Deep learning-based synthetic-CT generation in radiotherapy and PET: a review

Maria Francesca Spadea, Matteo Maspero, Paolo Zaffino, and Joao Seco

1 Department of Clinical and Experimental Medicine, University “Magna Graecia” of Catanzaro, 88100 Catanzaro, Italy
2 Department of Radiotherapy, Division of Imaging & Oncology, University Medical Center Utrecht, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands
3 Computational Imaging Group for MR diagnostics & therapy, Center for Image Sciences, University Medical Center Utrecht, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands
4 DKFZ German Cancer Research Center, Division of Biomedical Physics in Radiation Oncology, 69120 Heidelberg, Germany
5 Department of Physics and Astronomy, Heidelberg University, 69120 Heidelberg, Germany

* These authors equally contributed.

Version typeset: August 30, 2021

Abstract

Recently, deep learning (DL)-based methods for the generation of synthetic computed tomography (sCT) have received significant research attention as an alternative to classical ones. We present here a systematic review of these methods by grouping them into three categories, according to their clinical applications:
I) To replace CT in magnetic resonance (MR)-based treatment planning.
II) Facilitate cone-beam computed tomography (CBCT)-based image-guided adaptive radiotherapy.
III) Derive attenuation maps for the correction of positron emission tomography (PET).
Appropriate database searching was performed on journal articles published between January 2014 and December 2020.
The DL methods’ key characteristics were extracted from each eligible study, and a comprehensive comparison among network architectures and metrics was reported. A detailed review of each category was given, highlighting essential contributions, identifying specific challenges, and summarising the achievements. Lastly, the statistics of all the cited works from various aspects were analysed, revealing the popularity and future trends and the potential of DL-based sCT generation. The current status of DL-based sCT generation was evaluated, assessing the clinical readiness of the presented methods.

Authors to whom correspondence should be addressed. Email: j.seco@dkfz.de
Medical imaging’s impact on oncological patients’ diagnosis and therapy has grown significantly over the last decades\(^1\). Especially in radiotherapy (RT)\(^2\), imaging plays a crucial role in the entire workflow, from treatment simulation to patient positioning and monitoring\(^3,4,5,6\).

Traditionally, computed tomography (CT) is considered the primary imaging modality in RT. It provides accurate and high-resolution patient’s geometry, enabling direct electron density conversion needed for dose calculations\(^7\). X-ray based imaging, including planar imaging and cone-beam computed tomography (CBCT), are widely adopted for patient positioning and monitoring before, during or after the dose delivery\(^4\). Along with CT, positron emission tomography (PET) is commonly acquired to provide functional and metabolic information allowing tumour staging and improving tumour contouring\(^8\). Magnetic resonance imaging (MRI) has also proved its added value for tumours and organs-at-risk (OARs) delineation, thanks to its superb soft tissue contrast\(^9,10\).

To benefit from the complementary advantages offered by different imaging modalities, MRI is generally registered to CT\(^11\). However, residual misregistration and differences in patient set-up may introduce systematic errors that would affect the accuracy of the whole treatment\(^12,13\).

Recently, MR-only based RT has been proposed\(^14,15,16\) to eliminate residual registration errors. Furthermore, it can simplify and speed up the workflow, decreasing patient’s exposure to ionising radiation, which is particularly relevant for repeated simulations\(^17\) or fragile populations, e.g. children. Also, MR-only RT may reduce overall treatment costs\(^18\) and workload\(^19\). Additionally, the development of MR-only techniques can be beneficial for MR-guided RT\(^20\).

The main obstacle regarding the introduction of MR-only radiotherapy is the lack of tissue attenuation information required for accurate dose calculations\(^12,21\). Many methods have been proposed to convert MR to CT-equivalent representations, often known as synthetic CT (sCT), for treatment planning and dose calculation. These approaches are summarised in two specific reviews on this topic\(^22,23,24\), in site-specific reviews\(^18,25,26\) or broader review on MR-guided\(^27\) or proton therapy\(^28\).

Additionally, similar techniques to derive sCT from a different imaging modality have
been envisioned to improve the quality of CBCT\textsuperscript{29}. Cone-beam computed tomography plays a vital role in image-guided adaptive radiation therapy (IGART) for photon and proton therapy. However, due to the severe scatter noise and truncated projections, image reconstruction is affected by several artefacts, such as shading, streaking and cupping\textsuperscript{30,31}. For this reason, daily CBCT has not commonly been used for online plan adaptation. The conversion of CBCT-to-CT would allow accurate dose computation and improve the quality of IGART provided to the patients.

Finally, sCT estimation is also crucial for PET attenuation correction. Accurate PET quantification requires a reliable photon attenuation correction (AC) map, usually derived from CT. In the new PET/MRI hybrid scanners, this step is not immediate, and MRI to sCT translation has been proposed to solve the MR attenuation correction (MRAC) issue. Besides, standalone PET scanners can benefit from the derivation of sCT from uncorrected PET\textsuperscript{32,33,34}.

In the last years, the derivation of sCT from MRI, PET or CBCT has raised increasing interest based on artificial intelligence algorithms such as machine learning or deep learning (DL)\textsuperscript{35}. This paper aims to systematically review and summarise the latest developments, challenges and trends in DL-based sCT generation methods. Deep learning is a branch of machine learning, a field of artificial intelligence that involves using neural networks to generate hierarchical representations of the input data to learn a specific task without hand-engineered features\textsuperscript{36}. Recent reviews have discussed the application of deep learning in radiotherapy\textsuperscript{37,38,39,40,41,42,43}, and in PET attenuation correction\textsuperscript{34}. Convolutional neural networks (CNNs), which are the most successful models for image processing\textsuperscript{44,45}, have been proposed for sCT generation since 2016\textsuperscript{46}, with a rapidly increasing number of published papers on the topic. However, DL-based sCT generation has not been reviewed in details, except for applications in PET\textsuperscript{47}. With this survey, we aim at summarising the latest developments in DL-based sCT generation, highlighting the contributions based on the applications and providing detailed statistics discussing trends in terms of imaging protocols, DL architectures, and performance achieved. Finally, the clinical readiness of the reviewed methods will be discussed.
II. Material and Methods

A systematic review of techniques was carried out using the PRISMA guidelines. PubMed, Scopus and Web of Science databases were searched from January 2014 to December 2020 using defined criteria (for more details, see Appendix VII.). Studies related to radiation therapy, either with photons or protons and attenuation correction for PET, were included when dealing with sCT generation from MRI, CBCT or PET. This review considered external beam radiation therapy, excluding, therefore, investigations that are focusing on brachytherapy. Conversion methods based on fundamental machine learning techniques were not considered in this review, preferring only deep learning-based approaches. Also, the generation of dual-energy CT was not considered along with the direct estimation of corrected attenuation maps from PET. Finally, conference proceedings were excluded: proceedings can contain valid methodologies; however, the large number of relevant abstracts and incomplete report of information was considered not suitable for this review. After the database search, duplicated articles were removed and records screened for eligibility. A citation search of the identified articles was performed.

Each included study was assigned to a clinical application category. The selected categories were:

I  MR-only RT;

II  CBCT-to-CT for image-guided (adaptive) radiotherapy;

III  PET attenuation correction.

For each category, an overview of the methods was constructed in the form of tables. The tables were composed by capturing salient information of DL-based sCT generation approaches, which has been schematically depicted in Figure 1.

Independent of the input image, i.e. MRI, CBCT or PET, the chosen architecture (CNN) can be trained with paired or unpaired input data and different configurations. In this review, we define the following configurations: 2D (single slice, 2D, or patch, 2Dp) when training was performed considering transverse (tra), sagittal (sag) or coronal (cor) images;

---

1The tables presented in this review have been made publicly accessible at https://matteomaspero.github.io/overview_sct.
Figure 1: **Schematic of deep learning-based sCT generation study.** The input images/volumes, either being MRI (green), CBCT (yellow) or PET (red), are converted by a Convolutional Neural Network (CNN) into sCT. The CNN is trained to generate sCT similar to the target CT (blue). Several choices can be made in terms of network architecture, configuration, data pairing. After the sCT generation, the output image/volume is evaluated with image- and task-specific metrics.

2D+ when independently trained 2D networks in different views were combined during or after inference; multi-2D (m2D, also known as multi-plane) when slices from different views, e.g. transverse, sagittal and coronal, were provided to the same network; 2.5D when training was performed with neighbouring slices which were provided to multiple input channels of one network; 3D when volumes were considered as input (the whole volume, 3D, or patches, 3Dp). The architectures generally considered are introduced in the next section (II.A.). The sCTs are generated inferring on an independent test set the trained network or combining an ensemble (ens) of trained networks. Finally, the quality of the sCT can be evaluated with image-based or task-specific metrics (II.B.).

For each of the sCT generation category, we compiled tables providing a summary
of the published techniques, including the key findings of each study and other pertinent factors, here indicated: the anatomic site investigated; the number of patients included; relevant information about the imaging protocol; DL architecture, the configuration chosen to sample the patient volume (2D or 2D+ or m2D, 2.5D or 3D); using paired/unpaired data during the network training; the radiation treatment adopted, where appropriate, along with the most popular metrics used to evaluate the quality of sCT (see II.B.).

The year of publication for each category was noted according to the date of the first online appearance. Statistics in terms of popularity of the mentioned fields were calculated with pie charts for each category. Specifically, we subdivided the papers according to the anatomical region they dealt with: abdomen, brain, head & neck (H&N), thorax, pelvis and whole body; where available, tumour site was also reported. A discussion of the clinical feasibility of each methodology and observed trends follows.

The most common network architectures and metrics will be introduced in the following sections to facilitate the tables’ interpretation.

II.A. Deep learning for image synthesis

Medical image synthesis can be formulated as an image-to-image translation problem, where a model that maps input image (A) to a target image (B) has to be found. Among all the possible strategies, DL methods have dramatically improved state of the art. DL approaches mainly used to synthesise sCT belong to the class of CNNs, where convolutional filters are combined through weights (also called parameters) learned during training. The depth is provided by using multiple layers of filters. The training is regulated by finding the "optimal" model parameters according to the search criterion defined by a loss function ($\mathcal{L}$). Many CNN-based architectures have been proposed for image synthesis, with the most popular being the U-nets and generative adversarial networks (GANs) (see Figure 2). U-net presents an encoding and a decoding path with additional skip connections to extract and reconstruct image features, thus learning to go from domain A to B. In the most simple GAN architecture, two networks are competing. A generator (G) that is trained to obtain synthetic images (B') similar to the input set ($\mathcal{L}_G$), and a discriminator (D) that is trained to classify whether B' is real or fake ($\mathcal{L}_D$), improving G’s performances. GANs learn a loss that combines both the tasks resulting in realistic images. Given these
premises, many variants of GANs can be arranged, with U-nets being employed as a possible generator in the GAN framework. We will not detail all possible configurations since it is not the scope of this review, and we address the interested reader to\textsuperscript{54,55,56}. A particular derivation of GAN, called cycle-consistent GAN (cycle-GAN), is worth mentioning. Cycle-GANs opened the era of unpaired image-to-image translation\textsuperscript{57}. Here, two GANs are trained, one going from A to B', called forward pass (forw), and the second going from B' to A, called backwards pass (back), are adopted with their related loss terms (Figure 2 bottom right). Two consistency losses $\mathcal{L}_c$ are introduced, aiming at minimising differences between A and A' and B and B', enabling unpaired training.

Figure 2: Deep learning architectures used for image-to-image translation. In the most straightforward configurations (CNN and U-Net, top left and right, respectively), a single loss function between input and output images is computed. GANs (bottom) use more than one CNN and loss to train the generator’s performance (G). Cycle-GANs enable unsupervised learning by employing multiple GANs and cycle-consistency losses ($\mathcal{L}_{\text{cycle}}$).
II. MATERIAL AND METHODS

II.B. Metrics

An overview of the metrics used to assess and compare the reviewed publications’ performances is summarised in Table 1, subdivided in image similarity, geometric accuracy and task-specific as suggested in\(^ {58}\).

Table 1: **Overview of the most popular metrics** reported in the literature subdivided into image similarity, geometric accuracy, task-specific metrics, and category.

| Category                  | Metric                                                                 |
|---------------------------|------------------------------------------------------------------------|
| **Image similarity**      | M(A)E = \( \frac{\sum_i^n |CT_i - sCT_i|}{n} \), with \( n \)=voxel number in ROI; |
|                           | (R)MSE = \( (\sqrt{\frac{\sum_i^n (CT_i - sCT_i)^2}}) \)               |
|                           | PSNR = 10log\( \frac{MAX^2_{CT}}{MSE} \)                             |
|                           | SSIM = \( \frac{(2\mu_{sCT}\mu_{CT}+c_1)(2\sigma_{sCT, sCT}+c_2)}{(\sigma_{sCT}+\mu_{CT}+c_1)(\sigma_{sCT}+\mu_{CT}+c_2)} \) with |
|                           | \( c_1 = (k_1L)^2 \), \( c_2 = (k_2L)^2 \)                            |
|                           | \( \mu = \text{mean}, \sigma = \text{variance/covariance} \)            |
|                           | \( L = \text{dynamic range}, k_1 = 0.01 \) and \( k_2 = 0.03 \)        |
| **Geometry accuracy**     | DSC(Seg\(_{sCT}\), Seg\(_{sCT}\)) = \( 2\frac{Seg_{sCT}\cap Seg_{CT}}{Seg_{sCT}+Seg_{CT}} \) |
| **Task specific**         | DD = 100 \cdot \frac{D_{sCT}-D_{CT}}{D_{CT}} \%, with \( D=\text{dose} \); |
| CBCT-to-CT                | DPR = \% of voxel with \( DD < x \%) in an ROI                       |
|                           | GPR=\% of voxel with \( \gamma < 1 \) in an ROI                      |
|                           | DVH=difference of specific points in dose-volume histogram plot       |
| PET reconstruction        | \( \text{PET}_{|err|} = 100 \cdot \frac{|PET_{sCT}-PET_{CT}|}{PET_{CT}} \%) |

**Image similarity** The most straightforward way to evaluate the quality of the sCT is to calculate the similarity of the sCT to the ground truth/target CT on a voxel-wise basis. The calculation of voxel-based image similarity metrics implies that sCT and CT are aligned by translation, rigid (rig), affine (aff) or deformable (def) registrations. Widespread similarity metrics for this task are reported in Table 1 and include: mean (absolute) error (M(A)E), sometimes referred to as mean absolute prediction error (MAPE), peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). Other less common metrics are cross-correlation (CC) and normalised cross-correlation (NCC), along with the (root) mean
squared error ((R)MSE).

M(A)E and (R)MSE are relatively easy to compute as the average of the (absolute) difference and difference in quadrature over a defined region of interest. For both the metrics, lower values indicate better prediction accuracy for sCT. MAE and ME are often computed together to represent the random and systematic error, respectively. MSE and RMSE are used to give more weight to higher errors, thus understanding the impact of possible outliers. PSNR is the ratio between the maximum in an image and the intensity of the corrupting noise affecting the fidelity of its representation, calculated as MSE. PSNR evaluates the noise introduced in the CT synthesis relatively to the ground truth CT. SSIM is a more sophisticated metric developed to take advantage of the known characteristics of the human visual system perceiving the loss of image structure due to variations in lighting.

**Geometric accuracy** Along with voxel-based metrics, the geometric accuracy of the generated sCT can also be assessed by comparing corresponding segmented structures on CT and sCT, e.g. bones, fat, muscle, air and body. The segmentation can be performed manually but can also be automatic. In this context, the delineations are found after applying a threshold to CT and sCT and, if necessary, morphological operations on the obtained binary masks. The metrics for geometric accuracy are, therefore, generally the same used for a segmentation task. For example, the Dice similarity coefficient (DSC) is a common metric that assesses the accuracy of depicting specific tissue classes/structures. DSC is twice the ratio between the correctly classified voxel and all the voxels in the mask from CT and sCT \((\text{Seg}_{CT} \text{ and Seg}_{sCT})\). Additionally, metrics generally used to estimate the distance among segmentations can also be adopted as the Hausdorff distance (HD) or mean absolute surface distance, which measures two sets of contours’ maximum and average distance, respectively. Even if segmentation-based metrics are common, choosing the right metric for the specific task is a non-trivial task, as recently highlighted by Reinke et al. and should be assessed on an application basis.

Other image-based metrics can be subdivided according to the application and presented in the following sections’ appropriate sub-category.

**Task-specific metrics** In MR-only RT and CBCT-to-CT for adaptive RT, dose calculation accuracy on sCT is generally compared to CT-based in specific ROIs for dose calculations performed either for photon \((x)\) and proton \((p)\) RT.
The most common voxelwise-based metric is the dose difference (DD), calculated as the average dose ($D_{CT} - D_{sCT}$) in ROIs as the whole body, target or other structures of interest. The dose difference can be expressed as an absolute value (Gy) or relative (%), either to the prescribed dose, the maximum dose or the voxel-wise reference dose. The dose pass rate (DPR) is directly correlated to DD, and it is calculated as the percentage of voxels with DD< than a set threshold.

Gamma ($\gamma$) analysis allows combining dose and spatial criteria\textsuperscript{63}, and it can be performed either in 2D or 3D. Several parameters need to be set to perform $\gamma$-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 since they depend on the chosen parameters, dose grid size, and voxel resolution\textsuperscript{64,65}. Results of $\gamma$-analysis are generally expressed as gamma pass rate (GPR), counting the percentage of voxels with $\gamma < 1$ or the mean $\gamma$ in an ROI generally defined based on a threshold of the reference dose distribution.

Dose-volume histograms (DVHs) are one of the most diffused tools in the clinical routine\textsuperscript{66}. DVH summarises 3D dose distributions in a graphical 2D format offering no spatial information.

For the evaluation of sCT, generally, the differences among clinically relevant DVH points is reported.

In proton RT, range shift (RS) analysis is also performed. Here, the ideal range (known as the prescribed range) is defined as the depth at which the dose has decreased to 80% of the maximum dose, on the distal dose fall-off ($R_{80}$)\textsuperscript{67}. RS error (RSe) can be defined both as the absolute difference between the prescribed and the actual range ($RSe = R_{80CT} - R_{80sCT}$) and as relative RS (%RS) error, expressed as the shift in % relative to the prescribed range, along the beam direction\textsuperscript{68}

$$\%RS = \left| \frac{R_{80CT} - R_{80sCT}}{R_{80CT}} \right| \times 100$$ (1)

For sCT for PET attenuation correction, the relative error (signed PET_{err} and unsigned PET_{|err|}) of PET reconstruction is usually reported along with the difference in standard uptake values (SUV).
Please note that even if two papers calculate the same metric, differences could occur in the ROI where the metrics are calculated, making challenging performance comparisons. For example, MAE can be computed on the whole predicted volume, in a volume of interest or a cropped volume. In addition to that, the implementation of the metric computation can change. In gamma analysis, for example, different dose difference and distance to agreement criteria can be stated ($\gamma_{3\%},3\text{mm}$ ($\gamma_3$), $\gamma_{2\%},2\text{mm}$ ($\gamma_2$) and $\gamma_{1\%},1\text{mm}$ ($\gamma_1$)). Moreover, it can be calculated on ROI obtained from different dose thresholds and 2D or 3D algorithms. In the following sections, we will highlight the possible differences speculating on the impact.

III. Results

Database searching led to 91 records on PubMed, 98 on Scopus and 218 on Web of Science. After duplicates removal and content check, 83 eligible papers were found. Figure 3 summarises the number of articles published by year, grouped in 51 (61.4%), 15 (18.1%) and 17 (20.5%) for MR-only RT (category I), CBCT-to-CT for adaptive RT (category II), and sCT for PET attenuation correction (category III), respectively. The first conference paper appeared in 2016. Given that we excluded conference papers from our search, we found that the first work was published in 2017. In general, the number of articles increased over the years, except for CBCT-to-CT and sCT for PET attenuation correction, which was stable in the last years. Figure 3 shows that the brain, pelvis and H&N were the most popular anatomical regions investigated in DL-based sCT for MR-only RT, covering $\sim$80% of the studies. For CBCT-to-CT, H&N and pelvic regions were the most explored sites, being present in $>75\%$ of the works. Finally, for PET AC, H&N was investigated in the majority of the studies, followed by the pelvic region. Together, they covered $>75\%$ of the publications.

The total number of patients included in the analysis was variable, but most studies dealt with less than 50 patients for all three categories. The largest patient cohorts included 402 (I), 328 (II) and 193 patients (I), while the smallest studies included 10 patients and another 10 volunteers (I).

Most papers enrolled adult patients. Paediatric (paed) patients represent a more heterogeneous dataset for network training, and its feasibility has been investigated first for
III. RESULTS

Figure 3: (Top) Number of published articles grouped by application and year; (middle) pie charts of the anatomical regions investigated for each application; (bottom) bar plot of the publications binned per the total number of patients included in the study.
attenuation correction in PET\textsuperscript{74} (79 patients) and more recently for photon and proton RT\textsuperscript{75,76}.

All the models were trained to perform a regression task from the input to sCT, except for two studies where networks were trained to segment the input image into a pre-defined number of classes, thus performing a segmentation task\textsuperscript{77,78}.

In most of the works, training was implemented in a paired manner, with unpaired training investigated in 13/83 articles. Four studies compared paired against unpaired\textsuperscript{71,79,80,81}. Over all the three categories, 2D networks were the most common adopted. Specifically, 2D networks were used about 61\% of the times, 2D+ 6\%, 2.5D 10\%, and 3D configuration 24\%. In some studies, multiple configurations were investigated, for example\textsuperscript{79,82,83}. GANs were the most popular architectures (45-times), followed by U-nets (36) and other CNNs. Note that U-nest may be employed as generator of GANs, and that in this case, the architecture was categorised as GAN.

All the investigations employed registration between sCT and CT to evaluate the quality of the sCT, except for Xu et al.\textsuperscript{81} and Fetty et al.\textsuperscript{84}, where metrics were defined to assess the quality of the sCT in an unpaired manner, e.g. Frechet inception distance (FID).

Main findings are reported in Table 2 for studies on sCT for MR-only RT without dosimetric evaluations, in Table 3a, 3b for studies on sCT for MR-only RT with dosimetric evaluations, in Table 4 for studies on CBCT-to-CT for IGART, and in Table 5 for studies on PET attenuation correction. Tables are organised by anatomical site and tumour location where available. Studies investigating the independent training and testing of several anatomical regions are reported for each specific site\textsuperscript{70,81,85,86,87}. Works using the same network to train or test data from different scanners and anatomy are reported at the bottom of the table\textsuperscript{88,89}. Detailed results based on these tables are presented in the following sections subdivided for each category.

\section{III.A. MR-only radiotherapy}

The first work ever published in this category, and in among all the categories, was by Han in 2017, where he proposed to use a paired U-net for brain sCT generation. After one year, the first work published with a dosimetric evaluation was presented by Maspero et
al.\textsuperscript{90}, investigating a 2D paired GAN trained on prostate patients and evaluated on prostate, rectal and cervical cancer patients.

Considering the imaging protocol, we can observe that most of the MRIs were acquired at 1.5 T (51.9%), followed by 3 T (42.6%), and the remaining 6.5% at 1 T or 0.35/0.3 T. The most popular MRI sequences adopted depends on the anatomical site: T1 gradient recalled-echo (T1 GRE) for abdomen and brain; T2 turbo spin-echo (TSE) for pelvis and H&N. Unfortunately, for more than ten studies, either sequence or magnetic field were not adequately reported.

Generally, a single MRI sequence is used as input. However, eight studies investigated using multiple input sequences or Dixon reconstructions\textsuperscript{73,76,90,98,102,112,125} based on the assumption that more input contrast may facilitate sCT generation. A relevant aspect related to MRI is which kind on pre-processing is applied to the data before being fed to the network. Generally intensity normalisation techniques like z-score\textsuperscript{126}, percentile\textsuperscript{75,90} or range-based normalisation, histogram matching\textsuperscript{79,82,85,98} or linear rescaling were applied\textsuperscript{111,127}. However, techniques like bias field\textsuperscript{69,79,82,84,85,91,94,95,98,100,104,105,109,112,115,122}, intensity homogeneity\textsuperscript{69,79,82,84,85,91,94,95,98,100,104,105,109,112,115} were also applied to minimise inter-patient intensity variations.

Some studies compared the performance of sCT generation depending on the sequence acquired. For example, Massa et al.\textsuperscript{96} compared sCT from the most adopted MRI sequences in the brain, e.g. T1 GRE with (+Gd) and without Gadolinium (-Gd), T2 SE and T2 fluid-attenuated inversion recovery (FLAIR), obtaining the lowest MAE and highest PSNR for T1 GRE sequences with Gadolinium administration. Florkow et al.\textsuperscript{102} investigated how the performance of a 3D patch-based paired U-net was impacted by different combinations of T1 GRE images along with its Dixon reconstructions, finding that using multiple Dixon images is beneficial in the human and canine pelvis. Qi et al.\textsuperscript{123} studied the impact of combining T1 (±Gd) and T2 TSE, obtaining that their 2D paired GAN model trained on multiple sequences outperformed any model on a single sequence.

When focusing on the DL model configuration, we found that 2D models were the most popular ones, followed by 3D patch-based and 2.5D models. Only one study adopted a multi-2D (m2D) configuration\textsuperscript{106}. Three studies also investigated whether the impact of combining sCTs from multiple 2D models after inference (2D+) shows that 2D+ is beneficial compared to single 2D view\textsuperscript{75,111,122}. When comparing the performances of 2D against 3D models,
| Tumor site | MRI field | x-fold | conf | arch | MSE/MAE | SSIM | PSNR | FID | DSC | Patients | Reference |
|------------|-----------|--------|------|------|---------|------|------|-----|------|----------|-----------|
| Brain      | T1        | 1.5    | n.a. | 3D T2 | 2D GAN | 3D T2 | 3D T2 | 3D T2 | 3D T2 | 3D T2    | Mass2020  |
| Brain      | T1        | 1.5    | n.a. | 3D T2 | 2D GAN | 3D T2 | 3D T2 | 3D T2 | 3D T2 | 3D T2    | Mass2020  |
| Brain      | T1        | 1.5    | n.a. | 3D T2 | 2D GAN | 3D T2 | 3D T2 | 3D T2 | 3D T2 | 3D T2    | Mass2020  |
| Brain      | T1        | 1.5    | n.a. | 3D T2 | 2D GAN | 3D T2 | 3D T2 | 3D T2 | 3D T2 | 3D T2    | Mass2020  |
| Brain      | T1        | 1.5    | n.a. | 3D T2 | 2D GAN | 3D T2 | 3D T2 | 3D T2 | 3D T2 | 3D T2    | Mass2020  |

**Table 2: Overview of sCT methods for MR-only radiotherapy with some image-based evaluation.**

- **Patients:** n.a.
- **Reference:** n.a.
- **Reg:** n.a.
- **MRI:** n.a.
- **DI method:** n.a.
- **Image-similarity:** n.a.
- **Volunteers, not patients:** n.a.
- **Four segment CT into 5 classes:** n.a.
- **Multiple combinations of Dixon images was investigated but omitted here:** n.a.
- **Dataset from [1]:** n.a.
- **Abbreviations:** n.a.
- **arch:** architecture, GRE=gradient echo, (T)SE=(turbo) spin-echo Dixon reconstruction, LoO=leave-one-out, (RMSE, root) mean squared error, ME=mean error, DSC=Dice similarity coefficient, NCC=normalized cross-correlation.
Table 3a: Overview of MR-only radiotherapy with image-based and dose evaluation.

| Tissues      | Region with | Dice | MAE | OASSE | MAE | OASSE | MAE | OASSE | MAE | OASSE | MAE | OASSE |
|--------------|-------------|------|-----|-------|-----|-------|-----|-------|-----|-------|-----|-------|
| Brain        | Brain       |      |     |       |     |       |     |       |     |       |     |       |
| Abdomen      | Abdomen    |      |     |       |     |       |     |       |     |       |     |       |
| Liver        | Liver       |      |     |       |     |       |     |       |     |       |     |       |

Note: The table compares the performance of different DL methods for MR-only radiotherapy with image-based and dose evaluation. The metrics include Dice, MAE, and OASSE for different tissue regions (Brain, Abdomen, Liver) and regions with varying Dice scores. The results indicate the effectiveness and accuracy of the DL methods in predicting tumor volumes and dose distributions.
Table 3b: Overview sCT methods for MR-only radiotherapy with image-based and dose evaluation.

| Tumor site | Patients | MRI | DL method | Image-similarity | Dose | Reference |
|------------|----------|-----|------------|------------------|------|-----------|
| Prostate | 36 15 | 3 T2 TSE | 2D pair | U-net | def | 30±6 | ME tissues | x 0.16±0.09 99.4 ±0.09 <0.2Gy γγ1 | Chen2018 |
| Prostate | 39 | 4x 3 | 3D T2 | 2D pair | U-net | def | 34±8 | ME DSC | x -0.01±0.64 98.5±0.72 <3% γγ1 | Arabi2018 |
| Prostate | 17 | LoO 1.5 | T2 | 3Dp | GAN | def | 34±17 24±2 | NCC, bone dist, uniform | p 0.07±0.07 98±6 <1% NCC, γγ peak, γγ | Liu2018 |
| Prostate | 25 | 14 | 3 | 3D T2 TSE | 2D pair | U-net | def | 34±8 | tissues | x <1% γγ 99.2±1 | γγ1 | Largent2019 |
| Pelvis | 11 | 8 | 3 1.5 | T2 TSE | 2D pair | GAN | def | 49±6 | ME organs | x 0.7±0.4 99.2±10 | <1.5% | Bani2019 |
| Pelvis | 26 | 15 | 10+19 | 0.5 | 1.5/3 | 3D T2 | 2D pair | GAN | def | 41±4 31.4±1 | ME VSE bone | x ±1 | <1.5% | Fett2019 |
| Rectum | 39 | 14 | 0.35 | GRE | 2D pair | U-net | def | 54±12 | tissues | x+Ho < 0.5 99.0±0.72 | <1% γγ1 | Catmam2020 |
| Rectum | 46 | 46 | 0.35 | 1.5 | 3D T2 | 2D pair | GAN | def | 35±7 | ME bone | x <0.8 99.8±0.12 | <1% γγ1 | Bird2020 |
| H&N | 34 | 3x | 1.5 | 3D T2 TSE | 3Dp | GAN | def | 75±5 | ME DSC bone | x -0.07±0.22 95.6±2.96 | γγ1 | Dinkla2019 |
| H&N | 15 12 | 3 | T1 GRE | 2Dp | GAN | def | 68±2 | SSIM | p <0.5 | <0.5 | γγ1 | Klages2019 |
| H&N | 30 | 15 | 3 | T1+T2 TSE | 2D pair | GAN | def | 30±12 | 71±12 | 29±1.3 29.2±1.3 | SSIM DSC DRR | p 0.3±0.2 97.6±3.2 | γγ1 | Peng2020 |
| H&N | 33 | 5 | 3 | 3D T1 TSE | 2D pair | GAN | def | 30±9 101±8 | ME, DSC tissues | x 0.1±0.3 | 98.5±1 | <1.5% | beam depth | Thummel2020 |
| H&N | 27 | 3 | 3 | 3D T1 | 2D+ GAN | def | 65±4 | ME DSC | p ±0.2 | 94.5±3.4 | <1.5% | NTCP RS γγ | Thummel2020 |
| H&N | 27 | 18 | 3 | TD TSE | 2D pair | GAN | def | 94±11 103±15 | NCC | p <0.5 | 98.4±3.5 | γγ1 | Oberg2019 |

| Tumor site | MRI | DL method | Image-similarity | Dose | Reference |
|------------|-----|------------|------------------|------|-----------|
| Prostate | 32 | 27 | 3 | 3D T1 | 2D pair | GAN | def | 90±8 | ME | x 0.3±0.4 | 99.4±0.6 | γγ1 | Maspero2018 |
| Rectum | 32 | 18 | 1.5 | 1.5 | 3D T1 | 2D pair | GAN | def | 50±8 | ME | x 0.3±0.5 | 98.5±1.1 | γγ1 | Maspero2018 |
| Cervix | 32 | 14 | 1.5 | 1.5 | 3D T1 | 2D pair | GAN | def | 50±8 | ME | x 0.3±0.3 | 99.6±1.9 | γγ1 | Maspero2018 |

* Comparison with other architecture has been provided.

+ trained in 2D on multiple view and aggregated after inference.

Robustness to training size was investigated.

Multiple combinations (also Dixon reconstruction, where present) of the sequences were investigated but omitted.

m data from multiple centers; x: photon plan; p: proton plan.
Fu et al.\textsuperscript{82} found that a modified 3D U-net outperformed a 2D U-net; while Neppl et al.\textsuperscript{83} one month later published that their 3D U-net under-performed a 2D U-net not only on image similarity metrics but also considering photon and proton dose differences. These contradicting results will be discussed later. Paired models were the most adopted, with only ten studies investigating unpaired training.\textsuperscript{71,79,80,81,84,93,95,99,113,116} Interestingly, Li et al.\textsuperscript{80} compared a 2D U-net trained in a paired manner against a cycle-GAN trained in an unpaired manner, finding that image similarity was higher with the U-net. Similarly, two other studies compared 2D paired against unpaired GANs, achieving slightly better similarity and lower dose difference with paired training in the abdomen\textsuperscript{79} and H&N\textsuperscript{71}. Mixed paired/unpaired training was proposed by Jin et al.\textsuperscript{93} who found such a technique beneficial against either paired or unpaired training. Yang et al.\textsuperscript{95} found that structure-constrained loss functions and spectral normalisation ameliorated unpaired training performances in the pelvic and abdominal regions.

An interesting study on the impact of the directions of patch-based 2D slices, patch size and GAN architecture was conducted by Klages et al.\textsuperscript{122} who reported that 2D+ is beneficial against a single view (2D) training, overlapping/non-overlapping patches is not a crucial point, and that upon good registration training of paired GANs outperforms unpaired training (cycle-GANs).

If we now turn to the architectures employed, we can observe that GAN covers the majority of the studies (\textasciitilde 55\%), followed by U-net (\textasciitilde 35\%) and other CNNs (\textasciitilde 10\%). A detailed examination of different 2D paired GANs against U-net with different loss functions by Largent et al.\textsuperscript{117} showed that U-net and GANs could achieve similar image- and dose-base performances. Fetty et al.\textsuperscript{119} focused on comparing different generators of a 2D paired GAN against the performance of an ensemble of models, finding that the ensemble was overall better than single models being more robust to generalisation on data from different scanners/centres. When considering CNNs architectures, it is worth mentioning using 2.5D dilated CNNs by Dinkla et al.\textsuperscript{106} where the m2D training was claimed to increase the robustness of inference in a 2D+ manner, maintaining a big receptive field and a low number of weights.

An exciting aspect investigated by four studies is the impact of the training size\textsuperscript{69,71,75,95,125}, which will be further reviewed in the discussion section.

Finally, when considering the metric performances, we found that 21 studies reported only image similarity metrics, and 30 also investigated the accuracy of sCT-based dose

\section*{III. RESULTS}
\subsection*{III.A. MR-only radiotherapy}
calculation on photon RT (19), proton RT (8), or both (3). Two studies performed treatment planning, considering the contribution of magnetic fields\(^79,86\), which is crucial for MR-guided RT. Also, only four publications studied the robustness of sCT generation in a multiple centres\(^69,75,118,120\).

Overall, DL-based sCT resulted in DD on average <1\% and \(\gamma_{2\%,2\text{mm}}\) GPR>95\%, except for one study\(^124\). For each anatomical site, the metrics on image similarity and dose were not always calculated consistently. Such aspect will be detailed in the next section.

### III.B. CBCT-to-CT generation

CBCT-to-CT conversion via DL is the most recent CT synthesis application, with the first paper published in 2018\(^128\). Some of the works (5 out of 15) focused only on improving CBCT image quality for better IGRT\(^87,128,129,130,131\). The remaining 10 proved the validity of the transformation with dosimetric studies for photons\(^70,75,105,132,133,134,135\), protons\(^124\) and for both photons and protons\(^89,136,137\).

Only three studies investigated unpaired training\(^88,132,137\); in eleven cases, paired training was implemented by matching the CBCT and ground truth CT by rigid or deformable registration. In Eck et al.\(^70\), however, CBCT and CT were not registered for the training phase, as the authors claimed the first fraction CBCT was geometrically close enough to the planning CT for the network. Deformable registration was then performed for image similarity analysing. In this work, the quality of contours propagated to sCT from CT was compared to manual contours drawn on the CT to assess each step of the IGART workflow: image similarity, anatomical segmentation and dosimetric accuracy. The network, a 2D cycle GAN implemented on a vendor’s provided research software, was independently trained and tested on different sites, H&N, thorax and pelvis, leading to best results for the pelvic region.

Other authors studied training a single network for different anatomical regions. In Maspero et al.\(^88\), authors compared the performances of three cycle-GANs trained independently on three anatomical sites (H&N, breast and lung) vs a single trained with all the anatomical sites together, finding similar results in terms of image similarity.
Table 4: Overview sCT methods for adaptive radiotherapy with CBCT.

| Tumor site | Patients | DL method | Reg | Image-similarity | Plan | Dose | Reference |
|------------|----------|------------|-----|------------------|------|------|-----------|
|            | train val test |            | conf | arch | **MAE** | **PSNR** | **SSIM** | **others** | **DD** | **DPR** | **GPR** | **DVH** | others | |
| Pancreas   | 30       | 50x LoO 3D pair GAN | def | 56.9±3.8 28.8±2.5 71±3.3 | NCC SNU | x | <1Gy | Liu 2020<sup>10</sup> |
| Thorax     | 53       | 15 2D pair GAN | def | 94±42 | ME DSC HD tis | x | 76.7±17.3 93.8±5.9 | <2.6 73 | Eckl 2020<sup>20</sup> |
| Brain      | 21/20    | 5x LoO 3D pair GAN | rig | 13±2 18±5 37.5±2.2 30.7±3.7 | NCC SNU | No | | Harno 2019<sup>18</sup> |
| Pelvis     | 16/4     | 5x 2D pair U-net | def | 50.9 967 | SNU RMSE | No | | Kida 2019<sup>13</sup> |
| Prostate   | 27/8     | 8 2D pair U-net | def | 58 | ME | x | p | 99.9±0.3 80.5±5.7 | <±1.6% | Landry 2019<sup>19</sup> |
| Pelvis     | 16/4     | 2D pair GAN | rig | 87±5 | ME | x | p | 99.9±0.3 80.5±5.7 | <±1.6% | Kida 2019<sup>13</sup> |
| Pelvis     | 205/15   | 15 2D pair GAN | def | 82±5 | ME DSC HD tis | x | 88±9.3 98.5±17 | <1 | 73 | Eckl 2020<sup>10</sup> |
| Bladder    | 81/9     | 20 2D pair GAN | def | 29.9±4.9 30.7±1.4 85±3.4 | RMSE phantom | x | 98±17 96.3±3.6 | | Liang 2019<sup>14</sup> |
| Nasopharynx| 50/10    | 10 2D pair U-net | rig | 6±27 | ME organs | x | p 0.2±0.1 | 95.5±16 | <1% | Liu 2019<sup>13</sup> |
| Bladder    | 30/7     | 7 2D pair U-net | rig | 18.98 33.26 0.8911 | RMSE tissues | No | | Chen 2019<sup>12</sup> |
| Bladder    | 50/10    | 10 2D pair U-net | rig | 89.28 14.25 85 | SNU | No | | Yuan 2020<sup>27</sup> |
| Bladder    | 22/11    | 3x 2D pair U-net | def | 36.6 | ME DSC SNU | x | p <0.1±0.3 | 98±12 | 73 | Thurner 2020<sup>13</sup> |
| Bladder    | 30/14    | 14 2D pair GAN | def | 82.4±10.6 | ME tissues | x | 91±5.5 | <1Gy <1% | Banate 2020<sup>34</sup> |
| Bladder    | 25/15    | 15 2D pair GAN | def | 77±6 10.6 | ME DSC HD tis | x | 91±4.6 95.0±24 | <2.4 | 73 | Eckl 2020<sup>10</sup> |

* comparison with other architecture has been provided; 1 dose pass rate (DPR) 1% or γ<sub>1%,1mm</sub> = γ<sub>1</sub>; 2 DPR 2% or γ<sub>2%,2mm</sub> = γ<sub>2</sub>; DPR 3% or γ<sub>3%,3mm</sub> = γ<sub>3</sub>; + trained in 2D on multiple view and aggregated after inference; w different nets were trained and the different outputs were weighted to obtain final sCT; 4 robustness to training size was investigated; x: photon plan; p: proton plan.
Zhang et al.\textsuperscript{89} trained a 2.5D conditional GAN\textsuperscript{57} with feature matching on a large cohort of 135 pelvic patients. Then, they tested the network on additional 15 pelvic patients acquired with a different CT scanner and ten H&N patients. The network predicted sCT with similar MAE for both testing groups, demonstrating the potentialities to transfer pre-trained models to different anatomical regions. They also compared different GAN flavours and U-net finding the latter statistically worse than any GAN configuration.

Three works tested unpaired training with cycle-GANs\textsuperscript{88,132,137}. In particular, Liang et al.\textsuperscript{132} compared unsupervised training among cycle-GAN, DCGAN\textsuperscript{139} and PGGAN\textsuperscript{140} on the same dataset, finding the first to perform better both in terms of image similarity and dose agreement.

Considering the anatomical regions investigated, most of the studies dealt with H&N and pelvic regions. Liu et al.\textsuperscript{135} investigated CBCT-to-CT in the framework of breath-hold stereotactic pancreatic radiotherapy, where they trained a 3D patch cycle-GAN introducing an attention gate (AG)\textsuperscript{141} to deal with moving organs. They found that the cycle-GAN with AG performed better than U-net and cycle-GAN without AG. Moreover, the DL approach led to a statistically significant improvement in sCT vs CBCT, although some residual discrepancies were still present for this particular anatomical site.

### III.C. PET attenuation correction

DL methods for deriving sCT for PET AC have been published since 2017\textsuperscript{142}. Two possible image translations are available in this category: i) MR-to-CT for MR attenuation correction (MRAC), where 14 papers were found; ii) uncorrected PET-to-CT, with three published articles.

In the first case, most methods have been tested with paired data in H&N (9 papers) and the pelvic region (4 papers) except Baydoun et al.\textsuperscript{143} who investigated the thorax district. The number of patients used for training ranged between 10 and 60. Most of the MR images employed in these studies have been acquired directly through 3T PET/MRI hybrid scanners, where specific MR sequences, such as UTE (ultra-short echo time) and ZTE (zero time echo) are used to enhance short $T_2$ tissues, such as in the cortical bone and Dixon reconstruction is employed to derive fat and water images.

Leynes et al.\textsuperscript{142} compared the Dixon-based sCT vs sCT predicted by U-net receiving both
Dixon and ZTE. Results showed that DL prediction reduced the RMSE in corrected PET SUV by a factor of 4 for bone lesions and 1.5 for soft tissue lesions. Following this first work, other authors showed the improvement of DL-based AC over the traditional atlas-based MRAC proposed by the vendors\textsuperscript{74,143,144,145,146,147,148}, also comparing several network configurations\textsuperscript{149,150}.

Torrado et al.\textsuperscript{146} pre-trained their U-net on 19 healthy brains acquired with $T_1$ GRE MRI and, subsequently, they trained the network using Dixon images of colorectal and prostate cancer patients. They showed that pre-training led to faster training with a slightly smaller residual error than U-net weights’ random initialisation.

Pozaruk et al.\textsuperscript{149} proposed data augmentation over 18 prostate cancer patients by perturbing the deformation field used to match the MR/CT pair for feeding the network. They compared the performance of GAN with augmentation vs 1) Dixon based and 2) Dixon + bone segmentation from the vendor, 3) U-net with and 4) without augmentation. They found significant differences between the 3 DL methods and classic MRAC routines. GAN with augmentation performed slightly better than the U-net with/without augmentation, although the differences were not statistically relevant.

Gong et al.\textsuperscript{150} used unregistered MR/CT pair for a 3D patch cycle GAN, comparing the results vs atlas-based MRAC and CNN with registered pair. Both DL methods performed better than atlas MRAC in DSC, MAE and PET$_{err}$. No significant difference was found between CNN and cycle-GAN. They concluded that cycle-GAN has the potentiality to skip the limit of using a perfectly aligned dataset for training. However, it requires more input data to improve output.

Baydoun et al.\textsuperscript{143} tried different network configurations (VGG16\textsuperscript{157}, VGG19\textsuperscript{157}, and ResNet\textsuperscript{158}) as a benchmark with a 2D conditional GAN receiving either two Dixon input (water and fat) or four (water, fat, in-phase and opposed-phase). The GAN always performed better than VGG19 and ResNet, with more accurate results obtained with four inputs.

In the effort to reduce the time for image acquisition and patient discomfort, some authors proposed to obtain the sCT directly from diagnostic images, $T_1$- or $T_2$-weighted, both using images from standalone MRI scanners\textsuperscript{115,151,153} or hybrid machines\textsuperscript{78}. In particular, Bradshaw et al.\textsuperscript{78} trained a combination of three CNNs with $T_1$ GRE and $T_2$ TSE MRI (single sequence or both) to derive an sCT stratified in classes (air, water, fat and bone) which was
Table 5: Overview methods on sCT for PET AC.

| Region | Patients | MRI | DL method | Reg | Image-similarity | PET-related | Others | Reference |
|--------|----------|-----|------------|-----|------------------|-------------|--------|-----------|
|        |          |     |            |     |                  |             |        |           |
| Pelvis | 10       | 16  | 3'H        | Dixon ±2TE | 3Dp pair | U-net | def  | 15-F-FDG 68Ga-PSMA | RMSE SUV diff | Leynes2017142 |
| Pelvis | 15       | 4   | 3'H        | T1 GRE Dixon | 2D pair | U-net | def  | 15-F-FDG | µ-map diff | Torrado2019146 |
| Pelvis | 12       | 6   | 3'H        | T1 GRE T2 TSE | 3Dp pair | CNN | def  | 15-F-FDG | RMSE | Bradshaw201878 |
| Prostate | 18      | 10  | 3'H        | Dixon | 2D pair | GAN* | def  | 68Ga-PSMA | μ-map diff | Pozaruk2020149 |
| Head   | 30       | 10  | 1.5        | T1 GRE Gd | 2D pair | CNN | def  | .971±.005 | n.a. | -0.7±1.1 | Lin2018155 |
| Head   | 30 ⁰+6   | 8   | 1.5⁰+3⁰   | UTE     | 2D pair | U-net | def  | .76±.03 | .96±.01 | .88±.01 | Jang2018146 |
| H&N    | 12       | 8   | 5 ⁰       | Dixon ±2TE | 2D pair | U-net | rig  | 13.8±1.4 | 12.6±1.5 | .76±.04 | .80±.04 | Gong2018144 |
| Head   | 60       | 19  | 4          | mDixon +UTE | 3Dp pair | U-net | rig  | .90±.07 | 15-F-FET | biol tumor vol, SUV | Ladeogo201974 |
| Head   | 40       | 2   | 3          | T1 GRE   | 3Dp pair | GAN  | def  | 101±40 | 302±79 | 407±228 | 10±5 | .80±.07 | Arabi20191352 |
| Head   | 44       | 11  | 1.5        | T1 GRE   | 2.5D pair | U-net | rig  | U-CWAY | U-C.DASB | -0.49±1.7 | -1.52±.73 | Syst µ-map, kin anal | Spuhler2019153 |
| Head   | 23       | 47  | 3⁰        | ZTE      | 3Dp pair | U-net | def  | .81±.03 | 15-F-FDG | -0.2±.56 | Jac | Blanc-Durand2019147 |
| Head   | 32       | 4   | 3⁰        | Dixon*   | 3Dp pair | GAN* | def  | 15.8±2.4 | 74±.05 | 15-F-FDG | -1.0±13 | SUV | Gong2020a150 |
| Head   | 35       | 5   | 3          | mDixon UTE* | 2.5D pair | U-net | rig  | 10.9±0.01 | .87±.03 | 11-C-Pb 19F-MK201 | <2 | Gong2020h148 |
| Thorax | 14       | LoO | 3⁰        | Dixon*   | 2D pair | GAN* | def  | 67.45±9.89 | 15-F-Naf | PSNR SSIM RMSE | Baydoun2020141 |

Other than MR-based sCT

| Body   | 100      | 28  | PET, no att corrected | 2D pair | U-net | Y* | 111±16 | 94±.01 | 15-F-FDG | -0.6±2.0 | abs err | Liu2018155 |
| Body   | 80       | 39  | PET, no att corrected | 3Dp pair | GAN | Y* | 109±19 | 87±.03 | 15-F-FDG | <1.0 | NCC PSNR ME | Dong2019157 |

* comparison with other architecture has been provided; ⁵ data from another MRI sequence used as pre-training; ⁶ patients acquired with different scanner; ⁷ MRI data from hybrid PET/MRI scanner; ⁸ max in SUV max; ⁹ mean in SUV mean; ¹ in air or bowel gas; ² in the bony structures; ³ in the soft tissue; ⁴ in the fatty tissue; ⁵ in water; ¹ trained to segment the CT/sCT into classes; ² expressed in terms of Jaccard index and not DSC; ³ multiple combinations (also: Dixon reconstruction, where present) of the sequences were investigated but omitted; ⁴ intrinsically registered: PET-CT data; paed: paediatric.
compared with the scanner default MRAC output. The RMSE on PET reconstruction computed on SUV and was significantly lower with the deep learning method and $T_1/T_2$ input. However, recently, Gong et al.\textsuperscript{148} tested on a brain patient cohort a CNN with either $T_1$ or Dixon and multiple echo UTE (mUTE) as input finding that using mUTE outperformed $T_1$. Liu et al.\textsuperscript{151} trained a CNN to predict CT tissue classes from diagnostic 1.5 T $T_1$ GRE of 30 patients. They tested on ten independent patients of the same cohort, whose results are reported in table 5 in terms of DSC. Then, they predicted sCT for five patients acquired prospectively with a 3T MRI/PET scanner ($T_1$ GRE), and they computed the PET$_{err}$, resulting $<1\%$. They concluded that DL approaches are flexible and promising to be applied to heterogeneous datasets acquired with different scanners and settings.

DL methods have also been proposed to estimate sCT from uncorrected PET. Thanks to the more considerable number of single PET exams, these methods have been tested on the full-body acquisitions and larger patient populations (up to 100 for training and 39 for testing). Although the global MAE is higher than site-specific MR-to-CT studies (about 110HU vs 10-15 HU), PET$_{err}$ is below 1\% on average, demonstrating the validity of the approach for the scope of PET AC.

\section{IV. Discussion}

This review encompassed DL-based approaches to generate sCT from other radiotherapy imaging modalities, focusing on published journal articles. The research topic was earlier introduced at conferences in 2016\textsuperscript{46}. Since 2016, we have observed increasing interest in using DL for sCT generation. DL methods’ success is probably related to the growth of available computational resources in the last decade that allowed training large volume datasets\textsuperscript{50} achieving fast image translation, i.e., in the order of a few seconds\textsuperscript{159}. Fast image-to-image translation facilitates applying DL in clinical cases and demonstrates its feasibility for clinical scenarios. In this review, we considered three clinical purposes for deriving sCT from other image modality, which are discussed in the following:

\section{I MR-only RT}

The generation of sCT for MR-only RT with DL is the most populated category. Its 51 papers demonstrate the potential of using DL for sCT generation from MRI. Several training techniques and configurations have been proposed. For anatom-
ical regions, as pelvis and brain/H&N, high image similarity and dosimetric accuracy, i.e., dose differences < 1%, can be achieved for photon RT and proton therapy. In region strongly affected by motion\textsuperscript{160,161}, e.g. abdomen and thorax, the first feasibility studies seem to be promising\textsuperscript{76,79,86,116,125}. However, no study proposed the generation of DL-based 4D sCT yet, as from classical methods\textsuperscript{162}. An exciting application is the DL-based sCT generation for the paediatric population\textsuperscript{75,76}, which is considered more radiation-sensitive than an adult population\textsuperscript{163} and could enormously benefit from MR-only, especially when patients’ simulations are repeated\textsuperscript{19}.

The geometric accuracy of sCT needs to be thoroughly tested to enable the clinical adoption of sCT for treatment planning purposes, primarily when MRI or sCT are used to substitute CT for position verification purposes. So far, the number of studies that investigated such an aspect from DL-based sCT is still scarce. Only Gupta et al.\textsuperscript{110}, for the brain, and Olberg et al.\textsuperscript{125}, for breast cancer, have investigated this aspect assessing the accuracy of alignment based on CBCT and digitally reconstructed radiography, respectively. Future studies are required to strengthen the clinical use of sCT, especially considering that geometric accuracy has been already extensively investigated for sCT generated with classical methods for 3 T and below\textsuperscript{164,165,166}.

DL-based sCT generation in the context of MR-guided radiotherapy\textsuperscript{20,167,168,169,170,171} may reduce the treatment time, facilitating daily image guidance and plan adaptation based on sole MRI\textsuperscript{172,173}. For this application, the accuracy of dose calculation in the magnetic field’s presence must be assessed before clinical implementation. So far, the studies investigating this aspect are still few, e.g. for abdominal\textsuperscript{79} and pelvic tumours\textsuperscript{86} and only considered low magnetic fields. Recently, Groot Koerkamp et al.\textsuperscript{174} published the first dosimetric evaluation of DL-based sCT for high magnetic field MR-guided RT achieving dose differences < 1% for breast cases. The results are promising, but we advocate for further studies on additional anatomical sites and magnetic field strengths.

II CBCT-to-CT for image-guided (adaptive) radiotherapy. In-room CBCT imaging is widespread in photon and proton RT for daily patient set-up\textsuperscript{175}. However, CBCT is not commonly exploited for daily plan adaptation and dose recalculation due to the artefacts associated with scatter and reconstruction algorithms that affect the quality of the electron density predicted by CBCT\textsuperscript{176}. Traditional methods to cope with this
issue have been based on image registration\textsuperscript{177,178}, scatter correction\textsuperscript{179}, look-up-table to rescale HU intensities\textsuperscript{180} and histogram matching\textsuperscript{181}. DL’s introduction for converting CBCT to sCT has substantially improved image quality leading to faster results than image registration and analytical corrections\textsuperscript{138}. Speed is crucial for the translation of the method into the clinical routine. However, one of the problems arising in CBCT-to-CT conversion for clinical application is the different field of view (FOV) between CBCT and CT. Usually, the training is performed by registering, cropping and resampling the volume to the CBCT size, which is smaller than the planning CT. Nonetheless, for replanning purposes, the limited FOV may hinder calculating the plan to the sCT. Some authors have proposed to assign water equivalent density within the CT body contour for the missing information\textsuperscript{134}. In other cases, the sCT patch has been stitched to the planning CT to cover the entire dose volume\textsuperscript{88}. Ideally, appropriate FOV coverage should be employed when re-calculating the plan for online adaptive RT. Besides the dosimetric aspect, improved image quality may increase accuracy during image guidance for patient set-up and OAR segmentation. These are necessary steps for online adaptive radiotherapy, especially for anatomical sites prone to large movements, as speculated by Liu et al.\textsuperscript{135} in the framework of pancreatic treatments. CBCT-to-CT resulted in accurate dose calculations both for photon and proton radiotherapy. For proton RT, the set-up accuracy and dose calculation are even more relevant to avoid range shift errors that could jeopardise the benefit of treatment\textsuperscript{67}. Because there is an intrinsic error in converting HU to relative proton stopping power\textsuperscript{182}, it has been shown that deep learning methods can translate CBCT directly to stopping power\textsuperscript{183}. This approach has not been covered in this review, but it is an exciting approach that will probably lead to further investigations.

Interestingly, increasing the quality of CBCT can be tackled as an image-to-image translation problem and as an inverse problem, i.e. from a reconstruction perspective. Specifically, by having the raw data measurements (projections), DL could improve tomography. In this sense, many investigations have been proposed but considered out of the scope of this review. For the interested reader, we suggest the following resources\textsuperscript{184,185,186,187,188}. Currently, it is unclear whether formulating (CB)CT quality enhancement as a synthesis or reconstruction problem would be beneficial. First attempts showed that training convolutional networks for reconstruction enhanced their

IV. DISCUSSION
generalisation capability to other anatomy\textsuperscript{189}; however, research on such aspects is still ongoing.

III PET attenuation correction. The sCT in this category is obtained either from MRI or from uncorrected PET. In the first case, the work’s motivation is to overcome the current limitations in generating attenuation maps ($\mu$-maps) from MR images in MRI/PET hybrid acquisitions that miscalculated the bone contribution\textsuperscript{190}. In the second case, the limits to overcome are different: i) to avoid extra-radiation dose when the sole PET exam is required, ii) to avoid misregistration errors when standalone CT and PET machines are used, iii) to be independent of the MR contrast in MRI/PET acquisitions. Besides the network configuration, MRI used for the input, or the number of patients included in the studies, DL-based sCT have consistently outperformed current MRAC methods available on commercial software. The results of this review support the idea that DL-based sCT will substitute current AC methods, being also able to overcome most of the limitations mentioned above. These aspects seem to contradict the stable number of papers in this category in the last three years. Nonetheless, we have to consider that the recent trend has been to derive the $\mu$-map from uncorrected PET via DL directly. Because this review considered only image-to-CT translation, these works were not included, but they can be found in a recent review by Lee\textsuperscript{47}. However, it is worth mentioning a recent study from Shiri et al.\textsuperscript{191}, where the largest patient cohort ever (1150 patients split in 900 for training, 100 for validation and 150 for test) was used for the scope. Direct $\mu$-map prediction via DL is an auspicious opportunity that may direct future research efforts in this context.

Deep learning considerations and trends
The number of patients used for training the networks is quite variable, ranging from a minimum of 7 (in I)\textsuperscript{72} to a maximum of 205 (in II)\textsuperscript{70} and 242\textsuperscript{69} (in I). In most cases, the patient number is limited to the availability of training pairs. Data augmentation is performed as linear and non-linear transformation\textsuperscript{192} to increase the training accuracy, as demonstrated in Pozaruk et al.\textsuperscript{149}. However, few publications investigated the impact of increasing the training size\textsuperscript{69,71,75,125,131}, finding that image similarity increases when training up to fifty patients. This investigation can indicate the minimum amount of patients necessary to include in the training to achieve the state of the art performances. The optimal patient number may also
depend on the anatomical site and its inter-fraction and intra-fraction variability. Besides, attention should be dedicated to balancing the training set, as performed in\textsuperscript{69,75}. Otherwise, the network may overfit, as previously demonstrated for segmentation tasks\textsuperscript{193}.

GANs were the most popular architecture, but we cannot conclude that it is the best network scheme for sCT. Indeed, some studies compared U-net or other CNN vs GAN finding GAN performing statistically better\textsuperscript{89,143}; others found similar results\textsuperscript{149,150} or even worse performances \textsuperscript{80,148}. We can speculate that, as demonstrated by\textsuperscript{117}, a vital role is played by the loss function, which, despite being the effective driver for network learning, has been investigated less than the network architecture, as highlighted for image restoration\textsuperscript{194}. Another important aspect is the growing trend, except category III, in unpaired training (5 and 7 papers in 2019 and 2020, respectively). The quality of the registration when training in a paired manner influences the quality of deep learning-based sCT\textsuperscript{126}. In this sense, unpaired training offers an option to alleviate the need for well-matched training pairs. When comparing paired vs unpaired training, we observed that paired training leads to slightly better performances. However, the differences were not always statistically significant\textsuperscript{71,80,95}. As proposed by Yang et al.\textsuperscript{95}, unsupervised training decreases the semantic information from one domain to another\textsuperscript{95}. Such an issue may be solved by introducing a structure-consistency loss, which extracts structural features from the image defining the loss in the feature space. Yang et al.’s results showed improvements in this sense relative to other unsupervised methods. They also showed that pre-registering unpaired MR-CT further improves unsupervised training results, which can be an option when input and target images are available, but perfect alignment is not achievable. In some cases, unpaired training even demonstrated to be superior to paired training\textsuperscript{195}. A trend lately emerged is the use of architecture initially thought for unpaired training, e.g. cycle-GAN to be used for paired training\textsuperscript{87,94}.

Focusing on the body sites, we observed that most of the investigations were conducted in the brain, H&N and pelvic regions. Fewer studies are available for the thorax and the abdomen, representing a more challenging patient population due to the organ motion\textsuperscript{196}.

In MR-only RT, we found contradicting results regarding the best performing spatial configuration for the papers that directly compared 2D vs 3D training\textsuperscript{82,83}. It is undoubtedly clear that 2D+ increases the sCT quality compared to a single 2D views, as demonstrated in Spadea et al.\textsuperscript{111} and Maspero et al.\textsuperscript{75}; however, when comparing 2D against 3D training,

IV. DISCUSSION
patch size is a vital aspect. 3D deep networks require a more significant number of training parameters than 2D networks. For sCT generation, the approaches adopted have chosen to use patch size much smaller than the whole volume, probably hindering the contextual information considered. Generally, downsampling approaches have been proposed to increase the network’s perceptive field, e.g. for segmentation tasks, but they have not been applied to sCT generation. We believe this will be an exciting area of research.

For what concerns the latest development from the deep learning perspective, in 2018, Oktay et al. proposed a new mechanism, called attention gate (AG), to focus on target structures that can vary in shape and size. Liu et al. incorporated the AG in the generator of a cycle-GAN to learn organ variation from CBCT-CT pairs in the context of pancreas adaptive RT, showing that its contribution significantly improved the predictions compared to a network without AG. Other papers also adopted attention. Embedding has also been proposed to increase the network’s expressivity of the network and applied by Xiang et al. As AG’s mechanism is a way to focus the attention on specific portions of the image, it can potentially open the path for new research topics. In 2019, Schlemper and colleagues evaluated the AG for different tasks in medical image processing: classification, object detection, segmentation. So, in the online IGART, we can envision that such a mechanism could lead to multi-task applications, such as deriving sCT while delineating the structure of interests.

**Benefits and challenges for clinical implementations**

Deep learning-based sCT generations may reduce the need for additional or non-standard MRI sequences, e.g. UTE or ZTE. Avoiding additional sequences will shorten the total acquisition time, speed up the workflow, increasing patient throughput. As already mentioned, speed is particularly interesting for MR-guided RT and for adaptive RT in II, which is considered crucial for online correction. For what concern categories II and III, the generation of DL-based sCT possibly enables dose decreasing during imaging by reducing the need for CT in case of anatomical changes (in II) or by possibly diminishing the amount of radioactive material injected (in III).

Finally, it is worth commenting on the current status of the clinical adoption of DL-based sCT. We could not find that any of the methods considered are now clinically implemented and used. We speculate that this is probably related to the fact that the field is still relatively
young, with the first publications only from 2017 and that time for clinical implementations generally last years, if not decades\textsuperscript{200,201}. Additionally, as already mentioned, for categories I/II, the impact of sCT for position verification still needs to be thoroughly investigated. The implementation may also be more comfortable for category III if the methods would be directly integrated into scanners. In general, the involvement of vendors may streamline the clinical adoption of DL-based sCT. In this sense, we can report that vendors are currently active in evaluating their methods in research settings, e.g. for brain\textsuperscript{69}, pelvis\textsuperscript{120} in I, and for H&N, thorax and pelvis in II\textsuperscript{70}. In the last month, Palmer et al.\textsuperscript{202} also reported using a pre-released version of a DL-based sCT generation approach for H&N in MR-only RT. Another essential aspect that needs to be satisfied is the compliance to the currently adopted regulations\textsuperscript{203}, where vendors can offer vital support\textsuperscript{204,205}.

A key aspect of clinical implementation is the precise definition of a DL-based solution’s requirements before being accepted. If we consider the reported metrics, we cannot find uniform criteria for reporting. Multiple metrics have been defined, and it is not clear which region of interests they should be computed. For example, the image-based similarity was reported on the body contour or in tissues generally defined by different thresholds; for task-specific metrics, the methods employed are even more heterogeneous. For example, in I and II, gamma pass rates can be performed in 2D, 3D and different dose thresholds level have been employed, e.g. 10%, 30%, 50% or 90% of the prescribed or maximum dose. In III, the PET\textsubscript{err} can be computed either on SUV, max SUV or larger VOI, making it difficult to compare different network configurations’ performances. We think that this lack of standardisation in reporting the results is also detrimental to clinical adoption. A first attempt at revising the metrics currently adopted has been performed by Liesbeth et al.\textsuperscript{206}. However, this is still insufficient, considering the differences in how such metrics can be calculated and reported. In this sense, we advocate for consensus-based requirements that may facilitate reporting in future clinical trials\textsuperscript{207}. Also, no public datasets arranged in the form of grand challenges (https://grand-challenge.org/) are available to enable a fair and open evaluation of different approaches\textsuperscript{208}.

To date, four scientific studies have already investigated the performance of DL-based sCT in a multi-centre setting\textsuperscript{75,118,119,120}. These studies have been reported only for MR-only RT. Future work should focus on assessing the performance of DL-based sCT generation for II and III. On the contrary, investigations on sCT generation with classical methods

\section*{IV. DISCUSSION}
using multi-centre data are more diffuse for all the three categories\textsuperscript{26,209,210,211,212,213}. Of particular relevance when considering the generalisation of a DL model for sCT generation may be the application of transfer learning\textsuperscript{214,215}. Mainly, transfer learning may be exploited to facilitate fine-tuning a model pre-trained on a specific MRI contrast or CBCT image protocols; or generalise among multiple anatomies. No paper was found up to December 2020 investigating this aspect, but it could be an exciting research area. More recently, Li et al.\textsuperscript{216} showed that transfer learning facilitated training a DL model on different MRI contrasts for sCT generation.

The quality of sCT cannot be judged by a user, except when its quality is inferior. Therefore, software-based quality assurance (QA) procedures should be put in place. In general, having at disposal phantoms to verify the quality of the sCT may enable regular QA procedures, as for QA of CT\textsuperscript{217}. This would be relatively straightforward for II; however, in MR-based sCT, phantoms’ manufacturing is quite challenging due to the need for contrast in MRI and CT. Recently, the first phantoms have been proposed for such task\textsuperscript{218,219,220,221} showing the potential of additive manufacturing.

Alternatively, it would be relevant if a CNN could automatically generate a metric to assess the quality of sCTs, as, for example, already presented for automatic segmentation\textsuperscript{222}. In this sense, Bragman et al.\textsuperscript{223} introduce uncertainty for such a task by adopting a multi-task network and a Bayesian probabilistic framework. More recently, two other works proposed to use uncertainty either from the combination of independently trained networks\textsuperscript{75} or via dropout-based variational inference\textsuperscript{224}. So far, the field of uncertainty estimation with deep learning\textsuperscript{225} has been superficially touched for sCT generation. It would be interesting to see future work focusing on developing criteria for automatically identifying failure cases using uncertainty prediction. Patients with inaccurate synthetic CTs will be flagged for CT rescan or manual adjustment of the sCT if deemed feasible.

**Beyond sCT for radiotherapy**

We found other possible applications of DL-based image generation during the database search, which are beyond the categories mentioned so far or the radiotherapy application. For example, Kawahara et al.\textsuperscript{226} proposed to generate synthetic dual-energy CT from CT to assess the body material composition using 2D paired GANs. Also, commercial solutions start to be evaluated for the generation of DL-based sCT from MRI for lesion detection.
of suspected sacroiliitis\textsuperscript{227} or to facilitate surgical planning of the spine\textsuperscript{228}. An exciting application is also the generation of sCT to facilitate multi-modal image registration, as proposed by McKenzie et al.\textsuperscript{229}.

All the techniques of category I could be directly applied to MR-guided high-intensity focused ultrasound, where otherwise an additional CT would be required to plan the treatment properly\textsuperscript{230}.

Additionally, the methods here reviewed to generate sCT can be applied to translating other image modalities. Interesting examples in the RT realm are provided by Jiang et al.\textsuperscript{231}, who investigated using MRI-to-CT translation to increase the segmentation’s robustness. Kieselmann et al.\textsuperscript{232} generated synthetic MRI from CT to train segmentation networks that exploit the wealth of delineation on another modality. A detailed review of other image-to-image translation applications in radiotherapy has been recently compiled by Wang et al.\textsuperscript{49}.

V. Conclusion

Deep learning-based methods for sCT generation have been reviewed in the context of I) MR to replace CT in radiotherapy treatment planning, II) CBCT-based adaptive radiotherapy, and III) in generating attenuation maps for PET.

For each category, we presented a detailed comparison in terms of imaging protocols, DL architectures, and performances according to the most popular metrics reported. We found that DL-based sCT generation is an active and growing area of research. For several anatomical sites, e.g. H\&N/brain and pelvis, sCT seems feasible, with deep learning achieving dose difference to CT-based planning $< 1\%$ in the radiotherapy context and better performance for PET attenuation correction to the standard MRAC methods.

We can conclude that the deep learning-based generation of sCT has a bright future, with an extensive amount of research work being done on the topic. Further steps to spread DL-based sCT techniques into the clinic will be necessary to evaluate their generalisation among multiple centres and propose comprehensive commissioning and QA methods, to ensure treatment efficacy and patient safety.
VI. Acknowledgements

Matteo Maspero is grateful to prof.dr.ir. Cornelis (Nico) A.T. van den Berg, head of the Computational Imaging Group for MR diagnostics & therapy, Center for Image Sciences, UMC Utrecht, the Netherlands for the general support provided during this manuscript’s compilation.

VII. Conflict of interest

None of the authors has conflict of interests to disclose.

Appendix

The query used in selected databases - PubMed, Scopus and Web of Science - in the fields (Title/Abstract/Keywords) was the following (Figure 4):

(("radiotherapy") OR ("radiation therapy") OR ("proton therapy") OR ("oncology") OR ("imaging") OR ("radiology") OR ("healthcare") OR ("CBCT") OR ("cone-beam CT") OR ("PET") OR ("attenuation correction") OR ("attenuation map") AND ("synthetic CT") OR ("syntheticCT") OR ("synthetic-CT") OR ("pseudo CT") OR ("pseudoCT") OR ("pseudo-CT") OR ("virtual CT") OR ("virtualCT") OR ("virtual-CT") OR ("derived CT") OR ("derivedCT") OR ("derived-CT") OR ("sCT").AND ("deep learning") OR ("convolutional network") OR ("CNN") OR ("GAN") OR ("GANN") OR (artificial intelligence));
Figure 4: Schematic of the search inclusion/exclusion criteria adopted for this review selecting the time window, keywords, type of article, content and the three categories defined.

VIII. Acronyms and abbreviations

2Dp: 2D patches; 3Dp: 3D patches; AC: attenuation correction; aff: affine; AT: attention gate; back: backwards pass; CBCT: cone-beam computed tomography; CC: cross-correlation; CNNs: Convolutional neural networks; cor: coronal; CT: computed tomography; D: discriminator; DD: dose difference; def: deformable; DL: deep learning; DPR: dose pass rate; DSC: Dice similarity coefficient; DVH: dose-volume histogram; ens: ensemble; FID: Frechet inception distance; FLAIR: fluid-attenuated inversion recovery; forw: forward pass; FOV: field of view; G: generator; GANs: generative adversarial networks; Gd: Gadolinium; GPR: gamma pass rate; GRE: gradient recalled-echo; H&N: head & neck; IGART: image-guided adaptive radiation therapy; m2D: multi-2D; MAE: mean absolute error; MR: magnetic resonance; MRAC: magnetic resonance attenuation correction; MSE: mean squared error; mUTE: multiple echo UTE; NCC: normalised cross-correlation; OARS: organs-at-risk; p: proton; paed: paediatric; PET\textsubscript{err}: absolute error PET reconstruction; PET\textsubscript{err}: relative error PET reconstruction; PET: positron emission...
tomography; **PSNR**: peak signal-to-noise ratio; **QA**: quality assurance; **rig**: rigid; **RMSE**: root mean squared error; **ROI**: region-of-interest; **RS**: range shift; **RT**: radiotherapy; **sag**: sagittal; **sCT**: synthetic computed tomography; **SSIM**: structural similarity index measure; **SUV**: standard uptake values; **tra**: transverse; **TSE**: turbo spin-echo; **UTE**: ultra-short echo time; **VOI**: volume-of-interest; **x**: photon; **ZTE**: zero time echo; **µ−map**: attenuation maps.

**References**

1. J. Husband, R. H. Reznek, and J. E. Husband, *Imaging in oncology*, CRC Press, 2016.

2. L. Beaton, S. Bandula, M. N. Gaze, and R. A. Sharma, How rapid advances in imaging are defining the future of precision radiation oncology, Br J Cancer **120**, 779–790 (2019).

3. D. Verellen, M. De Ridder, N. Linthout, K. Tournel, G. Soete, and G. Storme, Innovations in image-guided radiotherapy, Nat Rev Canc **7**, 949–960 (2007).

4. D. A. Jaffray, Image-guided radiotherapy: from current concept to future perspectives, Nat Rev Clin Oncol **9**, 688 (2012).

5. J. Seco and M. F. Spadea, Imaging in particle therapy: state of the art and future perspective, Acta Oncol **54**, 1254–1258 (2015).

6. IAEA, *Radiotherapy in Cancer Care: Facing the Global Challenge*, Non-serial Publications, INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna, 2017.

7. J. Seco and P. M. Evans, Assessing the effect of electron density in photon dose calculations, Medical Physics **33**(2), 540–552 (2006).

8. M. Unterrainer et al., Recent advances of PET imaging in clinical Radiat Oncol, Radiat Oncol **15**, 1:15 (2020).

9. P. Dirix, K. Haustermans, and V. Vandecaveye, The value of magnetic resonance imaging for radiotherapy planning, **24**, 151–159 (2014).

10. M. A. Schmidt and G. S. Payne, Radiotherapy planning using MRI, Phys Med Biol **60**, R323 (2015).
11 S. Devic, MRI simulation for radiotherapy treatment planning., Med Phys 39, 6701 (2012).

12 T. Nyholm, M. Nyberg, M. G. Karlsson, and M. Karlsson, Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments, Radiat Oncol 4, 1–9 (2009).

13 K. Ulin, M. M. Urie, and J. M. Cherlow, Results of a multi-institutional benchmark test for cranial CT/MR image registration, Int J Radiat Oncol Biol Phys 77, 1584–1589 (2010).

14 B. A. Fraass, D. L. McShan, R. F. Diaz, R. K. Ten Haken, A. Aisen, S. Gebarski, G. Glazer, and A. S. Lichter, Integration of magnetic resonance imaging into radiation therapy treatment planning: i. technical considerations, Int J Radiat Oncol Biol Phys 13, 1897–908 (1987).

15 Y. K. Lee, M. Bollet, G. Charles-Edwards, M. A. Flower, M. O. Leach, H. McNair, E. Moore, C. Rowbottom, and S. Webb, Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone, Radiother Oncol 66, 203–216 (2003).

16 T. Nyholm and J. Jonsson, Counterpoint: Opportunities and Challenges of a Magnetic Resonance Imaging-Only Radiotherapy Work Flow, Semin Radiat Oncol 24, 175–80 (2014).

17 M. Kapanen, J. Collan, A. Beule, T. Seppälä, K. Saarilahti, and M. Tenhunen, Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate, Magn Reson Med 70, 127–35 (2013).

18 A. M. Owrangi, P. B. Greer, and C. K. Glide-Hurst, MRI-only treatment planning: benefits and challenges, Phys Med Biol 63, 05TR01 (2018).

19 M. Karlsson, M. G. Karlsson, T. Nyholm, C. Amies, and B. Zackrisson, Dedicated Magnetic Resonance Imaging in the Radiotherapy Clinic, Int. J. Radiat. Oncol. Biol. Phys. 74, 644–51 (2009).

20 J. J. Lagendijk, B. W. Raaymakers, C. A. Van den Berg, M. A. Moerland, M. E. Philippens, and M. Van Vulpen, MR guidance in radiotherapy, Phys Med Biol 59, R349 (2014).
21 J. H. Jonsson, M. G. Karlsson, M. Karlsson, and T. Nyholm, Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions, Radiat Oncol 5, 62 (2010).

22 J. M. Edmund and T. Nyholm, A review of substitute CT generation for MRI-only radiation therapy, Radiat Oncol 12 (2017).

23 E. Johnstone, J. J. Wyatt, A. M. Henry, S. C. Short, D. Sebag-Montefiore, L. Murray, C. G. Kelly, H. M. McCallum, and R. Speight, Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging-Only Radiation Therapy, Int J Radiat Oncol Biol Phys 100, 199–217 (2018).

24 B. Wafa and A. Moussaoui, A review on methods to estimate a CT from MRI data in the context of MRI-alone RT, Med Tech J 2, 150–178 (2018).

25 L. Kerkmeijer, M. Maspero, G. Meijer, J. van der Voort van Zyp, H. de Boer, and C. van den Berg, Magnetic Resonance Imaging only Workflow for Radiotherapy Simulation and Planning in Prostate Cancer, Clinic Oncol 30, 692–701 (2018).

26 D. Bird, A. M. Henry, D. Sebag-Montefiore, D. L. Buckley, B. Al-Qaisieh, and R. Speight, A Systematic Review of the Clinical Implementation of Pelvic Magnetic Resonance Imaging-Only Planning for External Beam Radiation Therapy, Int J Radiat Oncol Biol Phys 105, 479–492 (2019).

27 D. Thorwarth and D. A. Low, Technical Challenges of Real-Time Adaptive MR-Guided Radiotherapy, Front Oncol 11 (2021).

28 A. Hoffmann, B. Oborn, M. Moteabbed, S. Yan, T. Bortfeld, A. Knopf, H. Fuchs, D. Georg, J. Seco, M. F. Spadea, O. Jäkel, C. Kurz, and K. Parodi, MR-guided proton therapy: a review and a preview, Radiat Oncol 15 (2020).

29 V. T. Taasti, P. Klages, K. Parodi, and L. P. Muren, Developments in deep learning based corrections of cone beam computed tomography to enable dose calculations for adaptive radiotherapy, Physics and Imaging in Radiat Oncol 15, 77–79 (2020).

30 L. Zhu, J. Wang, and L. Xing, Noise suppression in scatter correction for cone-beam CT, Med Phys 36, 741–752 (2009b).
31 L. Zhu, Y. Xie, J. Wang, and L. Xing, Scatter correction for cone-beam CT in radiation therapy, Med Phys 36, 2258–2268 (2009c).

32 A. Mehranian, H. Arabi, and H. Zaidi, Vision 20/20: magnetic resonance imaging-guided attenuation correction in PET/MRI: challenges, solutions, and opportunities, Med Phys 43, 1130–1155 (2016).

33 I. Mecheter, L. Alic, M. Abbod, A. Amira, and J. Ji, MR Image-Based Attenuation Correction of Brain PET Imaging: Review of Literature on Machine Learning Approaches for Segmentation, Journal of Digital Imaging, 1–18 (2020).

34 C. Catana, Attenuation correction for human PET/MRI studies, Phys Med Biol 65, TR02 (2020).

35 Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature 521, 436–444 (2015).

36 I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, Number 2 in Adaptive Computation and Machine Learning, MIT press Cambridge, 2016.

37 P. Meyer, V. Noblet, C. Mazzara, and A. Lallement, Survey on deep learning for radiotherapy, Comp Biol Med 98, 126–146 (2018).

38 B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, Deep learning in medical imaging and radiation therapy, Med Phys 46, e1–e36 (2018).

39 I. Boon, T. A. Yong, and C. Boon, Assessing the Role of Artificial Intelligence (AI) in Clinical Oncology: Utility of Machine Learning in Radiotherapy Target Volume Delineation, Medicines 5, 131 (2018).

40 C. Wang, X. Zhu, J. C. Hong, and D. Zheng, Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future, Tech Canc Res Treat 18, 153303381987392 (2019).

41 L. Boldrini, J.-E. Bibault, C. Masciocchi, Y. Shen, and M.-I. Bittner, Deep Learning: A Review for the Radiation Oncologist, Front Oncol 9 (2019).
42. D. Jarrett, E. Stride, K. Vallis, and M. J. Gooding, Applications and limitations of machine learning in Radiat Oncol, Brit J Radiol 92, 20190001 (2019).

43. K. J. Kiser, C. D. Fuller, and V. K. Reed, Artificial intelligence in Radiat Oncol treatment planning: a brief overview, J Med Art Intel 2, 9–9 (2019).

44. A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, Adv Neur Inf Proc Syst 25, 1097–1105 (2012).

45. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, Med Image Anal 42, 60–88 (2017).

46. D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, Estimating CT image from MRI data using 3D fully convolutional networks, pages 170–178 (2016).

47. J. S. Lee, A review of deep learning-based approaches for attenuation correction in positron emission tomography, IEEE Transactions on Radiation and Plasma Medical Sciences (2020).

48. B. Yu, Y. Wang, L. Wang, D. Shen, and L. Zhou, Medical Image Synthesis via Deep Learning, pages 23–44, Springer International Publishing, Cham, 2020.

49. T. Wang, Y. Lei, Y. Fu, J. F. Wynne, W. J. Curran, T. Liu, and X. Yang, A review on medical imaging synthesis using deep learning and its clinical applications, J App Clin Med Phys (2020).

50. Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature 521, 436–444 (2015).

51. O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in International Conference on Medical image computing and computer-assisted intervention, pages 234–241, Springer, 2015.

52. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27, 2672–2680 (2014).
53 P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, Image-to-image translation with conditional adversarial networks, in *Proc IEEE CVPR*, pages 1125–1134, 2017.

54 X. Wu, K. Xu, and P. Hall, A survey of image synthesis and editing with generative adversarial networks, *Tsinghua Science and Technology* 22, 660–674 (2017).

55 A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, Generative adversarial networks: An overview, *IEEE Signal Processing Magazine* 35, 53–65 (2018).

56 X. Yi, E. Walia, and P. Babyn, Generative adversarial network in medical imaging: A review, *Medical image analysis* 58, 101552 (2019).

57 J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

58 C. Paganelli, G. Meschini, S. Molinelli, M. Riboldi, and G. Baroni, Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats, *Med Phys* 45, e908–e922 (2018).

59 Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Trans Imag Proc* 13, 600–612 (2004).

60 L. R. Dice, Measures of the amount of ecologic association between species, *Ecology* 26, 297–302 (1945).

61 D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, Comparing images using the Hausdorff distance, *IEEE Transactions on pattern analysis and machine intelligence* 15, 850–863 (1993).

62 A. Reinke et al., Common limitations of image processing metrics: A picture story, *arXiv preprint arXiv:2104.05642* (2021).

63 D. A. Low, Gamma dose distribution evaluation tool, *250*, 012071 (2010).

64 B. M. Clasie, G. C. Sharp, J. Seco, J. B. Flanz, and H. M. Kooy, Numerical solutions of the $\gamma$-index in two and three dimensions, *Physics in Medicine & Biology* 57, 6981 (2012).
M. Hussein, C. Clark, and A. Nisbet, Challenges in calculation of the gamma index in radiotherapy – Towards good practice, Phys Med 36, 1–11 (2017).

R. Drzymala, R. Mohan, L. Brewster, J. Chu, M. Goitein, W. Harms, and M. Urie, Dose-volume histograms, Int J Radiat Oncol Biol Phys 21, 71–78 (1991).

H. Paganetti, Range uncertainties in proton therapy and the role of Monte Carlo simulations, Phys Med Biol 57, R99 (2012).

G. Pileggi, C. Speier, G. C. Sharp, D. Izquierdo Garcia, C. Catana, J. Pursley, F. Amato, J. Seco, and M. F. Spadea, Proton range shift analysis on brain pseudo-CT generated from T1 and T2 MR, Acta Oncologica 57, 1521–1531 (2018).

E. A. Andres et al., Dosimetry-driven quality measure of brain pseudo Computed Tomography generated from deep learning for MRI-only radiotherapy treatment planning, Int J Radiat Oncol Biol Phys 108, 813–823 (2020).

M. Eckl, L. Hoppen, G. R. Sarria, J. Boda-Heggemann, A. Simeonova-Chergou, V. Steil, F. A. Giordano, and J. Fleckenstein, Evaluation of a cycle-generative adversarial network-based cone-beam CT to synthetic CT conversion algorithm for adaptive radiation therapy, Physica Medica 80, 308–316 (2020).

Y. Peng et al., Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning, Radiother Oncol 150, 217–224 (2020).

P. Qian, K. Xu, T. Wang, Q. Zheng, H. Yang, A. Baydoun, J. Zhu, B. Traughber, and R. F. Muzic, Estimating CT from MR Abdominal Images Using Novel Generative Adversarial Networks, J Grid Comp 18, 1–16 (2020).

K. Xu, J. Cao, K. Xia, H. Yang, J. Zhu, C. Wu, Y. Jiang, and P. Qian, Multichannel residual conditional GAN-leveraged abdominal pseudo-CT generation via Dixon MR images, IEEE Access 7, 163823–163830 (2019).

C. N. Ladefoged, L. Marner, A. Hindsholm, I. Law, L. Højgaard, and F. L. Andersen, Deep learning based attenuation correction of PET/MRI in pediatric brain tumor patients: Evaluation in a clinical setting, Frontiers in neuroscience 12, 1005 (2019).
M. Maspero, L. G. Bentvelzen, M. H. Savenije, F. Guerreiro, E. Seravalli, G. O. Janssens, C. A. van den Berg, and M. E. Philippens, Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy, Radiother Oncol 153, 197–204 (2020).

M. C. Florkow et al., Deep learning-enabled MRI-only photon and proton therapy treatment planning for paediatric abdominal tumours, Radiother Oncol 153, 220–227 (2020).

W. Jeon, H. J. An, J.-i. Kim, J. M. Park, H. Kim, K. H. Shin, and E. K. Chie, Preliminary Application of Synthetic Computed Tomography Image Generation from Magnetic Resonance Image Using Deep-Learning in Breast Cancer Patients, J Radiat Prot Res 44, 149–155 (2019).

T. J. Bradshaw, G. Zhao, H. Jang, F. Liu, and A. B. McMillan, Feasibility of deep learning–based PET/MR attenuation correction in the pelvis using only diagnostic MR images, Tomography 4, 138 (2018).

J. Fu, K. Singhrao, M. Cao, V. Yu, A. P. Santhanam, Y. Yang, M. Guo, A. C. Raldow, D. Ruan, and J. H. Lewis, Generation of abdominal synthetic CTs from 035 T MR images using generative adversarial networks for MR-only liver radiotherapy, Biom Phys Eng Express 6, 015033 (2020).

Y. Li, W. Li, J. Xiong, J. Xia, and Y. Xie, Comparison of Supervised and Unsupervised Deep Learning Methods for Medical Image Synthesis between Computed Tomography and Magnetic Resonance Images, BioMed Research International 2020 (2020).

L. Xu, X. Zeng, H. Zhang, W. Li, J. Lei, and Z. Huang, BPGAN: Bidirectional CT-to-MRI prediction using multi-generative multi-adversarial nets with spectral normalization and localization, Neural Networks 128, 82–98 (2020).

J. Fu, Y. Yang, K. Singhrao, D. Ruan, F.-I. Chu, D. A. Low, and J. H. Lewis, Deep learning approaches using 2D and 3D convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging, Med Phys 46, 3788–3798 (2019).
83 S. Neppl et al., Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans, Acta Oncol 58, 1429–1434 (2019).

84 L. Fetty, M. Bylund, P. Kuess, G. Heilemann, T. Nyholm, D. Georg, and T. Löfstedt, Latent space manipulation for high-resolution medical image synthesis via the StyleGAN, Zeits Med Phy 30 (2020).

85 L. Xiang, Q. Wang, D. Nie, L. Zhang, X. Jin, Y. Qiao, and D. Shen, Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image, Med Imag Anal 47, 31–44 (2018).

86 D. Cusumano et al., A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases, Radiother Oncol 153, 205–212 (2020).

87 J. Harms, Y. Lei, T. Wang, R. Zhang, J. Zhou, X. Tang, W. J. Curran, T. Liu, and X. Yang, Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography, Med Phys 46, 3998–4009 (2019).

88 M. Maspero, A. C. Houweling, M. H. Savenije, T. C. van Heijst, J. J. Verhoeff, A. N. Kotte, and C. A. van den Berg, A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer, Phys Imag Radiat Oncol 14, 24–31 (2020).

89 Y. Zhang, N. Yue, M.-Y. Su, B. Liu, Y. Ding, Y. Zhou, H. Wang, Y. Kuang, and K. Nie, Improving CBCT Quality to CT Level using Deep-Learning with Generative Adversarial Network, Med Phys (2020).

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

91 X. Han, MR-based synthetic CT generation using a deep convolutional neural network method, Med Phys 44, 1408–1419 (2017).
92 H. Emami, M. Dong, S. P. Nejad-Davarani, and C. K. Glide-Hurst, Generating synthetic CTs from magnetic resonance images using generative adversarial networks, Med Phys 45, 3627–3636 (2018).

93 C.-B. Jin, H. Kim, M. Liu, W. Jung, S. Joo, E. Park, Y. S. Ahn, I. H. Han, J. I. Lee, and X. Cui, Deep CT to MR synthesis using paired and unpaired data, Sensors 19, 2361 (2019).

94 Y. Lei, J. Harms, T. Wang, Y. Liu, H.-K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang, MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks, Med Phys 46, 3565–3581 (2019).

95 H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, J. L. Prince, and Z. Xu, Unsupervised MR-to-CT Synthesis Using Structure-Constrained CycleGAN, IEEE Trans Med Imaging 39, 4249–4261 (2020).

96 H. Massa, J. Johnson, and A. McMillan, Comparison of deep learning synthesis of synthetic CTs using clinical MRI inputs, Phys Med Biol 65, NT03 (2020).

97 Y. Wang, C. Liu, X. Zhang, and W. Deng, Synthetic CT generation based on T2 weighted MRI of nasopharyngeal carcinoma (NPC) using a deep convolutional neural network (DCNN), Front Oncol 9 (2019).

98 X. Tie, S.-K. Lam, Y. Zhang, K.-H. Lee, K.-H. Au, and J. Cai, Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients, Med Phys 47, 1750–1762 (2020).

99 V. Kearney, B. P. Ziemer, A. Perry, T. Wang, J. W. Chan, L. Ma, O. Morin, S. S. Yom, and T. D. Solberg, Attention-Aware Discrimination for MR-to-CT Image Translation Using Cycle-Consistent Generative Adversarial Networks, Radiol: Art Intel 2, e190027 (2020).

100 A. Largent et al., Head-and-Neck MRI-only radiotherapy treatment planning: From acquisition in treatment position to pseudo-CT generation, Cancer/Radiothérapie 24, 288–297 (2020).
P. Su, S. Guo, S. Roys, F. Maier, H. Bhat, E. Melhem, D. Gandhi, R. Gullapalli, and J. Zhuo, Transcranial MR Imaging–Guided Focused Ultrasound Interventions Using Deep Learning Synthesized CT, Am J Neurorad 41, 1841–1848 (2020).

M. C. Florkow et al., Deep learning–based MR-to-CT synthesis: The influence of varying gradient echo–based MR images as input channels.

A. Bahrami, A. Karimian, E. Fatemizadeh, H. Arabi, and H. Zaidi, A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI, Med Phys 47, 5158–5171 (2020).

Y. Liu et al., MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method, Phys Med Biol 64, 145015 (2019).

L. Liu, A. Johansson, Y. Cao, J. Dow, T. S. Lawrence, and J. M. Balter, Abdominal synthetic CT generation from MR Dixon images using a U-net trained with ‘semi-synthetic’ CT data, Phys Med Biol 65, 125001 (2020).

A. M. Dinkla, J. M. Wolterink, M. Maspero, M. H. Savenije, J. J. Verhoeff, E. Seravalli, I. Išgum, P. R. Seevinck, and C. A. van den Berg, MR-only brain radiation therapy: dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network, Int J Radiat Oncol Biol Phys 102, 801–812 (2018).

F. Liu, P. Yadav, A. M. Baschnagel, and A. B. McMillan, MR-based treatment planning in radiation therapy using a deep learning approach, J App Clin Med Phys 20, 105–114 (2019).

S. Kazemifar, S. McGuire, R. Timmerman, Z. Wardak, D. Nguyen, Y. Park, S. Jiang, and A. Owrangi, MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach, Radiother Oncol 136, 56–63 (2019).

G. Shafai–Erfani et al., MRI-based proton treatment planning for base of skull tumors, Int J Part Ther 6, 12–25 (2019).
D. Gupta, M. Kim, K. A. Vineberg, and J. M. Balter, 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 9, 964 (2019).

M. F. Spadea, G. Pileggi, P. Zaffino, P. Salome, C. Catana, D. Izquierdo-Garcia, F. Amato, and J. Seco, Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—application in brain proton therapy, Int J Radiat Oncol Biol Phys 105, 495–503 (2019).

Y. Koike, Y. Akino, I. Sumida, H. Shiomi, H. Mizuno, M. Yagi, F. Isohashi, Y. Seo, O. Suzuki, and K. Ogawa, Feasibility of synthetic computed tomography generated with an adversarial network for multi-sequence magnetic resonance-based brain radiotherapy, J Radiat Res 61, 92–103 (2020).

S. Kazemifar, A. M. Barragán Montero, K. Souris, S. T. Rivas, R. Timmerman, Y. K. Park, S. Jiang, X. Geets, E. Sterpin, and A. Owrangi, Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors, J App Clin Med Phys (2020).

S. Chen, A. Qin, D. Zhou, and D. Yan, U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning, Med Phys 45, 5659–5665 (2018).

H. Arabi, J. A. Dowling, N. Burgos, X. Han, P. B. Greer, N. Koutsouvelis, and H. Zaidi, Comparative study of algorithms for synthetic CT generation from MRI: consequences for MRI-guided radiation planning in the pelvic region, Med Phys 45, 5218–5233 (2018).

Y. Liu et al., Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning, Phys Med Biol 64, 205022 (2019).

A. Largent et al., Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning, Int J Radiat Oncol Biol Phys 105, 1137–1150 (2019).
K. N. B. Boni, J. Klein, L. Vanquin, A. Wagner, T. Lacornerie, D. Pasquier, and N. Reynaert, MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network, Phys Med Biol 65, 075002 (2020).

L. Fetty, T. Löfstedt, G. Heilemann, H. Furtado, N. Nesvacil, T. Nyholm, D. Georg, and P. Kuess, Investigating conditional GAN performance with different generator architectures, an ensemble model, and different MR scanners for MR-sCT conversion, Phys Med Biol 65, 5004 (2020).

D. Bird et al., Multicentre, deep learning, synthetic-CT generation for ano-rectal MR-only radiotherapy treatment planning, Radiother Oncol 156, 23–28 (2021).

A. M. Dinkla, M. C. Florkow, M. Maspero, M. H. Savenije, F. Zijlstra, P. A. Doornaert, M. van Stralen, M. E. Philippens, C. A. van den Berg, and P. R. Seevinck, Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network, Med Phys 46, 4095–4104 (2019).

P. Klages, I. Benslimane, S. Riyahi, J. Jiang, M. Hunt, J. O. Deasy, H. Veeraraghavan, and N. Tyagi, Patch-based generative adversarial neural network models for head and neck MR-only planning, Med Phys 47, 626–642 (2020).

M. Qi et al., Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy, Med Phys 47, 1880–1894 (2020).

A. Thummerer, B. A. de Jong, P. Zaffino, A. Meijers, G. G. Marmitt, J. Seco, R. J. Steenkikers, J. A. Langendijk, S. Both, and M. F. Spadea, Comparison of the suitability of CBCT-and MR-based synthetic CTs for daily adaptive proton therapy in head and neck patients, Phys Med Biol 65, 235036 (2020).

S. Olberg et al., Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy, Med Phys 46, 4135–4147 (2019).

M. C. Florkow, F. Zijlstra, L. G. Kerkmeijer, M. Maspero, C. A. van den Berg, M. van Stralen, and P. R. Seevinck, The impact of MRI-CT registration errors on deep learning-based synthetic CT generation, in Medical Imaging 2019: Image Processing, volume 10949, page 1094938, International Society for Optics and Photonics, 2019.
J. C. Reinhold, B. E. Dewey, A. Carass, and J. L. Prince, Evaluating the impact of intensity normalization on MR image synthesis, in *Medical Imaging 2019: Image Processing*, edited by E. D. Angelini and B. A. Landman, SPIE, 2019.

S. Kida, T. Nakamoto, M. Nakano, K. Nawa, A. Haga, J. Kotoku, H. Yamashita, and K. Nakagawa, Cone beam computed tomography image quality improvement using a deep convolutional neural network, *Cureus* **10** (2018).

L. Chen, X. Liang, C. Shen, S. Jiang, and J. Wang, Synthetic CT generation from CBCT images via deep learning, *Med Phys* **47**, 1115–1125 (2020).

S. Kida, S. Kaji, K. Nawa, T. Imae, T. Nakamoto, S. Ozaki, T. Ohta, Y. Nozawa, and K. Nakagawa, Visual enhancement of Cone-beam CT by use of CycleGAN, *Med Phys* **47**, 998–1010 (2020).

N. Yuan, B. Dyer, S. Rao, Q. Chen, S. Benedict, L. Shang, Y. Kang, J. Qi, and Y. Rong, Convolutional neural network enhancement of fast-scan low-dose cone-beam CT images for head and neck radiotherapy, *Phys Med Biol* **65**, 035003 (2020).

X. Liang, L. Chen, D. Nguyen, Z. Zhou, X. Gu, M. Yang, J. Wang, and S. Jiang, Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy, *Phys Med Biol* **64**, 125002 (2019).

Y. Li, J. Zhu, Z. Liu, J. Teng, Q. Xie, L. Zhang, X. Liu, J. Shi, and L. Chen, A preliminary study of using a deep convolutional neural network to generate synthesized CT images based on CBCT for adaptive radiotherapy of nasopharyngeal carcinoma, *Phys Med Biol* **64**, 145010 (2019).

A. Barateau et al., Comparison of CBCT-based dose calculation methods in head and neck cancer radiotherapy: from Hounsfield unit to density calibration curve to deep learning, *Med Phys* **47**, 4683–4693 (2020).

Y. Liu, Y. Lei, T. Wang, Y. Fu, X. Tang, W. J. Curran, T. Liu, P. Patel, and X. Yang, CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy, *Med Phys* (2020).
136 G. Landry, D. Hansen, F. Kamp, M. Li, B. Hoyle, J. Weller, K. Parodi, C. Belka, and C. Kurz, Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations [J], Phys Med Biol 64 (2019).

137 C. Kurz, M. Maspero, M. H. Savenije, G. Landry, F. Kamp, M. Pinto, M. Li, K. Parodi, C. Belka, and C. A. Van den Berg, CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation, Phys Med Biol 64, 225004 (2019).

138 A. Thummerer, P. Zaffino, A. Meijers, G. G. Marmitt, J. Seco, R. J. Steenbakkers, J. A. Langendijk, S. Both, M. F. Spadea, and A. C. Knopf, Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy, Phys Med Biol 65, 095002 (2020).

139 A. Radford, L. Metz, and S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434 (2015).

140 T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive growing of gans for improved quality, stability, and variation, arXiv preprint arXiv:1710.10196 (2017).

141 O. Oktay et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999 (2018).

142 A. P. Leynes, J. Yang, F. Wiesinger, S. S. Kaushik, D. D. Shanbhag, Y. Seo, T. A. Hope, and P. E. Larson, Direct pseudoCT generation for pelvis PET/MRI attenuation correction using deep convolutional neural networks with multi-parametric MRI: zero echo-time and dixon deep pseudoCT (ZeDD-CT), J Nuc Med, jnumed–117 (2017).

143 A. Baydoun et al., Dixon-based thorax synthetic CT generation using Generative Adversarial Network, Intelligence-Based Medicine 3, 100010 (2020).

144 K. Gong, J. Yang, K. Kim, G. El Fakhri, Y. Seo, and Q. Li, Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images, Phys Med Biol 63, 125011 (2018).

145 H. Jang, F. Liu, G. Zhao, T. Bradshaw, and A. B. McMillan, Deep learning based MRAC using rapid ultrashort echo time imaging, Med Phys 45, 3697–3704 (2018).
A. Torrado-Carvajal, J. Vera-Olmos, D. Izquierdo-Garcia, O. A. Catalano, M. A. Morales, J. Margolin, A. Soricelli, M. Salvatore, N. Malpica, and C. Catana, Dixon-VIBE deep learning (DIVIDE) pseudo-CT synthesis for pelvis PET/MR attenuation correction, Journal of nuclear medicine 60, 429–435 (2019).

P. Blanc-Durand, M. Khalife, B. Sgard, S. Kaushik, M. Soret, A. Tiss, G. El Fakhri, M.-O. Habert, F. Wiesinger, and A. Kas, Attenuation correction using 3D deep convolutional neural network for brain 18F-FDG PET/MR: Comparison with Atlas, ZTE and CT based attenuation correction, PloS one 14, e0223141 (2019).

K. Gong, P. K. Han, K. A. Johnson, G. El Fakhri, C. Ma, and Q. Li, Attenuation correction using deep learning and integrated UTE/multi-echo Dixon sequence: evaluation in amyloid and tau PET imaging, Eur J Nucl Med Mol Imaging, 1–11 (2020).

A. Pozaruk, K. Pawar, S. Li, A. Carey, J. Cheng, V. P. Sudarshan, M. Cholewa, J. Grummet, Z. Chen, and G. Egan, Augmented deep learning model for improved quantitative accuracy of MR-based PET attenuation correction in PSMA PET-MRI prostate imaging, Eur J Nucl Med Mol Imaging (2020).

K. Gong, J. Yang, P. E. Larson, S. C. Behr, T. A. Hope, Y. Seo, and Q. Li, MR-based attenuation correction for brain PET using 3D cycle-consistent adversarial network, IEEE Transactions on Radiation and Plasma Medical Sciences (2020).

F. Liu, H. Jang, R. Kijowski, T. Bradshaw, and A. B. McMillan, Deep learning MR imaging–based attenuation correction for PET/MR imaging, Radiology 286, 676–684 (2018).

H. Arabi, G. Zeng, G. Zheng, and H. Zaidi, Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI, Eur J Nucl Med Mol Imaging 46, 2746–2759 (2019).

K. D. Spuhler, J. Gardus, Y. Gao, C. DeLorenzo, R. Parsey, and C. Huang, Synthesis of patient-specific transmission data for PET attenuation correction for PET/MRI neuroimaging using a convolutional neural network, J Nucl Med 60, 555–560 (2019).
F. Liu, H. Jang, R. Kijowski, G. Zhao, T. Bradshaw, and A. B. McMillan, A deep learning approach for 18 F-FDG PET attenuation correction, EJNMMI physics 5, 1–15 (2018).

X. Dong, T. Wang, Y. Lei, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang, Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging, Phys Med Biol 64, 215016 (2019).

K. Armanious, T. Hepp, T. Küstner, H. Dittmann, K. Nikolaou, C. La Fougère, B. Yang, and S. Gatidis, Independent attenuation correction of whole body [18 F] FDG-PET using a deep learning approach with Generative Adversarial Networks, EJNMMI research 10, 1–9 (2020).

K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

J. J. Van Dyk, editor, The Modern Technology of Radiation Oncology, volume 4, Medical Physics Publisher, 2020.

B. Stemkens, E. S. Paulson, and R. H. Tijssen, Nuts and bolts of 4D-MRI for radiotherapy, Phys Med Biol 63, 21TR01 (2018).

C. Paganelli et al., MRI-guidance for motion management in external beam radiotherapy: current status and future challenges, Phys Med Biol 63, 22TR03 (2018).

J. N. Freedman, H. E. Bainbridge, S. Nill, D. J. Collins, M. Kachelrieß, M. O. Leach, F. McDonald, U. Oelfke, and A. Wetscherek, Synthetic 4D-CT of the thorax for treatment plan adaptation on MR-guided radiotherapy systems, Phys Med Biol 64, 115005 (2019).

T. R. Goodman, A. Mustafa, and E. Rowe, Pediatric CT radiation exposure: where we were, and where we are now, Pediatric Radiol 49, 469–478 (2019).
A. Walker, P. Metcalfe, G. Liney, V. Batumalai, K. Dundas, C. Glide-Hurst, G. P. Delaney, M. Boxer, M. L. Yap, J. Dowling, D. Rivest-Henault, E. Pogson, and L. Holloway, MRI geometric distortion: Impact on tangential whole-breast IMRT, J App Clin Med Phys 17, 7–19 (2016).

C. Gustafsson, F. Nordström, E. Persson, J. Brynolfsson, and L. E. Olsson, Assessment of dosimetric impact of system specific geometric distortion in an MRI only based radiotherapy workflow for prostate, Phys Med Biol 62, 2976–2989 (2017).

M. Maspero, M. D. Tyyger, R. H. Tijssen, P. R. Seevinck, M. P. Intven, and C. A. van den Berg, Feasibility of magnetic resonance imaging-only rectum radiotherapy with a commercial synthetic computed tomography generation solution, Phys Imag Radiat Oncol 7, 58–64 (2018).

J. J. Lagendijk, B. W. Raaymakers, A. J. Raaijmakers, J. Overweg, K. J. Brown, E. M. Kerkhof, R. W. van der Put, B. Härdemark, M. van Vulpen, and U. A. van der Heide, MRI/linac integration, Radiother Oncol 86, 25–29 (2008).

B. G. Fallone, The rotating biplanar linac–magnetic resonance imaging system, 24, 200–202 (2014).

S. Mutic and J. F. Dempsey, The ViewRay system: magnetic resonance–guided and controlled radiotherapy, 24, 196–199 (2014).

P. J. Keall et al., The Australian magnetic resonance imaging–linac program, 24, 203–206 (2014).

D. A. Jaffray, M. C. Carlone, M. F. Milosevic, S. L. Breen, T. Stanescu, A. Rink, H. Alasti, A. Simeonov, M. C. Sweitzer, and J. D. Winter, A facility for magnetic resonance–guided radiation therapy, 24, 193–195 (2014).

D. Winkel, G. H. Bol, P. S. Kroon, B. van Asselen, S. S. Hackett, A. M. Werensteijn-Honingh, M. P. Intven, W. S. Eppinga, R. H. Tijssen, L. G. Kerkmeijer, H. C. de Boer, S. Mook, G. J. Meijer, J. Hes, M. Willemsen-Bosman, E. N. de Groot-van Breugel, I. M. Jürgenliemk-Schulz, and B. W. Raaymakers, Adaptive radiotherapy: The Elekta Unity MR-linac concept, Clin Transl Radiat Oncol 18, 54–59 (2019).
W. A. Hall et al., The transformation of radiation oncology using real-time magnetic resonance guidance: A review, Eur J Cancer 122, 42–52 (2019).

M. L. G. Koerkamp, Y. J. M. de Hond, M. Maspero, C. Kontaxis, S. Mandija, J. E. Vasmel, R. K. Charaghvandi, M. E. P. Philippens, B. van Asselen, H. J. G. D. van den Bongard, S. S. Hackett, and A. C. Houweling, Synthetic CT for single-fraction neoadjuvant partial breast irradiation on an MRI-linac, Phys Med Biol xx, xxxxx (2021).

J. Boda-Heggemann, F. Lohr, F. Wenz, M. Flentje, and M. Guckenberger, kV cone-beam CT-based IGRT, Strahlen Onkol 187, 284–291 (2011).

U. V. Elstrøm, L. P. Muren, J. B. Petersen, and C. Grau, Evaluation of image quality for different kV cone-beam CT acquisition and reconstruction methods in the head and neck region, Acta Oncologica 50, 908–917 (2011).

M. Peroni, D. Ciardo, M. F. Spadea, M. Riboldi, S. Comi, D. Alterio, G. Baroni, and R. Orecchia, Automatic segmentation and online virtualCT in head-and-neck adaptive radiation therapy, Int J Radiat Oncol Biol Phys 84, e427–e433 (2012).

C. Veiga, J. Alshaikhi, R. Amos, A. M. Lourenço, M. Modat, S. Ourselin, G. Royle, and J. R. McClelland, Cone-beam computed tomography and deformable registration-based “dose of the day” calculations for adaptive proton therapy, Int J Part Ther 2, 404–414 (2015).

Y.-K. Park, G. C. Sharp, J. Phillips, and B. A. Winey, Proton dose calculation on scatter-corrected CBCT image: Feasibility study for adaptive proton therapy, Med Phys 42, 4449–4459 (2015).

C. Kurz, R. Nijhuis, M. Reiner, U. Ganswindt, C. Thieke, C. Belka, K. Parodi, and G. Landry, Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images, Radiat Oncol 11, 1–9 (2016).

K. Arai et al., Feasibility of CBCT-based proton dose calculation using a histogram-matching algorithm in proton beam therapy, Physica Medica 33, 68–76 (2017).

C. Gomà, I. P. Almeida, and F. Verhaegen, Revisiting the single-energy CT calibration for proton therapy treatment planning: a critical look at the stoichiometric method, Phys Med Biol 63, 235011 (2018).
J. Harms, Y. Lei, T. Wang, M. McDonald, B. Ghavidel, W. Stokes, W. J. Curran, J. Zhou, T. Liu, and X. Yang, Cone-beam CT-derived relative stopping power map generation via deep learning for proton radiotherapy, Med Phys 47, 4416–4427 (2020).

D. C. Hansen, G. Landry, F. Kamp, M. Li, C. Belka, K. Parodi, and C. Kurz, ScatterNet: A convolutional neural network for cone-beam CT intensity correction, Med Phys 45, 4916–4926 (2018).

G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, Image Reconstruction is a New Frontier of Machine Learning, IEEE Trans Med Imag 37, 1289–1296 (2018).

G. Wang, Y. Zhang, X. Ye, and X. Mou, Machine Learning for Tomographic Imaging, IOP Publishing, 2019.

Y. Li, K. Li, C. Zhang, J. Montoya, and G.-H. Chen, Learning to Reconstruct Computed Tomography Images Directly From Sinogram Data Under A Variety of Data Acquisition Conditions, IEEE Trans Med Imag 38, 2469–2481 (2019).

A. K. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen, Learning with known operators reduces maximum error bounds, Nat Machine Intell 1, 373–380 (2019).

K. Lønning, P. Putzky, J.-J. Sonke, L. Reneman, M. W. Caan, and M. Welling, Recurrent inference machines for reconstructing heterogeneous MRI data, Med Image Anal 53, 64–78 (2019).

D. Izquierdo-Garcia, S. J. Sawiak, K. Knesaurek, J. Narula, V. Fuster, J. Machac, and Z. A. Fayad, Comparison of MR-based attenuation correction and CT-based attenuation correction of whole-body PET/MR imaging, European journal of nuclear medicine and molecular imaging 41, 1574–1584 (2014).

I. Shiri, H. Arabi, P. Geramifar, G. Hajianfar, P. Ghafarian, A. Rahmim, M. R. Ay, and H. Zaidi, Deep-JASC: joint attenuation and scatter correction in whole-body 18 F-FDG PET using a deep residual network, European Journal of Nuclear Medicine and Molecular Imaging (2020).
192 C. Shorten and T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6, 1–48 (2019).

193 Z. Li, K. Kamnitsas, and B. Glocker, Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation, in International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 402–410, Springer, 2019.

194 Z. Hang, G. Orazio, F. Iuri, and K. Jan, Loss Functions for Neural Networks for Image Processing, CoRR abs/1511.08861 (2015).

195 J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Işgum, Deep MR to CT synthesis using unpaired data, in Int Work SASHIMI, pages 14–23, Springer, 2017.

196 A. Rehman and F. G. Khan, A deep learning based review on abdominal images, Multimedia Tools and Applications , 1–32 (2020).

197 S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, 3D deep learning on medical images: a review, Sensors 20, 5097 (2020).

198 K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, DeepMedic for brain tumor segmentation, in International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries, pages 138–149, Springer, 2016.

199 J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Med Image Anal 53, 197–207 (2019).

200 P. Keeling, J. Clark, and S. Finucane, Challenges in the clinical implementation of precision medicine companion diagnostics, Expert review of molecular diagnostics 20, 593–599 (2020).

201 J. Bertholet et al., Patterns of practice for adaptive and real-time radiation therapy (POP-ART RT) part II: Offline and online plan adaption for interfractional changes, Radiother Oncol 153, 88–96 (2020).
E. Palmér, A. Karlsson, F. Nordström, K. Petruson, C. Siversson, M. Ljungberg, and M. Sohlin, Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic resonance imaging only workflow for head and neck radiotherapy, Phys Imag Radiat Oncol 17, 36–42.

Council of European Union, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, 2017, http://data.europa.eu/eli/reg/2017/745/oj.

C. Fiorino, M. Guckenberger, M. Schwarz, U. A. van der Heide, and B. Heijmen, Technology-driven research for radiotherapy innovation, Mol Oncol 14, 1500–1513 (2020).

R. Beckers, Z. Kwade, and F. Zanca, The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics, Phys Med 83, 1–8 (2021).

V. Liesbeth et al., Overview of artificial intelligence-based applications in radiotherapy: recommendations for implementation and quality assurance, Radiother Oncol (2020).

X. Liu, S. C. Rivera, D. Moher, M. J. Calvert, and A. K. Denniston, Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension, Brit Med J 370 (2020).

J. A. Dowling and J. Korhonen, MR-Only Methodology, in MRI for Radiotherapy, pages 131–151, Springer International Publishing, 2019.

J. Teuho, J. Johansson, J. Linden, A. E. Hansen, S. Holm, S. H. Keller, G. Delso, P. Veit-Haibach, K. Magota, V. Saunavaara, T. Tolvanen, M. Teras, and H. Iida, Effect of Attenuation Correction on Regional Quantification Between PET/MR and PET/CT: A Multicenter Study Using a 3-Dimensional Brain Phantom, J Nuc Med 57, 818–824 (2016).

J. J. Wyatt, J. A. Dowling, C. G. Kelly, J. McKenna, E. Johnstone, R. Speight, A. Henry, P. B. Greer, and H. M. McCallum, Investigating the generalisation of an
atlas-based synthetic-CT algorithm to another centre and MR scanner for prostate MR-only radiotherapy, Phys Med Biol 62, N548–N560 (2017).

211 E. Persson, C. Gustafsson, F. Nordström, M. Sohlin, A. Gunnlaugsson, K. Petruson, N. Rintelä, K. Hed, L. Blomqvist, B. Zackrisson, T. Nyholm, L. E. Olsson, C. Siversson, and J. Jonsson, MR-OPERA: A Multicenter/Multivendor Validation of Magnetic Resonance Imaging–Only Prostate Treatment Planning Using Synthetic Computed Tomography Images, Int J Radiat Oncol Biol Phys 99, 692–700 (2017).

212 P. Greer, J. Martin, M. Sidhom, P. Hunter, P. Pichler, J. H. Choi, L. Best, J. Smart, T. Young, M. Jameson, T. Afinidad, C. Wratten, J. Denham, L. Holloway, S. Sridharan, R. Rai, G. Liney, P. Raniga, and J. Dowling, A Multi-center Prospective Study for Implementation of an MRI-Only Prostate Treatment Planning Workflow, Front Oncol 9 (2019).

213 G. Loi, M. Fusella, C. Vecchi, S. Menna, F. Rosica, E. Gino, N. Maffei, E. Menghi, A. Savini, A. Roggio, L. Radici, E. Cagni, F. Lucio, L. Strigari, S. Strolin, C. Garibaldi, C. Romanò, M. Piovesan, P. Franco, and C. Fiandra, Computed Tomography to Cone Beam Computed Tomography Deformable Image Registration for Contour Propagation Using Head and Neck, Patient-Based Computational Phantoms: A Multicenter Study, Pract Radiat Oncol 10, 125–132 (2020).

214 S. J. Pan and Q. Yang, A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering 22, 1345–1359 (2010).

215 V. Cheplygina, M. de Bruijne, and J. P. Pluim, Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis, Med Image Anal 54, 280–296 (2019).

216 W. Li, S. Kazemifar, T. Bai, D. Nguyen, Y. Weng, Y. Li, J. Xia, J. Xiong, Y. Xie, A. Owrangi, and S. Jiang, Synthesizing CT images from MR images with deep learning: model generalization for different datasets through transfer learning, 7, 025020 (2021).

217 S. Mutic, J. R. Palta, E. K. Butker, I. J. Das, M. S. Huq, L.-N. D. Loo, B. J. Salter, C. H. McCollough, and J. Van Dyk, Quality assurance for computed-tomography sim-
ulators and the computed-tomography-simulation process: report of the AAPM Radiation Therapy Committee Task Group No. 66, Med Phys 30, 2762–2792 (2003).

218 R. R. Gallas, N. Hünemohr, A. Runz, N. I. Niebuhr, O. Jäkel, and S. Greilich, An anthropomorphic multimodality (CT/MRI) head phantom prototype for end-to-end tests in ion radiotherapy, Zeitsch Mediz Phys 25, 391–399 (2015).

219 N. Niebuhr, W. Johnen, G. Echner, A. Runz, M. Bach, M. Stoll, K. Giske, S. Greilich, and A. Pfaffenberger, The ADAM-pelvis phantom—an anthropomorphic, deformable and multimodal phantom for MRgRT, Phys Med Biol 64, 04NT05 (2019).

220 K. Singhrao, J. Fu, H. H. Wu, P. Hu, A. U. Kishan, R. K. Chin, and J. H. Lewis, A novel anthropomorphic multimodality phantom for MRI-based radiotherapy quality assurance testing, Med physics 47, 1443–1451 (2020).

221 E. Colvill et al., Anthropomorphic phantom for deformable lung and liver CT and MR imaging for radiotherapy, Phys Med Biol 65, 07NT02 (2020).

222 X. Chen, K. Men, B. Chen, Y. Tang, T. Zhang, S. Wang, Y. Li, and J. Dai, CNN-based quality assurance for automatic segmentation of breast cancer in radiotherapy, Front Oncol 10 (2020).

223 F. J. Bragman, R. Tanno, Z. Eaton-Rosen, W. Li, D. J. Hawkes, S. Ourselin, D. C. Alexander, J. R. McClelland, and M. J. Cardoso, Uncertainty in multitask learning: joint representations for probabilistic MR-only radiotherapy planning, in International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 3–11, Springer, 2018.

224 M. Hemsley, B. Chugh, M. Ruschin, Y. Lee, C.-L. Tseng, G. Stanisz, and A. Lau, Deep Generative Model for Synthetic-CT Generation with Uncertainty Predictions, in International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 834–844, Springer, 2020.

225 M. Abdar et al., A review of uncertainty quantification in deep learning: Techniques, applications and challenges, arXiv preprint arXiv:2011.06225 (2020).
226 D. Kawahara, A. Saito, S. Ozawa, and Y. Nagata, Image synthesis with deep convolutional generative adversarial networks for material decomposition in dual-energy CT from a kilovoltage CT, Comp Biol Med 128, 104111 (2020).

227 L. B. Jans, M. Chen, D. Elewaut, F. Van den Bosch, P. Carron, P. Jacques, R. Wittoek, J. L. Jaremko, and N. Herregods, MRI-based synthetic CT in the detection of structural lesions in patients with suspected sacroiliitis: comparison with MRI, Radiol, 201537 (2020).

228 V. E. Staartjes, P. R. Seevinck, W. P. Vandertop, M. van Stralen, and M. L. Schröder, Magnetic resonance imaging-based synthetic computed tomography of the lumbar spine for surgical planning: a clinical proof-of-concept, Neurosurgical Focus 50, E13 (2021).

229 E. M. McKenzie, A. Santhanam, D. Ruan, D. O’Connor, M. Cao, and K. Sheng, Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge, Med Phys 47, 1094–1104 (2020).

230 F. Siedek, S. Y. Yeo, E. Heijman, O. Grinstein, G. Bratke, C. Heneweer, M. Puesken, T. Persigehl, D. Maintz, and H. Grüll, Magnetic Resonance-Guided High-Intensity Focused Ultrasound (MR-HIFU): Technical Background and Overview of Current Clinical Applications (Part 1), RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren 191, 522–530 (2019).

231 J. Jiang, Y.-C. Hu, N. Tyagi, P. Zhang, A. Rimner, J. O. Deasy, and H. Veeraraghavan, Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets, Med Phys 46, 4392–4404 (2019).

232 J. P. Kieselmann, C. D. Fuller, O. J. Gurney-Champion, and U. Oelfke, Cross-modality deep learning: Contouring of MRI data from annotated CT data only, Med Phys (2020).