medicine & Biology 2020;65:125001. https://doi.org/10.1088/1361-6560/ab8cd2.
[43] Qian P, Xu K, Wang T, Zheng Q, Yang H, Baydoun A, et al. Estimating CT from MR Abdominal Images Using Novel Generative Adversarial Networks. J Grid Computing 2020;18:211–26. https://doi.org/10.1007/s10723-020-09513-3.
[44] Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. In: Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, et al., editors. Deep Learning and Data Labeling for Medical Applications, vol. 10008, Cham:

ACCEPTED MANUSCRIPT - CLEAN COPY

Springer International Publishing; 2016, p. 170–8. https://doi.org/10.1007/978-3-319-46976-8_18.
[45] 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.
[46] Chen S, Qin A, Zhou D, Yan D. Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. Medical Physics 2018;45:5659–65. https://doi.org/10.1002/mp.13247.
[47] Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-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.
[48] Florkow MC, Zijlstra F, Willemsen K, Maspero M, Berg CAT van den, Kerkmeijer LGW, et al. Deep learning–based MR-to-CT synthesis: The influence of varying gradient echo–based MR images as input channels. Magnetic Resonance in Medicine 2020;83:1429–41. https://doi.org/10.1002/mrm.28008.
[49] Florkow MC, Zijlstra F, M.d LGWK, Maspero M, Berg CAT van den, Stralen M van, et al. The impact of MRI-CT registration errors on deep learning-based synthetic CT generation. Medical Imaging 2019: Image Processing, vol. 10949, International Society for Optics and Photonics; 2019, p. 1094938. https://doi.org/10.1117/12.2512747.
[50] 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. Medical Physics 2019;46:3788–98. https://doi.org/10.1002/mp.13672.
[51] 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. International Journal of Radiation Oncology*Biology*Physics 2019;105:1137–50. https://doi.org/10.1016/j.ijrobp.2019.08.049.
[52] Liu Y, Lei Y, Wang Y, Shafai-Erfani G, Wang T, Tian S, et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. Phys Med Biol 2019;64:205022. https://doi.org/10.1088/1361-6560/ab41af.
[53] Stadelmann JV, Schulz H, Heide UA van der, Renisch S. Pseudo-CT image generation from mDixon MRI images using fully convolutional neural networks. Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 10953, International Society for Optics and Photonics; 2019, p. 109530Z. https://doi.org/10.1117/12.2512741.
[54] Bahrami A, Karimian A, Fatemizadeh E, Arabi H, Zaidi H. A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI. Med Phys 2020;47:5158–71. https://doi.org/10.1002/mp.14418.
[55] 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. Radiotherapy and Oncology 2021;156:23–8. https://doi.org/10.1016/j.radonc.2020.11.027.
[56] 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 era using a conditional generative adversarial network. Physics in Medicine & Biology 2020. https://doi.org/10.1088/1361-6560/ab7633.
[57] Nyholm T, Svensson S, Andersson S, Jonsson J, Sohlin M, Gustafsson C, et al. MR and CT data with multiobserver delineations of organs in the pelvic area—Part of the Gold Atlas project. Medical Physics 2018;45:1295–300. https://doi.org/10.1002/mp.12748.
[58] Fetty 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. https://doi.org/10.1088/1361-6560/ab857b.

ACCEPTED MANUSCRIPT - CLEAN COPY